UNIVERSITAT DE BARCELONA

Measuring and reducing cognitive load in map-based data visualisations

An analysis of communication mechanisms to improve decision-making processes in three different contexts: climate, media and administration

Luz Calvo Flores

Doctoral Dissertation

Luz Calvo Flores

Advisors:
Mario Pérez-Montoro, PhD
Fernando Cucchietti, PhD

Measuring and reducing cognitive load in map-based data visualisations

An analysis of communication mechanisms to improve decision-making processes in three different contexts: climate, media and administration.

UNIVERSITAT DE BARCELONA

2023

Universitat de Barcelona
PhD program in Information Communication

Faculty of Information and Audiovisual Media

Acknowledgements

Words cannot express my gratitude to my advisors, Mario Pérez-Montoro and Fernando Cucchietti, for their patience and feedback. I also could not have undertaken this journey without Marta Terrado and Isadora Christel who generously provided invaluable knowledge and expertise. Additionally, this endeavour would not have been possible without the generous support from Steve Watts, who helped me improve the readability of my publications. I am also grateful to users and participants who kindly and unselfishly collaborate to evaluate my research.

Lastly, I would be remiss in not mentioning my friends and family, especially my partner, Alberto Almajano, to support me during these years. His belief in me has kept my spirits and motivation high during this process. I would also like to thank Charo P., Mario P. and Nadia K. for all the emotional support and unconditional love. You are always on my mind.

Abstract (English - Spanish)

Maps are one of the oldest and most versatile visual representations, still in use today in many fields, not only as military or terrain analysis tools (as their origins were) but also as information storage, for educational, administrative, scientific and climatic purposes. They are a familiar metaphor for most readers and are present in magazines, reports, digital media, press and television. Their familiarity and popularity often make them prime candidates to act as a framework for data representation, when communicating social, climate and political data, among others. While the use of maps is often the most obvious choice when presenting data in a visual format, sometimes their use to support the visualisation of complex data in infographics and visual tools is not the best alternative, especially where decision making is key. Through three distinct scenarios: climate, media and public administration, we will measure the impact on cognitive load of our users by analysing uncertainty, spatio-temporal and multi-categorical data respectively. We use different metrics, as well as qualitative and quantitative methodologies to analyse the effectiveness of geospatial visualisations in decision making tools. We examine ways to reduce the cognitive load to improve the user experience and understanding of the data presented. For the first scenario (climate), we simplify the visual information channels in order to shorten the response time and comprehension when visualising uncertainty. In the second scenario (media), we use different maps depending on the users' objectives, in order to reinforce the detection of trends over time. For the third scenario (administration), we customise the visualisations, allowing the user to dive into more specific data to improve comparison among several categories. The common objective of these three scenarios is favouring decision making.

Los mapas son una de las representaciones visuales más antiguas y polivalentes, que aún en nuestros días siguen vigentes en múltiples ámbitos, no ya sólo militares o como herramientas para analizar el terreno (tal y como fueron sus orígenes) si no también como almacenamiento de información, con fines educativos, administrativos, científicos y climáticos. Constituyen una metáfora familiar para la mayoría de los lectores y están presentes en revistas, informes, medios digitales, prensa y televisión. Su uso y popularidad los convierten a menudo en los principales candidatos para actuar como marco de representación de datos, a la hora de comunicar datos sociales, climáticos o políticos, entre otros. Aunque el uso de mapas suele ser la opción más obvia a la hora de presentar los datos en un formato visual, a veces su uso para apoyar la visualización de datos complejos en infografías y herramientas visuales no es la mejor alternativa, especialmente cuando la toma de decisiones es clave. A través de tres escenarios distintos: investigación climática, los medios de comunicación y la administración pública, mediremos el impacto en la carga cognitiva de nuestros usuarios analizando la incertidumbre de datos predictivos climáticos, datos socioeconómicos y datos multi categóricos respectivamente. Utilizamos diferentes métricas, así como metodologías cualitativas y cuantitativas para analizar la eficacia de las visualizaciones geo-espaciales utilizadas en herramientas diseñadas para la toma de decisiones. Examinamos, a su

vez, formas de reducir la carga cognitiva para mejorar la experiencia del usuario y la comprensión de los datos presentados. En el primer escenario (investigación clima), simplificamos los canales de información visual para acortar el tiempo de respuesta al visualizar la incertidumbre. En el segundo escenario (medios de comunicación), comparamos la eficiencia de diferentes mapas para reforzar la detección de tendencias a lo largo del tiempo. En el tercer escenario (administración), personalizamos las visualizaciones, permitiendo al usuario bucear en datos más específicos para mejorar la comparación entre varias categorías. El objetivo común de estos tres escenarios es favorecer la toma de decisiones.

*This dissertation is presented in the form of a **compendium of publications**, complying with the requirements of the PhD program Information and Communication in terms of authorship, type of publication and languages. The texts of each article, the category of the journals and the identifiers of each of the four publications obtained from this study are included at the end of the document (more details are included in the Introduction section).*

Figures

Tables

Index

1.Introduction

In recent decades visualising data effectively has become more and more essential. This is partly due to the ever increasing rise of Big Data and the need to understand behavioural patterns and to convey large volumes of complex information. The problem with the visual representation of complex data is that it is often left to data visualisation experts or data scientists who encourage the use of visual information channels without taking into account the cognitive limitations of the audience, or their lack of familiarity with visualisations.

Visual representations involving the use of maps to understand geospatial data are very familiar to the audience, as they have been used for centuries (during war, and for economic or strategic purposes) and continue to be used frequently in the print and general media.

However, despite their familiarity, the use of maps brings with it a number of challenges (space use, time representation, overlapping, among others), which calls into question their effectiveness in assisting the users in understanding the content and facilitating decision-making.

In addition to these problems, we face a diversity of standards and a lack of clear guidelines in the visual representation of complex data on maps for multiple scenarios (climate, media, administration) where, paradoxically, communication between institutions or broadcasters and the people who consume this information is key.

Data visualisation, until now seen as a tool for analysis, needs to take on a new dimension in which it acts as an information architecture environment in itself. It is no longer only a question of facilitating comprehension and decision-making, but also of guaranteeing and offering the minimum and necessary knowledge, at each step of the search process when accessing more detailed or in-depth information.

In this study we examine the complexity and limitations presented by the visualisation of geospatial data and the mechanisms which reduce cognitive load on our users, in order to favour decision-making. We analyse three different contexts: the representation of uncertainty in climate data, temporary changes in socio-economic data and lastly, the representation of multi-categorical data in the field of administration. At the same time, we review good practices commonly applied in these disciplines and provide a list of recommendations that can serve as guidance for climate service providers aiming to co-develop more effective service visualisations.

1.1 Doctoral Program directives

We remind you that this thesis is a compendium of publications from the UB Doctoral Programme in Information and Communication. The requirements in this modality include the following requirements: The publication of three articles of which:

- At least two articles must be listed in WoS or Scopus.
- In at least two articles the candidate must appear as first author, one as second author.
- At least two articles must be written in English.

Of which we fulfilled all conditions:

- **Four articles appearing in WoS or Scopus (article 1,2,3 and 4).**
- **Two articles as first author (article 1 and 4), two articles as second author (articles 2 and 3).**
- **Four articles in English (article 1,2,3 and 4).**

The following is a list of the articles published, as well as the characteristics of the journals:

Article 1
Type of contribution: Article
Article type: Indexed at Web of Science present and JCR (SCIE, SSCI)
Authors (sorted by signature): Luz Calvo, Isadora Christel, Marta Terrado, Mario Pérez-Montoro, Fernando Cucchietti
Title: Users' Cognitive Load: A Key Aspect to Successfully Communicate Visual Climate Information
Year: 2021
Journal (title, volume, pages): Bulletin of the American Meteorological Society, v. 103 (1), p. 1-42
ISSN: 0003-0007 EISSN: 1520-0477
Category: METEOROLOGY & ATMOSPHERIC SCIENCES
Quartil: 1
Impact factor (IF): 9,116
Category rank: 6/94

Quality indicators:
For the past 100 years, the American Meteorological Society (AMS) has been publishing the world's premier research in the atmospheric, oceanic, and hydrologic sciences—from high-impact journals to award-winning books to a long-standing series of meteorological monographs. With outstanding author support and post-publication marketing efforts, AMS remains the publisher of choice for the weather, water, and climate community. Since 1919, AMS has been publishing some of the top research in the atmospheric, oceanic, and hydrologic sciences. Today, we publish more than 37,000 pages per year across 12 scientific journals and a series of monographs. Peer reviewed and highly

respected, our journals are consistently ranked near the top of their fields in impact factor (IF). The Bulletin of the American Meteorological Society is the flagship magazine of AMS and publishes articles of interest and significance for the weather, water, and climate community as well as news, editorials, and reviews for AMS members.

Summary:
The visual communication of climate information is one of the cornerstones of climate services. It often requires the translation of multidimensional data to visual channels by combining colours, distances, angles, and glyph sizes. However, visualisations including too many layers of complexity can hinder decision-making processes by limiting the cognitive capacity of users, therefore affecting their attention, recognition, and working memory. Methodologies grounded on the fields of user-centred design, user interaction, and cognitive psychology, which are based on the needs of the users, have a lot to contribute to the climate data visualisation field. Here, we apply these methodologies to the redesign of an existing climate service tool tailored to the wind energy sector. We quantify the effect of the redesign on the users' experience performing typical daily tasks, using both quantitative and qualitative indicators that include response time, success ratios, eye-tracking measures, user perceived effort, and comments, among others. Changes in the visual encoding of uncertainty and the use of interactive elements in the redesigned tool reduced the users' response time by half, significantly improved success ratios, and eased decision-making by filtering non relevant information. Our results show that the application of user-centred design, interaction, and cognitive aspects to the design of climate information visualisations reduces the cognitive load of users during tasks performance, thus improving user experience. These aspects are key to successfully communicating climate information in a clearer and more accessible way, making it more understandable for both technical and nontechnical audiences.

Article 2
Type of contribution: Article
Article type: Indexed at Web of Science present and JCR (SCIE, SSCI)
Authors (sorted by signature): Marta Terrado, Luz Calvo, Isadora Christel
Title: Towards more effective visualisations in climate services: good practices and recommendations

Year: 2022
Journal (title, volume, pages): Climatic Change, v. 172(1), p. 1-2
ISSN: 0165-0009 **eISSN:** 1573-1480
Category: METEOROLOGY & ATMOSPHERIC SCIENCES
Quartil: 1
Impact factor (IF): 5,174
Category rank: 23/94

Quality indicators:
Springer is a leading global scientific, technical and medical portfolio, providing researchers in academia, scientific institutions and corporate R&D departments with quality content through innovative information, products and services. Springer has one of the strongest STM and HSS eBook collections and archives, as well as a comprehensive range of hybrid and open access journals and books under the SpringerOpen imprint. Springer is a part of Springer Nature, a global publisher that serves and supports the research community. Springer Nature aims to advance discovery by publishing robust and insightful science, supporting the development of new areas of research and making ideas and knowledge accessible around the world.

Summary:
Visualisations are often the entry point to information that supports stakeholders' decision- and policy-making processes. Visual displays can employ either static, dynamic or interactive formats as well as various types of representations and visual encodings, which differently affect the attention, recognition and working memory of users. Despite being well-suited for expert audiences, current climate data visualisations need to be further improved to make communication of climate information more inclusive for broader audiences, including people with disabilities. However, the lack of evidence-based guidelines and tools makes the creation of accessible visualisations challenging, potentially leading to misunderstanding and misuse of climate information by users. Taking stock of visualisation challenges identified in a workshop by climate service providers, we review good practices commonly applied by other visualisation-related disciplines strongly based on users' needs that could be applied to the climate services context. We show how lessons learned in the fields of user experience, data visualisation, graphic design and psychology make useful recommendations for the development of more effective climate service visualisations.

Article 3
Type of contribution: Article
Article type: Indexed at Web of Science present and JCR (SCIE, SSCI)
Authors (sorted by signature): Marta Terrado, **Luz Calvo,** Dragana Bojovich, Isadora Christel
Title: Current Practice in Climate Service Visualization: Taking the Pulse of the Providers' Community
Year: 2022
Journal (title, volume, pages): Bulletin of the American Meteorological Society, v. 8, p. 28-37
ISSN: 0003-0007 EISSN: 1520-0477
Category: METEOROLOGY & ATMOSPHERIC SCIENCES
Quartil: 1
Impact factor (IF): 9,116
Category rank: 6/94

Quality indicators:
For the past 100 years, the American Meteorological Society (AMS) has been publishing the world's

premier research in the atmospheric, oceanic, and hydrologic sciences—from high-impact journals to award-winning books to a long-standing series of meteorological monographs. With outstanding author support and post-publication marketing efforts, AMS remains the publisher of choice for the weather, water, and climate community. Since 1919, AMS has been publishing some of the top research in the atmospheric, oceanic, and hydrologic sciences. Today, we publish more than 37,000 pages per year across 12 scientific journals and a series of monographs. Peer reviewed and highly respected, our journals are consistently ranked near the top of their fields in impact factor (IF). The Bulletin of the American Meteorological Society is the flagship magazine of AMS and publishes articles of interest and significance for the weather, water, and climate community as well as news, editorials, and reviews for AMS members.

Summary:
Effective visualisations of climate information that can be easily understood by non–climate experts are strongly needed. At present, the absence of a common or standardised visualisation approach for climate services results in the application of different practices by climate service providers, sometimes leading to users' misinterpretation or misuse of climate information. In this report, we analyse the outputs from a workshop with climate service providers that had the aim to identify current practices and challenges faced when developing visualisations in the field of climate services. The analysis of the results obtained depicted the current status of the climate services visualisation field, identified the main lessons learned by different projects, and highlighted challenges that required further research efforts. Insights obtained from the analysis are valid for climate service providers worldwide.

Article 4
Type of contribution: Article
Article type: Indexed at Web of Science present and JCR (SCIE, SSCI)
Authors (sorted by signature): Luz Calvo, Mario Pérez-Montoro, Fernando Cucchietti
Title: Measuring the effectiveness of Static Maps to Communicate Changes over time
Year: 2022
Journal (title, volume, pages): IEEE TRANSACTIONS ON VISUALISATION AND COMPUTER GRAPHICS, p. 1-13
ISSN: 1077-2626 **eISSN:** 1941-0506
Category: COMPUTER SCIENCE, SOFTWARE ENGINEERING
Quartil: 1
Impact factor (IF): 5,226
Category rank: 13/110

Quality indicators: IEEE Transactions on Visualization and Computer Graphics publishes papers on subjects related to computer graphics, information and scientific visualisation, visual analytics, and virtual and augmented reality, focusing on theory, algorithms, and research-inspiring applications. Each author enjoys the advantage of becoming part of IEEE's stellar reputation and

125-year history of innovation. Past and present members include nearly two dozen Nobel Prize-winning innovators. Today, IEEE has over 400,000 professional and student members who will shape technology for years to come. Researchers rely on IEEE for trusted information. Eleven million downloads each month is strong evidence that the IEEE Xplore® digital library provides authors with the “findability” they seek. Consistent quality and high demand make IEEE publications trusted sources for researchers in corporations, academia, and government.
IEEE journals and conference proceedings receive nearly three times the patent citations of competing publishers, which gives IEEE authors a distinct mark of credibility within their field of interest.IEEE maintains active partnerships with abstracting and indexing providers such as Elsevier, Clarivate, ProQuest, IET, and NLM to maximise the discovery of author works. As a member of CrossRef, IEEE Xplore articles include outbound reference links to other publisher content registered in the CrossRef system. IEEE Xplore is indexed by Google™, allowing Google search results to include links to IEEE Xplore.

Summary:
Both in digital and print media, it is common to use static maps to show the evolution of values in various regions over time. The ability to communicate local or global trends, while reducing the cognitive load on readers, is of vital importance for an audience that is not always well versed in map interpretation. This study aims to measure the efficiency of four static maps (choropleth, tile grid map and their banded versions) to test their usefulness in presenting changes over time from a user experience perspective. We first evaluate the effectiveness of these map types by quantitative performance analysis (time and success rates). In a second phase, we gather qualitative data to detect which type of map favours decision-making. On a quantitative level, our results show that certain types of maps work better to show global trends, while other types are more useful when analysing regional trends or detecting the regions that fit a specific pattern. On a qualitative level, those representations which are already familiar to the user are often better valued despite having lower measured success rates.

2. Previous work

In this section, we review previous literature related to the three scenarios and use cases approached in this thesis: uncertainty in climate visualisation, temporary changes over time when communicating socio-economic data in the media, and multi-categorical data analysis for the administration, in order to identify advances in data visualisation, as well as unresolved issues in each of the fields.

2.1. The complexity of climate data representation

The accessibility to climate information has implications for how society makes the best use of scientific knowledge to adapt to climate change (Harold et al., 2016). For decades, climate service providers have faced the challenge of how to best communicate climate-related data and information, together with its inherent uncertainty, in an easy to understand way for both expert and non-expert users (Kaye et al., 2012; Lorenz et al., 2015). This has resulted in service providers often using visual representations to supply climate knowledge tailored to users' decision-making processes (Gerst et al., 2020a; Taylor et al., 2015a).

Visualisations of complex climate data tend to prioritise solutions that take into account the greatest combination of variables and dimensions, e.g., colour, size, position, opacity of the glyphs used to represent a value. This complexity calls for formats paying special attention to visual encoding, which encompasses the translation of multidimensional data into visual elements on a chart or map representation. Visual encoding is useful in the sense that it allows us to convey a higher amount of information in a single visualisation (Grainger et al., 2016; Lloyd, 1997). However, it rarely takes into account whether the information needs to be displayed all at the same time, with certain visual aesthetics, or if it will be too complex to be interpreted correctly (Cleveland & McGill, 1985; Sager et al., 2007). Indeed, aesthetics might be worth considering if striving to create something memorable, that helps to raise awareness about a specific scientific challenge (Borkin et al., 2013). However, an attractive visualisation cannot qualify as effective unless it accurately conveys something meaningful or credible reinforcing the target message of the graph (Holmes, 1984; Kosara, 2013).

According to (Stephens et al., 2012), for a visualisation to be effective, it is important to consider a balance between information density, robustness (the representation of scientific confidence and consensus), and the relevance of the information to the user's needs. Although visualising climate forecast uncertainties and associated probabilities have been thought to increase user trust (Joslyn & LeClerc, 2012; Roulston et al., 2006), it does not automatically lead to better decisions (Greis et al., 2015). This is especially critical when the visual elements used to represent uncertainty compete with the limited cognitive resources of the users such as attention, recognition, and working memory (Antifakos et al., 2004; McInerny et al., 2014). Visualisations requiring a high cognitive load, that is, involving a high amount of working memory resources and attention mechanisms, can have negative effects on users' understanding and learning and impact negatively on their ability to complete a task or make an informed decision (Cairo, 2012; Few, 2009; McInerny et al., 2014). In such cases, the use of interactive elements can offer mechanisms for progressively dosing the information that is to be shown (Bertin, 2010; Veras Guimarães, 2019; Ware et al., 2002; Yoghourdjian et al., 2018).

Methodologies grounded in the fields of user-centred design (UCD), human-computer interaction (HCI), and cognitive psychology have much to contribute to the field of climate data visualisation (Bevington et al., 2019; Christel et al., 2018). UCD techniques involve the input of users throughout the design process in order to create highly usable visualisation tools based on their needs (B. M. Davis et al., 2020; Dong et al., 2008; Yucong et al., 2019). HCI makes use of interactive elements to allow users to decide what and when to see, or to show only those values which are relevant for a given task. This facilitates decision-making, returns control to users, and allows them to discard non relevant information at any given time (Gerharz & Pebesma, 2009; Lau & Vande Moere, 2007; Ware, 2012). Within the framework of UCD methodologies, technologies such as the eye-tracker have been used to quantify and analyse visual patterns, attention, and cognitive aspects (Fu et al., 2016; Krejtz et al., 2018). Adding cognition and perception (i.e processes about how humans acquire knowledge, understanding and interpretation) can help detect and solve initial design problems. Rather than just favouring visual exploration, these disciplines offer increasingly effective methods to develop and evaluate visualisation systems that explicitly take into account real-world user requirements (Block, 2013).

Tools and learnings from the UCD field have already been applied to the visualisation of climate information and climate services. For instance, (Argyle et al., 2017) showed how incorporating usability evaluation into the design of decision support systems can improve the efficiency, effectiveness, and user experience of a weather forecasting application. Other studies have similarly applied UCD methodologies to reinforce the design of climate information websites, apps, and prototypes (Khamaj et al., 2019; Ling et al., 2015; Oakley & Daudert, 2016). Design elements have also been introduced in the development of climate services to increase their usability, in particular for the renewable energy sector (Christel et al., 2018). On the other hand, cognitive and psychological sciences have been used in the visualisation of climate data. Some examples are the use of cognitive psychology methods to help make information provided by IPCC graphs more accessible to expert and non-expert audiences (Harold et al., 2016) and improve users' task performance (Hegarty et al., 2010). Differences in the interpretation of climate graphs between experienced and inexperienced users have been explored elsewhere (Atkins & Mcneal, 2018; Gerst et al., 2020a), both for climate change variables and for temperature and precipitation forecasts.

2.2. Lack of standards in climate visualisation

The need for standards in climate services is a topic that has recently gained momentum. Guidelines for the quality management of climate services have been proposed by the World Meteorological Organisation

(WMO, 2018) and broad standards for climate change adaptation and mitigation also exist (ISO 9000 and 14,000 standards families). In addition, some basic recommendations on communicating forecast uncertainty have been published (M. Davis et al., 2016; WMO, 2018), but there is no standardised approach guiding the development of the visual component of climate services (See Fig. 1 and Fig. 2). This lack of guidance has favoured the proliferation of different practices, often resulting in stakeholders spending more time trying to understand the approach used for representing data than focusing on the interpretation of the information itself. This lack of standards has also been reported to lead to misunderstanding or misuse of climate information (C. D. Hewitt et al., 2021). Due to the diversity of content and stakeholders, determining a one-size-fits-all visualisation practice for climate services would be a daunting task, if not impossible.

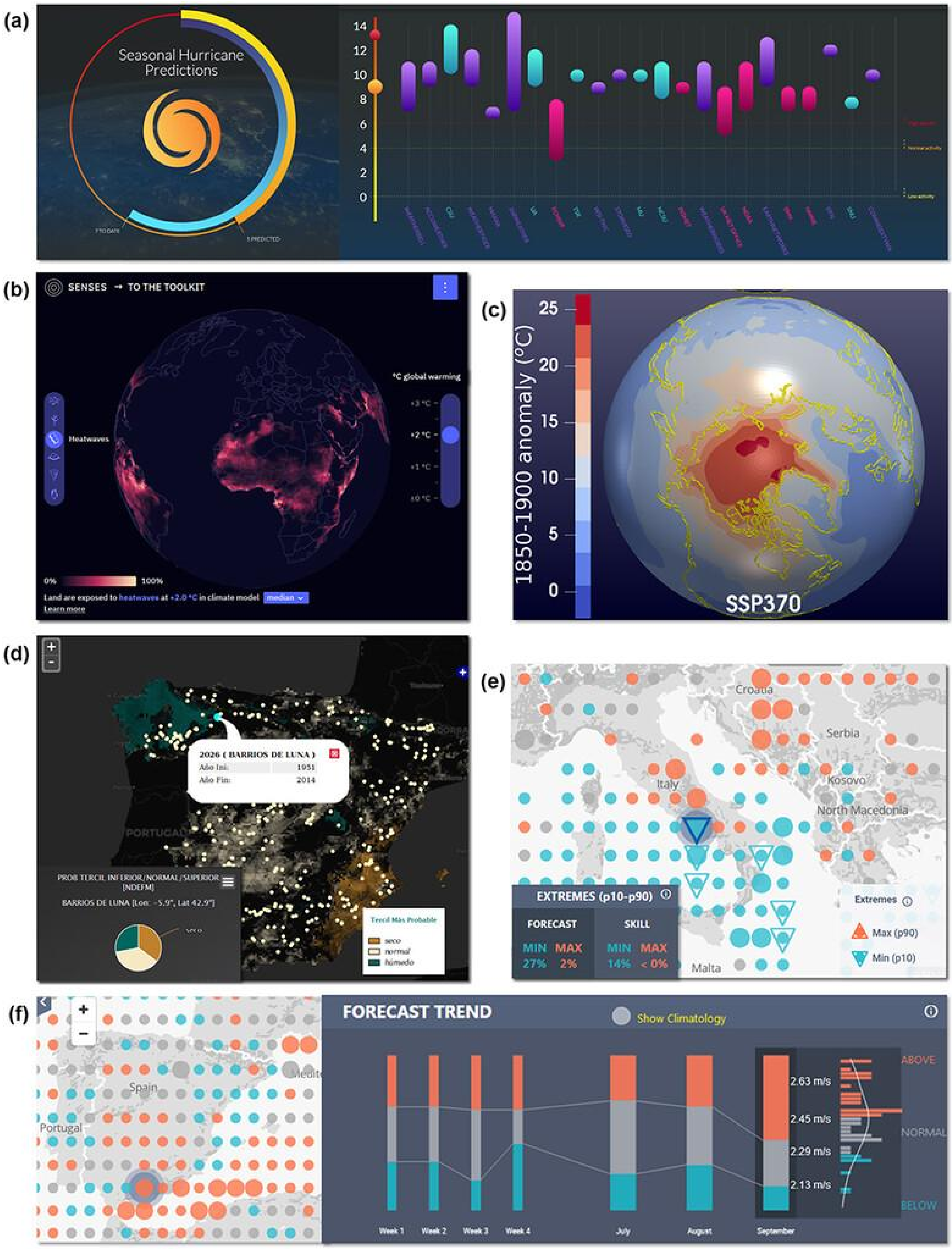


Fig. 1: Representation of first-order uncertainty in different climate service visualisations: (a) average forecast value, (b) average forecast value for different scenarios, (c) anomalies, (d) most likely category (e.g., terciles), (e) probability of different categories (e.g., extreme events probabilities), and (f) probability distribution function. Source projects: (a) Seasonal hurricane predictions platform (*www.seasonalhurricanepredictions.org*), (b) SENSES (www.senses-project.org; Auer et al. 2021), (c) eClimViz (Walton et al. 2021), (d) MEDSCOPE, and (e),(f) S2S4E (*www.s2s4e-dst.bsc.es*).

2.2.1. Lacks of standards in uncertainty data visualisation

As a standardised mapping approach to represent climate uncertainty is lacking, different techniques have been applied for this purpose, often resulting in users spending more time trying to understand the mapping approach than focusing on the interpretation of the presented information itself (Kaye et al., 2012). In general, users' familiarity with a type of data visualisation has been found to play a significant role in the process of reading and making sense of maps and graphs (Lorenz et al., 2015). Commonly used representations of climate data include choropleths, heat maps, and line charts (Taylor et al., 2015b). Although familiar to users, these elements have limitations when used to communicate climate data to decision-makers in a way that is transparent, understandable, and that does not lead to a false sense of certainty. An example of such limitations is seen when communicating climate predictions from the next two weeks up to a few decades into the future. A characteristic of such predictions is that they are probabilistic, meaning they provide information on the probability of a certain climate outcome to occur (e.g., winds below or above a threshold, not optimal for the energy production). In addition, climate predictions are often given in the form of large amounts of data covering the whole globe, and their quality (i.e., level of success of a prediction against observationally based information) depends on the specific location and time (Kaye et al., 2012). Both aspects, probabilities and forecasts quality (referred to as skill by the climate science community), add complexity to the visual communication and can eventually compromise the understanding of climate predictions by users (Bonneau et al., 2014; Terrado et al., 2019).

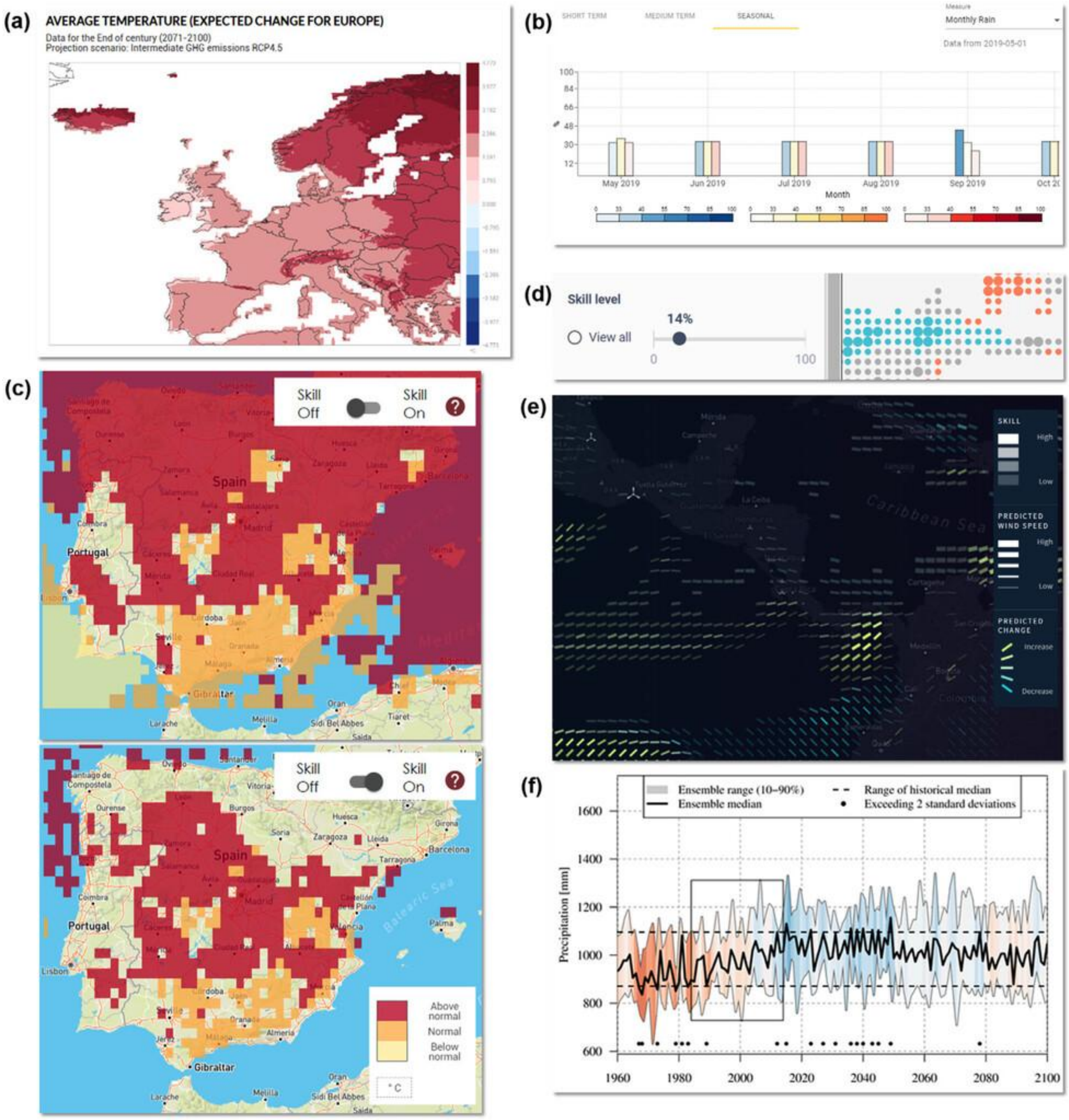


Fig. 2: Representation of uncertainty in different climate service visualisations: (a) forecast information shown, no indication of uncertainty, (b) high-uncertainty information not shown (e.g., replaced by a reference), no indication of uncertainty, (c) possibility to hide/show high-uncertainty information (e.g., activation of a mask), (d) possibility to hide/show information with different levels of uncertainty (e.g., filter by threshold), (e) uncertainty integrated in the visualisation through visual encoding, and (f) uncertainty represented as an additional parameter (e.g., uncertainty range). Source projects: (a) C3S Press Data Portal (https://climate.copernicus.eu/press-data-portal), (b) VISCA (Marcos-Matamoros et al. 2020), (c) MED-GOLD (https://dashboard.med-gold.eu), (d) S2S4E (www.s2s4e-dst.bsc.es), (e) Project Ukko (www.project-ukko.net), and (f) CIREG.

2.3. Map particularities: Choropleths and cartograms in the media.

Choropleth maps are a widely known representation and are common in media communications because of their popularity with target audiences. The use of easy-to-understand representations to illustrate articles helps in understanding the news items which they accompany, as well as improving the perception of reliability (Dovbysh, 2020; Esteves & Neves, 2021; Spinde et al., 2020). However, despite their familiarity for many audiences, choropleth maps present some associated problems such as "dark-is-more bias" (i.e ranking of colour lightness perception), the "area-size bias" (i.e small areas are less dominant than larger ones) and the "data-classification effect" (the established classification intervals used in detecting patterns) (Armstrong & Xiao, 2018; Schiewe, 2019b).

Tile grid maps, a kind of cartogram (see Fig. 3), are another widely used representation which effectively communicates broader trends and patterns summarising data (Besançon et al., 2020). They represent the different areas of the map making use of uniform size and shapes (often squares) and are arranged close to their real-world positions. The key to cartogram design is to use distortion or metaphors to change the size, shape, boundaries and/or geographical representation of regions, depending on statistical results or following a conceptual convention, but in a way which keeps the map recognisable to the user (Keim et al., 2002).

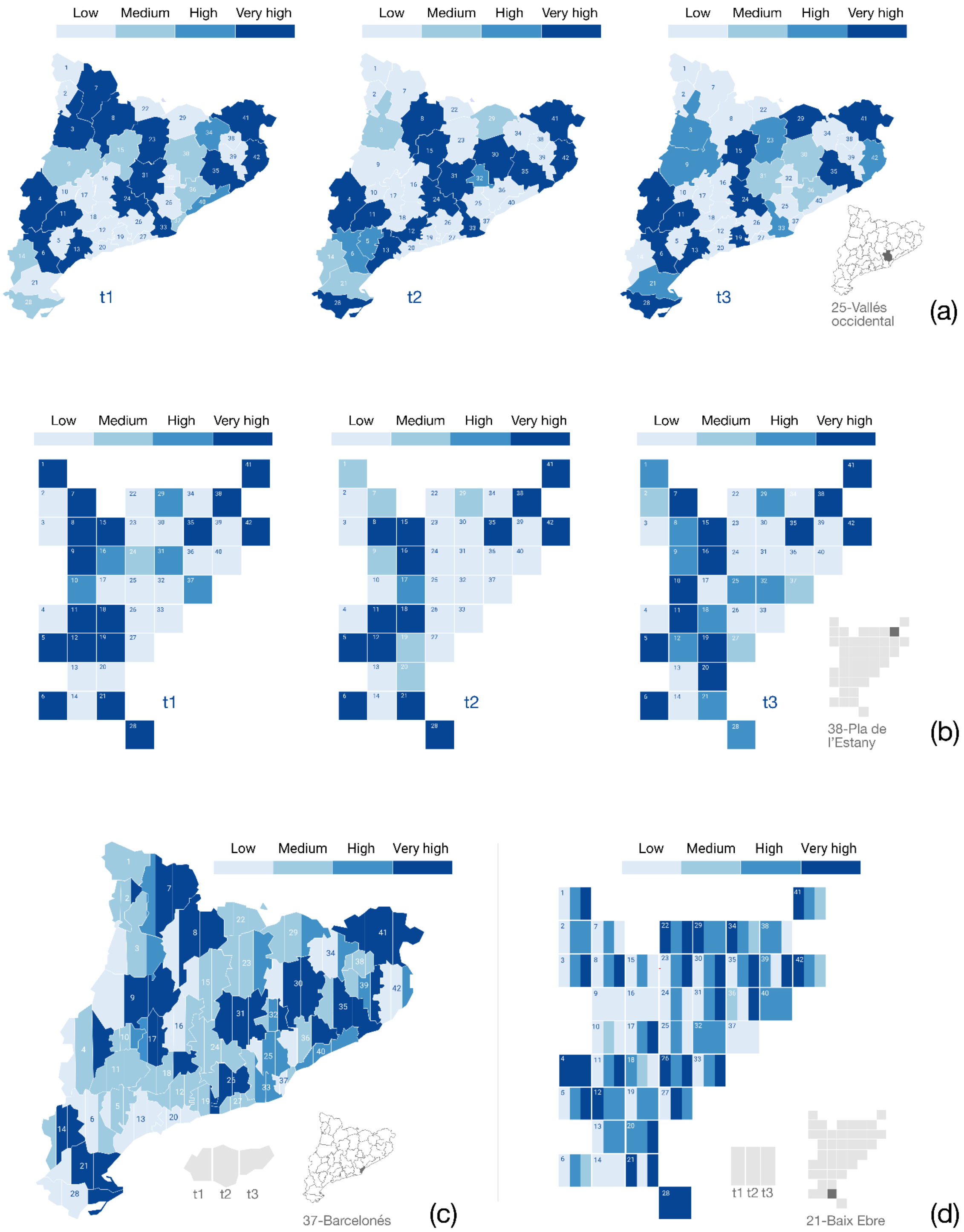


Fig. 3: Visualisations evaluated in the second experiment of this research: First, small multiple maps using Choropleth (a) and Tile grid map (b). Second, their banded versions, unifying in the same map three different time steps: Banded choropleth (c) and Banded tile grid map (d).

While less common, tile grid maps and their variants (Datavizcatalogue, 2018) offer an improvement over choropleth maps, and some other cartograms: for example, in solving the problem of distortion resulting from “area-size bias” (Nusrat et al., 2018; Nusrat & Kobourov, 2016; Ueland & Vincent, 2004). However, we must take into account some of their known limitations or weak points: lookalike aspects (they must bear a resemblance to the original map), topology and inaccuracy (neighbouring areas must be maintained), and misdirection (relative positions between neighbouring areas must be close to reality) in order to avoid misinterpretation (Langton and Solymosi, 2019; Wongsuphasawat, n.d.). In addition, not all geographies can be easily represented in a conceptual fashion, while at the same time maintaining their capabilities to be identified without resorting to the use of legends or labels (Griffin, 1980).

With regards to problems concerning recognition, Fotheringham points out that they can be especially difficult to interpret without resorting to the use of additional information (such as geographical elements) that helps users to locate or identify a specific region (Fotheringham et al., 2000). In the same way, another drawback is that they are often used as a creative solution instead of as a cartographic projection used to analyse a particular problem (Tobler, 2004).

To illustrate this, we can analyse the representation of the map of London using blocks of the same size to represent each neighbourhood. This visualisation takes advantage of The River Thames, using its meanders to provide a visual clue within the structure of the map as a metaphor that helps the audience to identify the city of London (Cheshire & Uberti, 2014). Nevertheless, we must bear in mind that these representations may not be easily identified by people who are not familiar with the geography of the area. Therefore, it is necessary to use tooltips and legends (see Fig. 8) to ensure the correct identification of the regions on the map (Engebretsen & Kennedy, 2020).

Depending on the map and its topography, we must also take into account human perception rules such as figure-ground, depth, readability, identification, and balance, which may affect the information that users obtain through them (Board & Taylor, 1977). The specific characteristics of the data (i.e. number of categories to show, magnitudes of the values of the different regions or areas, etc.) also force and limit us to using specific visualisations linked to a use-context that may require a dedicated or specific visualisation (Aigner et al., 2011).

Apart from the map representation itself, other factors such as visual search patterns and cognition must be considered in order to measure the efficiency of the alternatives discussed above. The analysis of visual search patterns on maps is a complicated process as it involves complex cognitive operations, which include the detection of relationships and the identification of different visual encodings (location, size, colour, shape, etc.). The perception of maps requires the identification of a greater number of symbols in a greater range of variability, in which short-term memory is insufficient to cover the global analysis of a task (Swoboda & Vighi, 2016) (Ciołkosz-Styk, n.d.).

2.4. Communicating changes over time using static maps

If we focus on the problems related to the representation of changes over time, the multiple static map strategy juxtaposes two or more maps to provide a simultaneous visual comparison in specific units of time or snapshots. A possible solution could be to employ dynamic strategies to present the maps in a temporal sequence or through an evolution of a geographical pattern using various dynamic sequences to favour memory processes (Moon et al., 2019b), (Slocum et al., 2006), (Araya et al., 2019). However, these techniques have limitations: the constraints of the human visual system (change blindness, foveal and peripheral attention) suggest that humans often do not easily perceive changes within dynamic graphics (Goldsberry & Battersby, 2009), (Fish et al., 2011). Hence the importance of taking into account these barriers of limited visual and cognitive user processing capabilities when designing time-related map representations. Studies on the cognitive effects and differences between the use of static and dynamic maps from a perception perspective, lead us to solutions such as providing users with instructions and the division of maps into fragments to help them focus their attention on the relevant areas (Harrower, 2007).

Recent studies by (Du et al., 2018) proposed a novel visualisation technique which attempts to solve the time dynamic sequencing issue: The banded choropleth map (see Fig. 3.c), which divides sub-regions of a map into partitions of the same area size and uses each partition to represent a different time step. In its favour, the technique makes better use of space than representations of small multiples and performs better than animations, where it would have been necessary to consume more memory and cognitive resources. However, it inherits the aforementioned problems associated with its precursor, the choropleth map.

2.5. Visualising multicategorical data of public opinion

When visualising survey and opinion results, the choice of the most suitable chart often depends on the question type (numeric, likert, multiple choice...) (J. Liu et al., 2018; Wexler, 2016) and the user expertise level (Liao et al., 2021), so the charts used to represent the data gathered must be clear and precise and also familiar to the users (M. Hewitt, 2016). Depending on the target audience, the use of familiar visualisations such as bar charts, line charts, donut charts or stacked bars is recommended (Nash, 2011) but they are inefficient when comparing values of multiple charts (Kozak et al., 2015). In the case of scales with more than two categories, it is advisable to use diverging colour scales (instead of the classic representation using qualitative colours), to indicate a higher or lower priority of the results presented (Petrillo et al., 2011; Pirrone, 2020). The use of less conventional representations such as waffle charts, dot charts, radar charts or lollipop charts is no longer effective as more than two categories are represented simultaneously (Nash, 2011; Wexler, 2016).

Effectively communicating multicategorical data containing more than three dimensions is a challenge, not only from a visualisation point of view but also from a user experience perspective (Sarkar, 2018). The most frequent existing alternatives combine different charts including information on the relationships between them (Cui et al., 2012; Marcus et al., 2011) and often involve multiple visual properties such as colour, size, position, shape, and the use of patterns and textures on multiple layers. However, when the number of categories is high, many of these representations lose their effectiveness and require evaluation from a perceptual point of view (Kucher et al., 2018).

Moreover, visualisation of multicategorical data is also closely linked to two key user tasks: classification and comparison, therefore knowledge of user expectations and their search patterns is essential when proposing effective, significant and easy-to-understand charts (Chan, n.d.; Tamaazousti et al., 2017).

2.6. The role of visualisation tools for public administration

The influence of public opinion in the state and administrative policy design process is one of the most urgent goals in achieving effective policies (Fernández-Prados et al., 2019a; Rachynska, 2020). The knowledge and analysis of public opinion implies, in the first place, knowing the events that affect it and the demand for actions by citizens towards public administrations. This knowledge allows us to learn of their interests, their level of trust in the administration and the mechanisms of education that contribute to direct and moderate the social response. Obtaining reliable data increases the effectiveness of the communicative dialogue between authorities and civil institutions (Rachynska, 2020).

Despite the strong relationship between social research and public policy, the most prominent international surveys have little or no presence in public policies in some countries (Fernández-Prados et al., 2019a).

People's way of thinking, behaviour patterns and innovative management ideas, have a high potential in the application in the field of public administration. Access to this data is critical to many aspects of society.

However, analysing this big data and how to use data visualisation in tools that allow us to show and optimise social phenomena are the challenges that government administrations face. The existing data visualisation tools often do not allow us to analyse, filter, compare or interact with data at a deeper level. Analysis of specific cases can provide transparency about the use of data visualisation and technology, in order to promote more sustainable practices (Jia et al., 2017).

There are numerous tools for the visualisation and analysis of open and statistical data, but they require a level of advanced technical knowledge or they lack simple, intuitive interfaces for understanding the information, and do not favour decision-making (Tambouris et al., 2017; Wheeldon & K. Ahlberg, 2021).

3. Purpose and objectives

Having reviewed these scenarios and their respective particularities, we identified a number of aspects that needed to be estimated and improved within an overall purpose, focused on the cognitive aspects associated with data visualisation in three different contexts: climate, media and administration.

3.1. Purpose

The main purpose of this study is to measure the impact of the reduction in cognitive load through different metrics (quantitative and qualitative) in the decision making process for different use contexts using map-based representations.

A. Use case: Representation of **predictive** data. *Sector: Climate.*
B. Use case: Representation of **time-varying socio-economic data**. *Sector: Communication and print media.*
C. Use case: **Multi-categorical public opinion data**. *Sector: Public administration.*

3.2. Objectives

In order to accomplish the purpose of this study, we addressed the following objectives:

- Analyse the impact of cognitive load reduction on the representation of uncertainty in climate data.
- Understand the challenges facing data visualisation in the climate domain.
- Establish a set of best practices based not only on the user experience, data visualisation or design, but also take into account the cognitive aspect.

- Evaluate the efficiency and difficulty in interpreting global and regional patterns of socio-economic data using different static maps.
- Design-Optimise the visualisation of spatio-temporal data by means of an alternative representation that favours the identification of trends.
- Analyse the effects that highly customisable visualisations have in reducing the cognitive load in the decision-making processes.

4.Methodology

Here, we will explain the main methodologies used to address our objectives (see section 4.2). Some of these methodologies are transversal to all of the objectives, while others focus instead on measuring the cognitive load of users by analysing visualisations through very specific scenarios. At the end of this section you will be able to consult a table (see Table 1) with all of the methodologies in order to compare their use across the three studies, as well as the benefits of using a particular methodology in order to reach objectives.

4.1. Cross-cutting methodologies:

4.1.1. Systematic review of scientific literature - Searching and Reading

We used systematic methods to find, select, and synthesise all available previous literature with regards to cognitive aspects in the field of data visualisation applied in the map representation, in order to establish a starting point for the research.

Our research question was: *Is complex data visualisation on maps effective for common scenarios (climate, media, administration) with regards to the audience and the cognitive load required to interpret patterns and to favour decision-making?*

We established a protocol that involved background information, the research objectives cited in the previous section (see Section 3.2) and methodologies.

We searched for relevant previous literature and resources in the following sources:

- **Databases:** Web of Science, Scopus and Google Scholar.
- **Handsearch:** Specific conferences and proceedings and the references list and citations of the articles
- **Experts:** Significant authors, specialised websites, online articles and blogs.
- **Selection criteria:** A first selection based on titles/abstracts and conclusions. Then a second review based on the full text: We read the titles, abstracts, main content and conclusions of the studies that

were identified in this three-scenario research. We excluded any study that did not refer to the research questions, sometimes after having read the full manuscript or other some specific sections that were ruled out definitively. We read the complete articles to decide if any more studies needed to be included in the selection criteria.

We also gathered data about the results, discussion ideas, methodologies applied, and outcomes.

4.2. Main methodologies for Objective 1: Simplifying uncertainty in climate data

To compare the cognitive load required to represent the uncertainty of two visualisations (the original and the redesigned tool), we combined quantitative (objective in terms of results and timing) and qualitative (more revealing) methodologies to understand behaviours and the underlying reasons behind the quantitative metrics:

4.2.1. Sample

To reach this objective, we used the same sample of participants in all the methodologies (quantitative and qualitative). A sample of 20 people was tested in order to detect 95% of problems (Faulkner, 2003; Virzi, 1992). The sample comprised 50% of men and 50% of women aged between 22 and 50, with a low-to-medium knowledge in data visualisation, which entailed being familiar with common charts or basic maps but without expertise in visualisation of climate or uncertainty information. The recruited participants were students and administrative staff in academia from the human resources, communication, and finance departments. Participants were asked not to have consumed stimulant substances that could have affected the test results.

4.2.2. **User Testing (Quantitative)**

Quantitative analysis was conducted through a two-task user testing session to be completed with both tools, Project Ukko and the redesign.

The development of this user testing session followed well-established recommendations of planning, moderation, and analysis (Faulkner, 2003; Nielsen, 1992) . The user testing session was moderated by an expert and recorded to measure time and analyse responses, after the test(Holtzblatt et al., 2004a). Participants were asked to perform two tasks (see section 4.2.6) with each of the tools, based on two typical daily work activities of the intended users (Anderson et al., 2011; Block, 2013, p. 201).

To compare and validate the results from the point of view of perception and cognitive load, we assessed the changes applied to the original tool (Project Ukko) to create the redesigned tool(S2S4E) and determine if they were fully effective for a general public and not only for an expert audience. To avoid interference in the comparison of cognitive load measurements between both visual representations, not all the improvements included in the S2S4E tool were applied in this study (e.g., colour blindness palette, labelling improvement and customizable elements). We conducted an experiment with non-experts to determine if the users' task performance when using Project Ukko had improved after redesigning the tool (taking into account user-centred design, visualisation techniques and changing the interaction of some interactive elements).

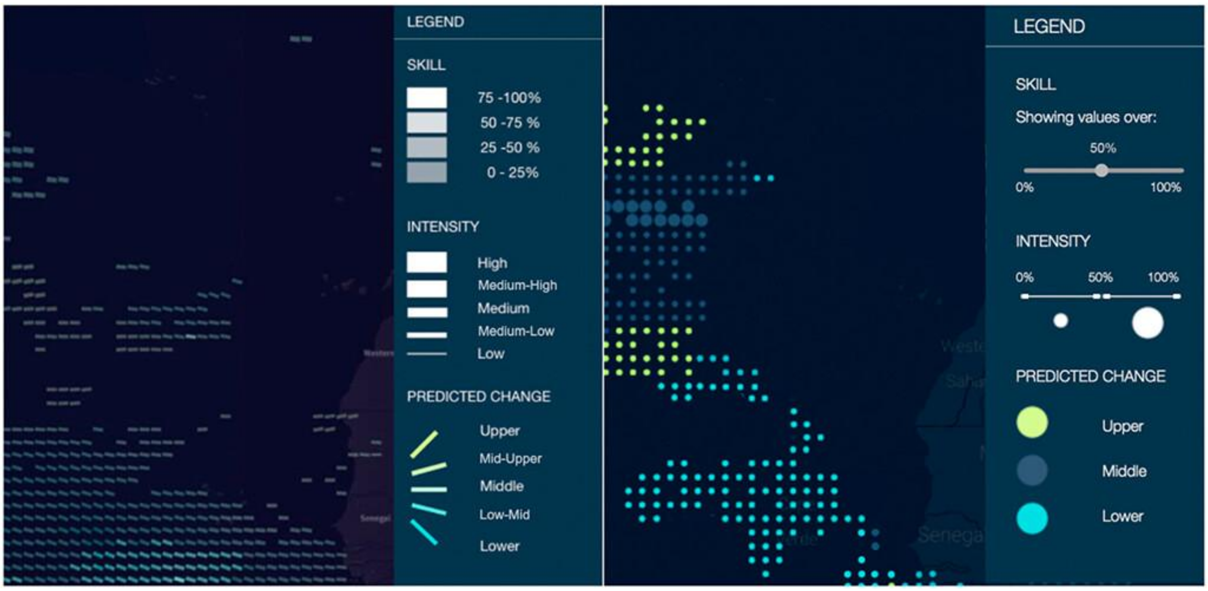


Fig. 4: (left) Project Ukko and (right) the redesigned tool based on user requirements, research and visual encoding techniques. Applied changes: filtering out glyphs under a certain prediction skill threshold; simplification of the intensity representation from five to only two sizes; and simplification of the predicted change colour scale down from five to three colours, where a similar colour to the background colour is used for values close to the historical average and more distinguishable colours are used for values higher and lower than the historical average. This was done because the middle values were not of interest to potential users.

4.2.3. Bipolar Laddering (Qualitative)

We used the bipolar laddering pocket methodology (qualitative indicator 8), which is a reduced version of the bipolar laddering technique. It consists of asking the user about three positive and three negative aspects of both tools (instead of the 10 aspects requested in the extended modality). Then the users rate (from 1 to 10) the aspects mentioned regarding their importance (for the positive ones) or their severity (for the negative ones). By evaluating users' subjective opinions using a scale from 1 to 10 we are able to better quantify input results that are qualitative, which helps identify the key aspects to work with or prioritise the problems that need to be solved (Pifarré & Tomico, 2007a).

4.2.4. Two-question quiz (Qualitative)

We also implemented a brief two-question quiz (Schrepp et al., 2017) to determine which visualisation needed more mental effort, and which was the tool preferred by participants for decision-making processes:

- Subjective effort perception (i.e., user perception on which tool was easier to use while performing the tasks (two-question quiz).
- Preference between both tools regarding the decision-making process (two-question quiz).

4.2.5. Eye Tracker (Quantitative)

The number of fixations, the fixation duration, and the number of visual paths to legend (quantitative indicators 3, 4, and 5) were monitored using an eye-tracker (GP3 Eye tracker of gazepoint), a sensor-based device that measures eye positions (i.e., point of regard) or eye movement.

4.2.6. Test environment and tasks

The same tasks were applied for user testing (See section 5.2.1) and Eye-tracker (see section 5.2.4) Participants were provided with some context about the real-world conditions in which the tasks would take place. In addition, all the information and conditions necessary to carry out the tasks were presented in the statements, visualisation, and captions shared with participants, for them to be able to perform the tasks without having an expert or climate science background. It was simply necessary to identify the requested areas or properties visually (M. C. Trivedi, 2012). Task 1 was aimed to test if the tool allows good detection of areas with particular conditions while task 2 was aimed to test the differentiability of glyph representation at specific locations.

- **Task 1 statement:** Locate or identify an area on the map that is appropriate to the location of a wind power plant. The area must meet a series of conditions: For Project Ukko, the suitable area should have a prediction skill over 50%, high or medium-high intensity, and an upper or mid-upper predicted change in wind speed (Fig. 4, left panel). For the redesigned version, the area should have a skill over 50%, intensity over 50%, and upper predicted change (Fig. 4, right panel). [This task required users to be able to identify at least one of the areas that met the conditions specified.]

- **Task 2 statement:** Identify aloud the conditions that occur in the points included in the highlighted area on the map in terms of skill, intensity, and predicted change. [This task required users to be able to identify the characteristics of two kinds of glyphs contained in a certain map area].

Also, the sequence of tasks was randomly presented to the users to avoid biases derived from the learning acquired during the completion of the tasks (Li et al., 2013).

5.2.7 Quantitative metrics

For both task 1 and task 2, we measured the cognitive load and task performance from a quantitative perspective using the following indicators. We gathered different metrics through the *User testing* methodology and the *Eye-tracker*:

1. Success **rates** when completing a task, including total or partial success (Ellis & Dix, 2006; Freitas et al., 2002; Winckler et al., 2004) *(User Testing).*
2. **Response time** when completing a task *(User Testing).*
3. Number of **fixations** (i.e., number of gaze points located very close in space, when the eyes are locked toward an object) *(Eye-tracker).*
4. **Fixation duration** (i.e., period of time allocated to a fixation (Majooni et al., 2018; Wang et al., 2014) *(Eye-tracker).*
5. Number of **accesses to legend** for the completeness of tasks (Klingner et al., 2008; Pretorius et al., 2005) *(Eye-tracker).*

Two first metrics were collected from the user Testing technique (see section 5.2.1) and Eye-tracker measurements (see section 5.2.4).

5.2.8 Statistical Analysis

To investigate whether there were differences between tools or tasks, a chi-squared test was performed for indicator 1 and a two-way analysis of variance (ANOVA) was performed for indicators 2 to 5 using R software (Core Team, 2018).

During the performance of tasks 1 and 2, we promoted the thinking-aloud technique to identify and detect usability problems (Nielsen, 1992; Olmsted-Hawala et al., 2010).

4.3. Main methodologies carried out for Objectives 2 and 3: Challenges facing data visualisation in the climate domain & Best practices in climate data visualisation

In order to identify the main problems in the representation of climate data, it was necessary to know the various practices, opinions and barriers of climate experts. To do so, we combined knowledge from different disciplines to carry out various co-creative methodologies.

4.3.1. Sample

A total of 25 participants attended the workshop and co-creative sessions, representing 22 projects working in climate services, including EU H2020 and ERA4CS projects and a few national projects and private contracts. Participants were from different European countries, including Spain, Belgium, Germany, Italy, the Netherlands, Ireland, Norway, Portugal and the UK.

4.3.2. Workshop (Qualitative)

This study takes as a departing point the challenges identified in an interactive visualisation workshop organised in the framework of a Coordination and Support Action on climate services (Terrado et al., 2022). The aim of the workshop was to monitor current practices applied by climate services providers when developing visualisations and identify issues that need to be tackled for moving towards the development of more effective visualisations. The workshop consisted in a 2.5-h interactive session in breakout groups.

4.3.3. Co-creative sessions (Qualitative)

The purpose of these sessions were to convene target users and bring them into the design process. We co-created services and specific alternatives, investigated how communities work, and understood how to orient a specific solution. This was also likely to adopt a practice or service that it helped create, and also gain valuable insight into all facets of the final solution. Co-creative sessions were iterative to discuss, evaluate and improve the different alternatives proposed.

4.4. Methodologies Objective 4: Evaluate the efficiency of static maps

The methodologies used were aimed at measuring the different types of static maps for different tasks, not only from a quantitative point of view but also from a user's point of view: which maps favoured or hindered each of the tasks posed.

4.4.1. Sample

For this study we used a convenience sample of 32 people, composed of 15 men and 17 women between 22-50 years of age, with low-medium map visualisation knowledge. All the participants were familiar with choropleth maps, but none of them had any knowledge about the other representation types. The recruited participants were students, researchers and administrative staff from the fields of human resources, communication and accounting. Before the test sessions, users were asked not to have consumed stimulant substances that could have affected their performance and the test results. The participants were residents of Catalonia, 25% of them of foreign origin but long term residents (over 10 years) and familiar with the geographical distribution of Catalonia (see Table 7 of Annex I). The participants took part in the experiment one at a time in a private room, in order to ensure a disturbance free environment and at the start of business hours (from 9 to 12 h), complying with the protocols required for COVID-19 prevention. The duration of each test was about 15-20 minutes. The test sessions were carried out over the course of three weeks. All participants completed the experiment and no data was eliminated. The participants did not receive any remuneration or compensation for their participation.

4.4.2. **User Testing (Quantitative)**

We developed a simple online tool to guide the participants through the experiment, presenting each of the tasks, and which was controlled by a moderator who assisted the participants during the test. This assistance consisted of answering any queries related to the task statements but not in helping to solve the tasks themselves or giving any clues on the solutions (Babich, 2020). The sessions were also recorded in order to review the results in terms of quantitative data (time and success rates) and qualitative data (opinions and valuations). Participants were informed that times and results, together with their comments, would be collected.

4.4.2.1. Test environment and Taks

Our objective was to evaluate the effectiveness of choropleth and tile grid maps, and their banded versions, to show changes over time. We chose the tile grid map as an alternative to the choropleth as it overcomes some of its limitations (as mentioned in the previous sections)(Schiewe, 2019a). We aimed to conduct an experiment that covered the most common and possible tasks related to changes over time (see also Annex I, Table 1). The final tasks set out in the test were taken from a list of all the combinations for the three variables included in the study (region, time step, and value) and then the tasks which did not involve any changes over time were discarded (see also Fig.1, Fig.2 and Fig. 3 of Annex I).

Once these questions had been set, we reviewed them with potential users in order to define which questions and tasks they wanted to answer and perform while working with the maps. We also discussed the intended use of each selected type of task and a representative statement for each one. As a result, the final selected tasks were the following (see also Annex I, Table 6):

1. **Task 1 statement:** Detect trends within a specific region. Task statement: Could you describe the evolution of COVID-19 new cases for region number n (region name) over time (t1, t2, t3)?
2. **Task statement:** Could you name five regions that have remained stable in terms of number of new cases throughout these three periods of time?
3. **Task 3 statement:** Determine the global trend (all regions) over time. Statement: Could you determine what the global trend for all of Catalonia was over time?

4.4.2.2. Metrics

1. **Success rates** when completing a task, including total or partial success (Freitas et al. 2002; Winckler et al. 2004; Ellis and Dix 2006).
2. **Response time** when completing a task.
3. Number of times rated as the best / worst representation

4.4.3. Thinking aloud technique (Qualitative)

Participants were also encouraged to share their thoughts on the positive or negative aspects of each map, as well as any other observations while performing the task.

This helped us to understand some answers, mistakes or why users valued some maps more positively than others (McDonald et al., 2019; Michela & McDonald, 2020).

This approach of analysing each task individually as they are completed, favours review quality by using short term memory, which allows users to provide comments with the tasks still fresh in their minds, instead of having a questionnaire at the end of the test (Kittur et al., 2007). At the same time, this approach has no effect on the quantitative metrics measured (completion times and success rates).

4.4.4. Best / Worst rated (Quantitative)

After finishing the three tasks, we required the users to review the four representations (one after another) involved in each task. Users were asked about which representation favoured decision making and facilitated the resolution of the task (Hassenzahl & Sandweg, 2004).

4.4.5. Statistical analysis

We used ANOVA F-tests with significance level =.05 to carry out the statistical analysis for the time results together with a Tukey post-hoc test to reveal the most significant pairwises. We used a Chi-square test for independence with α=.05 to evaluate categorical variables in the success rate results. The within-subject independent variables were the four map representations. The dependent measurements are the completion time and success rate. The null hypothesis was that the representation type does not affect completion times or success rates. When the probability of the null hypothesis (p-value) is less than 0.05 (or, equivalently the F-value is greater than the critical F-value, Fcr), the null hypothesis is rejected.

4.5. Methodologies Objectives 4 and 5: New Maps design

New ways of visualising data change over time, require analysing the alternatives proposed to date, the problems they present, and evaluating possible new alternatives. The ideation of new ways of visualisation requires a vision from different points of view.

4.5.1. Ideation

Based on the previous literature review, different sketches and tests were proposed to design an alternative representation to the existing maps (Choropleth, Cartograms and Banded choropleth). The result was a simplification of the Banded choropleth map to a new alternative that would correct the problems associated with the areas, with the aim of testing its effectiveness with respect to the other alternatives. These sketches were carried out first with the sketch technique and in more advanced stages with the use of prototyping tools (Camburn et al., 2017).

4.6. Methodologies Objective 6: Highly customisable visualisations

For the design of a visualisation tool focused on solving specific problems, we need to understand and solve the needs of the users, without forgetting the ease of use of the available functionalities.

4.6.1. Surveys

With the help of an initial **survey**, we explored the general needs related to the information that would be relevant to show via the tool and the questions that had to be answered through the visualisation (see Appendices A1 and A2 to check the final variables and demographics included) (Fernández-Prados et al., 2019b).

4.6.2. Interviews

Based on individual **interviews** (Yang et al., 2016), it was possible to delve into more specific government needs when planning social actions and help identify target groups (those groups which are less engaged with environmental issues), motivate their participation, and/or find new ways for them to collaborate in actions which favour the environment.

4.6.3. Co-creative sessions

With the help of the collaborative tool MURAL, several possible solutions were identified and discussed in two online co-creation sessions. In these sessions, possible representations, combinations of different graphs and maps, and navigation for the final tool were discussed. The various alternatives were evaluated and potential users discussed the pros and cons of each alternative and possible new solutions and functionalities arose (Kim & Hullman, n.d.; Veer et al., 2009; Wassink et al., 2009).

These sessions contributed to a more detailed analysis of the problem, the common daily tasks initially described by potential users, the usefulness of the data selected, as well as the questions to be answered through the visualisation. They also allowed us to co-create sketches, wireframes and interactive prototypes, as well as to define the main visualisation interactions and navigation. These prototypes were used during the evaluation stage (Gill et al., 2011; Moon et al., 2019a).

4.6.4. Exploratory Data analysis with R

Exploratory Data Analysis (EDA) refers to the process of performing initial overview on data so as to discover patterns, anomalies and check some initial hypothesis and assumptions with the help of different statistics and visual representations (Nasser et al., 2006). EDA also serves to rule out certain final visualisations based on the morphology of the data and to guarantee that the visualisation selected to communicate our findings favours the exploration, understanding and exploitation of them (Rao et al., 2021). In this study we used R to visualise the different charts to verify that they were the most suitable to solve the requirements of the end users. These R files are available at (git@github.com:lcalvofl/public_opinion_tool.git).

4.6.5. Service design

To address the domain and visualisation challenges and favour a sustainable development, a service design approach was adopted to guide the tool creation process (Blevis, 2007; Prendeville & Bocken, 2017). This methodology included three main stages (see Fig. 5):

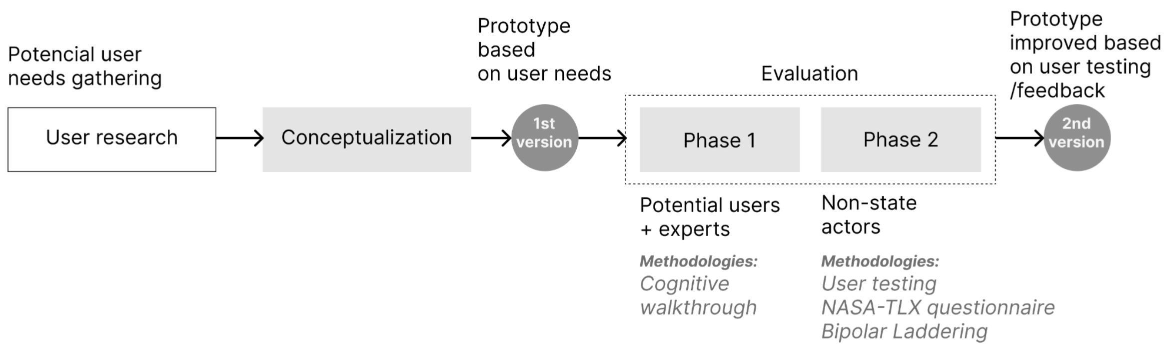


Fig. 5: Service design process including User research, conceptualization and evaluation. The qualitative evaluation stage was carried out in two different phases using different users and methodologies in each phase.

In the initial stage of **research**, requirements of potential users were collected as well as the identification and selection of the relevant data needed from the Eurobarometer. Once the research stage had been finalised, the **conceptualization** stage was carried out, in which possible solutions were explored. At the same time, an assessment of currently available tools was carried out to explore alternative solutions. Lastly, in the **evaluation** stage, and through qualitative metrics, the effectiveness of the tool was evaluated through a two-step process: Firstly with potential and expert users. Secondly, with the non-state actors in order to guarantee the usability of the tool (Holtzblatt et al., 2004a). In this stage, the insights gathered during the qualitative evaluation were also analysed in order to improve the final tool design (see Fig. 6).

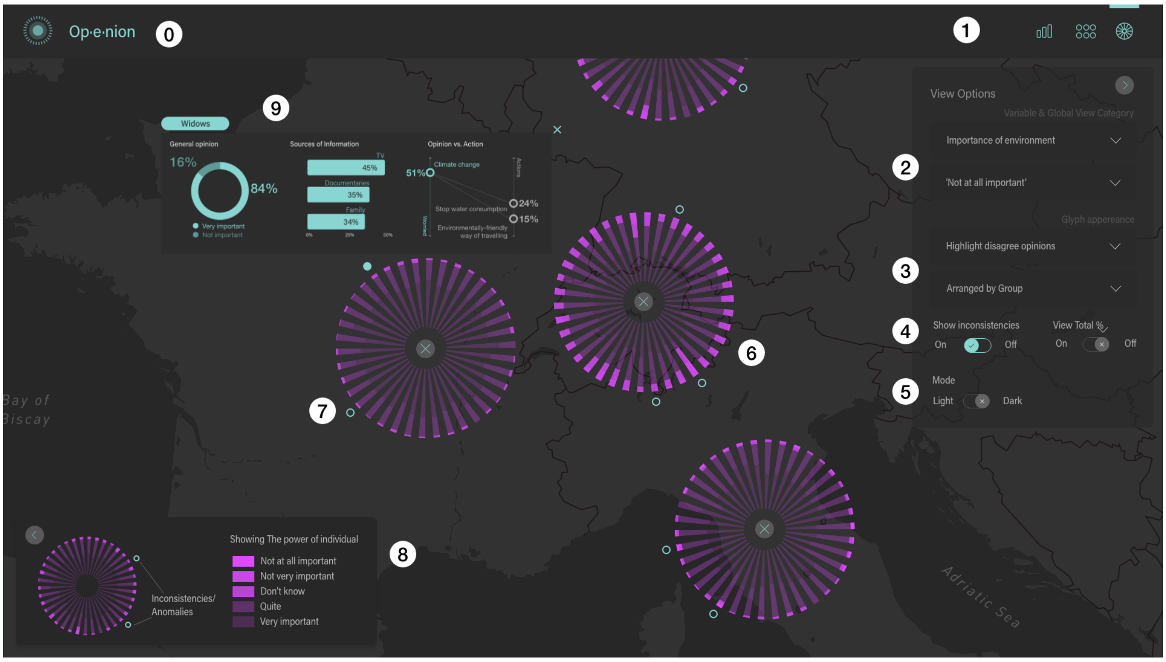


Fig. 6: Detailed view of the socio-demographic groups showing information inconsistencies of the tool in its dark version. On the right side we see the filter menu in which we select the variable to show (2), the type of answer (2) and the information we want to highlight (3). In the main navigation (top) we see the three main options: Summary, Small multiples and Explore options(1). Clicking in one of the inconsistency dots (7), we access the inconsistency detail panel (9). Specific legends are shown for each of the different views (8).

4.6.5.1. Evaluation : Cognitive walkthrough with experts

The first phase was carried out with **five potential users** who had participated in determining requirements. The profiles corresponded to city council staff, public entities, and freelancers with previous experience in public administration. Given that some of these profiles do not work with tools at a country level, the evaluation of functionalities was undertaken from an urban perspective, transferring the visualisation to a possible municipal use, for neighbourhood and district management.

The sessions were carried out using the **Cognitive Walkthrough** methodology (Allendoerfer et al., 2005; Wilson, 2014), through which the participants could explore the tool freely, discover and comment on the tool, as well as suggest improvements in terms of functionality. Employing the aforementioned methodology, a second round of individual sessions was held with experts in the areas of Data Visualisation (three participants), User Experience (two participants) and Graphic Design (one participant). In these sessions, the comments, in the case of the visualisation experts, focused on specific aspects of graphical representation, as well as the types of graphs used in the tool and their efficiency in effectively communicating the particularities in the data (Prümper et al., 1991). The user experience and graphic design experts shared their opinions on information architecture, interaction, usability and colour theory (Sauer et al., 2010a). Evaluation: User testing with non-state actors

In this second phase of the evaluation, we sought to validate the ease of use of the tool with the non-state actors in order to detect problems related to the comprehension of the visualisations and their ability to aid comparison and decision-making (Donker & Reitsma, 2007; Faulkner & Wick, 2005; Gerardo, 2007).

The convenience sample consisted of twenty participants (eleven men and nine women of ages between 25 and 66) with only a basic knowledge of data visualisation (Alroobaea and Mayhew, 2014). The participants had not received any previous training and were volunteers from the following fields: administration, marketing and communication environmental consultancies staff and independent activists. These users were required to perform the sessions remotely via the Zoom application due to COVID-19 limitations.

The participants were instructed not to take stimulants in the eight hours prior to the test, so as not to alter their cognitive abilities during the performance of the tasks (McLellan et al., 2016).

Through the Guerrilla user testing methodology, we gave a brief explanation to provide the necessary context to the participants ((Dr.) Munesh Trivedi & Khanum, 2012), and then we asked them to solve five tasks presented in logical order of search and exploration. The tasks set out in the test aimed to reproduce the common daily tasks of potential users that were gathered during the user research stage (see Section 4.1). We asked the users to explore the prototype using the thinking-aloud technique (Hertzum et al., 2015), while explaining to the researcher exactly what they were thinking and what their main doubts were. In addition, we asked them questions to check their understanding of some specific aspects of the prototype.

Task1: Identify which countries have less interest in environmental issues (General overview).

Task2: Detect less committed socio-demographic groups for one country (Detailed view by country) .

Task3: Compare the opinion (level of agreement, importance, etc.) towards a certain question from one socio-demographic group for multiple countries (Socio-demographic group comparison view) .

Task4: Detect and analyse inconsistencies between concern and action (Inconsistencies view).

Task5: Analyse summary visualisation ranking charts for less committed groups and countries (Summary view).

4.6.6. NASA-TLX based questionnaire

After the Guerrilla user testing, two additional activities were carried out in order to quantify the severity of the problems detected as well as the perceived cognitive load.

First, users completed a questionnaire on perceived difficulty for each task to assess cognitive load based on the NASA-TLX questionnaire categories (Cao et al., 2009b), where aspects related to perceived difficulty and cognitive resource use were evaluated.

- Mental Demand or level of concentration required.
- Physical Demand (not applicable for the type of tasks we are dealing with in this study).
- Temporal Demand (NASA-TLX).
- Performance demand (NASA-TLX) or the level of confidence while performing the task.
- Perceived Effort (NASA-TLX) or perceived difficulty while performing the task.
- Frustration Level (NASA-TLX).

4.6.7. Bipolar Laddering

Second, for insight gathering we used the Bipolar Laddering methodology, in which users comment on the positive and negative aspects encountered while performing the tasks during the Guerrilla user testing and rate them from 1 to 10 to assess their usefulness or severity in a more precise fashion. Additionally, the number of times that a specific aspect was mentioned by users gave us an idea of its weight or relevance (Pifarré & Tomico, 2007b).

- Number of positive and negative aspects (Bipolar Laddering)
- Severity or usefulness level for each aspect gathered (Bipolar Laddering)

Table 1. Summary of Methodologies and metrics

	Objective 1	Objective 2	Objective 3	Objective 4	Objective 5	Objective 6
Interviews						●
Ideation					●	
User Testing	●			●		●
Workshops & Co-creative sessions	●	●	●			●
Systematic literature review	●	●	●	●		●
Statistical analysis	●			●		
Cognitive walkthrough						●
Think aloud technique	●			●		●
Exploratory data analysis (EDA)						●
Questionnaires / NASA TLX	●			●		●
Bipolar laddering	●					●
Eye-Tracker	●					

Summary of section 3: Methodology
The main objective of the methodologies applied(apart from those focused on user needs research) was to measure the impact of cognitive load in the context of data visualisation. For that, we used a main quantitative methodology, the user test, focused on measuring the completion time and outcome of the tasks. Depending on the type of experiment, we include other quantitative methodologies such as the eye-tracker, specialised in measuring cognitive load through metrics such as pupillometry, number of fixations and the average fixation duration. To explain, understand and contrast the results of the user test and the eye-tracker, it was essential to use qualitative methodologies. NASA-TLX was used to analyse the cognitive load perceived by our users. The Think-aloud technique (or the gathering of user feedback) was used to understand the barriers and reasons behind the quantitative results. Finally, the Bipolar Laddering methodology, which tries to quantify the favourable and unfavourable aspects of the tools, helped us to measure the perceived value of the visualisation functionality.
El principal objetivo de las metodologías aplicadas (aparte de las centradas en la investigación de las necesidades de los usuarios) fue la de medir el impacto de la carga cognitiva en el contexto de la visualización de datos. Para ellos usamos una metodología principal cuantitativa, el test de usuarios, enfocado en medir el tiempo de completitud y resultado de las tareas. Dependiendo del tipo de experimento, incluímos otras metodologías cuantitativas como el eye-tracker, especializado en medir la carga cognitiva a través de métricas como la pupilometría, el número de fijaciones y la duración media de la fijación. Para explicar, entender y contrastar los resultados del test con usuarios y el eye-tracker, era imprescindible el uso de metodologías cualitativas como NASA-TLX para analizar la carga cognitiva percibida por nuestros usuarios, el Think-aloud technique / Comentarios libres de la tarea para ayudar a comprender las barreras y motivos detrás de los resultados cuantitativos y finalmente la metodología de Bipolar Laddering, que trata de cuantificar los aspectos favorables y desfavorables de la visualización y su magnitud o impacto en la experiencia de uso.

5. Results by objective

For each of our objectives we gathered different results based on the mentioned methodologies. Based on these results we were able to establish conclusions about the cognitive load associated with specific scenarios and how to minimise it.

5.1. Objective 1: Simplifying uncertainty in climate data

We addressed the problem of uncertainty visualisation from a wider perspective, going beyond traditional statistical techniques, linking to the analysis of advanced visualisation techniques.

5.1.1. Quantitative analysis

The assessment of success rates (indicator 1) indicated that the number of successfully completed tasks was significantly better when using the redesigned tool than when using Project Ukko, X2 (9, N = 20) = 60.6, p < 0.001 (see Fig. 7a and 7b). In the case of Project Ukko, only 15.8% of the participants successfully completed task 1 and 21.1% task 2. In contrast, with the redesigned tool, task success was much higher, reaching 97.4% and 68.4% for task 1 and task 2, respectively. Although a higher proportion of failures and abandonments occurred during task 1, especially with Project Ukko, completion of task 2 showed various cases of partial success (i.e., identified just one of the two types of glyphs presented) for both tools, reaching 31.6% with the redesigned tool.

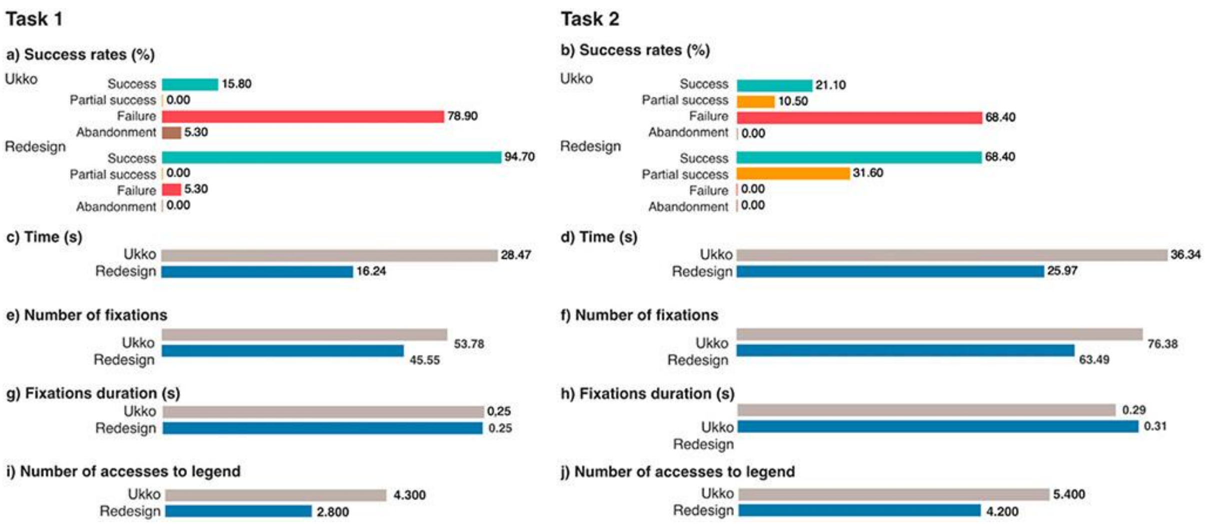


Fig. 7: (a),(b) Comparison of participants' success rates, (c),(d) average time to solve the tasks, (e),(f) number of fixations, (g),(h) fixations duration, and (i),(j) number of accesses to legend when using Project Ukko and the redesigned tool for tasks 1 and 2, respectively.

The average time to solve a task (indicator 2) was significantly lower for task 1 compared to task 2 (p = 0.001) and for the redesigned tool when compared to Project Ukko (p < 0.001) (see Fig. 7c and Fig. 7d).

The eye-tracker measurements of the number of fixations (indicator 3) was significantly lower for task 1 compared to task 2 (p = 0.024) but showed no significant differences between Project Ukko and the redesigned tool (p = 0.234) (see Fig. 7e and 7f). The fixation duration (indicator 4) did not show a significant difference between tasks (p = 0.061) or between both tools (p = 0.651) (see Fig. 7g and 7h). The number of accesses to legend (indicator 5) followed a similar trend as observed for the indicator of response time, with Project Ukko showing higher numbers than the redesigned tool (p = 0.017) and significantly lower values for task 1 when compared with task 2 (p = 0.021) (see Fig. 7i and 7j).

5.1.2. Qualitative analysis

In terms of perceived effort (indicator 6), a much bigger effort was perceived for Project Ukko (89.9% of participants) than the redesigned version (10.5% of participants). Likewise, regarding the tool preferred for decision-making (indicator 7), 89.50% of the participants stated that they would choose the redesigned tool as their working tool for daily tasks decision-making.

Table 2. Participants' positive and negative comments for Project Ukko and the redesigned tools with the number of participants that mentioned a particular aspect (freq.) and average rate of its importance/severity (avg.)

Project Ukko					
Positive	**Freq.**	**Avg.**	**Negative**	**Freq.**	**Avg.**
Easy to distinguish extreme areas	7	7.8	Difficult to distinguish opacity	9	8
A very detailed version with lots of information	3	6.6	Mixing width and brightness is too complex	9	8.22
Static legends are more traditional	3	6.6	Thin glyphs with low visibility are impossible to distinguish	7	7.57
Visually attractive	2	7.5	Using the combination of two variables (color and slope) for prediction change is too complex	4	6.25
			Slopes are confusing, they usually are used to show wind direction	3	7.33
			Mixing too many categories increases complexity	3	6.33
			Overwhelming representation	2	7
			Terrain not visible enough	1	6
			Slope value (of glyphs) is difficult to measure	1	5
			Legend is confusing	1	5
Total	15	7.33		41	7.39
Redesign					
Positive	**Freq.**	**Avg.**	**Negative**	**Freq.**	**Avg.**
Shapes and sizes are easier to identify	8	8.25	Small dark glyphs (showing average values) have a color similar to the background	5	5.6
The representation is very clear	6	9.3	The descriptive labels of the skill slider (using mathematical terms) may not be clear to all audiences.	3	6.66
Easy to distinguish between glyphs	5	8.4	Too basic to represent predicted change	1	6
Skill filtering is very useful	4	9.25	By simplifying some of the categories, we lose information	1	6
High contrast	2	9.5	Terrain not visible enough	1	6
Easy location of an area	2	9	All the three filters or descriptive labels could have been interactive (such as color category)	1	4
Simple categorization	2	8			
Total	29	8.75		12	5.83

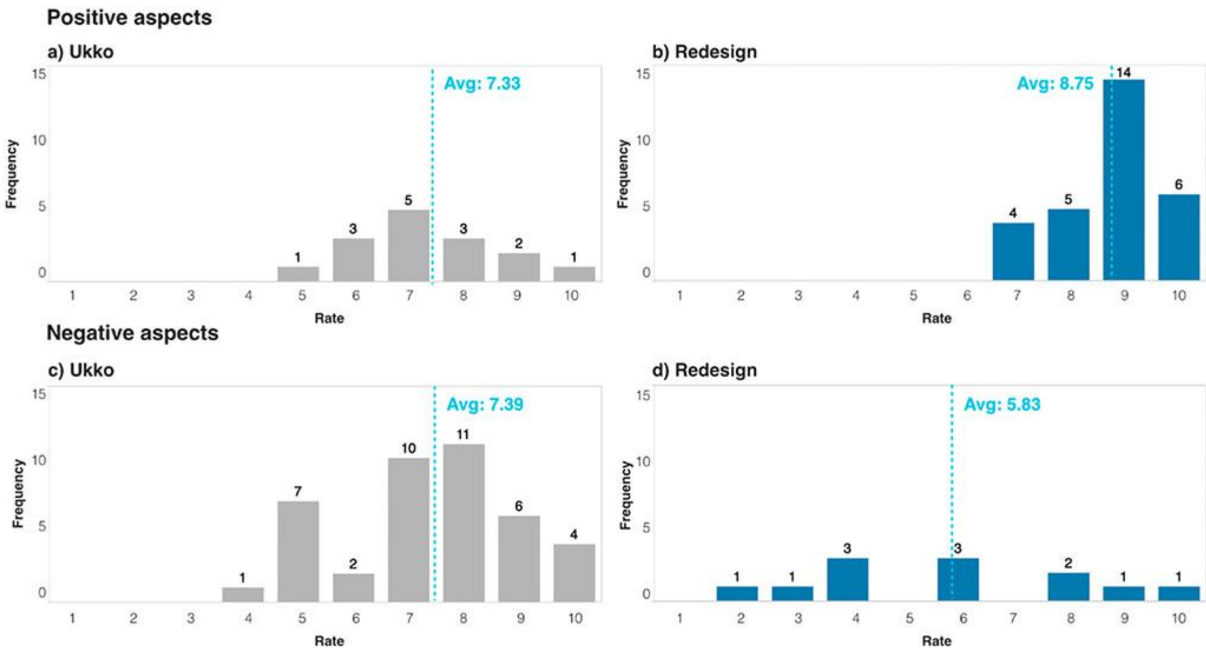


Fig. 8: Frequency histograms of participants' ratings of positive and negative aspects mentioned for (a),(c) Project Ukko and(b),(d) the redesigned tool. All the mentioned aspects (n = 97) receive a score from 1 to 10. Positive aspects are rated from 10 (very positive) to 1 (less positive) and negative aspects are rated from 10 (very serious) to 1 (slightly serious).

By analysing the results of the bipolar laddering pocket technique (indicator 8) we found that the number of positive aspects mentioned by participants was just 15 for Project Ukko against 42 mentioned for the redesigned tool (see Fig. 8 and Table 2). Regarding the negative aspects, 32 and 12 aspects were pointed out for Project Ukko and the redesigned tool, respectively. On average, positive aspects were rated with an average score of 7.3 for Project Ukko and 8.75 for the redesign. Conversely, negative aspects were rated with a greater severity for Project Ukko, with an average score of 7.4, than for the redesigned tool, with an average score of 5.8.

5.2. Objective 2: Challenges when visualising climate data

5.2.1. Qualitative Analysis

Through the qualitative methodologies carried out, we detected a series of barriers faced by climate experts when representing uncertainty: Different terminologies and lack of standards, as well as the lack of interactive mechanisms to dose the information presented.

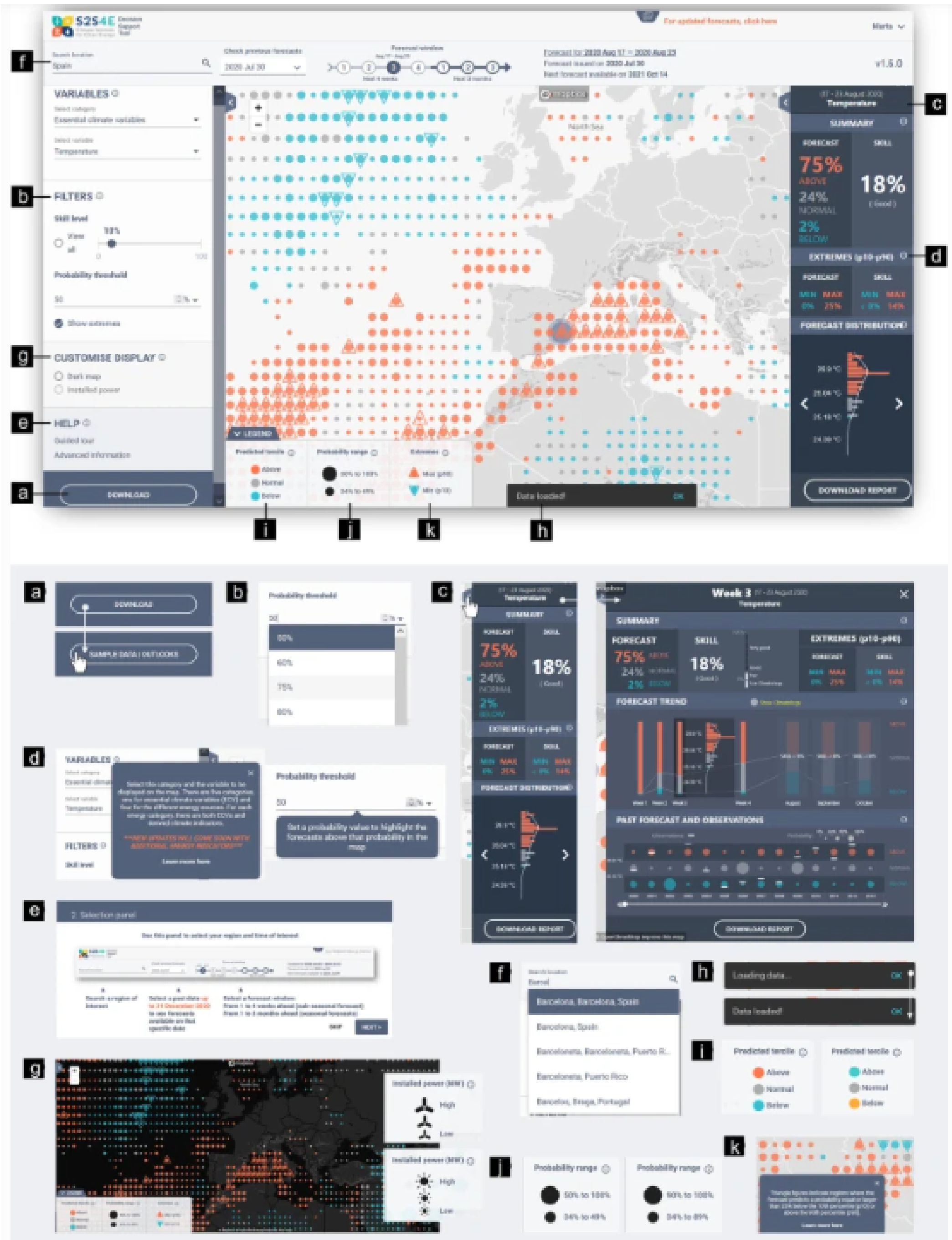


Fig. 9: Interactive decision support tool for the energy sector that considers aspects from different disciplines (source: S2S4E 2020). In the upper panel, highlighted features that are further expanded in the lower panel, exemplify particular aspects explained in the paper. **User experience aspects:** (a) button reflecting available actions; (b) possibility to filter information for skill, probabilities and extremes; basic panel that can be expanded into advanced panel; (d) tooltips and hyperlinks; (e) help documentation section; (f) search location; (g) customisation options; (h) feedback of system status. **Visualisation design aspects:** (g) use of intuitive patterns and conventions, (i) and (j) glyph map variables representation (e.g. temperature, precipitation) with circles of changing size and colour and use of contrasting colour hues in a dynamic legend. **Graphic design aspects:** (i) typography with good readability and use of colour- blind-friendly palettes. **Psychological aspects:** (k) use of triangle symbols for extremes to enhance attention, (i) simple visual encoding.

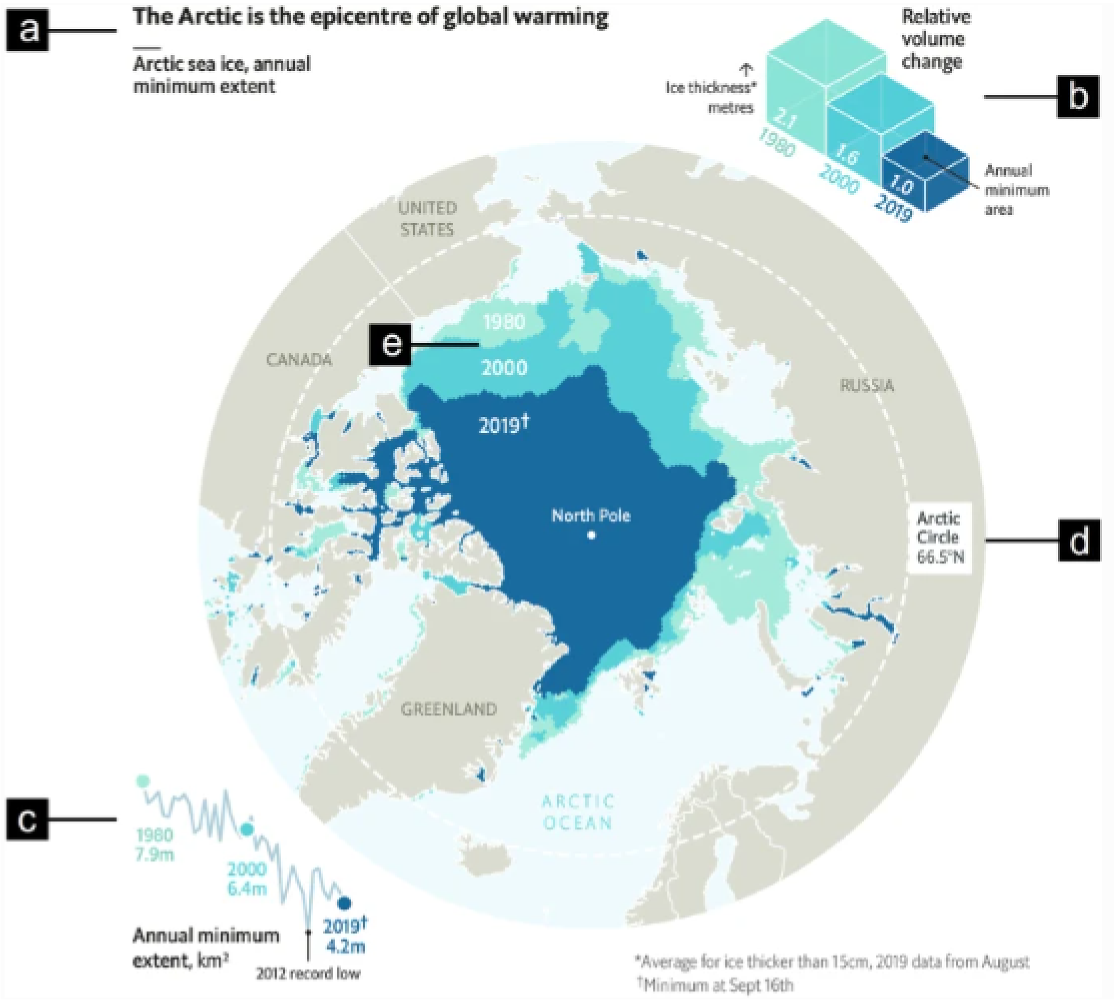


Fig. 10: Static representation of observed sea ice annual minimum extension and relative volume change that considers aspects from different disciplines (source: The Economist 2019). Highlighted features exemplify particular aspects explained in the paper. **User experience aspects:** (a, b) and (c) clear positioning of titles and legends. **Data visualisation design aspects:** (d) use of size and position to convey hierarchy, (e) visualisation embedded into narrative context, (b, c) and (e) integration of different types of representations such as infographic, time series and region map, (d) use of soft hues in the base map to emphasise colour scale values, (e) use of single-colour inverted sequential scheme to represent ice, i.e. the darker the colours, the lower the ice values. **Graphic design aspects:** (b, c) and (e) consistent use of colours across different visualisation types. Psychological aspects: (b, c) and (e) use of redundancy in the legend to enhance understanding, (e) ice cube metaphor to enhance memorability.

5.2.1.1 A plethora of approaches to represent uncertainty in climate data.

Different ways of visualising uncertainty were identified according to the visualisation purpose, the target audience, and the time scale of the information provided. This applied to both first- and second-order uncertainty (Spiegelhalter et al., 2011; Taylor et al., 2015a). Whereas first-order uncertainty refers to information on the likelihood of an event happening according to a particular forecast (i.e., probabilities or risk), second-order uncertainty refers to “uncertainty about the uncertainty” or ignorance (i.e., skill or spread), and exists because forecasts are not able to capture all the factors influencing the climate. Information on first- and second-order uncertainty can either be integrated in the same visual representation or presented separately using two different visualisations. Regarding first-order uncertainty, while some projects opted for showing the mean or median value of the modelled results without an indication of its probability of occurrence (Fig. 1.a and Fig. 1.b), other projects decided to display the information as anomalies (Fig. 1.c), that is, the variation of a variable relative to the climatological normal or long-term average. Additional options to represent first-order uncertainty used by climate services projects included showing information through a number of categories, such as *terciles* or quintiles, either indicating the most likely category (Fig. 1.d) or reporting the probability of the different categories to occur (Fig. 1.e). An alternative option consisted in providing the probability distribution function (Fig. 1.f), which gives an overview of the different amounts of change in a climate hazard and their respective likelihoods for a single point in time and a specific geographical area.

Even though showing second-order uncertainty was generally considered an exercise of transparency, some workshop participants considered that this information could be overwhelming or confusing for some users. For this reason, this information was not generally provided by climate services projects. When second-order uncertainty was not presented, some projects used scenario approaches displaying average values for different pathways, capturing the various plausible descriptions of how the system and/or its driving forces may develop in the future (Fig. 2a). In other cases, projects decided to keep particular climate information hidden from the visualisation when the level of uncertainty was too high to use this information meaningfully in decision-making (Fig. 2b). In such cases, applied practices ranged from not providing any information about uncertainty, to replacing uncertain forecasts by a reference value (e.g., climatology). Further alternatives consisted in giving users the possibility to hide or show high-uncertainty information (Fig. 2c) or allowing them to select a specific uncertainty threshold should they have an idea of the level of uncertainty they were ready to bear (Fig. 2d). Some projects also integrated second-order uncertainty through visual encoding (e.g., through transparency) (Fig. 2e) or showed it as a range in the plot, be it the full ensemble range, the standard deviation, confidence intervals, or the signal-to-noise ratio (Fig. 2f).

5.2.1.2 Use of interactive elements.

A clear trend toward developing visualisations that allow the user to interact with the different elements was observed, in line with the assumption that understanding of information can be improved through greater interaction (Yi et al., 2007). The progressive disclosure of information, which aims at the initial simplification of information followed by the possibility to reveal additional options and content, was identified as a commonly applied technique in the field of climate service visualisation, which also grants users a more active role (Bostrom et al., 2008; Spiegelhalter et al., 2011). However, in the case of particular types of services (e.g., dashboards), participants pointed out that users had explicitly indicated their need to access all resources simultaneously. For particular formats that allow low or no interactivity (e.g., factsheets, newsletters or bulletins, direct advice), it was also indicated that they could be effectively used. In the end, understanding when and how to integrate interactivity requires careful considerations of both users' requirements and tool's functionality.

5.2.1.3 Differences in terminology used by scientific and stakeholder communities.

The climate services community involved in the workshop identified more than 25 technical terms commonly used in the field of climate science that are confusing or not well understood by stakeholders outside academia. The more frequently repeated terms were "skill," "anomaly," "reliability," "uncertainty," "percentile," "ensemble," and "model" (Fig. 9). Project representatives also mentioned that the use of conventions such as "likely" or "unlikely" or the distinction between temporal forecasting scales (e.g., hindcasts, climate predictions, climate projections) can make sense in the context of climate science, but that stakeholders are not aware of such distinctions (Fagen-Ulmschneider, n.d.). Participants agreed that more resources should be put in place to overcome the terminology barrier, both among different academic disciplines and between academia and stakeholders, since the same term can be differently understood by these groups. Discussions indicated that terminology should be adapted when possible, even if it involves compromising scientific precision. Otherwise, explanations in lay language should be offered. Although this may not be straightforward and can induce some tensions during the coproduction process, overall, it will prevent wasting time discussing complex terminology concepts and will allow stakeholders to focus on the interpretation of the information. The use of glossaries and thesauruses emerged from the discussions as a good practice to try to find a common ground between the different communities. These tools can also include use case examples to illustrate the term in the context of the target users. Using elements such as tooltip hints, defining technical terms, and hyperlinks to additional explanation were also mentioned to help build a greater understanding.

5.2.1.4 Taking the vernacular language of target audiences into account.

An additional factor for misinterpretation arises when the native language of stakeholders is different from that of the producer of the climate information (WMO, 2018). Many of the projects involved in the workshop were run at the European or multicountry scale. To adhere to users' preferences, such scales require a multi-language approach that can become a challenge for the development of visualisations. Therefore, English was a dominant language used in the visualisations developed by the different projects. However, the proportion of climate services in local languages may be substantially higher when moving to the national, regional, and local scales. Despite many climate service visualisations being available in English, project representatives mentioned that, when needed, engagement activities with stakeholders were conducted in local languages to ensure understanding. For that, summary documents, including user guides and illustrative figures, were developed in stakeholders' languages. Related to the use of figures, one of the participants mentioned that the translation of text labels in visuals tends to be more time consuming than translating text explanations. Participants also mentioned that having visualisations in local languages was particularly needed for specific terminology lying in the traditional knowledge domain of local and indigenous communities (e.g., Arctic regions, Pacific Islands). Such terms, that for instance can refer to the characteristics of the local climate, often lack a translation in other languages (e.g., snow types in polar regions; see (Eira et al., 2013)). Discussions suggested that the definition of an appropriate language should be considered as part of the coproduction of a climate service. This involves tailoring information to match the language in which intended users are accustomed to working as well as the consideration of other elements that enhance usability (Miraz et al., 2016).

5.3. Objective 3: Best practices in climate data visualisation

The results of the qualitative analysis carried out to identify the challenges when visualising climate data in summary reveal the need of include different disciplines in the definition of representation standards:

5.3.1. Qualitative analysis

A number of heuristics (i.e. broad rules of thumb) and good practices from the fields of user experience, data visualisation design, graphic design and psychology are presented in the form of recommendations in Table 3. These recommendations not only refer to the visualisation itself but also apply to the entire climate service interface, which encompasses other elements of the climate service product. The recommendations try to cover a range of possible situations faced by climate service providers when developing visualisations to be used by stakeholders. However, this list must be understood in a flexible way, and climate service providers should reflect on the adequacy of the different recommendations in their contexts and take their possibilities and resources into account. Also, depending on the type of climate service provided, its purpose, the target user and the decision at hand, some recommendations will simply not be relevant or applicable. The analysis of the effectiveness of these recommendations in particular cases is available in the literature and is out of the scope of the present review. Although recommendations are divided according to the different disciplines and aspects considered in this work, the boundaries among such divisions are porous and allow for some overlapping. We obtained a list of recommendations and ideated examples in the context of climate services, both described throughout the text and displayed in Fig. 9 (interactive tool with sub-seasonal and seasonal forecasts for the energy sector) and Fig. 10 (static representation with observations of sea ice extent and volume).

Table 3: Recommendations from other disciplines and aspects that can improve the efficiency of climate services visualisations

Discipline	Aspects	Recommendation
User experience	User-centred design (UCD)	- Gather user requirements using participatory approaches - Keep information simple and digestible - Involve users from the initial analysis to the final evaluation stage - Explore multiple visualisation options with users - Assess user performance in quantitative and qualitative terms
	Interaction design	- Include interactive elements and visual features appropriate for users' skills in an intuitive way (e.g. select, zoom, sort and filtering options) - Ensure that interactions have a visible response for users - Use progressive disclosure of information when appropriate - Include hyperlinks and tooltips for contextualisation - Consider multi-device visualisation options
	Personalisation	- Allow users to customise the display reflecting individual preferences
	Error prevention and system status	- Avoid ambiguous visual encoding of key information - Use selectors rather than open fields to minimise mistakes - Use indicators of system status - Use confirmation questions
	Information architecture	- Have a clear hierarchy of the information (e.g. take into account fields' dependencies, navigation) - Adapt terminology and language to the user's context

		- Use terminology consistently (e.g. avoid using different terms for the same concept) - Reduce jargon and ambiguous terms - Add tooltips and links to glossaries and extended descriptions - Ensure that help documentation is easy to find
Data visualisation design	Visual information display	- Use chart size and position to convey hierarchy and relevance (e.g. in dashboards or multiple simultaneous displays) - Use patterns and conventions that users are accustomed to - Embed the visualisation in a narrative context
	Type of representation	- Choose the optimal and familiar representations (e.g. bars, heatmaps) according to the objective of the visualisation - Favour representations with references to the real world
	Visual encoding	- Choose visual codification carefully, since some options might be difficult to interpret - Use visual codification that allows discrimination between relevant and non-relevant values - Make data visualisation inclusive (e.g. use contrast, colour-blind friendly palettes, clear fonts) - Select a colour palette that matches the nature of the data to be represented (e.g. qualitative, sequential, divergent) - Take into account the meaning associated to colour (e.g. use blue for wet or cold conditions)
	Labels/legends	- Include in the legend all relevant information required to interpret the visualisation

		- Locate legend in a visible area and, when possible, integrate it in the visualisation - Facilitate labels' reading
Graphic design	Design aspects	- Use appropriate design to support and improve user experience (e.g. colours, typography, display of information) - Be consistent in the visual language used in the visualisation, graphic elements and styles
Psychology	Perception and cognition	- Provide no more than the necessary information, but sufficient to ease user's understanding - Reduce the complexity of the information presented - Reduce spatial distance between similar elements and between these elements and the legend or caption - Use predictability, let the user successfully foresee the result of an interaction
	Pre-attentive processes and attention	- Use pre-attentive processing elements (i.e. indicating where to look first) - Reduce the use of elements that distract attention - Use visual metaphors coherently
	Memorability	- Use visualisation elements to reinforce memorability

5.4. Objectives 4 and 5: Evaluate the efficiency of static maps & New map designs for specific needs

Through the evaluation of the different types of static maps, we collected quantitative and qualitative metrics. The qualitative metrics helped us to understand the quantitative results, with the objective of detecting which were the most significant factors that hindered the interpretation of the data.

5.4.1. Quantitative results

For the quantitative analysis we present the response times and the result for each type of map:

5.4.1.1. Detecting trends within a specific region (Task 1)

Regarding the time needed to perform the task, we found statistically significant differences depending on the map type ($f(3)=10.63$, $p < 0.001$). A Tukey post-hoc test revealed what the most significant pairwise differences were between the banded tile grid map and the tile grid map, with an average difference of ($p<0.001$) and choropleth and tile grid map, with an average difference of ($p<0.001$) (for more details see Fig. 11 in the main text and Table 2 of Annex I).

Values on success rates ($X2\ (1, N = 32) = 3.3231$, p-value = 0.7673) indicated that the results obtained were not statistically significant (see Fig. 11 in the main text and Table 2 of Annex I).

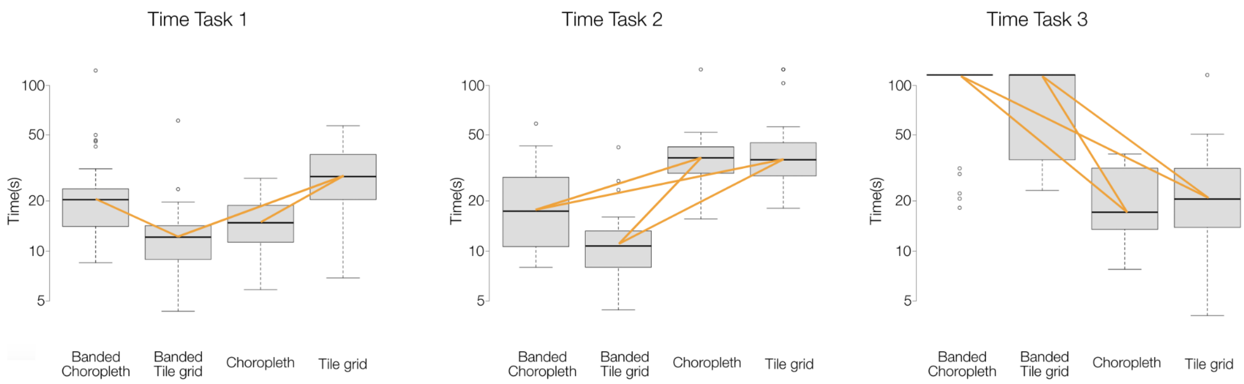


Fig. 11. Boxplots showing the distribution for the results related to completion time for all representations in the three tasks presented. Orange lines show the statistically significant differences between maps.

In this case, the H1 hypothesis was ruled out, as the banded versions were expected to work more efficiently (completion times and success rates) than the individual maps (due to the visual distance between maps). In the case of success rates, we observed no significant difference between the four representations. In terms of completion time, they showed that the banded tile grid map and choropleth performed significantly better in terms of completion times (for further details see Table 2, Table 3, and Figure 5 of Annex I).

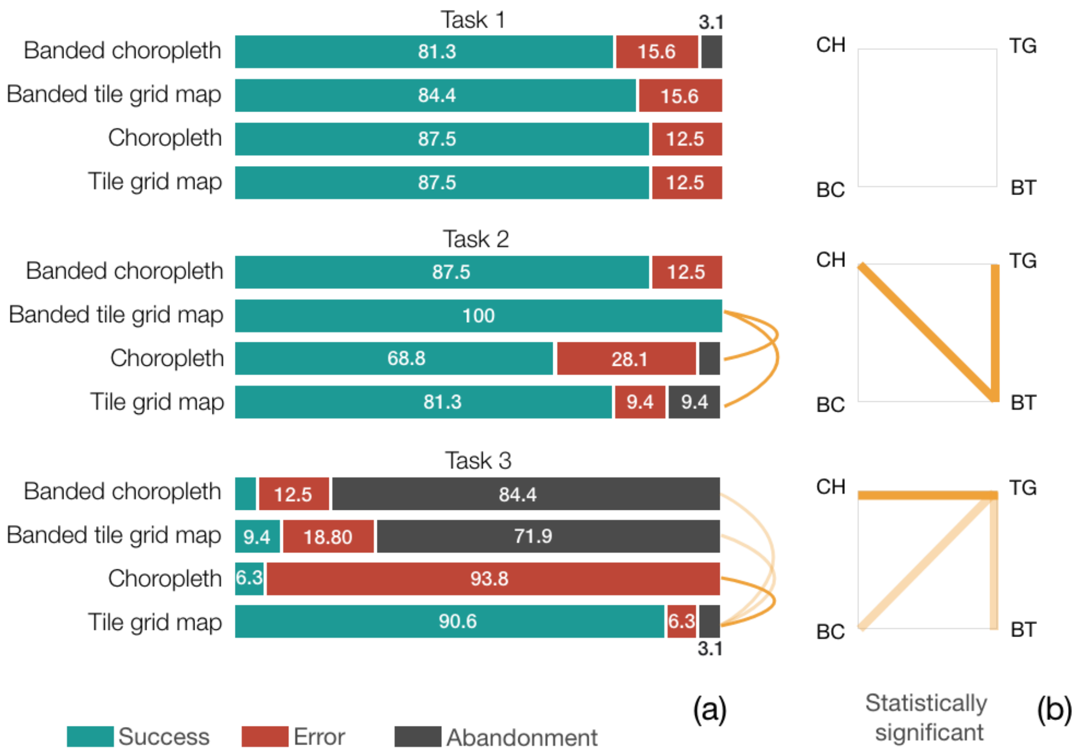


Fig. 12 (a) Success, failure and abandonment rates by task and type of map. (b) Relationships of significant statistical differences between representations. Opacity indicates a higher or lower level of significance.

5.4.1.2. Identifying the regions that meet certain conditions (Task 2)

Regarding the time needed to perform the task, we found statistically significant differences depending on the map type ($f(3)=21.6$, $p < 0.001$). A Tukey post-hoc test revealed significant pairwise differences between the banded versions and the individual versions. The same happened with the success rate (see Table 2 and Table 3 of Annex I).

In this case, the H2 hypothesis is fulfilled since the differences in terms of completion times and success rates are significantly better for the banded versions.

5.4.1.3. Determining global trend (Task 3)

Regarding the time needed to perform the task, we found statistically significant differences depending on the map type ($f(3)=73.86$, $p < 0.001$). A Tukey post-hoc test revealed significant pairwise differences between the banded versions and the individual versions (see Fig. 11 in the main text and Table 2 of the Annex I). The same happened with the success rates, which were also significantly better for the tile grid map than for the choropleth (see Fig 12 of the main text and Table. 3 of Annex I).

The H3 hypothesis is also confirmed. Applying the time correction in the case of abandonment results, we observed significant differences between the individual maps and the banded versions. Moreover, and regarding the success rates, there were also significant differences observed between the choropleth and the tile grid map, the latter obtaining better success rates. The banded versions presented the highest abandonment rates (84.4 and 71.9%).

5.4.2. Qualitative results

After carrying out each of the tasks, users cast their vote for the representation which best favoured the task (helping in decision-making, simplicity, level of comfort while performing the task) and the representation which least favoured the task. They also shared with us the problems encountered together with general feedback while performing the tasks for the different maps.

5.4.2.1. Detecting trends within a specific region (Task 1)

For Task 1, the responses were more balanced, with the best valued representations being the banded maps (banded tile grid map, followed by banded choropleth) compared to the individual representations (See Fig. 13) . The reasons given by users regarding the difficulty of some maps were related to the detection or location of the selected region and the interference (shape and colour) caused by the surrounding regions. These comments are analysed in more detail in the Discussion section.

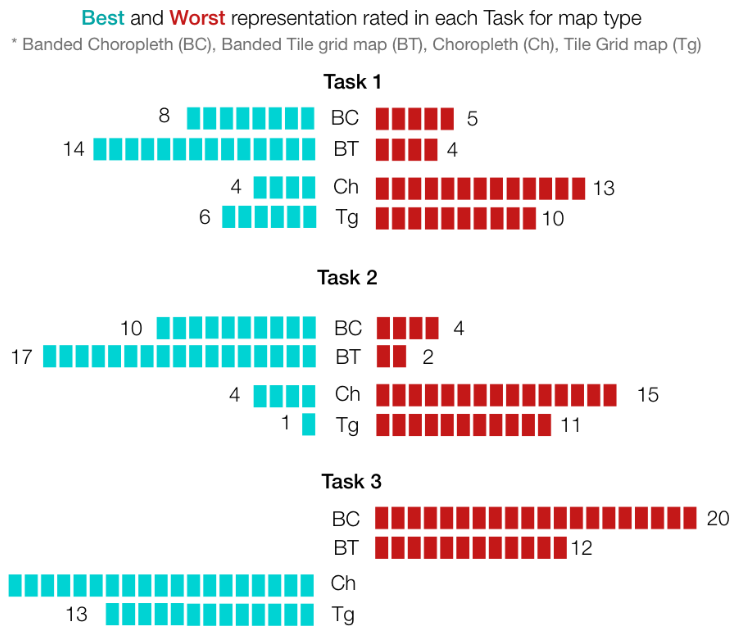


Fig. 13: Most and least favourable maps (number of positive and negative votes) to solve the task and which favour decision making.

5.4.2.2. Identifying the regions that meet certain conditions (Task 2)

In Task 2, the results were much more polarised than in Task 1, and the differences between the banded (the best valued) and the individual representations (once again, the worst valued) increased substantially (see Fig. 13). The banded representations not only received the highest evaluation but also received hardly any negative votes. Conversely, the individual representations received hardly any positive votes.

During the test, many of the users commented on the difficulty of the task, and the advantages of using the banded versions compared to the individual ones (where more visual travel is required).

5.4.2.3. Determining the global trend (Task 3)

For Task 3, the "usefulness" of individual maps was considered unanimous: users valued the banded representations very negatively as they did not favour the resolution of the task at all.

Interestingly, focusing solely on evaluations towards individual maps, users valued the choropleth map as the most favourable for carrying out the task (See Fig. 13), due to its familiarity (despite having a high error rate). For specific comments see Table 5 of Annex I.

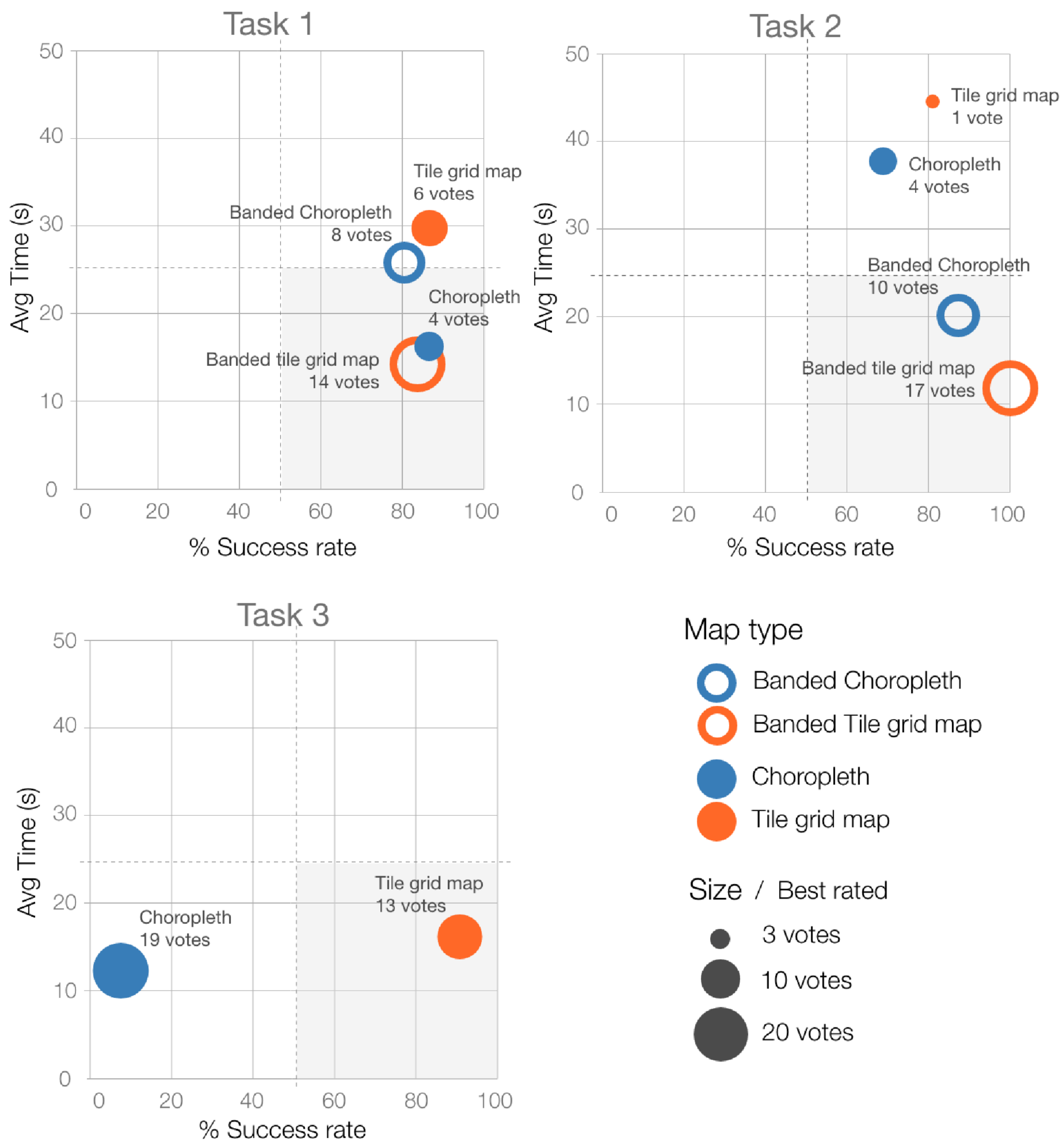


Fig. 14: In his graph we have combined four of the most significant variables, using a scatter plot representation. For the three tasks, we compare: success rate (x axis), time spent to accomplish the task (y axis), number of times a graph was voted as the most effective (size), and the type of map (colour and glyph type or shape). The larger glyphs indicate the best valued map representations. The grey area highlights the most favourable results: highest success rates and lowest completion times.

5.5. Objective 6: Highly customisable visualisations

Results while designing and evaluating highly customizable visualisation tools are required quantitative and qualitative metrics, however unlike other objectives, such as objective one (uncertainty representation) and three (static maps) we don't have any other tool to compare with, as the current tool functionalities are not available in any other tool available for public opinion analysis.

5.5.1. Quantitative results

Qualitative analysis, the same way that in other objectives was presented through time and result metrics. Results showed how potential users valued aspects such as the ability to compare all socio-demographic groups at the same time for different countries. Sections such as the summary view or grid view with sorting options were also highly valued. Among the representation views that were considered the most intuitive by users were the comparison of the situation of a particular socio-demographic group for several countries and the visualisation of inconsistencies, as they favour decision-making and save users time and effort.

Evaluation with non-state actors showed the following results for each of the tasks: Task 1 (general overview) was the task with the highest error (30%) and abandonment rates (10%), and the highest completion time (44.4 sec). For Task 2 (detailed view), we obtained 24.5 seconds for completion time and 25% error rate. Task 3 presented 28.3 seconds for completion time and 10% of error rate. Tasks 4 and 5 (analysis of inconsistencies and summary view) had very short completion times (14.7 and 10.8 seconds respectivaley) and 100% success rate (see Fig. 15).

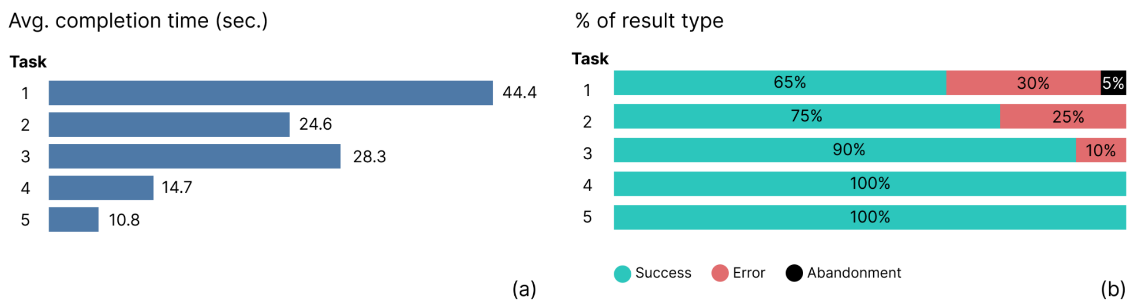


Fig 15. Completion time for the usability test (a). Percentage of result rates (b).

5.5.2. **Qualitative results**

Qualitative analysis, the same way that in other objectives was used to reinforce and understand the quantitative metrics obtained. In this specific case, evaluation was carried out with two different kinds of users, potential users and the general audience (non-state actors).

5.5.2.1. Evaluation with potential users and experts

Results gathered from the Cognitive Walkthrough methodology carried out show that potential users valued aspects such as the ability to compare all socio-demographic groups at the same time for different countries. Sections such as the summary view or small-multiples view with sorting options were also highly valued. Among the representation views which were considered most intuitive by users, were the comparison of the situation of a particular socio-demographic group for several countries and the visualisation of inconsistencies, as they favour decision-making and save users time and effort.

5.5.2.2. Evaluation with non-state actors

From the NASA-TLX results, we identified the level of visualisation difficulty linked to each of the tasks, based on the different metrics involved. The visualisation of task one, general overview visualisation (see Fig.16.a), required a high level of mental demand, temporal demand, frustration and perceived effort. For the second task the cognitive load was moderate, and low for the last three tasks since values for mental demand, temporal demand, frustration level and perceived effort were under 5 (see Fig. 16). We will analyse these results in section 5 together with the comments from users and the design decisions to improve these issues (see also Annex II, sections C and C.1).

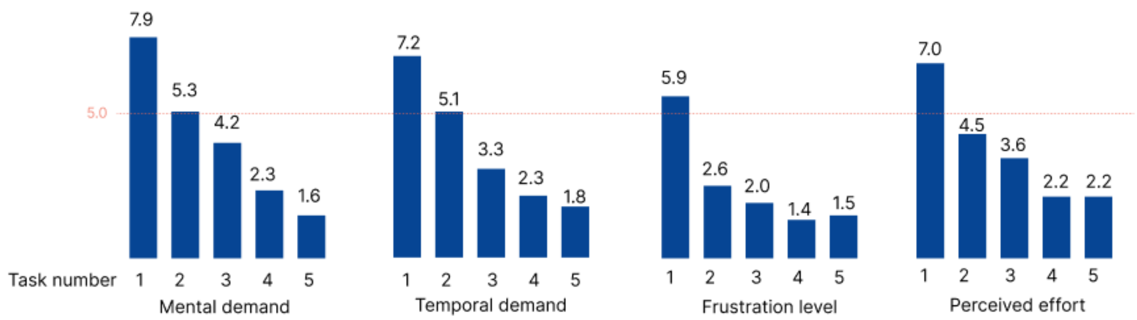


Fig. 16: Average values of questionnaire results on perceived difficulty (based on NASA-TLX categories) for the five tasks for different metrics from Mental effort to Perceived effort. In the case of the first four metrics, a higher value implies major problems during the resolution of the task. Tasks 1 and 2 present higher values in the first four metrics and lower values in the last metric which means a higher cognitive load or difficulty in performing these tasks (Cao et al., 2009a).

The positive and negative comments gathered from the Bipolar laddering technique, together with their frequency and valuation, are presented in Table D.1 in Annex D. For positive aspects, the higher values are better while, for negative aspects, the lower values indicate less severity. Positive aspects totalled 63 with a mean score of 8.4 and negative aspects totalled 29 with a lower mean score of 4.6.

Among the negative aspects, two stand out given their frequency (mentioned 7 and 6 times) and their corresponding level of severity (8.2 and 4.4 respectively). They were the difficulty of the general overview visualisation (showing the average opinions by country) (see Fig. 17) and the need to improve some aspects of the legends to make them more intuitive and understandable. In the case of the best valued positive aspects, the following stand out: The flexibility and high level of customization to analyse socio-demographic opinion and the simplicity in how the inconsistencies were presented.

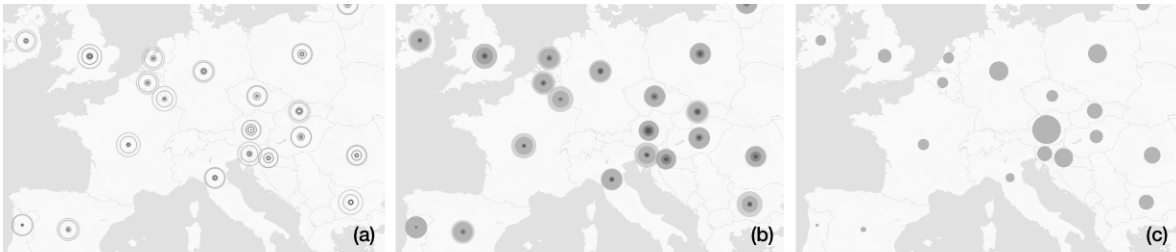


Fig. 17: Initial view of the tool through three possible representations for variable X: (a) Concentric circles proposed before the user evaluation phase that were found to be difficult to interpret. (b) Alternative proposal employing concentric areas, after the user test phase. These two options present the response percentages for all opinions simultaneously. (c) Bubble map representing just one opinion per country ('Not at all important') which was also suggested during the co-creative sessions.

Summary of Section 5: Results
In the case of objective 1, simplifying the visual encoding to favour decision making in the redesign of a climate data visualisation tool, we managed to reduce the percentage of error from 90% to 10%, as well as reduce the completion times by half. At the same time, the tool went from being a mere exploration tool to favour decision making and day-to-day tasks according to the comments and evaluations received from the users. In relation to the second and third objectives of this dissertation, and due to the results of the qualitative methodologies, a series of standards for the representation of climate data were established. These best practices and guidelines were established from co-creation sessions with different experts from various disciplines (climate, data visualisation, UX, graphic design and psychology). Objectives 4 and 5, focused on assessing the effectiveness of the use of static maps such as choropleths and cartograms (as well as their banded versions) showed that although each map facilitates a specific type of task (banded maps favour comparative tasks at a regional level and small multiples favour the detection of global trends and patterns), the percentage of users who claimed to be unsure of their answers was too high. Objective 6, which dealt with the impact of visualisation customization, has favourable results both qualitatively and quantitatively, and users highlighted its potential in the case of decision making. However, for some of the tasks the perceived cognitive load was moderate.
En el caso del objetivo 1, simplificar la codificación visual para favorecer la toma de decisiones en el rediseño de una herramienta de visualización de datos climáticos, implicaba jugar con uno de los aspectos más complejos y recurrentes en el ámbito de la predicción climática: La incertidumbre (geolocalizados punto-punto). En este caso la herramienta propuesta tuvo diferencias altamente significativas, logramos reducir el % de error del 90 al 10%, así como los tiempos de completitud a la mitad. Al mismo tiempo la herramienta pasa de ser una mera ... de exploración para favorecer la toma de decisiones y las tareas del día a día según los comentarios y valoraciones recibidos por los usuarios. En relación al segundo y tercer objetivo de la tesis, y gracias a las metodologías cualitativas se llegaron a establecer una serie de estándares para la representación de datos climáticos. Estas buenas prácticas, obtenidas a partir de sesiones de co-creación con diferentes perfiles expertos en diversas disciplinas (Clima , visualización de datos, UX, diseño gráfico y psicología. Los objetivos 4 y 5, enfocados en la valoración de la eficacia del uso de mapas estáticos como coropletas y cartogramas, así como sus versiones a bandas, demostraron que a pesar de que cada uno facilita un tipo específico de tarea (los mapas banded favorecen las tareas comparativas a nivel de región) y los pequeños múltiples en la detección de tendencias y patrones globales. No obstante, el porcentaje de usuarios que afirman no estar seguros de su respuesta es demasiado elevado. El objetivo 6, que plantea el impacto de la customización de la visualización tiene resultados favorables tanto a nivel cualitativo, como cuantitativo, y los usuarios destacan su potencial en caso de toma de decisiones, no obstante para alguna de las tareas la carga cognitiva percibida es moderada.

6. Discussion and Conclusions

We will now present the discussion and conclusions about the results for each of our objectives. Finally, we will present the common conclusions drawn from this study.

6.1. Objective 1: Simplifying uncertainty in climate data.

Considering user requirements when developing climate data visualisations is key to improve decision-making. Moreover, the simplification of a complex visualisation through changes in visual encoding and interactivity often increases efficiency. Here we used quantitative and qualitative indicators to assess whether the redesign of the Project Ukko tool, taking into account user requirements, visual encoding and interactivity, enhances communication, users' cognitive capacity, and translates into a better task performance.

When comparing the experience of participants with both Project Ukko and the redesigned tool, the quantitative indicators of success rate and response time when completing a task (indicators 1 and 2) demonstrate that the changes made to the redesigned tool increased the rate of success or partial success and allowed participants to perform the tasks faster. It is necessary to highlight that, in the case of Ukko, the success rate was extremely low for task 1. This was due to an incorrect identification of the area, which was done based on its brightness (high skill values), but that did not meet the minimum requirements of thickness (high intensity values). From a usability point of view, this is commonly defined as a false success, since users believe they have performed the task correctly, when in fact the answer was wrong (Brink, 2001).

Qualitative methods such as the bipolar laddering and two-question quiz applied in this study are based on subjective user assessment (e.g., experience, feelings, intuition, opinions), which is often perceived as less reliable than other quantitative methods (Szafir & Szafir, 2016). However, they are well-established practices that support and complement quantitative analysis and have been widely applied (Lim et al., 2019; Navarro et al., 2020). These methods provide useful insights to understand user's preferences and the positive and negative aspects that intervene in the performance and efficiency of users to perform day-to-day tasks.

In general, the changes in shape and size (i.e., the visual encoding of the information) as well as the reduction of categories were highly rated by the participants, enhancing the clarity and ease of use of the redesigned tool. In the case of Ukko, a large number of participants were not able to identify the categories associated with a glyph (skill, intensity, predicted change) nor the exact category, even when they tried to compare them to nearby glyphs. This was because thin lines, which were combined with opacity (a visual encoding that affects visibility), made colour detection more difficult. Sometimes the participants were unable to identify the exact thickness of a group of glyphs despite the many accesses to the legend, which led them to randomly select a category to avoid abandoning the task. Overall good practices in data visualisation indicate that opacity, combined with colour and reduced thickness, can make graphical interpretation worse (Dastani, 2002; Jenny & Kelso, 2007; Ware, 2012). Also, the use of the slope of the glyphs was perceived as a negative aspect by participants. This is confirmed by visual encoding good practices, indicating that changes in slope are more difficult to be interpreted, especially for nominal data (Alexandre & Tavares, 2010; Munzner, 2014). In addition, in the context of wind data visualisation, the use of slopes tends to be related to wind direction (Powers et al., 2017), as was also pointed out by some of the participants during the test. Therefore, using slope to display the wind predicted change can be counterintuitive for users.

Only two sizes and three colours were combined in the redesigned tool. This favours the detection of the areas of interest because the visual encoding does not create competition between variables (in this case, wind intensity and predicted change) (Iliinsky & Steele, 2011; Riveiro et al., 2008). Additionally, glyphs below a specific skill threshold (which would be discarded in a decision-making process) can be hidden from the display, allowing the user to focus on feasible options. This has been recognized as an effective means to reduce user memory workload and enhance task performance (Hegarty, 2011).

The most frequently mentioned negative aspect of the redesigned tool referred to the colour chosen for the middle prediction category ("Small dark glyphs, showing average values, have a similar colour to the background"). Users found the colour too similar to the background, therefore lacking sufficient contrast. However, this was decided on purpose, since target users are more interested in situations that depart from normal (i.e., upper and lower than average values), since these are the ones in which they need to take action (Kohlhammer & Zeltzer, 2004). Therefore, by choosing a colour similar to the background for non relevant values, we reduced visual noise in the representation. On the other hand, values of upper and lower predicted change (indicating wind conditions above and below normal) use a green-yellow colour hue, which stimulates more photoreceptors in the human eye and hence are easily detected by users (UNSW, 2015).

Despite colour being a crucial element of visual encoding, we did not include further changes in the colour scale used in the test to compare Project Ukko and its redesigned visualisation. This was decided in order to focus the analysis in the visual encoding and interactivity aspects and to avoid a major change between both tools that could bias the results of the test. However, in order to improve the accessibility of any climate service, colour-blindness should be taken into account in colour choices for visual representations (Light & Bartlein, 2004). In the case of Project Ukko, even using colour-blind-friendly scales, the combination of colour, opacity, and certain widths reduced the effective perception by colour-blind people. In this sense, the S2S4E Decision Support tool (*S2S4E*, 2020) already included changes in the colour palette to improve its accessibility by taking into account colour-blindness aspects (see Fig. 4).

Some changes applied in the redesign of Project Ukko were also related to the use of interactive filters to dose or personalise information. The skill filter in the redesigned version of the tool allows one to explore the uncertainty associated with the predicted change focusing just on the relevant data for the user (i.e., values below a preferred skill threshold are hidden). In the same way, the intensity slider allows the user to establish a preferred threshold, giving more visual presence (larger glyph size) to the values above this threshold. This capacity of filtering and personalising was highlighted as a strong positive aspect of the redesigned tool. One of the users even suggested adding interactivity to the colour legend, to be able to filter by the predicted change. Indeed, interactivity allows users to consume information step by step, explore particular aspects of complex datasets, and display relevant information in their own world view (Beddington, 2011; McInerny et al., 2014). This is thus a highly recommended feature for online climate services taking into account that there may be limits to how useful interactive visualisations are if the viewers do not have the required skills to interact with the presented information (diSessa, 2004).

The second task proposed to participants was more challenging than the first task regardless of the tool used. The classification of two types of glyphs proposed in task 2 took more time to complete than the identification of an area of interest in task 1. This was especially remarkable for Ukko, with a higher number of categories competing at the same time for visual attention (Alhadad, 2018; Munzner, 2014). Also, the number of times that participants needed to check the legend (indicator 5) was larger for Project Ukko. The difference between Project Ukko and the redesigned tool is probably related to Project Ukko's negative aspects linked to problems for understanding the legend, which had a more complex visual encoding. The combination of different categories in Project Ukko was also considered as a negative aspect (e.g., "overwhelming representation," "mixing too many categories increases complexity") in contrast to positive aspects of the redesigned tool linked with the simplicity of the representation (e.g., "the representation is very clear," "easy to distinguish between glyphs"). This would also explain the low success rate of task 1 for Ukko where, despite having more access to the legend, the area selected by participants did not meet the requirements of the statement in terms of skill, intensity and predicted change.

The difference in the purpose of task 1 and task 2 (identification versus classification, respectively), might also be the reason why participants needed to check the legend more often at the second task. Overall, the obtained number of accesses to the legend (combined with a simpler visualisation) suggests a reduction in the cognitive load of the participants during the completion of the tasks with the redesigned tool as they could retain the legend better and therefore reduce the number of times they had to check it.

Regarding the fixation duration, the quantitative indicator behaved almost equally between tasks and between tools, with durations ranging between 0.25 and 0.31 s. The number of fixations are the number of times a user pays attention to a certain point or area of interest on the screen. According to available bibliography, a longer fixation duration may indicate a bigger cognitive load during task performance (Andrzejewska & Skawińska, 2020; Duchowski, 2007; Klingner et al., 2008; Krejtz et al., 2018; Ooms et al., 2015) or that users have found more interesting elements to fix their attention for a longer time, without necessarily implying a greater difficulty or cognitive load (Andrzejewska & Skawińska, 2020; Klingner et al., 2008; Krejtz et al., 2018; Ooms et al., 2012).

Advanced brain monitoring tools, such as electroencephalograms and eye-tracker measures of pupillometry, can be useful to further study cognitive load (Anderson et al., 2011; Jiang et al., 2014; Keskin et al., 2020). It would be interesting to test them with other eye-tracker models to explore if this could be due to the accuracy of the model used.

The perceived effort during the task performance (indicator 6) identifies Project Ukko as being more complex to use than the redesigned tool, which was indicated as the tool preferred by 90% of the participants tested in this work (indicator 7). This contributes to the hypothesis that by eliminating or simplifying visual encodings non relevant to a target action or task and increasing interactivity, we favour decision-making.

The results and opinions of the bipolar laddering (indicator 8) clearly supported the previous indicators. A total of 29 positive comments were obtained for the redesigned tool, compared to 15 for the original Ukko tool (Table 2). This was also confirmed by the higher average score obtained for the redesign (8.75) when compared to Project Ukko (7.33). In the same way, the redesigned tool received fewer negative comments (12 compared to 41 for Ukko) and they were less serious (obtained scores of 5.83 for the redesign against 7.33 for Ukko). The most serious aspects associated with Ukko referred to the difficulty to differentiate the categorization of glyphs due to the combination of encoding through colour, intensity, and thickness, often making participants unable to identify the corresponding category. These aspects had a very high frequency, 25 comments with notable severity and a obtained score of 8.

Regarding the negative comments obtained for the redesigned tool, five participants referred to the colour similarity of some glyphs to the background of the tool (values close to the mean in terms of predictive change). However, after explaining the reasons for this design decision (reduce visual noise by attenuating non relevant points), all users found the change appropriate. Another negative aspect was related to the label terminology of the skill slider, which included mathematical terms that may not be clear to all audiences. When delving into the reasons for the negative assessment, the participant clarified that the control seems useful, but that the texts used in the labels could be clearer or more intuitive.

Paradoxically, some positive comments received for Ukko, referred to its higher density of information, greater detail in the representation of predicted change (five categories instead of the three categories in the redesigned tool) and visual appeal. However, although users who mentioned these aspects believed that these characteristics could be valuable in a context where exploration was the objective of visualisation, they did not favour clarity or decision-making.

The behavioural decision-making literature shows how people often struggle to understand particular climate terminology. There is often a mismatch between the understanding of concepts such as probabilities or uncertainty between experts and non-experts. Although visualising forecast uncertainties and associated probabilities is thought to increase users' trust, it does not automatically lead to better decisions.

Our results identify relevant aspects that can improve user experience and reduce cognitive load and that are worth considering when designing climate data visualisations. These include choosing representations and categories tailored to specific decisions, avoiding visual encoding that interferes with users' perception of the represented forms, and offering interactive elements that allow users to filter non relevant information or highlight relevant information for the decision at hand. In the redesign of Project Ukko we included all these changes at the same time. Hence, we cannot empirically establish the relative influence of each of these individual aspects in the overall reduction of the users' cognitive load. Nevertheless, we demonstrate that all these aspects can help reduce cognitive load, favour decision-making, and thus improve the overall user experience with a climate service.

In future works, it would be interesting to delve into the weight of each of the implemented actions (simplifying the number of categories, avoiding redundant visual encoding, customising the visualisations based on user needs through interactive controls) in the total reduction of the cognitive load.

Likewise, analysing the implemented changes in the context of the final tool, in combination with other improvements not assessed in the framework of this study (redesign of the navigation, colour-blind aspects, customization, and levels of detail available), can highlight additional benefits. This would delve further into the visual communication of climate information.

Our study highlights that, when combining techniques and knowledge from different disciplines [climate science, design, user-centred design (UCD), user interaction, and cognitive psychology], we are able to find better solutions for the visualisation of climate data, especially when aimed at supporting decision-making. In addition, we identify a clear need for co-design and increased empirical testing of the resulting products. We recommend information providers and tool designers in the field of climate services to collaborate more with end users throughout the whole design process to identify what is effective and to leverage the knowledge and well-established techniques from non climate related disciplines that have a lot to offer.

6.2. Objective 2: challenges in climate data visualisation domain.

Data visualisations, like charts, graphs, and maps, make it easy for many audiences to identify and understand patterns in climate data. But when not done properly, they can exclude audiences with visual or cognitive disabilities or those that lack appropriate background knowledge. This work analyses the current status of the climate service visualisation field and identifies challenges to be tackled for the development of more effective visualisations. The main challenges include the advancement of the climate services field toward a real transdisciplinary approach by effectively involving other disciplines and stakeholders in the visualisation co production process, a better co-evaluation of visualisations, a more effective representation of uncertainty in climate data, and bringing the terminology and language closer to those used by target audiences. For the development of more effective climate service visualisations, the climate science field may benefit from advances in other disciplines with a well-founded tradition, such as user experience, data visualisation, graphic design, or psychology, which are strongly based on stakeholders' needs. Only by including the expertise from other disciplines will climate service visualisations be able to build trust, prevent misuse of climate knowledge, and boost the uptake of climate information by society. This is a necessary step to move toward the codevelopment of common and agreed guidance and best practices necessary to achieve coherent and effective visualisations in climate services.

6.3. Objective 3: Best practices in climate data visualisation

The visualisation taxonomy covers a large variety of static and dynamic visualisation types (e.g. charts, graphs, maps) used across social and scientific domains. In the field of climate services, visualisations are used to facilitate readers' identification and understanding of patterns in climate data. However, when not appropriately co-produced, visualisations can exclude readers with visual or cognitive disabilities or those that lack appropriate background knowledge to correctly interpret climate information. This review paper aims to answer the need raised by climate service providers involved in a previous visualisation workshop, of having a list of recommendations to improve visualisations in climate services. For that, in this work, we review good practices commonly applied in other disciplines like user experience, data visualisation design, graphic design and psychology that can be a useful asset for the development of more effective visualisations, given the current lack of guidelines. In addition, as literature published by visualisation-related disciplines is rarely accessed by climate services providers, this list of recommendations and associated publications facilitates the access to this domain knowledge. The separate presentation of user experience, data visualisation design, graphic design and psychology disciplines in the present work obeys only to criteria aimed at facilitating the information flow for readers. However, practices from these fields are rarely applied in isolation, as reflected in the two examples included in this study. Thus, apart from being heavily grounded on the user needs, these disciplines complement each other and are all required for the development of visualisations of climate data and services. This aligns with the feedback provided by four experts from the mentioned disciplines, who highlighted that many of the aspects and recommendations provided in Table 3 could not be only mentioned under one particular discipline, and clearly advocates for the need of transdisciplinary approaches for the co-production of visualisations in climate services.

The recommendations provided in this work highlight useful considerations overlooked by the climate services community, often unaware of the good practices in fields beyond climate science. The list of recommendations is comprehensive yet generic, spanning different target groups (including experts and non-experts in climate science), needs and situations. However, not all recommendations are applicable or relevant to every case and, when relevant, it will be ultimately the task of the climate service providers to balance their possibilities and costs of implementing a recommendation against the potential benefits on the visualisation's usability, accessibility and effectiveness. Moreover, the application of some recommendations can be limited by a lack of the appropriate expertise in the research team or the need for a timely delivery of the results. Nevertheless, we argue that these recommendations can be the departure point for climate service providers willing to increase the impact of visualisations or interested in delving into more specialised literature on the topic.

The lack of guidance for the visualisation of climate information resonates with the discussions on the need for standards for climate services, especially in a time of unprecedented climate impacts affecting society. There is now more than ever an increasing demand for quality-assured climate services that are fit to support mitigation and adaptation strategies to climate change and variability. At this point, intuition is not enough guarantee that a visualisation of climate data works as intended by the climate science community; therefore, testing visualisations through application of quantitative or qualitative assessment methods is required. We show that only by including the expertise from disciplines beyond climate science will the climate services field be able to move towards the co-production of more effective and inclusive visualisations. This will contribute to build trust in science for society, facilitate the appropriate use of climate information and finally boost the uptake of climate services by decision-making and policy actors.

6.4. Objective 4: Evaluate the efficiency of static maps

In the analysis of temporal changes in geographical representations, the optimization of cognitive resources is essential when it comes to favouring comparison, the detection of changes and the identification of patterns (Schiewe, 2019c; Swoboda & Vighi, 2016). We used quantitative and qualitative indicators to assess the differences between small multiple representations and banded representations (see Fig. 3).

6.4.1. Detecting trends within a specific region

According to user feedback, Task 1 was the easiest to solve. All four maps or representations obtained very similar results in terms of success rates. However, the banded tile grid map and the choropleth showed significant differences in completion times when compared to the other two maps (see Fig. 12). Contrary to what we expected, we did not observe a marked advantage for both banded representations although they were the best valued by users. (see Fig. 13).

The comments regarding the individual representations were based on the difficulty in comparing the values in each time step for a certain region, due to the distance between maps: “With the individual maps the process is the following, you locate the region in each of the three maps and then identify (or compare) the changes in the three regions whereas in the combined (banded) maps everything is in the same place.” “It requires a little more effort because you have to compare the regions in all three maps”. “The comparison is more direct in the banded versions”. When evaluating these comments we had to pose the question: Why did choropleths obtain better times than banded choropleths despite not being the best rated or receiving the most favourable comments? This can be explained by analysing the users’ observations on the banded choropleth: "Depending on the colour of the bands in surrounding regions, sometimes it is more difficult to establish the boundaries. This does not happen, however, with the bands in the tile grid map" (Hillyard & Münte, 1984).

Similar comments were made for the simple choropleth: "The characteristic shape often helps to identify a region on the three maps, but other times, when the nearby regions share the same colour, it is no longer as simple" (Brewer, 1994; Healey, 1996; X. Liu et al., 2018).

Two users further pointed out that there is not much difference in the results due to the simplicity of the task but this could be changed by increasing the complexity (number of regions to compare) of the task: "Banded versions work better, but there is not much difference as the task statement only refers to one region". "There is not much difference as to which map works best. We are only looking at the evolution of one region. If we were asked to observe the changes of two or three regions, the banded versions would behave much more efficiently than the traditional ones (referring to choropleth and tile grid map)".

6.4.2. Identifying regions that meet certain conditions

In this task, we observed significant differences (completion time and success rates) between the banded versions and the individual versions (the former obtaining shorter completion times and higher success rates). Furthermore, the subjective evaluations were more favourable for the banded versions (see Fig. 14). "It is much easier to reach a conclusion with the banded version, all the information is in the same place, and you don't waste time comparing between maps." (Gesù & Starovoitov, 1999; Kraak et al., 2014).

The comments shared by users helped us to understand the way in which they solved this task and processed the visual information: "In the case of the individual maps (choropleth and tile grid map) the search is a bit random, you choose a specific region in t1 and check if it has not changed for t2 and t3". "You start with the most easily locatable regions (clearer positions) on the map. When you find one that meets the requirements, you pick another one at random, and so on". "Sometimes the comparison can be bidirectional: you pick a region at random at t3 and check that it has the same colour for t2 and t1".

Some users highlighted aspects such as the shape and location of the regions (repeating comments they made on Task 1): "It is easier to visually locate regions with a characteristic shape (with choropleths) or in an easy-to-locate position (with tile grid maps). Positions at the boundaries are easier to find than the interior ones, in which you have to consult the labels for each square (region) to ensure that it is the correct one."

Another user also commented on the high cognitive load associated with Task 2 and individual representations, including factors such as attention: “You are so focused on reviewing regions and jumping from map to map, that you no longer remember if you are repeating any of those regions mentioned”. “How many do I have? Have I already repeated this?” “The task is much more comfortable and straightforward to solve in the case of banded maps”.

To finish the analysis of Task 2, we want to comment on the simplicity of the pattern proposed for the search, which is “regions that remain stable over time”: “It is a simple pattern that is called for: regions that do not change over time. This implies that they are monochrome regions in the banded versions and regions of the same colour in the individual maps. However, it gets a lot more complicated for the individual representations. It requires a lot of concentration, and being aware of changes, jumping from map to map. With a more complex pattern (a particular combination of t1, t2, t3) it would be slightly more difficult for the banded versions but much more difficult for the individual ones.”

6.4.3. Determining the global trend

In this task, the banded maps showed a high abandonment rate: Users abandoned the task when they realised the answer was going to be simply guesswork: “I have no idea, I would be unable to answer the question. I have no idea how the trend changes”. “I would be unable to say it, no matter how long I spent on it”. “The only solution would be to count the sub-regions for the time steps, and I could spend the whole morning doing that”. “If I had to answer, it would be random”. “With the banded representations it is virtually impossible to determine the trend. If I had given an answer, it would have been because I was guessing”. “The graph has too much information, it is impossible to determine the trend, with all the bands and sub-regions”. These comments lead us to rule out the banded maps as a suitable option for communicating global trends.

Although users were able to solve the task using both the choropleth and the tile grid map, the success rates for these representations were completely different: the choropleth obtained just 6.3% success when compared to the tile grid map which obtained 90.6% success. Moreover, users valued the choropleth more positively (as the most helpful and effective map for solving the task) due to its familiarity, without taking into consideration the problem of area-size bias: "I have found it much easier with choropleths." “Without a doubt, the representation that best favours decision-making is the choropleth. It is a very familiar representation that is used a lot in this type of tasks”. “The easiest representation for me was the choropleth.

I'm used to them". This is what we refer to as a 'false success', which is when the user thinks that they have solved the task correctly but in fact, they are mistaken (Collins, n.d.; Experience, n.d.). Only in some cases users were able to identify the area-size bias problem related to choropleth maps with comments such as: "The Choropleth is the clearest one to show the global trend. Wait! Wait! Ah! No... If the trend is related to the number of regions, as the choropleth has different area sizes, I don't really know the answer", "I think that the trend is increasing. No, sorry. If we are referring to the number of regions instead of the area, I would say that it is stable or decreasing".

Despite the good results in the case of tile grid maps, users were not as confident in their answers when compared to using choropleths: "I think the pattern is increasing, but I am not sure".
It is noteworthy that for Task 3 the choropleth was the best valued map due to its familiarity, but this was not the case for Task 1 and Task 2. This may be explained if we suppose that when the perceived effectiveness (or the difficulty) is similar for each map, we tend to value more favourably what is more familiar to us (in this case the choropleth map), but when the perceived effectiveness between the maps is very clear, we tend to choose the representation that is proven to work more effectively. (Hansen & Wänke, 2009; Martens & Fox, 2007; Zissman & Neimark, 1990).

6.4.4. General conclusions

As a final summary and reviewing each of the studied maps, we conclude that the choropleth map, despite being the most familiar and best valued by users, does not communicate the global trend well due to the problem of area-size bias (Schiewe, 2019c). Depending on the colours of the surrounding regions, it can also present problems in identifying regional boundaries (Stewart & Kennelly, 2010).

Tile grid maps are a good alternative to choropleth maps for communicating global trends over multiple time steps (Task 3) . However, for the rest of the tasks, they present a hurdle due to the visual path between maps (or time steps). The same is also true of choropleth representations (Goldberg & Helfman, 2010).

Due to its region boundaries, the Banded choropleth map (see Fig. 14a) hinders the identification or the location of an area, as well as the differentiation of each of the bands (or time steps). It is also a poor option for showing global trends easily (Zhang & Maciejewski, 2017).

Banded tile grid maps (see Fig. 14b) are similar to banded choropleths, as they improve the boundary or contour problems, but they require legends to identify a certain region and are also a poor option for showing global trends. (Nusrat et al., 2018; Stewart & Kennelly, 2010).

We describe the evaluation of two static representations together with their banded versions for communicating changes over time. We measure completion times and success rates while the participants resolve habitual tasks common to their speciality fields. In addition, we analyse the participants' subjective opinions for each representation in terms of effectiveness. We also examine the general comments gathered in terms of the strong and weak points for each representation, as well as the problems encountered while carrying out the tasks.

The quantitative and qualitative metrics showed significant differences between the maps and their effectiveness, depending on the task being performed. The banded tile grid map stood out as the best option in tasks associated with detecting trends in one (or more) regions as well as when identifying regions that fit a certain pattern or condition (Task 1 and Task 2).

When communicating global changes over time (Task 3), we saw that both banded versions would need to be discarded because of the difficulty in interpreting them and the high abandonment rates recorded. The choropleth map, while familiar to the audience, is not a suitable option due to the high error rates (8.6 %) caused by the area-size bias problem. The only viable candidate, the tile grid map, is also open to debate. It obtained the highest success rates (90.6%) and optimal completion times, but the lack of confidence shown by the participants while resolving Task 3, leads us to advise against its use.

The fact that users expressed doubts when establishing the global trend while using tile grid maps, and preferred another, less effective representation (choropleth), leads us to conclude that no static map included in this study is sufficiently effective for the communication of global trends over time. The use of new representations such as banded maps or even tile grid maps is therefore not recommended, as users may not be familiar with them and there is a risk that the information presented will not be understood by a general audience.

However, in more specialised scenarios, the tile grid map is preferable to the traditional choropleth, because users will be well versed in interpreting these kinds of maps.

Given that the familiarity of certain representations affects the evaluation of their perceived effectiveness, and that choropleth maps are the most used in non-interactive alternatives, we believe that the alternatives (tile grid map and both banded versions) should be studied further as well as the learning curve required by the audience to guarantee their comprehension.

While it is unlikely that a single evaluation study can be conducted to analyse all possible scenarios, we believe that this study can be a useful starting point, while providing guidelines for future map variants for use in the media together with their possible risks: Banded versions facilitate the task of comparison but have their own inherent limitations. The tile grid map shows promise against traditional choropleths, but is unfamiliar to a non-expert audience. For this reason, further study is required to measure the possible effectiveness in media communication scenarios or in aiding with decision making.

6.5. Objective 5: New map designs for specific needs

The design of new visualisations involves an ideation process and a validation process. During the ideation process, it is advisable to include different types of profiles with different backgrounds (visualisation design, graphic design, user experience, researchers from different fields, psychology, etc) (Christel et al., 2018). At the same time, it was advisable to carry out this design phase in a co-creative way and in different sessions, allowing us to study and discuss the pros and cons of each of the alternatives from various perspectives.

Similarly, once we reached the final design, a validation phase was necessary to evaluate the usability of the final visualisation. In most cases the effectiveness of the proposed solution will be subject to specific objectives and tasks, so it is convenient to offer alternatives in which the user can adapt the visualisation to their needs (J. Liu et al., 2022; Synnott et al., 2012)

In the case of the use of static maps, although the proposed design (strip cartograms) is the best option for several of the frequent tasks carried out by users, user opinions lead us to believe that the use of maps does not favour the identification of patterns over time.

6.6. Objective 6: Highly customisable visualisations

We used service design to identify potential users' needs and the main questions to be answered with the data available. Through the design process, we also discover the value of how highly customization in the visualisations can affect cognitive aspects and favour decision-making processes. In this respect, the user interface design can be a strategic step, giving us a “big picture” of the challenge (Holtzblatt et al., 2004b; Stickdorn & Hormess, 2016; Wassink et al., 2009). The service design process gives methodologies to redefine how problems are tackled, identifying opportunities and barriers and offering broader solutions, always from a user-centred design approach and also letting the user play an important role in the co-creation of the tool.

The final evaluation with the non-state actors was to verify whether the requirements, functionalities and design choices made sense beyond those made by potential users and experts. Clearly, most of the proposed solutions would meet the needs of potential users, as they themselves participated in the co-creation sessions. Similarly, experts gave their point of view on problems that potential users and non-expert audiences may not have detected (Sauer et al., 2010b).

For this reason, the test with the non-state actors ensures that the final design of the tool is suitable for an audience with similar needs but without expert knowledge in data visualisation (Äijö and Mantere, 2015; Gerst et al., 2020b; Nasiritousi et al., 2016).

6.6.1. General overview

This first task to evaluate the global view, was the most difficult for both expert users and the non-state actors. When simultaneously displaying the total results by country for a given variable, we used concentric circles to represent the response percentages (see Fig. 17a). This representation favoured comparison against other alternatives (Kozak et al., 2015) and was very effective against the overlap problem. However, users had difficulty in understanding and comprehending the data. After the evaluation stage, we explore other design alternatives for the global view, as concentric areas (See Fig. 17b) this time obtaining favourable comments. In addition, a new form of simple representation with a single response category was added to the options view (see Fig. 17c). Instead of visualising all responses at the same time, the user selects the one they are interested in ('Very important', 'Strongly agree', etc.). This approach, using bubble charts, was simple, clear and intuitive and also provided a coherent transition to a more detailed view.

Following these suggested improvements, it was possible to simplify the control panel and the step-by-step selection of the different visualisation options: first, the variable to be analysed, and then which category to sort by (See Fig. 6.2 and Fig. 6.3).

6.6.2. Detailed view by country

The way these graphs are represented is flexibly adapted to the changing needs of the question to be answered through visualisation. Sometimes potential users need to know the least committed groups (in order to target these groups more effectively) or the most committed groups, being able to highlight different types of answers. In terms of interaction, the user can choose which countries to view and at what zoom level. This avoids glyph overlap problems on the map and allows the user to determine how much information they want to compare at any given time. Thus, we avoid a cognitive overload in our audience and provide a greater sense of control (see Fig. 18 and Fig. 19) .

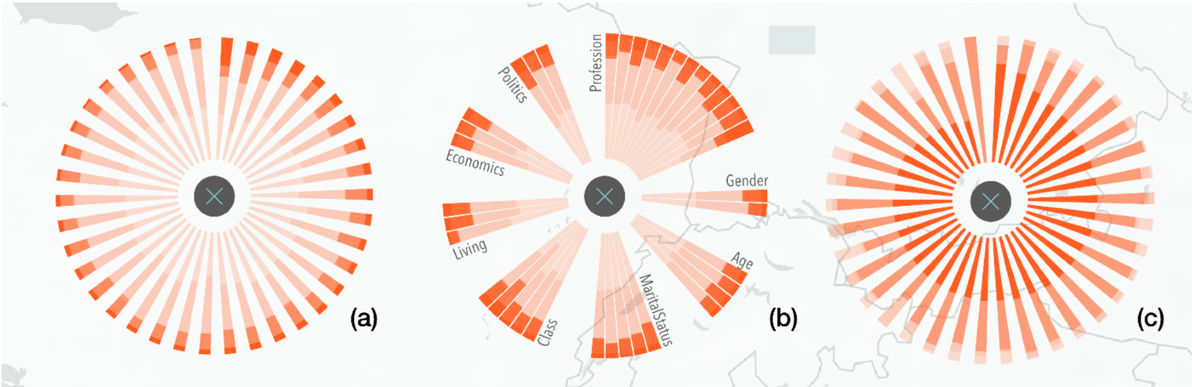


Fig. 18: Three versions of the visualisation using radial stacked bars with detailed information on the responses of socio-demographic groups to the variable 'The role you can play as an individual'. (a) Sorted by the most negative answer 'Totally disagree', (b) sorted by group categories, and (c) sorted by the most positive answer 'Strongly agree'.

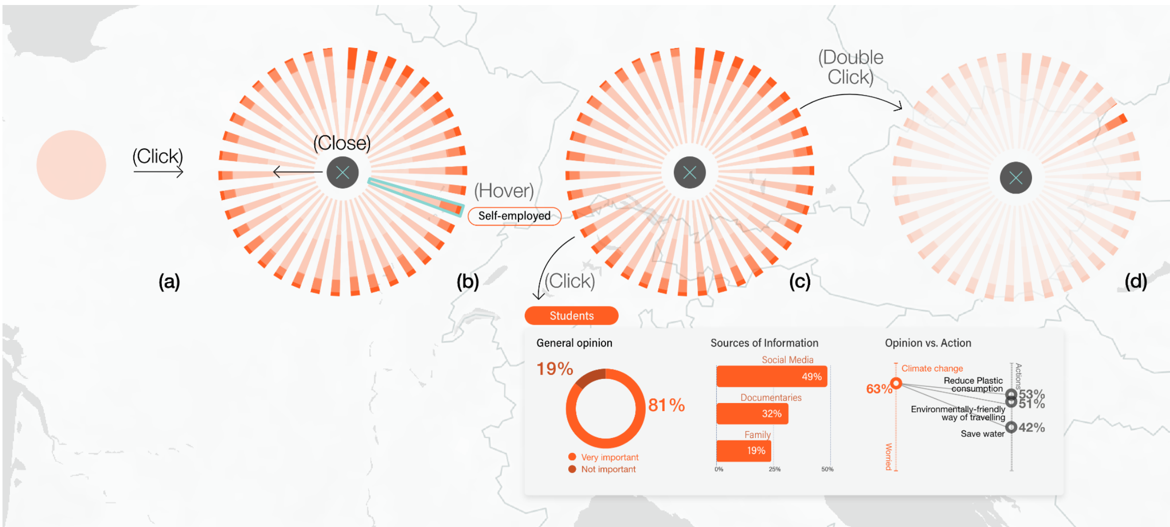


Fig. 19: Accessing different levels of detail through interaction: Clicking on the global view (a), we access the socio-demographic details (b). Hovering over each group we view the name of each group (b), and when clicking on one of the groups we access detailed information about rates, sources of information and attitudes also called inconsistency view (c). With a double click on one of the groups, we enter the group comparison view (d).

Both experts and non-state actors commented on the high density of information provided and the possibilities for analysis that it offered. However, the latter stated that it required an adaptation period due to the quantity of information shown. One aspect of note in this view was most of the participants had problems in understanding why it was necessary to highlight less favourable opinions. Experts in particular complained about the use of more intense colours to represent lower values, given that best practice visualisation rules dictate the opposite. This design decision, which was obvious to potential users as they had established this requirement, was based on the need to highlight the least committed groups and countries. These are the ones to be targeted by environmental awareness policies and campaigns. However, the tool also makes it possible to highlight the most environmentally committed groups. In this case, potential users had stated that this functionality was also necessary to be able to investigate the reasons why certain policies and social criteria are successful. Once the reasons for this duality (highlighting the lack or existence of commitment) were explained, the participants approved the design decision.

In terms of visualisation, users highlighted the large amount of information presented, but in a gradual and controlled manner: "As you explore, you can hide and unhide the opinion in detail by groups, for the countries that interest you. There is a lot of information but at all times, you can control what you want or do not want to see". However, other participants needed a brief time for analysis because they were not familiar with the representation: "The breakdown by socio-demographic group/country is surprising the first time you access the information. There is a lot of information. However, once you understand the distribution of the circular plots, it is very easy to understand. I can not think of a better map representation, as it optimises the space very well through the proposed interaction". This leads us to conclude that the introduction of the onboarding process helps to improve the user experience for novice users.

6.6.3. Summary views

The summary views were another of the most valued visualisations by the management profiles, as they allow for the analysis of the most salient information (See Fig. 20.c). They are also useful as they can be incorporated into printed reports. One user declared that "It shows, in a very direct way, the most relevant things to take into account, less committed groups from different countries. It does all the work for you and also allows you to select what you want to highlight".

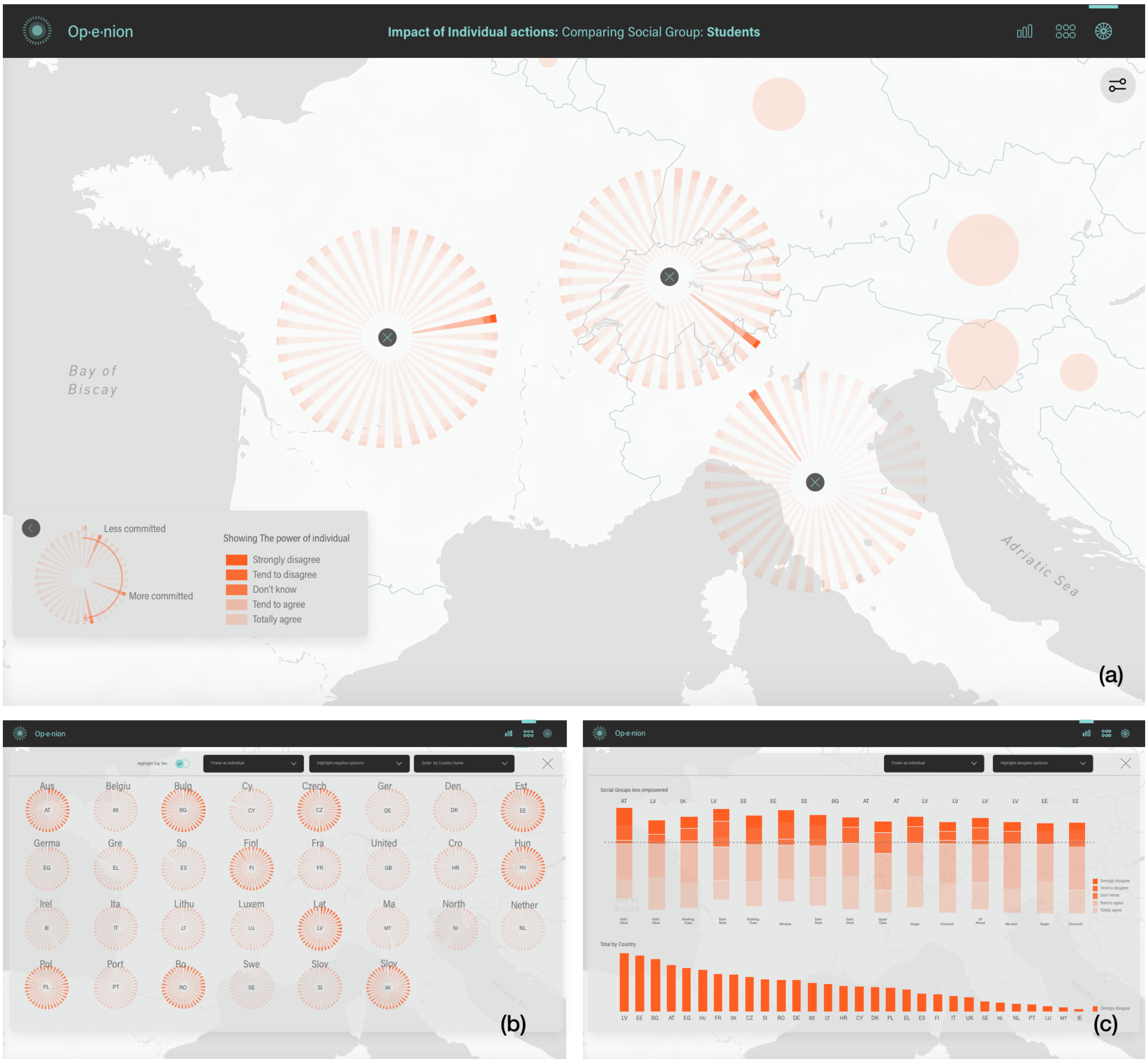


Fig. 20: (a) Level of commitment comparison between countries, highlighting a certain socio-demographic group ('Students'). Using a 'clock face' interpretation, commitment levels are reflected in the 'time' shown. Higher levels of commitment are represented as later times. (b) Small-multiples view highlighting the top ten countries with higher values for the category 'Totally disagree'. (c) Summary View. Top: top groups less committed to the environment of the European Union. Bottom: The 30 countries less committed to the environment (sorted by negative opinions).

6.6.4. The use of maps in DST

The visualisation of information on maps facilitates the analysis of complex data in a graphical and easy to understand way, providing a more realistic vision of the conditions that must be evaluated and are used in various fields such as public administration, and in the administration of transport, infrastructure , health, or social services (Mazăreanu, 2013; Relia et al., 2018, p.). From a data visualisation perspective, experts questioned the use of maps (following the approach used in the summary view). However, potential users preferred the presence of maps, especially in a municipal context of use providing more well known geographic metaphors and offering more freedom to explore and compare nearby or adjacent areas. Maps, even presenting many challenges at a spatial level for representing multi-categorical data, are very familiar for the audience and solve the geographic component in a natural way (Fuchs & Schumann, 2004).

6.6.5. Flexibility and customization

In addition to the validations carried out by expert users and the non-state actors, tests with potential users showed that the tool's high level of customization allowed them to focus on different aspects while prioritising the relevant information at the same time. One of these potential users declared: "The tool allowed me to highlight what I needed at all times. When I visualise opinions, for example about concern for the environment, I can highlight the most positive or the most negative responses, and the visualisation totally changes its appearance. This personalization helps me understand what I need at all times."

The tool allows for information comprehension through the use of different sorting options based on user needs, highlighting certain values over others, simplifying the way users analyse and compare socio-demographic group inconsistencies and attitudes, and favouring understanding and decision making. Making the necessary information available and in the appropriate format for the development of the task, reduces the sensation of frustration and perceived effort, increasing the level of confidence.

Often, public opinion datasets contain information on different socio-demographic groups and multiple variables. This high information density requires advanced visualisation tools to understand patterns and trends. However, the visualisation tools available to organisations and institutions are very complex or not very flexible, so it is necessary to resort to ad-hoc tools.

The valuations obtained on the proposed solution through service design, indicate the usefulness of the methodology employed to co-create visual tools adapted to specific needs. The validation phase allows for the identification of problems that can be solved in the early stages of design. The double review process allows for the evaluation of the tool from various points of view: First, from a functionality perspective, potential users and experts identified problems at the levels of visualisation and user experience. Second, we reproduced common tasks to be evaluated by non-expert users, to guarantee that the proposed solution could be user-friendly for administrative staff.

The tool allows for analysing citizen opinion not only at a general level, but also in relation to the opinions of different socio-demographic groups. It could be key for understanding social attitudes as to allow public institutions to as well as the problems and barriers faced by each group in order to deal with the lack of action and promote more suitable and personalised campaigns.

Therefore, the design of visualisations that favour the understanding of complex, multi-categorical and data is still a challenge for different fields, especially for non-scientific audiences without solid knowledge of data visualisation.

In scenarios where there are multiple user profiles with very diverse needs, employing a user-centred design (focusing solely on the most common functionalities or requirements of the majority) would not be sufficient. The customization of the tool must take on a new dimension and not limit itself to a basic level of analysis, but offer high flexibility and allow us to meet specific needs.

In summary, the main highlights of the knowledge acquired in the design process of this tool are the following:

1. Understanding the concerns, actions taken, and information channels used for different socio-demographic groups, can be a first step to delving into the true reasons why citizens do not actively participate in certain campaigns. A precarious economic situation or vested interests, can help us understand different behaviours and tackle social barriers from a new perspective.
2. As a part of user-centred design, validation with experts and the non-state actors can help us rethink the point of view of potential users, which are often overly focused on their own particular needs. This is a valuable step in the methodology as the detection of problems when using a tool can lead us to redefine and improve the initial approach and at the same time improving the sustainability of the final product development (Trischler et al., 2018).
3. Visualising multi-categorical data must be easy and flexible, adapting to the changing or specific needs of potential users (administrations and public institutions). In these cases, generalising functionalities suitable for the majority of users is not enough, and it is necessary to combine usability with a high level of customization of the available views: A user will be able to see the information in a way that facilitates the task they want perform at a given time, and according to specific needs.
4. Different visualisations can target different profiles and objectives inside the administrative staff and are closely related to the search patterns of users: Exploration and analytical tasks (used by analysts,

data experts, sociologists, etc.) versus tasks where reaching conclusions are needed, such as summaries and rankings, frequently needed by managers.

5. Providing powerful functionality can lead to an additional level of demand from users: Advanced features will allow them to make better use of the tool and identify new specific needs which, in turn, leads to more new features. This results from an iterative process of constant improvement.

Due to the complexity and specific nature of policy and management, it is advisable to develop flexible tools to meet user expectations or fully cover their needs.

Facilitating the analysis of public opinion and obtaining knowledge that favour decision-making will help administrative bodies to develop more efficient and user-centric campaigns.

6.7. Final conclusions and objectives reached

🟢 Achieved 🟠 Partially achieved 🔴 Not achieved

Table 4: Objectives review and accomplishment

Objective	**Achieved**
Reduce the visual density in the representation of climate data (uncertainty) to its minimum expression: only using the information necessary for decision making has a direct impact on the user's cognitive load and performance: reducing error (from 90 to 10%) and reducing task completion time (50%)	🟢
The lack of standards when representing climate data of different types is a problem for the scientific community and a hindrance in the communication and interpretation of results.	🟢
The representation of climate data should follow a set of guidelines that should include not only disciplines such as user experience or data visualisation, but also psychological aspects such as perception and cognition.	🟢
The interpretation of trends at the regional level is improved with the use of banded maps. Likewise the identification of regions that meet a certain pattern. However, the global trend over time is by no means effective with static maps despite their widespread use in media and reports since the cognitive load (given the presence of too many stimuli) is too high in these cases.	🟢
Despite basing the design on the existing (most used) representations and their limitations and strengths, it is not always possible to find a design that favours the interpretation of patterns at the regional-global level.	🟠
When visualising multi-categorical data for changing needs, the visualisation has to be dosed and customised. It should offer visualisations that present the information in a diverse way adapting to the user's needs and skills. The visualisation must answer specific questions that, according to their level of depth, must be answered progressively to favour both the search for information and the reduction of the cognitive load, and at the same time, to favour decision making. Data visualisation in this case takes on the dimension of information architecture, and not only that of a tool for visual analysis of data.	🟢

Summary of section 6: Discussion & Conclusions

In our first objective (to reduce the cognitive load in the climate representation of uncertainty), the effects of simplifying the visual encoding, in terms of results, time and user satisfaction, were clear. When redesigning or simplifying a visualisation, we must take into account two aspects, the purpose of the visualisation (exploratory, analytical, decision making, etc.) and the reduction or subtraction of visual information not relevant to the task.

In the establishment of climate-related guidelines, the collaboration of climate experts, together with user experience, graphic design, and visualisation professionals, was key to reaching our objectives. It was also crucial to take into account psychological aspects such as perception, cognition and the cognitive load of our users in a given context of use. Another important aspect that arose was the necessity (of climate centres and institutions) to disseminate findings and best practices, with the aim of promoting their use and encouraging constant improvement.

In the case of static maps, when it comes to communicating global patterns over time, none of the representations under study are recommendable as users completed the tasks without the certainty that their answers were correct, which affected decision-making. While some maps excel over others in specific tasks, our recommendation is to find an alternative representation to the static map or to use it only in cases where the temporal pattern follows a clear spatial direction. It is worth noting the weight given by users to the familiarity of some charts and how it affects the perception of their performance in completing the tasks. A high percentage of users felt that the choropleth map was more effective in tasks where times and success rates were well below other types of maps such as cartograms and banded cartograms.

Finally, the effects of customization together with intuitive design of the interaction are key when it comes to dosing information in visual information architectures. The user has just enough information to draw conclusions or make decisions, reaching the necessary level of detail through the visualisation. The general approach of designing tools for the majority or guaranteeing a minimum of satisfaction for a percentage of users does not always fit. Often, the needs of a single user profile may change depending on the type of task or context. These specific situations may involve a radically different way of displaying visual information.

Nuestro primer objetivo, reducir la carga cognitiva en la representación de la incertidumbre climática, quedaron claros los efectos de la simplificación de la codificación de la visualización tanto en resultados, tiempos y satisfacción de los usuarios. A la hora de rediseñar o simplificar una visualización, debemos tener en cuenta dos aspectos, el objetivo de la visualización (exploratorio, analítico, toma de decisiones, etc.) y reducción o sustracción de la información visual no relevante para la tarea.

En el establecimiento de estándares relacionados con clima queda clara la importancia de establecer una serie de buenas prácticas que sean discutidas y establecidas como fruto de la colaboración no solo de expertos en clima, expertos en experiencia de uso y diseño gráfico, visualización, si no a su vez tener en cuenta aspectos psicológicos como percepción, cognición y carga cognitiva de nuestros usuarios en un determinado contexto de uso. Otro aspecto clave será la colaboración entre centros e instituciones para difundir los hallazgos y buenas prácticas, con el objetivo de fomentar su uso y favorecer una constante mejora en los mismos.

En el caso de los mapas estáticos, a la hora de comunicar patrones globales a lo largo del tiempo ninguna de las representaciones bajo estudio resultan recomendables dado que los usuarios completan las tareas sin la seguridad de que sus respuestas sean correctas, cosa que afecta a la toma de decisiones. Si bien algunos mapas sobresalen por encima de otros en tareas específicas, nuestra recomendación es encontrar un representación alternativa al mapa estático o bien utilizarlo únicamente en casos en los que el patrón temporal sigue una clara dirección espacial. Cabe destacar el peso que los usuarios otorgan a la familiaridad del gráfico y cómo afecta a la percepción de su desempeño en la realización de las tareas. Un alto porcentaje de usuarios consideró que el mapa *coropletas* era más eficaz en tareas en las que los tiempos y los porcentajes de éxito estaban muy por debajo de otro tipo de mapas como cartogramas y cartogramas a bandas.

Finalmente los efectos de la customización junto con un diseño de interacción intuitivo son claves a la hora de dosificar la información en arquitecturas de información visuales. El usuario dispone de la información justa y necesaria para extraer conclusiones o tomar decisiones, llegando al nivel de detalle necesario a su vez a través de la visualización. La línea general de diseñar herramientas para la mayoría o garantizando un mínimo de satisfacción para un porcentaje de usuarios no siempre tiene cabida. Muchas veces las necesidades de un único perfil de usuario pueden ser cambiantes dependiendo del tipo de tarea o contexto, y en situaciones concretas puede implicar una manera radicalmente distinta de mostrar la información visual.

7.Limitations and Next Steps

We will now review the limitations, aspects to be taken into account and next steps related to some of the proposed objectives, such as when analysing the difficulty interpreting trends with static maps and the methodology used, the effect of customization on decision making processes and the aspects to take into account when using the eye-tracker technology to measure cognitive load.

7.1. When evaluating the efficiency and difficulty in interpreting global and regional patterns of socio-economic data using different static maps.

We focused this cartographic evaluation on the representation most used in the media to communicate social data over time: the choropleth. We also included in this analysis its simplest conceptual version (the tile grid map) and the banded versions of both. There are other tile grid map variants that we did not consider (heat tile grid maps, bar tile grid maps, full-width tile grid maps, etc.) (Datavizcatalogue, 2018).

Similarly, when trying to cover the spectrum of possible tasks and possible statements, we limited ourselves to a particular subset of common tasks and simple statements for task setup.

There are also numerous limitations when choosing the possible geographic areas, when representing data through cartograms, such as size, distance between regions and maps, level of detail (continent, country, region ...), and the number of time steps.

Another limitation, which we have already mentioned, is that the scenarios of the proposed tasks were not based on real data. This was due to the fact that we had to present multiple different values in order to avoid learning bias between the different maps for the same task.

Regarding the number of time steps selected, we assumed that the results obtained would be negatively impacted by increasing the time steps. That is, the difficulty for n + 1 would be equal to, or greater than for n time steps due to the increase in visual stimuli (Rensink, 2021). Additionally, in the case of banded representations, the number of time steps must be limited so as to permit sufficient width for the bands and guarantee the correct colour interpretation or the values they represent (especially for colour-blind users) (Buckley, 2013; Lupa et al., 2017) (see examples of reports with different number of time steps in Table 8 of Annex I).

With regards to the results obtained, we believe that it would be interesting to perform a new study employing real, large datasets, which would allow us to play with different snapshots (of time and values) and, at the same time, prevent learning bias in the user testing. The use of multiple geographies in the same test would also be advisable.

It would be recommendable to explore other possible tasks, in addition to those included in this study, or measuring the results for more complex scenarios is also recommendable. For example, adding the analysis for more than one region to Task 1, or adding more than one type of condition or pattern to Task 2.

In the introduction, as well as in the section dealing with time steps and their limitations, we make reference to dynamic or interactive mechanisms to make up for the limitations of static maps. Although these alternatives do have advantages and are often used for representations of multiple time steps, they can present the same problems as static maps from a cognitive load perspective. For this reason, their effectiveness should be questioned when seeking to communicate trends and changes over time or when aiding decision making processes (see Fig 1, Fig.2 and Fig.3 of Annex I) .

Another aspect of interest to be analysed, is the role of familiarity in the perception of effectiveness together with the influence of similarity in the maps presented, and also the subjective difficulty of the tasks.

7.2. When analysing the effects of the use of highly customisable visualisations in reducing the cognitive load in decision-making processes.

The tool presented is a prototype focused on solving the needs of a specific group, however, scalability aspects were taken into account, considering the possibility of including new variables to expand the information to be displayed in the future (see Annex II, Table A.1).

With the implementation of the final tool, it is also recommended to include metrics systems in order to identify the most used views or the time to complete certain tasks, in order to find new functionalities or improve existing functionalities.

As we have already mentioned, potential users were not only useful in gathering needs, but also in defining and identifying relevant aspects in the data. The data from which this prototype was developed were the Eurobarometer data, which in this case, contemplates total information by socio-demographic groups (see Annex II, section A). But not what percentage of each group belongs to the remaining groups, i.e. we know the percentage of engaged students, but not what percentage of them belong to the remaining categories, i.e. what percentage live in the city, in the village or in the countryside. When collecting the new data for urban settings, we intend to keep this information, so we must guarantee its anonymity, even at the minimum granularity we intend to reach (e.g., district, neighbourhood or sub-neighborhood).

7.3. How to interpret Eye-tracker metrics when analysing cognitive load

One of the most representative metrics measured with eye-tracker for the first scenario (Climate) was fixation duration and the number of fixations. The number of fixations are the number of times a user pays attention to a certain point or area of interest on the screen. According to available bibliography, a longer fixation duration may indicate a bigger cognitive load during task performance (Duchowski 2007; Ooms et al. 2014; Andrzejewska and Skawińska 2020; Klingner et al. 2008; Krejtz et al. 2018) or that users have found more interesting elements to fix their attention for a longer time, without necessarily implying a greater difficulty or cognitive load (Henderson and Ferreira 2004; Klingner et al. 2008; Ooms et al. 2012; Krejtz et al. 2018; Andrzejewska and Skawińska 2020) (see also Figure 4 of Annex I).

8.References

Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). Visualization Aspects. In W. Aigner, S. Miksch, H. Schumann, & C. Tominski (Eds.), *Visualization of Time-Oriented Data* (pp. 69–103). Springer. https://doi.org/10.1007/978-0-85729-079-3_4

Äijö, R., & Mantere, J. (2015). *Are non-expert usability evaluations valuable?* Helsinki University of Technology.

Alexandre, D., & Tavares, J. (2010). Introduction of Human Perception in Visualization. *International Journal of Imaging*, *4*.

Alhadad, S. S. J. (2018). Visualizing Data to Support Judgement, Inference, and Decision Making in Learning Analytics: Insights from Cognitive Psychology and Visualization Science. *Journal of Learning Analytics*, *5*(2), 60-85-60–85. https://doi.org/10.18608/jla.2018.52.5

Allendoerfer, K., Aluker, S., Panjwani, G., Proctor, J., Sturtz, D., Vukovic, M., & Chen, C. (2005). *Adapting the cognitive walkthrough method to assess the usability of a knowledge domain visualization*. https://doi.org/10.1109/INFVIS.2005.1532147

Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., & Silva, C. T. (2011). A User Study of Visualization Effectiveness Using EEG and Cognitive Load. *Computer Graphics Forum*, *30*(3), 791–800. https://doi.org/10.1111/j.1467-8659.2011.01928.x

Andrzejewska, M., & Skawińska, A. (2020). Examining Students' Intrinsic Cognitive Load During Program Comprehension – An Eye Tracking Approach. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (pp. 25–30). Springer International Publishing. https://doi.org/10.1007/978-3-030-52240-7_5

Antifakos, S., Schwaninger, A., & Schiele, B. (2004). Evaluating the Effects of Displaying Uncertainty in Context-Aware Applications. In N. Davies, E. D. Mynatt, & I. Siio (Eds.), *UbiComp 2004: Ubiquitous Computing* (pp. 54–69). Springer. https://doi.org/10.1007/978-3-540-30119-6_4

Araya, V. P., Pietriga, E., & Bezerianos, A. (2019). A Comparison of Visualizations for Identifying Correlation over Space and Time. *IEEE Transactions on Visualization and Computer Graphics*, *26*, 375–385. https://doi.org/10.1007/s00779-018-1120-y

Argyle, E. M., Gourley, J. J., Flamig, Z. L., Hansen, T., & Manross, K. (2017). Toward a User-Centered Design of a Weather Forecasting Decision-Support Tool. *Bulletin of the American Meteorological Society*, *98*(2), 373–382. https://doi.org/10.1175/BAMS-D-16-0031.1

Armstrong, M. P., & Xiao, N. (2018). Retrospective Deconstruction of Statistical Maps: A Choropleth Case Study. *Annals of the American Association of Geographers*, *108*(1), 179–203. https://doi.org/10.1080/24694452.2017.1356698

Atkins, R., & Mcneal, K. (2018). Exploring Differences Among Student Populations During Climate Graph Reading Tasks: An Eye Tracking Study. *Journal of Astronomy & Earth Sciences Education (JAESE)*, *5*, 85–114. https://doi.org/10.19030/jaese.v5i2.10219

Babich, N. (2020). Usability Testing: Moderated vs Unmoderated. In *Medium*. https://medium.com/thinking-design/usability-testing-moderated-vs-unmoderated-adbccc37404b

Beddington, J. (2011). *Blackett review of high impact low probability events* [PhD Thesis].

Bertin, J. (2010). *Semiology of Graphics—Diagrams, Networks, Maps*. 456.

Besançon, L., Cooper, M., Ynnerman, A., & Vernier, F. (2020). An Evaluation of Visualization Methods for Population Statistics Based on Choropleth Maps. *ArXiv:2005.00324 [Cs]*. http://arxiv.org/abs/2005.00324

Bevington, K., Ansari, S., Bilotta, R., Courtright, A. B., & Lang, A. (2019). Improving Climate Data Visualization and Information Dissemination on the U.S. Drought Portal through User Experience (UX) Design. *AGU Fall Meeting Abstracts*, *33*. http://adsabs.harvard.edu/abs/2019AGUFMPA33C1103B

Blevis, E. (2007). Sustainable interaction design: Invention & disposal, renewal & reuse. *Proceedings of the*

SIGCHI Conference on Human Factors in Computing Systems - CHI '07, 503. https://doi.org/10.1145/1240624.1240705

Block, G. (2013). *Reducing cognitive load using adaptive uncertainty visualization* [PhD]. Nova Southeastern University.

Board, C., & Taylor, R. M. (1977). Perception and Maps: Human Factors in Map Design and Interpretation. *Transactions of the Institute of British Geographers*, *2*(1), 19–36. JSTOR. https://doi.org/10.2307/622191

Bonneau, G.-P., Hege, H.-C., Johnson, C. R., Oliveira, M. M., Potter, K., Rheingans, P., & Schultz, T. (2014). Overview and State-of-the-Art of Uncertainty Visualization. In C. D. Hansen, M. Chen, C. R. Johnson, A. E. Kaufman, & H. Hagen (Eds.), *Scientific Visualization* (pp. 3–27). Springer London. https://doi.org/10.1007/978-1-4471-6497-5_1

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2306–2315. https://doi.org/10.1109/TVCG.2013.234

Bostrom, A., Anselin, L., & Farris, J. (2008). Visualizing Seismic Risk and Uncertainty. *Annals of the New York Academy of Sciences*, *1128*(1), 29–40. https://doi.org/10.1196/annals.1399.005

Brewer, C. A. (1994). Color Use Guidelines for Mapping and Visualization. *Modern Cartography Series*, 123–147. https://doi.org/10.1016/B978-0-08-042415-6.50014-4

Brink, T. (2001). *Usability for the Web—1st Edition*. Morgan Kaufmann; https://www.elsevier.com/books/usability-for-the-web/brinck/978-1-55860-658-6

Buckley, A. (2013, August). *Guidelines for the Effective Design of Spatio-Temporal Maps*.

Cairo, A. (2012). *The Functional Art: An introduction to information graphics and visualization* (Edición: 01). New Riders Publishing.

Camburn, B., Viswanathan, V., Linsey, J., Anderson, D., Jensen, D., Crawford, R., Otto, K., & Wood, K.

(2017). Design prototyping methods: State of the art in strategies, techniques, and guidelines. *Design Science*, *3*. https://doi.org/10.1017/dsj.2017.10

Cao, A., Chintamani, K., Pandya, A., & Ellis, R. (2009a). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods*, *41*, 113–117. https://doi.org/10.3758/BRM.41.1.113

Cao, A., Chintamani, K., Pandya, A., & Ellis, R. (2009b). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods*, *41*, 113–117. https://doi.org/10.3758/BRM.41.1.113

Chan, W. (n.d.). *A Survey on Multivariate Data Visualization*. 29.

Cheshire, J., & Uberti, O. (2014). *London: The Information Capital*. Penguin Books. https://www.oliveruberti.com/the-information-capital

Christel, I., Hemment, D., Bojovic, D., Cucchietti, F., Calvo, L., Stefaner, M., & Buontempo, C. (2018). Introducing design in the development of effective climate services. *Climate Services*, *9*, 111–121. https://doi.org/10.1016/j.cliser.2017.06.002

Ciołkosz-Styk, A. (n.d.). *The visual search method in map perception research*. *4*(1), 10.

Cleveland, W. S., & McGill, R. (1985). Graphical Perception and Graphical Methods for Analyzing Scientific Data. *Science*, *229*(4716), 828–833. https://doi.org/10.1126/science.229.4716.828

Collins, L. (n.d.). *How to analyse usability testing results*. Retrieved August 27, 2021, from https://info.webusability.co.uk/blog/how-to-analyse-usability-testing-results

Core Team, R. (2018). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org

Cui, W., Qu, H., Zhou, H., Zhang, W., & Skiena, S. (2012). Watch the Story Unfold with TextWheel: Visualization of Large-Scale News Streams. *ACM Transactions on Intelligent Systems and Technology*, *3*(2), 20:1-20:17. https://doi.org/10.1145/2089094.2089096

Dastani, M. (2002). The Role of Visual Perception in Data Visualization. *Journal of Visual Languages & Computing*, *13*, 601–622. https://doi.org/10.1006/jvlc.2002.0235

Datavizcatalogue. (2018, June 27). Chart Combinations: Tile Grid Maps - Dataviz Catalogue Blog. *The Data Visualisation Catalogue Blog*. http://datavizcatalogue.com/blog/chart-combinations-tile-grid-maps/

Davis, B. M., Dickerson, K., & Gillmore, S. C. (2020). User Centered Design Strategies for Improving Visualization of Sensor Data in Rotorcraft Cockpit Displays for Degraded Visual Environment Operations. In T. Ahram & C. Falcão (Eds.), *Advances in Usability and User Experience* (pp. 131–141). Springer International Publishing. https://doi.org/10.1007/978-3-030-19135-1_13

Davis, M., Lowe, R., Steffen, S., Doblas-Reyes, F., & Rodó, X. (2016). Barriers to Using Climate Information: Challenges in Communicating Probabilistic Forecasts to Decision-Makers. In J. L. Drake, Y. Y. Kontar, J. C. Eichelberger, T. S. Rupp, & K. M. Taylor (Eds.), *Communicating Climate-Change and Natural Hazard Risk and Cultivating Resilience: Case Studies for a Multi-disciplinary Approach* (pp. 95–113). Springer International Publishing. https://doi.org/10.1007/978-3-319-20161-0_7

diSessa, A. A. (2004). Metarepresentation: Native Competence and Targets for Instruction. *Cognition and Instruction*, *22*(3), 293–331. https://doi.org/10.1207/s1532690xci2203_2

Dong, F., Ghinea, G., & Chen, S. Y. (2008). *User Centered Design for Medical Visualization*. IGI Global. www.igi-global.com/book.aspx?ref=user-centered-design-medical-visualization&titleid=1018&f=e-book&f=e-book

Donker, A., & Reitsma, P. (2007). Young children's ability to use a computer mouse. *Computers & Education*, *48*(4), 602–617. https://doi.org/10.1016/j.compedu.2005.05.001

Dovbysh, O. (2020). Automating the News: How Algorithms are Rewriting the Media. *Digital Journalism*, *0*(0), 1–3. https://doi.org/10.1080/21670811.2020.1777883

Du, Y., Ren, L., Zhou, Y., Li, J., Tian, F., & Dai, G. (2018). Banded choropleth map. *Personal and Ubiquitous Computing*, *22*(3), 503–510. https://doi.org/10.1007/s00779-018-1120-y

Duchowski, A. (2007). *Eye Tracking Methodology | SpringerLink*. https://link.springer.com/book/10.1007/978-1-84628-609-4

Eira, I. M. G., Jaedicke, C., Magga, O. H., Maynard, N. G., Vikhamar-Schuler, D., & Mathiesen, S. D. (2013). Traditional Sámi snow terminology and physical snow classification—Two ways of knowing. *Cold Regions Science and Technology*, *85*, 117–130. https://doi.org/10.1016/j.coldregions.2012.09.004

Ellis, G., & Dix, A. (2006). An explorative analysis of user evaluation studies in information visualisation. *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, 1–7. https://doi.org/10.1145/1168149.1168152

Engebretsen, M., & Kennedy, H. (Eds.). (2020). *Data Visualization in Society*. Amsterdam University Press. https://doi.org/10.5117/9789463722902

Esteves, S., & Neves, M. (2021). Are Charts Going Digital? The Case of Data Visualization on Portuguese Media. In T. Ahram, R. Taiar, K. Langlois, & A. Choplin (Eds.), *Human Interaction, Emerging Technologies and Future Applications III* (pp. 404–409). Springer International Publishing. https://doi.org/10.1007/978-3-030-55307-4_61

Experience, W. L. in R.-B. U. (n.d.). *Prioritizing Web Usability: Book by Jakob Nielsen and Hoa Loranger*. Nielsen Norman Group. Retrieved August 27, 2021, from https://www.nngroup.com/books/prioritizing-web-usability/

Fagen-Ulmschneider, W. (n.d.). *Perception of Probability Words*. Retrieved January 25, 2023, from https://waf.cs.illinois.edu/visualizations/Perception-of-Probability-Words/

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, *35*(3), 379–383.

https://doi.org/10.3758/BF03195514

Faulkner, L., & Wick, D. (2005). Cross-user analysis: Benefits of skill level comparison in usability testing. *Interacting with Computers*, *17*(6), 773–786. https://doi.org/10.1016/j.intcom.2005.04.004

Fernández-Prados, J. S., Cuenca-Piqueras, C., & González-Moreno, M. J. (2019a). International public opinion surveys and public policy in Southern European democracies. *Journal of International and Comparative Social Policy*, *35*(2), 227–237. https://doi.org/10.1080/21699763.2018.1535997

Fernández-Prados, J. S., Cuenca-Piqueras, C., & González-Moreno, M. J. (2019b). International public opinion surveys and public policy in Southern European democracies. *Journal of International and Comparative Social Policy*, *35*(2), 227–237. https://doi.org/10.1080/21699763.2018.1535997

Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis* (1st edition). Analytics Press.

Fish, C., Goldsberry, K. P., & Battersby, S. (2011). Change Blindness in Animated Choropleth Maps: An Empirical Study. *Cartography and Geographic Information Science*, *38*(4), 350–362. https://doi.org/10.1559/15230406384350

Fotheringham, A., Brunsdon, C., & Charlton, M. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis* (Edición: First). SAGE Publications Ltd.

Freitas, C. M. D. S., Luzzardi, P. R. G., Cava, R. A., Winckler, M., Pimenta, M. S., & Nedel, L. P. (2002). On evaluating information visualization techniques. *Proceedings of the Working Conference on Advanced Visual Interfaces*, 373–374. https://doi.org/10.1145/1556262.1556326

Fu, B., Noy, N. F., & Storey, M.-A. (2016). Eye tracking the user experience – An evaluation of ontology visualization techniques. *Semantic Web*, *8*(1), 23–41. https://doi.org/10.3233/SW-140163

Fuchs, G., & Schumann, H. (2004). Visualizing abstract data on maps. *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004.*, 139–144. https://doi.org/10.1109/IV.2004.1320136

Gerardo, J. L. S. (2007). *The effectiveness of novice users in usability testing* [University of Oslo]. https://www.duo.uio.no/handle/10852/9681

Gerharz, L. E., & Pebesma, E. J. (2009). *Usability of interactive and non-interactive visualisation of uncertain geospatial information*. 9.

Gerst, M. D., Kenney, M. A., Baer, A. E., Speciale, A., Wolfinger, J. F., Gottschalck, J., Handel, S., Rosencrans, M., & Dewitt, D. (2020a). Using Visualization Science to Improve Expert and Public Understanding of Probabilistic Temperature and Precipitation Outlooks. *Weather, Climate, and Society*, *12*(1), 117–133. https://doi.org/10.1175/WCAS-D-18-0094.1

Gerst, M. D., Kenney, M. A., Baer, A. E., Speciale, A., Wolfinger, J. F., Gottschalck, J., Handel, S., Rosencrans, M., & Dewitt, D. (2020b). Using Visualization Science to Improve Expert and Public Understanding of Probabilistic Temperature and Precipitation Outlooks. *Weather, Climate, and Society*, *12*(1), 117–133. https://doi.org/10.1175/WCAS-D-18-0094.1

Gesù, V., & Starovoitov, V. (1999). Distance-based functions for image comparison. *Pattern Recognition Letters*, *20*, 207–214. https://doi.org/10.1016/S0167-8655(98)00115-9

Gill, C., Shim, S., & Sanders, L. (2011). PROTOTYPES AS INQUIRY, VISUALIZATION AND COMMUNICATION. *13th International Conference on Engineering and Product Design Education*, 672–677. https://doi.org/10.13140/2.1.1161.3447

Goldberg, J., & Helfman, J. (2010). Comparing Information Graphics: A Critical Look at Eye Tracking. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/2110192.2110203

Goldsberry, K., & Battersby, S. (2009). Issues of Change Detection in Animated Choropleth Maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*. https://doi.org/10.3138/carto.44.3.201

Grainger, S., Mao, F., & Buytaert, W. (2016). Environmental data visualisation for non-scientific contexts:

Literature review and design framework. *Environmental Modelling & Software*, *85*, 299–318. https://doi.org/10.1016/j.envsoft.2016.09.004

Greis, M., Ohler, T., Henze, N., & Schmidt, A. (2015). Investigating Representation Alternatives for Communicating Uncertainty to Non-experts. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2015* (pp. 256–263). Springer International Publishing. https://doi.org/10.1007/978-3-319-22723-8_21

Griffin, T. L. C. (1980). Cartographic transformation of the thematic map base. *Cartography*, *11*(3), 163–174. https://doi.org/10.1080/00690805.1980.10438102

Hansen, J., & Wänke, M. (2009). Liking What's Familiar: The Importance of Unconscious Familiarity in the Mere-Exposure Effect. *Social Cognition*, *27*(2), 161–182. https://doi.org/10.1521/soco.2009.27.2.161

Harold, J., Lorenzoni, I., Shipley, T. F., & Coventry, K. R. (2016). Cognitive and psychological science insights to improve climate change data visualization. *Nature Climate Change*, *6*(12), 1080–1089. https://doi.org/10.1038/nclimate3162

Harrower, M. (2007). The Cognitive Limits of Animated Maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*. https://doi.org/10.3138/carto.42.4.349

Hassenzahl, M., & Sandweg, N. (2004). From mental effort to perceived usability: Transforming experiences into summary assessments. *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 1283–1286. https://doi.org/10.1145/985921.986044

Healey, C. G. (1996). Choosing effective colours for data visualization. *Proceedings of Seventh Annual IEEE Visualization '96*, 263–270. https://doi.org/10.1109/VISUAL.1996.568118

Hegarty, M. (2011). The Cognitive Science of Visual-Spatial Displays: Implications for Design. *Topics in Cognitive Science*, *3*(3), 446–474. https://doi.org/10.1111/j.1756-8765.2011.01150.x

Hegarty, M., Canham, M., & Fabrikant, S. (2010). Thinking About the Weather: How Display Salience and

Knowledge Affect Performance in a Graphic Inference Task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*, 37–53. https://doi.org/10.1037/a0017683

Hertzum, M., Borlund, P., & Kristoffersen, K. B. (2015). What Do Thinking-Aloud Participants Say? A Comparison of Moderated and Unmoderated Usability Sessions. *International Journal of Human–Computer Interaction*, *31*(9), 557–570. https://doi.org/10.1080/10447318.2015.1065691

Hewitt, C. D., Guglielmo, F., Joussaume, S., Bessembinder, J., Christel, I., Doblas-Reyes, F. J., Djurdjevic, V., Garrett, N., Kjellström, E., Krzic, A., Costa, M. M., & Clair, A. L. S. (2021). Recommendations for Future Research Priorities for Climate Modeling and Climate Services. *Bulletin of the American Meteorological Society*, *102*(3), E578–E588. https://doi.org/10.1175/BAMS-D-20-0103.1

Hewitt, M. (2016, December 1). *11 Design Tips for Visualizing Survey Results*. Visage. https://visage.co/design-tips-for-visualizing-survey-results/

Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics*, *36*(2), 185–198. https://doi.org/10.3758/BF03202679

Holmes, N. (1984). *Designer's Guide to Creating Charts and Diagrams* (2nd Edition). Watson-Guptill.

Holtzblatt, K., Wendell, J. B., & Wood, S. (2004a). *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. Elsevier.

Holtzblatt, K., Wendell, J. B., & Wood, S. (2004b). *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. Elsevier.

Iliinsky, N., & Steele, J. (2011). *Designing Data Visualizations*. O'Reilly.

Jenny, B., & Kelso, N. V. (2007). Color Design for the Color Vision Impaired. *Cartographic Perspectives*, *58*, 61–67. https://doi.org/10.14714/CP58.270

Jia, Q.-H., Li, C., & Gao, D. L. (2017). *Study on the Application of Data Visualization in Public Administration*. 21–26. https://doi.org/10.2991/sschd-17.2017.4

Jiang, X., Atkins, M., Tien, G., Bednarik, R., & Zheng, B. (2014). *Pupil Responses during Discrete Goal-directed Movements*. https://doi.org/10.1145/2556288.2557086

Joslyn, S. L., & LeClerc, J. E. (2012). Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of Experimental Psychology: Applied*, *18*(1), 126–140. https://doi.org/10.1037/a0025185

Kaye, N. R., Hartley, A., & Hemming, D. (2012). Mapping the climate: Guidance on appropriate techniques to map climate variables and their uncertainty. *Geoscientific Model Development*, *5*(1), 245–256. https://doi.org/10.5194/gmd-5-245-2012

Keim, D. A., North, S. C., Panse, C., & Schneidewind, J. (2002). Efficient cartogram generation: A comparison. *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.*, 33–36. https://doi.org/10.1109/INFVIS.2002.1173144

Keskin, M., Ooms, K., Dogru, A. O., & De Maeyer, P. (2020). Exploring the Cognitive Load of Expert and Novice Map Users Using EEG and Eye Tracking. *ISPRS International Journal of Geo-Information*, *9*(7), 429. https://doi.org/10.3390/ijgi9070429

Khamaj, A., Kang, Z., & Argyle, E. (2019). Users' Perceptions of Smartphone Weather Applications' Usability. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2216–2220. https://doi.org/10.1177/1071181319631098

Kim, Y.-S., & Hullman, J. (n.d.). *User-driven Expectation Visualization: Opportunities for Personalized Feedback*. 3.

Kittur, A., Chi, H., & Suh, B. (2007). *Crowdsourcing for Usability: Using Micro-Task Markets for Rapid , Remote , and Low-Cost User Measurements*.

Klingner, J., Kumar, R., & Hanrahan, P. (2008). *Measuring the Task-Evoked Pupillary Response with a Remote Eye Tracker*. https://doi.org/10.1145/1344471.1344489

Kohlhammer, J., & Zeltzer, D. (2004). Towards a visualization architecture for time-critical applications.

Proceedings of the 9th International Conference on Intelligent User Interfaces, 271–273. https://doi.org/10.1145/964442.964502

Kosara, R. (2013). InfoVis Is So Much More: A Comment on Gelman and Unwin and an Invitation to Consider the Opportunities. *Journal of Computational and Graphical Statistics*, *22*(1), 29–32. https://doi.org/10.1080/10618600.2012.755465

Kozak, M., Hartley, J., Wnuk, A., & Tartanus, M. (2015). Multiple Pie Charts: Unreadable, Inefficient, and Over-Used. *Journal of Scholarly Publishing*, *46*, 282–289. https://doi.org/10.3138/jsp.46.3.05

Kraak, M.-J., Köbben, B., & Tong, Y. (2014). Integrated Time and Distance Line Cartogram: A Schematic Approach to Understand the Narrative of Movements. *Cartographic Perspectives*, *77*, 7–16. https://doi.org/10.14714/CP77.1234

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE*, *13*(9), e0203629. https://doi.org/10.1371/journal.pone.0203629

Kucher, K., Paradis, C., & Kerren, A. (2018). The State of the Art in Sentiment Visualization. *Computer Graphics Forum*, *37*(1), 71–96. https://doi.org/10.1111/cgf.13217

Langton, S., & Solymosi, R. (2019). *Cartograms, hexograms and regular grids: A commentary on minimising misrepresentation in spatial data visualisations* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/t6agd

Lau, A., & Vande Moere, A. (2007). Towards a Model of Information Aesthetics in Information Visualization. *2007 11th International Conference Information Visualization (IV '07)*, 87–92. https://doi.org/10.1109/IV.2007.114

Li, Y., Hollender, N., & Held, T. (2013). *Task Sequence Effects in Usability Tests*.

Liao, X., Nazir, S., Zhou, Y., Shafiq, M., & Qi, X. (2021). User Knowledge, Data Modelling, and Visualization: Handling through the Fuzzy Logic-Based Approach. *Complexity*, *2021*, e6629086.

https://doi.org/10.1155/2021/6629086

Light, A., & Bartlein, P. (2004). The End of the Rainbow? Color Schemes for Improved Data Graphics. *Eos, Transactions American Geophysical Union*, *85*. https://doi.org/10.1029/2004EO400002

Lim, K., Selamat, A., Alias, R., Krejcar, O., & Fujita, H. (2019). Usability Measures in Mobile-Based Augmented Reality Learning Applications: A Systematic Review. *Applied Sciences*, *9*, 2718. https://doi.org/10.3390/app9132718

Ling, C., Hua, L., Karstens, C. D., Stumpf, G. J., Smith, T. M., Kuhlman, K. M., & Rothfusz, L. (2015). A Comparison Between WarnGen System and Probabilistic Hazard Information System for Severe Weather Forecasting. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1791–1795. https://doi.org/10.1177/1541931215591387

Liu, J., Saquib, N., Chen, Z., Kazi, R. H., Wei, L.-Y., Fu, H., & Tai, C.-L. (2022). *VCoach: A Customizable Visualization and Analysis System for Video-based Running Coaching* (arXiv:2204.08805). arXiv. https://doi.org/10.48550/arXiv.2204.08805

Liu, J., Tang, T., Wang, W., Xu, B., Kong, X., & Xia, F. (2018). A Survey of Scholarly Data Visualization. *IEEE Access*, *6*, 19205–19221. https://doi.org/10.1109/ACCESS.2018.2815030

Liu, X., Hu, Y., North, S., & Shen, H.-W. (2018). CorrelatedMultiples: Spatially Coherent Small Multiples With Constrained Multi-Dimensional Scaling. *Computer Graphics Forum*, *37*(1), 7–18. https://doi.org/10.1111/cgf.12526

Lloyd, R. (1997). *Spatial Cognition: Geographic Environments*. Springer Science & Business Media.

Lorenz, S., Dessai, S., Forster, P. M., & Paavola, J. (2015). Tailoring the visual communication of climate projections for local adaptation practitioners in Germany and the UK. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2055), 20140457. https://doi.org/10.1098/rsta.2014.0457

Lupa, M., Szombara, S., Chuchro, M., & Chrobak, T. (2017). Limits of Colour Perception in the Context of

Minimum Dimensions in Digital Cartography. *ISPRS International Journal of Geo-Information*, *6*(9), 276. https://doi.org/10.3390/ijgi6090276

Majooni, A., Masood, M., & Akhavan, A. (2018). An eye-tracking study on the effect of infographic structures on viewer's comprehension and cognitive load. *Information Visualization*, *17*(3), 257–266. https://doi.org/10.1177/1473871617701971

Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., & Miller, R. C. (2011). Twitinfo: Aggregating and visualizing microblogs for event exploration. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 227–236. https://doi.org/10.1145/1978942.1978975

Martens, M. H., & Fox, M. R. J. (2007). Do familiarity and expectations change perception? Drivers' glances and response to changes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *10*(6), 476–492. https://doi.org/10.1016/j.trf.2007.05.003

Mazăreanu, V. P. (2013). Using geographical information systems as an information visualization tool. A case study. *Analele Ştiinţifice Ale Universităţii »Alexandru Ioan Cuza« din Iaşi. Ştiinţe Economice*, *60*(1), 139–146.

McDonald, J. K., Rich, P. J., & Gubler, N. B. (2019). The Perceived Value of Informal, Peer Critique in the Instructional Design Studio. *TechTrends*, *63*(2), 149–159. https://doi.org/10.1007/s11528-018-0302-9

McInerny, G. J., Chen, M., Freeman, R., Gavaghan, D., Meyer, M., Rowland, F., Spiegelhalter, D. J., Stefaner, M., Tessarolo, G., & Hortal, J. (2014). Information visualisation for science and policy: Engaging users and avoiding bias. *Trends in Ecology & Evolution*, *29*(3), 148–157. https://doi.org/10.1016/j.tree.2014.01.003

McLellan, T. M., Caldwell, J. A., & Lieberman, H. R. (2016). A review of caffeine's effects on cognitive, physical and occupational performance. *Neuroscience & Biobehavioral Reviews*, *71*, 294–312. https://doi.org/10.1016/j.neubiorev.2016.09.001

Michela, E., & McDonald, J. K. (2020). Relationships, Feedback, and Student Growth in the Design Studio: A Case Study. In B. Hokanson, G. Clinton, A. A. Tawfik, A. Grincewicz, & M. Schmidt (Eds.), *Educational Technology Beyond Content: A New Focus for Learning* (pp. 183–192). Springer International Publishing. https://doi.org/10.1007/978-3-030-37254-5_16

Miraz, M. H., Excell, P. S., & Ali, M. (2016). User interface (UI) design issues for multilingual users: A case study. *Universal Access in the Information Society*, *15*(3), 431–444. https://doi.org/10.1007/s10209-014-0397-5

Moon, E., Schneider, S. M., Ferronato, P., Salah, M., Ruecker, S., & McDonagh, D. (2019a). The Role of Empathic Design Research in the Prototyping of Interactive Visualizations. *The Design Journal*, *22*(sup1), 1515–1527. https://doi.org/10.1080/14606925.2019.1594974

Moon, E., Schneider, S. M., Ferronato, P., Salah, M., Ruecker, S., & McDonagh, D. (2019b). The Role of Empathic Design Research in the Prototyping of Interactive Visualizations. *The Design Journal*, *22*(sup1), 1515–1527. https://doi.org/10.1080/14606925.2019.1594974

Munzner, T. (2014). *Visualization Analysis and Design* (1st edition). A K Peters/CRC Press.

Nash, J. (2011). *Plotting Likert and Other Rating Scales*. 9.

Nasiritousi, N., Hjerpe, M., & Linnér, B.-O. (2016). The roles of non-state actors in climate change governance: Understanding agency through governance profiles. *International Environmental Agreements: Politics, Law and Economics*, *16*(1), 109–126. https://doi.org/10.1007/s10784-014-9243-8

Nasser, A., Hamad, D., & Nasr, C. (2006). Visualization Methods for Exploratory Data Analysis. *2006 2nd International Conference on Information & Communication Technologies*, *1*, 1379–1384. https://doi.org/10.1109/ICTTA.2006.1684582

Navarro, I., de Reina, O., Fonseca, D., Gomez, M., & Ferrer, Á. (2020). *Virtual Reality Using Smart-Devices in Educational Frameworks: Case Study - Museum Casa Batlló* (pp. 629–641).

https://doi.org/10.4018/978-1-7998-1757-4.ch036

Nielsen, J. (1992). *Advances in Human-computer Interaction (Vol. 3)* (H. R. Hartson & D. Hix, Eds.; pp. 69–82). Ablex Publishing Corp. http://dl.acm.org/citation.cfm?id=159433.159440

Nusrat, S., Alam, M. J., & Kobourov, S. (2018). Evaluating Cartogram Effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, *24*(2), 1077–1090. https://doi.org/10.1109/TVCG.2016.2642109

Nusrat, S., & Kobourov, S. (2016). The State of the Art in Cartograms. *ArXiv:1605.08485 [Cs]*. http://arxiv.org/abs/1605.08485

Oakley, N. S., & Daudert, B. (2016). Establishing Best Practices to Improve Usefulness and Usability of Web Interfaces Providing Atmospheric Data. *Bulletin of the American Meteorological Society*, *97*(2), 263–274. https://doi.org/10.1175/BAMS-D-14-00121.1

Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010). Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2381–2390. https://doi.org/10.1145/1753326.1753685

Ooms, K., Coltekin, A., De Maeyer, P., Dupont, L., Fabrikant, S., Incoul, A., Kuhn, M., Slabbinck, H., Vansteenkiste, P., & Van der Haegen, L. (2015). Combining user logging with eye tracking for interactive and dynamic applications. *Behavior Research Methods*, *47*(4), 977–993. https://doi.org/10.3758/s13428-014-0542-3

Ooms, K., De Maeyer, P., Fack, V., Van Assche, E., & Witlox, F. (2012). Interpreting maps through the eyes of expert and novice users. *International Journal of Geographical Information Science*, *26*(10), 1773–1788. https://doi.org/10.1080/13658816.2011.642801

Petrillo, F., Spritzer, A., Freitas, C., & Pimenta, M. (2011). *Interactive analysis of Likert scale data using a multichart visualization tool* (p. 365).

Pifarré, M., & Tomico, O. (2007a). Bipolar laddering (BLA): A participatory subjective exploration method on user experience. *Proceedings of the 2007 Conference on Designing for User EXperiences, DUX'07*, 2. https://doi.org/10.1145/1389908.1389911

Pifarré, M., & Tomico, O. (2007b). Bipolar laddering (BLA): A participatory subjective exploration method on user experience. *Proceedings of the 2007 Conference on Designing for User EXperiences, DUX'07*, 2. https://doi.org/10.1145/1389908.1389911

Pirrone, A. (2020, June 4). *Seven Different Ways to Display Likert Scale Data*. Medium. https://medium.com/nightingale/seven-different-ways-to-display-likert-scale-data-d0c1c9a9ad59

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., Gochis, D. J., Ahmadov, R., Peckham, S. E., Grell, G. A., Michalakes, J., Trahan, S., Benjamin, S. G., Alexander, C. R., Dimego, G. J., Wang, W., Schwartz, C. S., Romine, G. S., … Duda, M. G. (2017). The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. *Bulletin of the American Meteorological Society*, *98*(8), 1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1

Prendeville, S., & Bocken, N. (2017). Sustainable Business Models through Service Design. *Procedia Manufacturing*, *8*, 292–299. https://doi.org/10.1016/j.promfg.2017.02.037

Pretorius, M., Calitz, A. P., & Greunen, D. van. (2005). *The added value of eye tracking in the usability evaluation of a network management tool*.

Prümper, J., Frese, M., Zapf, D., & Brodbeck, F. (1991). Errors in computerized office work: Differences between novice and expert users. *ACM SIGCHI Bulletin*, *23*, 63–66. https://doi.org/10.1145/122488.122498

Rachynska, O. (2020). Influence of Public Opinion on the Optimization of Communicative Interaction in the Public Administration Sphere. *Public Administration Aspects*, *8*(2), Article 2. https://doi.org/10.15421/152024

Rao, A. S., Vardhan, B. V., & Shaik, H. (2021). Role of Exploratory Data Analysis in Data Science. *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, 1457–1461. https://doi.org/10.1109/ICCES51350.2021.9488986

Relia, K., Akbari, M., Duncan, D., & Chunara, R. (2018). Socio-spatial Self-organizing Maps: Using Social Media to Assess Relevant Geographies for Exposure to Social Processes. *Proceedings of the ACM on Human-Computer Interaction*, *2*. https://doi.org/10.1145/3274414

Rensink, R. A. (2021). Visualization as a stimulus domain for vision science. *Journal of Vision*, *21*(8), 3–3. https://doi.org/10.1167/jov.21.8.3

Ribecca, S. (2018, June 27). Chart Combinations: Tile Grid Maps - Dataviz Catalogue Blog. *The Data Visualisation Catalogue Blog*. http://datavizcatalogue.com/blog/chart-combinations-tile-grid-maps/

Riveiro, M., Falkman, G., & Ziemke, T. (2008). Improving maritime anomaly detection and situation awareness through interactive visualization. *2008 11th International Conference on Information Fusion*, 1–8.

Roulston, M. S., Bolton, G. E., Kleit, A. N., & Sears-Collins, A. L. (2006). A Laboratory Study of the Benefits of Including Uncertainty Information in Weather Forecasts. *Weather and Forecasting*, *21*(1), 116–122. https://doi.org/10.1175/WAF887.1

S2S4E Decision Support Tool. (2020). https://s2s4e.eu/dst

Sager, L., Grier, R., Jackson, C., Levchuk, Y., & Stelzer, E. (2007). *Visualization techniques for revealing uncertainty to decision makers*.

Sarkar, D. (DJ). (2018, December 12). Effective Visualization of Multi-Dimensional Data—A Hands-on Approach. *Medium*. https://medium.com/swlh/effective-visualization-of-multi-dimensional-data-a-hands-on-approach-b48f36a56ee8

Sauer, J., Seibel, K., & Rüttinger, B. (2010a). The influence of user expertise and prototype fidelity in usability tests. *Applied Ergonomics*, *41*(1), 130–140. https://doi.org/10.1016/j.apergo.2009.06.003

Sauer, J., Seibel, K., & Rüttinger, B. (2010b). The influence of user expertise and prototype fidelity in usability tests. *Applied Ergonomics*, *41*(1), 130–140. https://doi.org/10.1016/j.apergo.2009.06.003

Schiewe, J. (2019a). Empirical Studies on the Visual Perception of Spatial Patterns in Choropleth Maps. *KN - Journal of Cartography and Geographic Information*, *69*(3), 217–228. https://doi.org/10.1007/s42489-019-00026-y

Schiewe, J. (2019b). Empirical Studies on the Visual Perception of Spatial Patterns in Choropleth Maps. *KN - Journal of Cartography and Geographic Information*, *69*(3), 217–228. https://doi.org/10.1007/s42489-019-00026-y

Schiewe, J. (2019c). Empirical Studies on the Visual Perception of Spatial Patterns in Choropleth Maps. *KN - Journal of Cartography and Geographic Information*, *69*(3), 217–228. https://doi.org/10.1007/s42489-019-00026-y

Schrepp, M., Thomaschewski, J., & Hinderks, A. (2017). Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(Regular Issue). https://www.ijimai.org/journal/bibcite/reference/2634

Skaarhoj, K. (1998, 2023). *GESIS - Leibniz Institute for the Social Sciences*. https://www.gesis.org/en/eurobarometer-data-service/survey-series/candidate-countries-eb/samplin g-fieldwork

Slocum, T. A., Robeson, S. H., & Egbert, S. L. (2006). Traditional Versus Sequenced Choropleth Maps/An Experimental Investigation. *Cartographica: The International Journal for Geographic Information and Geovisualization*. https://doi.org/10.3138/CG7N-0158-1537-6177

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing Uncertainty About the Future. *Science*, *333*(6048), 1393. https://doi.org/10.1126/science.1191181

Spinde, T., Hamborg, F., Donnay, K., Becerra, A., & Gipp, B. (2020). Enabling News Consumers to View and Understand Biased News Coverage: A Study on the Perception and Visualization of Media Bias. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 389–392. https://doi.org/10.1145/3383583.3398619

Stephens, E. M., Edwards, T. L., & Demeritt, D. (2012). Communicating probabilistic information from climate model ensembles—Lessons from numerical weather prediction. *WIREs Climate Change*, *3*(5), 409–426. https://doi.org/10.1002/wcc.187

Stewart, J., & Kennelly, P. (2010). Illuminated Choropleth Maps. *Annals of the Association of American Geographers*, *100*, 513–534. https://doi.org/10.1080/00045608.2010.485449

Stickdorn, M., & Hormess, M. E. (2016). *This is Service Design Doing—Book / School / Methods*. https://www.thisisservicedesigndoing.com/

Swoboda, E., & Vighi, P. (2016). *Early Geometrical Thinking in the Environment of Patterns, Mosaics and Isometries*. Springer.

Synnott, J., Chen, L., Nugent, C. D., & Moore, G. (2012). Flexible and customizable visualization of data generated within Intelligent Environments. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 5819–5822. https://doi.org/10.1109/EMBC.2012.6347317

Szafir, D. A., & Szafir, D. (2016). *Cognitive Load in Visualization: Myths & Misconceptions*. 1. https://doi.org/10.475/123 4

Tamaazousti, Y., Le Borgne, H., & Hudelot, C. (2017). MuCaLe-Net: Multi Categorical-Level Networks to Generate More Discriminating Features. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5282–5291. https://doi.org/10.1109/CVPR.2017.561

Tambouris, E., Kalampokis, E., & Tarabanis, K. (2017). Visualizing Linked Open Statistical Data to Support Public Administration. *Proceedings of the 18th Annual International Conference on Digital Government Research*, 149–154. https://doi.org/10.1145/3085228.3085304

Taylor, A. L., Dessai, S., & de Bruin, W. B. (2015a). Communicating uncertainty in seasonal and interannual climate forecasts in Europe. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2055), 20140454. https://doi.org/10.1098/rsta.2014.0454

Taylor, A. L., Dessai, S., & de Bruin, W. B. (2015b). Communicating uncertainty in seasonal and interannual climate forecasts in Europe. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2055), 20140454. https://doi.org/10.1098/rsta.2014.0454

Terrado, M., Calvo, L., Bojovic, D., & Christel, I. (2022). Current Practice in Climate Service Visualization: Taking the Pulse of the Providers' Community. *Bulletin of the American Meteorological Society*, *103*(3), E828–E837. https://doi.org/10.1175/BAMS-D-21-0194.1

Terrado, M., Lledó, L., Bojovic, D., St. Clair, A. L., Soret, A., Doblas-Reyes, F. J., Manzanas, R., San-Martín, D., & Christel, I. (2019). The Weather Roulette: A Game to Communicate the Usefulness of Probabilistic Climate Predictions. *Bulletin of the American Meteorological Society*, *100*(10), 1909–1921. https://doi.org/10.1175/BAMS-D-18-0214.1

Tobler, W. (2004). Thirty-Five Years of Computer Cartograms. *Annals of the Association of American Geographers*, *94*(1), 58–73. JSTOR.

Trischler, J., Pervan, S. J., Kelly, S. J., & Scott, D. R. (2018). The Value of Codesign: The Effect of Customer Involvement in Service Design Teams. *Journal of Service Research*, *21*(1), 75–100. https://doi.org/10.1177/1094670517714060

Trivedi, (Dr.) Munesh, & Khanum, A. (2012). *Role Of Context In Usability Evaluations: A Review*. https://doi.org/10.5121/acij.2012.3208

Trivedi, M. C. (2012). Role Of Context In Usability Evaluations: A Review. *Advanced Computing: An International Journal*, *3*(2), 69–78. https://doi.org/10.5121/acij.2012.3208

Ueland, J., & Vincent, P. (2004). Mapping Florida: Beyond The Choropleth Map. *The Florida Geographer*, *35*. https://journals.flvc.org/flgeog/article/view/77916

UNSW. (2015). Colour mixing and colour vision. In *Physclips*. https://www.animations.physics.unsw.edu.au/jw/light/colour-mixing.htm

Veer, G. C. van der, Hennipman, E.-J., & Oppelaar, E.-J. (2009). User Centered Design Patterns for Visualization. *Software Engineering*.

Veras Guimarães, R. (2019). *Visual encoding quality and scalability in information visualization*. http://ir.library.dc-uoit.ca/handle/10155/1024

Virzi, R. A. (1992). Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Human Factors*, *34*(4), 457–468. https://doi.org/10.1177/001872089203400407

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems*, *62*, 1–10. https://doi.org/10.1016/j.dss.2014.02.007

Ware, C. (2012). *Information Visualization: Perception for Design*. Elsevier.

Ware, C., Purchase, H., Colpoys, L., & McGill, M. (2002). Cognitive Measurements of Graph Aesthetics. *Information Visualization*, *1*(2), 103–110. https://doi.org/10.1057/palgrave.ivs.9500013

Wassink, I., Kulyk, O., Dijk, B. van, Veer, G. van der, & Vet, P. van der. (2009). Applying a User-centered Approach to Interactive Visualisation Design. In R. Liere, T. Adriaansen, & E. Zudilova-Seinstra (Eds.), *Trends in Interactive Visualization: State-of-the-Art Survey* (pp. 175–199). Springer. https://doi.org/10.1007/978-1-84800-269-2_8

Wexler, S. (2016). *Visualizing Survey Data*.

Wheeldon, J., & K. Ahlberg, M. (2021, July 15). *Visualizing Social Science Research*. SAGE Publications Inc. https://us.sagepub.com/en-us/nam/visualizing-social-science-research/book235388

Wilson, C. (2014). Chapter 4—Cognitive Walkthrough. In C. Wilson (Ed.), *User Interface Inspection Methods* (pp. 65–79). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-410391-7.00004-X

Winckler, M. A., Palanque, P., & Freitas, C. M. D. S. (2004). Tasks and scenario-based evaluation of information visualization techniques. *Proceedings of the 3rd Annual Conference on Task Models and Diagrams*, 165–172. https://doi.org/10.1145/1045446.1045475

WMO. (2018). Guidelines on Quality Management in Climate Services. *Climate Services*, *1221*, 53.

Wongsuphasawat, K. (n.d.). *Quality Metrics for Tile Grid Maps*. 2.

Yang, H., Tang, K., Liu, X., Xiao, L., Xu, R., & Kumara, S. (2016). A user-centred approach to information visualisation in nano-health. *International Journal of Bioinformatics Research and Applications*, *12*(2), 95. https://doi.org/10.1504/IJBRA.2016.077122

Yi, J. S., Kang, Y. ah, Stasko, J., & Jacko, J. A. (2007). Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1224–1231. https://doi.org/10.1109/TVCG.2007.70515

Yoghourdjian, V., Archambault, D., Diehl, S., Dwyer, T., Klein, K., Purchase, H. C., & Wu, H.-Y. (2018). Exploring the Limits of Complexity: A Survey of Empirical Studies on Graph Visualisation. *ArXiv:1809.00270 [Cs]*. http://arxiv.org/abs/1809.00270

Yucong, Ye, Sauer, F., Ma, K.-L., Aditya, K., & Chen, J. (2019). A User-centered Design Study in Scientific Visualization Targeting Domain Experts. *ArXiv:1903.12349 [Cs]*. http://arxiv.org/abs/1903.12349

Zhang, Y., & Maciejewski, R. (2017). Quantifying the Visual Impact of Classification Boundaries in Choropleth Maps. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1), 371–380. https://doi.org/10.1109/TVCG.2016.2598541

Zissman, A., & Neimark, E. (1990). The Influence of Familiarity on Evaluations of Liking and Goodness of Several Types of Music. *The Psychological Record*, *40*(4), 481–490. https://doi.org/10.1007/BF03399535

9. Annexes

9.1. Annex I

Table 1 Task selection based on the combinations of the three variables involved: time, region and value. Those combinations which did not involve changes over time (the fixed time variable) were discarded (shown in grey). The column with a 'yes' indicates a fixed value for the variable and is defined in the 'conditions' column.

id	time	region	value	conditions	Possible questions
–	Yes	–	–	t=n	Which value predominates in the whole map in (t_n)?
1	–	Yes	–	region ='A'	What changes occurred in region 'A' over time?
2	–	–	Yes	value='x'	Which region or regions followed a certain pattern (remain stable, increase, decrease, etc.) over time?
–	Yes	Yes	–	t=n region='A'	What is the value of region 'A' at (t_n)?
1b	–	Yes	Yes	region='A' value='x'	For region 'A', which period of time had a x ('Low') value?
–	Yes	–	Yes	t=n,value='x'	(Which regions/How many regions) had a x ('Low') value?
3	–	–	–	–	What was the trend for the whole map over time?
–	Yes	Yes	Yes	t=n, region='A' value='x'	Confirmation question: Does region 'A' meet value x at t_n ? Answer(Yes/No)

Table 2 Statistical results for completion time

Time analysis

Task 1

Anova

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
map	3	5332	1777.3	10.63	2.85e-06 ***
Residuals	124	20723	167.1		

Tukey results

	diff	lwr	upr	p adj
banded_tile_grid-banded_choropleth	-11.576250	-19.992827	-3.159673	**0.0027181**
choropleth-banded_choropleth	-9.595625	-18.012202	-1.179048	0.0185823
tile_grid_map-banded_choropleth	3.894063	-4.522515	12.310640	0.6249272
choropleth-banded_tile_grid	1.980625	-6.435952	10.397202	0.9278302
tile_grid_map-banded_tile_grid	15.470312	7.053735	23.886890	**0.0000279**
tile_grid_map-choropleth	13.489688	5.073110	21.906265	**0.0003229**

Task 2

Anova

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
map	3	21145	7048	21.6	2.53e-11 ***
Residuals	124	40455	326		

Tukey results

	diff	lwr	upr	p adj
banded_tile_grid-banded_choropleth	-8.24281	-20.002364	3.516739	0.266375
choropleth-banded_choropleth	**17.05656**	**5.297011**	**28.81611**	**0.001383**
tile_grid_map-banded_choropleth	**23.84437**	**12.08482**	**35.60392**	**0.000003**
choropleth-banded_tile_grid	**25.29937**	**13.53982**	**37.05892**	**0.0000008**
tile_grid_map-banded_tile_grid	**32.087188**	**20.327636**	**43.846739**	**0.00000**
tile_grid_map-choropleth	6.787813	-4.971739	18.547364	0.4386920

Task 3

Anova

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
map	3	226658	75553	73.86	<2e-16 ***
Residuals	124	126838	1023		

Tukey results

	diff	lwr	upr	p adj
banded_tile_grid-banded_choropleth	-11.457500	-32.27990	9.364903	0.4813345
choropleth-banded_choropleth	**-91.70250**	**-112.52490**	**-70.880097**	**0.0000000**
tile_grid_map-banded_choropleth	**-87.17281**	**-107.99522**	**-66.350409**	**0.0000000**
choropleth-banded_tile_grid	**-80.24500**	**-101.06740**	**-59.422597**	**0.0000000**
tile_grid_map-banded_tile_grid	**-75.715312**	**-96.53772**	**-54.892909**	**0.0000000**
tile_grid_map-choropleth	4.529688	-16.29272	25.352091	0.9418171

Table 3: Statistical results for success rates

Success rates analysis
Task 1 X-squared = 3.3231, df = 6, p-value = 0.7673
Tile grid map - Choropleth X-squared = 0, df = 1, p-value = 1 Tile grid map - Banded tile grid map X-squared = 0, df = 1, p-value = 1 Tile grid map - Banded choropleth X-squared = 1.1852, df = 2, p-value = 0.5529 Choropleth - Banded choropleth X-squared = 1.1852, df = 2, p-value = 0.5529 Banded tile grid map - Choropleth X-squared = 0, df = 1, p-value = 1 Banded tile grid map - Banded choropleth X-squared = 1.0189, df = 2, p-value = 0.6008
Task 2 – X-squared = 18.426, df = 6, p-value = 0.005251
Chi-squared pair results
Banded choropleth - Banded tile grid map X-squared = 2.4, df = 1, p-value = 0.1213 Banded choropleth - Tile grid map X-squared = 3.2169, df = 2, p-value = 0.2002 Choropleth - Banded choropleth X-squared = 3.6431, df = 2, p-value = 0.1618 Choropleth - Banded tile grid map **X-squared = 11.852, df = 2, p-value = 0.002669** Choropleth - Tile grid map X-squared = 4.3333, df = 2, p-value = 0.1146 Tile grid map - Banded tile grid map **X-squared = 6.6207, df = 2, p-value = 0.0365**
Task 3 – X-squared = 159.51, df = 6, p-value < 2.2e-16
Banded choropleth - Banded tile grid map

X-squared = 159.51, df = 6, p-value < 2.2e-16

Banded choropleth - Tile grid map
X-squared = 50.943, df = 2, p-value = 8.668e-12

Choropleth - Banded choropleth
X-squared = 47.216, df = 2, p-value = 5.588e-11

Choropleth - Banded tile grid map
X-squared = 39.2, df = 2, p-value = 3.075e-09

Choropleth - Tile grid map
X-squared = 49.016, df = 2, p-value = 2.271e-11

Tile grid map - Banded tile grid map
X-squared = 43.292, df = 2, p-value = 3.975e-10

Table 4: Information on the confidence intervals (CI) for each of the representations

Task 1	Task 2	Task 3
Choropleth 15.1 [13.2 – 17.0]	**Choropleth** 37.4 [31.1 – 43.6]	**Choropleth** 11.8 [8.8 – 14.7]
Tile grid map 28.6 [24.3 – 32.8]	**Tile grid map** 44.1[33.8 – 54.5]	**Tile grid map** 16.3[8.8 – 23.8]
Banded choropleth 24.7[17.4 – 32.0]	**Banded choropleth** 20.3[16.2 – 24.5]	**Banded choropleth** 103.5[89.4 – 117.5]
Banded tile grid map 13.1[9.7 – 16.5]	**Banded tile grid map** 12.0 [9.5–14.6]	**Banded tile grid map** 92.1[75.6 – 108.4]

Table 5: Most significant comments gathered during the qualitative review. Tg = Tile grid map, Ch = Choropleth, BC = Banded choropleth, BT = Banded tile grid map.

Task	Map	Comment
1	Tg	*It is difficult to locate a certain area on the map, you need the help of the legend.*
1	Tg	*It is more complicated when the county is inside* (in the central part) *the map.*
1	Tg	*I had to double check to verify the colour* (referring to one of the time steps) *in the legend.*
1	Tg-Ch	*The Tile grid map and the Choropleth require a little more effort because you have to compare the regions on the three maps.*
1	BC-BT	*The comparison is obviously more direct in the banded versions.*
1	BC-BT	*I think that the banded versions are more useful, now we only compare one region, but when more than region comparison is needed, the difference will be clearer.*
1	Ch	*Choropleths are more familiar, it is used everywhere, but I think the other maps are better to solve the task* (referring to banded maps).
1	Tg-Ch	*Normally we want to compare the trend of more than one region, and it would be much more complex for small multiple maps.*
1	BC	*The characteristic shape often helps to identify one region on all three maps, but other times, when regions nearby share the same colour it is not as straightforward.*
1	BC	*Depending on the colour of the bands in the surrounding regions, it is sometimes more difficult to establish the boundaries. However, this is not the case with the Banded Tile grid map.*
1	BC-BT	*We are only observing the trend of a region. If we were asked to observe the changes of two or three regions simultaneously, the banded versions would behave much more efficiently than the traditional ones* (referring to the choropleth and tile maps).
1	Ch-Tg	*With the small multiple maps the process is as follows: you would locate the region on each of the three maps and then identify or compare changes in the three maps, while in the combined* (banded) *maps, everything is in the same place.*
1	BT	*It is easy to distinguish, because of the proximity. It is automatic.*
1	BC	It is easier than having to compare between the three maps, but sometimes the shape of the regions is confusing depending on the colours of the surrounding regions.
1	BC-BT	*Both maps* (referring to banded versions) *are a good choice. The Tile grid map is very useful once you get used to it.*
1	BT	*Once the region is located, it is very easy to determine the change.*
1	BT	*It is a very simple, very intuitive task. I insist, for me, the best of the representations* (referring to Banded tile graph).
1	Ch	*It is the one with which I feel most comfortable and the one that favours the task more.*

1	BT	*This map* (referring to the Banded tile grid map) *is very intuitive. Very easy.*
1	BT	*I find it very difficult to locate or identify a region with this type of map.*
1	Ch	I like this representation better because it is the one I know better and it is more familiar to me.
1	All / BC-BT	*The task was quite easy in general, but I think the maps with the bands* (referring to the banded maps) *are slightly easier to read.*
1	Ch	*They are very similar but I like this one* (referring to Choropleth) *better because I know it better.*
1	BC	*This representation is more favourable than the differentiated ones* (referring to small multiples) *as you spend more time comparing maps. If you had to compare more than one region, it would be even more obvious.*
2	Tg	*My head is going to explode, I can no longer remember which regions I have mentioned and which I have not. These separate representations* (small multiples) *complicate the task a lot.*
2	Ch	*The choropleth is more complicated, despite the fact that the search pattern* (stable) *is simple.*
2	Tg-Ch	*In the case of small multiple maps* (choropleth and tile grid map) *the search is a bit random, you choose a particular region at t1 and check whether or not it has changed at t2 and t3.*
2	Tg-Ch	*We start with the regions that are most easily located* (clearer positions) *on the map. When you find one that meets the requirements, you pick another one at random, and so on.*
2	Tg	*The tile grid map makes the task more difficult than the choropleth because it is more difficult, even with the numbers, to locate the internal regions.*
2	Ch	*It is slightly better than the tile grid map.*
2	BC-BT	*It is much easier to reach a conclusion with the banded versions, all the information is in the same place, and you don't waste time comparing maps.*
2	BC-BT	*As I said in the previous task, the difference between the banded map and the small multiple map would be greater when looking for more difficult patterns rather than stable ones* (comparing both banded versions).
2	Ch	*It is easier to visually locate the regions with a characteristic shape (with choropleths) or in an easy-to-locate position (with grid maps).* *Positions at boundaries are easier to find than interior ones.*
	Tg	*External regions are easier to find than the interior ones, where you have to consult the labels of some of the regions to make sure it is the correct one.*
2	Ch–Tg / BT-BC	*Have I repeated this* (region) *already? The task is* *much more convenient and easier to solve in the case of the banded maps.*

2	Ch–Tg	*You are so focused on reviewing regions and jumping from map to map, that you can no longer remember if you are repeating any of the mentioned regions.*
2	BC-BT	*The exploration is quite easy with both representations, at least when comparing with the small multiple maps. They need more visual paths between time steps or maps.*
2	Ch-Tg	*Sometimes the comparison can be bidirectional: you choose a region at random in t3 and check that it has the same colour in t2.*
2	BT	*It is very easy to solve the task with this representation. From my point of view, the most favourable.*
3	Ch	*It is very easy to solve given its familiarity, for me, it is the simplest and the one that works the best.*
3	Tg	*In this case, it is not as confusing as the choropleth, as the areas of the regions are equal.*
3	Tg / BT-BC	*It is the easiest one* (referring to Tile grid map), *the banded versions are useless when identifying the trend.*
3	Tg	*I think it is ... stable? Or increase slightly? Anyway, it is the clearest for me, as the areas have the same size.*
3	Tg	*I think it increases slightly. It seems to me the clearest representation, as the areas of different sizes do not differ.*
3	BC-BT	*Impossible to tell in the case of these* (referring to the banded versions). *Maybe the second one is a bit clearer?* (referring to the banded tile grid map)... *but... anyway... I have preferred to abandon this task, than to say something random.*
3	Ch	*It is by far the simplest and the one I know best.*
3	BC-BT	*I have no idea. I would be unable to identify the trend. The only solution would be to count the sub-regions for the time steps, and I could spend the whole day doing that.*
3	Tg / Ch	*I think it favours the task, but I prefer the choropleths because I am more familiar with them.*
3	BC	*I could be analysing the map and I wouldn't come to any conclusions. My answer would be random.*
3	Tg	*I would say that the pattern is stable, but I am not very sure.*
3	BT-BC	*I couldn't tell you which is the worst* (referring to both banded types), *I think the banded choropleth, but both are completely useless for solving it.*
3	BT	*I think the pattern remains stable, but I am not sure.*
3	BC-BT	*If I had to answer, it would be random.*
3	BC-BT	*The graph has too much information, it is impossible to determine the trend.*
3	BC-BT	*This is a mesh, it's impossible to know the answer without checking each of the regions one by one.*

3	Ch	*Without doubt, the representation that best favours decision-making is the choropleth. It is a very familiar representation to me.*
3	Tg	*I think the pattern remains stable, but I am not sure.*
3	Ch	*The choropleth is the clearest one to show the global trend. Wait! No! If it is related to the number of regions, I think that the trend is not increasing.*
3	BC-BT	*Did any of the participants solve the task for these two maps? I couldn't anyway. I would say it is impossible just by looking at them.*

Table 6: Task statement, success and error definitions, possible answers and time measurement description for Task-based questions (quantitative evaluation).

Task based questions *(Quantitative Evaluation)*				
Task	Success	Error/failure	Possible answers	Time measurement
1: Determine the trend for a specific region over time. *Statement: Could you describe the evolution of COVID-19 new cases for region number n (region name) over time (t1, t2, t3)?*	The user correctly identifies the trend.	–The user does not correctly identify the trend. –The user solves the task needing more than 2 minutes.	– Increase – Decrease – Stable – Stable then decrease *(or increase)* – Increase *(or decrease)* then stable – Increase then decrease – Decrease then increase. **(See also Fig. 2 in this Annex)**	The time is measured from the time the visualisation is presented, to the time the user starts to talk.
1b: Can you identify the exact values in each time step?	This task is required (out of time) to check the validity of the previous answer. (See Fig.2 in this Annex I)		Values corresponding to each time step (t1,t2,t3): – 'Low' – 'Medium' – 'High' – 'Very High' *i.e:* *'High',* *'Medium','Low'*	*No time measured*
2: Identify the regions that follow	The user correctly identifies five	–The user identifies less than five	Identify 5 (of the 9 available regions) with	Time measured from the

a specific pattern. Statement: Could you name five regions that have remained stable in terms of the number of new cases throughout these three periods of time?	regions with a stable trend (same colour for the three time steps –t1,t2,t3–)	regions, or identifies the wrong regions that do not follow the stable trend. –The user solves/finishes the task needing more than 2 minutes.	this pattern. *Region numbers are between [1,42]. i.e:12, 25, 3, 6, 34* *A stable pattern involves the same colour for all three time steps (t1,t2,t3).*	presentation of the map, until the user gives the fifth (non-repeated) region in the map (correct or incorrect).
3: Determine the global trend (all regions) over time. *Statement: Could you determine the global trend for all Catalonia (over time) as increasing, decreasing or stable?* *You can abandon the task if you consider it is impossible to solve through the visualisation or your answer would be random.*	The user correctly identifies the trend (increasing, decreasing, stable).	–The user incorrectly identifies the trend. –The user identifies the trend, needing more than 2 minutes. –The user abandons the task.	– Increase – Decrease – Stable – I abandon the task. **The pattern is always **stable** (same number of regions for each value), but playing with different random regions (using the biggest or smallest areas) to perform the transition. (See Fig. 3 of this Annex)*	The time is measured from the time the visualisation is presented, to the time the user gives the answer.

Table 7: List of the 42 regions of Catalonia with the area in Km2

Region	***Area (km²)***
20 Big regions	
Noguera	1784.10
Alt Urgell	1,447.50
Segrià	1396.40
Pallars Sobirà	1378.00
Alt Empordà	1357.40
Pallars Jussà	1343.20

Osona	1245.10
Berguedà	1185.30
Bages	1092.20
Baix Ebre	1002.60
Solsonès	1001.10
Selva	994.9
Ripollès	956.6
Anoia	866.3
Ribera d'Ebre	827.1
Garrigues	797.7
Terra Alta	743.3
Montsià	735.5
Vallès Oriental	734.5
Garrotxa	734.5
2 Medium regions	
Segarra	722.8
Baix Empordà	701.8
20 Small regions	
Baix Camp	697.3
Conca de Barberà	650.2
Aran	633.5
Alt Penedès	592.7
Vallès Occidental	583.1
Urgell	579.6
Gironès	575.6
Cerdanya	546.6
Alt Camp	538.2
Priorat	498.7
Baix Llobregat	486.2
Alta Ribagorça	426.8
Maresme	398.6
Moianès	337.9

Tarragonès	319.2
Pla d'Urgell	305.2
Baix Penedès	296.4
Pla de l'Estany	262.8
Garraf	185.1
Barcelonès	145.8

Table 8: List of the most relevant reports reviewed, time steps used with the static maps presented, together with the area of coverage.

Subject	***n Steps***	***Area - Region***
Development / Economics	2	EU
Global trends 2030 (NIC)	2	Global
Economy / Unemployment	2	Country
Unemployment	2-3	Country
Assessing green gentrification	2	City
Gentrification	7	(Article) City
Public policies	3	City
Climate Trends 2020	2-3	EU
Temperatures	2-4	EU
Precipitations	3	EU
Lake surface temperatures	4	EU
Atmospheric circulation	4	EU
Late spring frost	2-3	EU
Mediterranean summer extremes	2-3	EU
Floodings	2	EU
Climate change and agriculture	2-3-4	Country - USA
Climate change and social vulnerability	2	Country - USA
Inequality	4	Country
Evolution humanitarian mapping	2	Global
Uganda poverty index	2-3	Country
Territorial scenarios Baltic region (population)	2	Group of countries

COVID-19 America	3	Continent
COVID-19 cases Madagascar	2	Country
World social protection	12	Global
Trade commercial interaction	2	Groups of countries
Air traffic	2	City
Migration central-eastern europe	2	Groups of countries

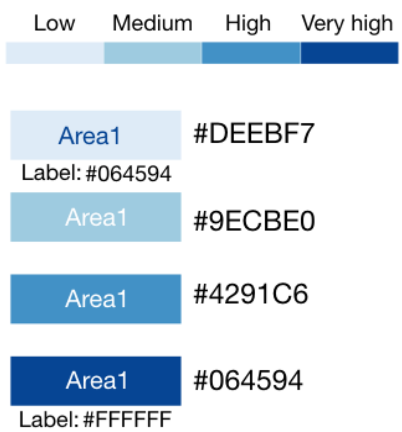


Figure 1: The colour values presented in the different regions/chorograms were: '#deebf7', '#9ecbe0', '#4291c6', '#064594' from lowest to highest. Label colours were presented in white ('#ffffff') for regions with darker backgrounds ('#9ecbe0', '#4291c6', '#064594'), and dark blue ('#064594') for regions with the lightest background ('#deeb7f') in order to guarantee sufficient contrast and optimum readability.

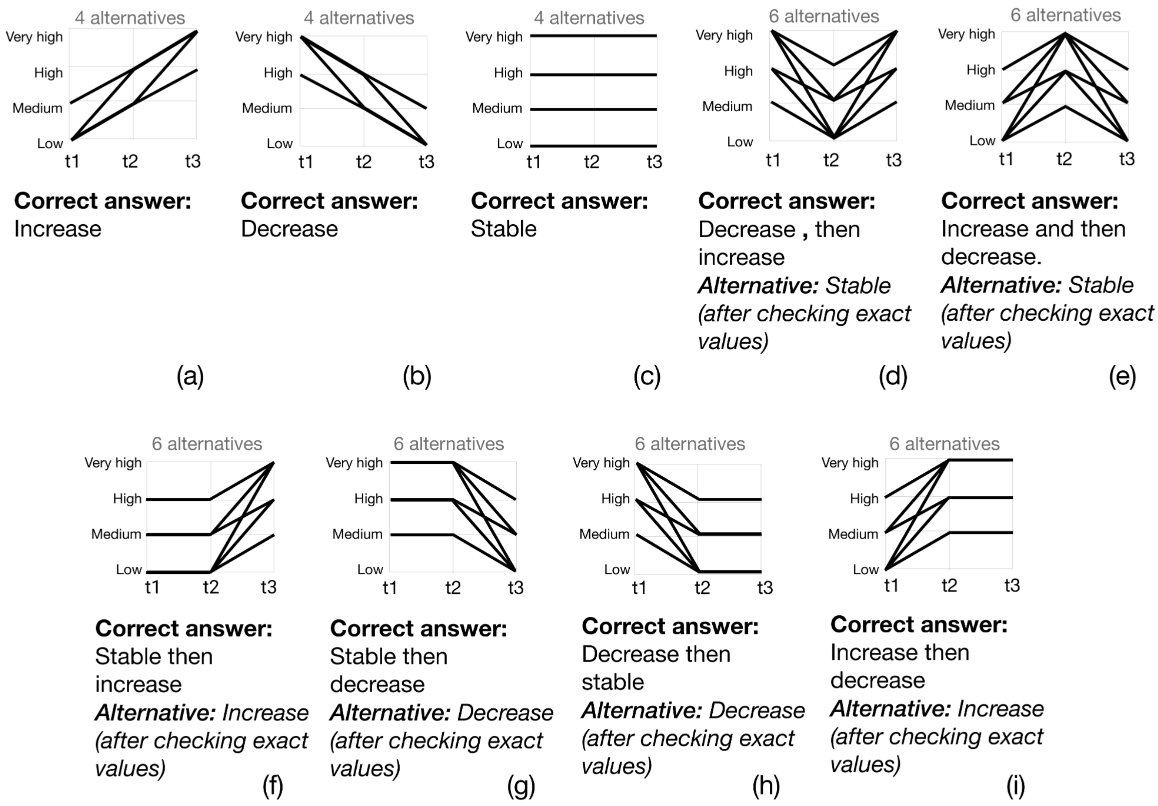


Figure 2: Possible random values for the selected region in Task 1 and the valid answers in each of possible random values. In the case of [d-i] scenarios, the user can answer in a global manner so checking the exact values for each time step (t1,t2,t3) is needed.

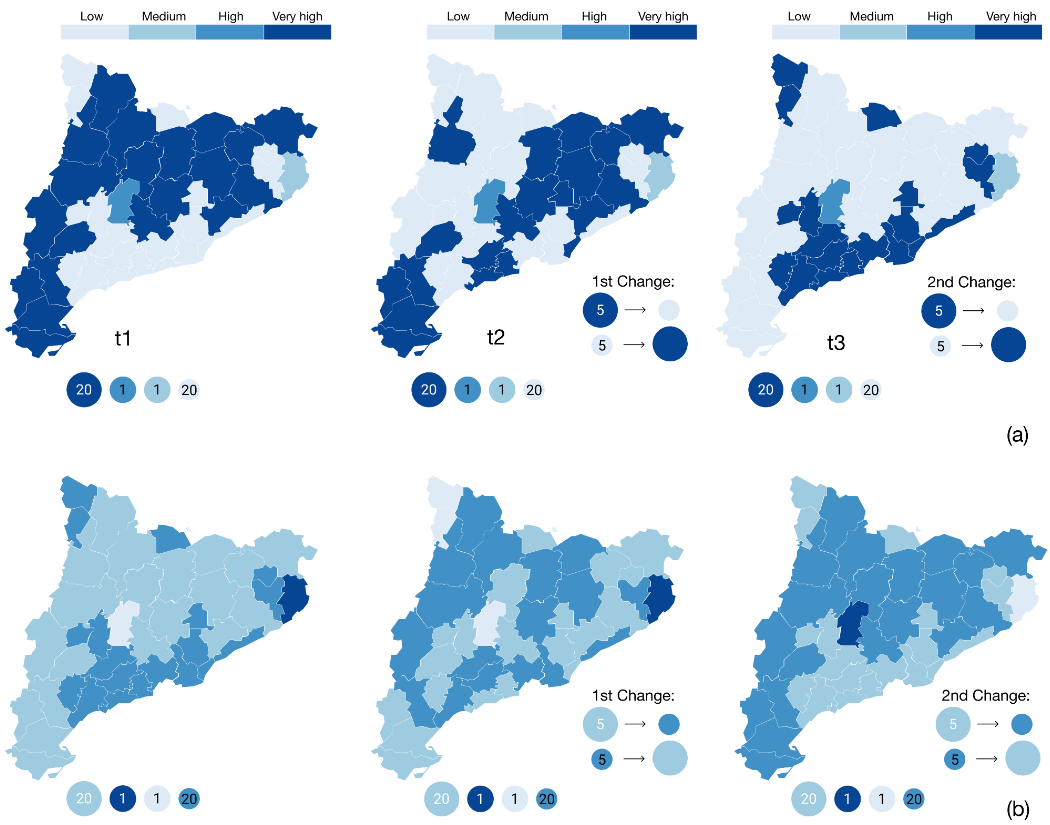


Figure 3: Exemplary random distribution for small multiple Choropleth. The transitions randomly swap blocks of five regions, selecting between the largest and smallest areas to accentuate the area-bias of Choropleth maps. The medium-area regions remain stable or interchange the values. However, the number of regions for each category is always the same, to maintain a pattern of stability over time (t1,t2,t3). In the first case (a) the pattern seems to be decreasing, and in the second (b) increasing, however in both cases it is stable.

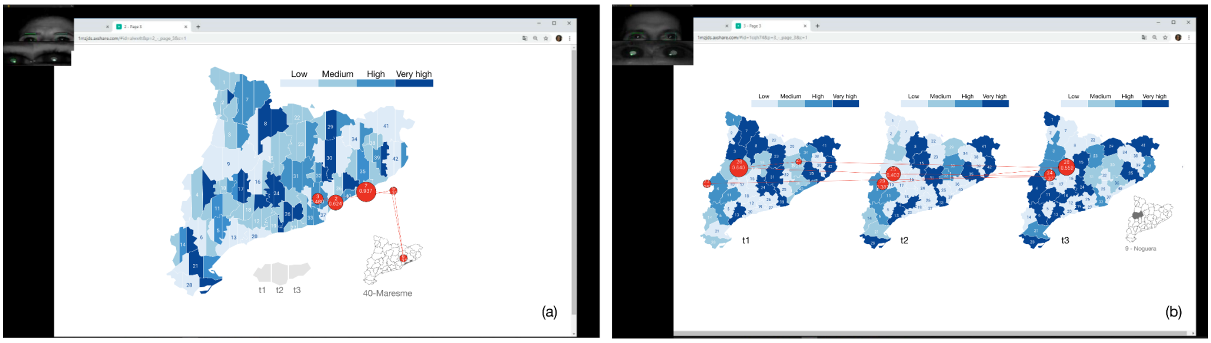


Figure 4: Two different scatter patterns for Task 1. In the case of banded choropleth (a) the visual path is concentrated in the region of interest. In the case of small multiple choropleth the visual pattern jumps from map to map making the comparison more difficult.

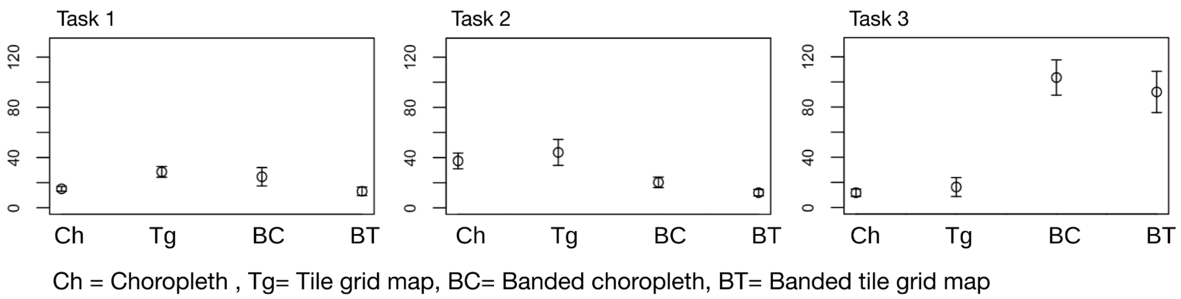


Figure 5: Plot of confidence intervals for the three tasks and the four maps.

9.2. Annex II

A. Detailed Eurobarometer description

Eurobarometer is a series of open datasets published by the European Union on public opinion on different topics of public interest. Specifically, our dataset referred to the opinions of EU citizens on environmental issues collected in 2017. The data is presented in Excel format (.xls), with one file for each member state. Each file contains several tabs, with information related to a specific question on the environment. There are forty-one survey questions (or tabs) in total, showing the total rates for each country but also broken down into thirty-six different socio-demographic groups (see Table A.3 of the Annex II).

These socio-demographic groups are based on certain characteristics of the population (gender, age, family situation, employment situation, place of residence, etc.).

We are therefore faced with multivariate/multidimensional data (answers to several questions) which is also multi-categorical. The categorization aspect in this case is double: On the one hand the thirty-six categories themselves, and on the other hand, the five possible answers based on two Likert scale possible answers: ('Very important', 'Fairly important', 'Not too important', 'Not at all important', 'Don't know' or in other cases with 'Strongly agree', 'Quite agree', 'Disagree', 'Totally disagree' and 'Don't know'.

The 'Don't know' category (which some users considered the most negative and others considered neutral) retains its negative character, and is located in the centre of the scale, with medium opacity.
Some of the questions posed, referred to the level of importance given to the environment by the respondents, or which aspects were of greatest interest or concern to them ('endangered species and ecosystems', 'floods and droughts', 'pollution' or 'climate change'). In turn, respondents were asked about the actions they carry out to cooperate in conserving the environment and with what frequency, the effects on health, and the level of involvement that private companies, institutions, governments and even the European Union itself should have.

The sampling chosen for the Eurobarometer surveys was based on a random selection of points according to the stratification of the resident national population in terms of metropolitan, urban and rural areas, and proportional to the size of the population (to guarantee total coverage of the country) and its population density. (Skaarhoj, 1998).

The Eurobarometer survey answers were presented in two languages: French and English.

The variables are represented in table A1. We group them by category according to their area of scope (citizen empowerment, impact on health and environment, responsibility and EU policy obligations). Information about the type of answer is also included (See Table A.1). Some other information, distributed in different tabs referred to the lack of actions by citizens or inconsistencies between concerns and actions: For example, some respondents claimed to be concerned about specific environmental issues (pollution, drinking water, plastic production, climate change, etc.), but, in spite of these claims, did not carry out any corrective actions to improve the situation relative to what they were concerned about. They were classified by potential users in ten categories that related concerns expressed and actions with their impact in each category. This information is included in Table A.1 and A.2.

Table A.1 Variables included

Empowerment	Type of answer (Very important - Not at all important)
How important is protecting the environment to you personally?	
Impact	Type of answer (Totally agree - Totally disagree)
– As an individual, you can play a role in protecting the environment in your country. – Environmental issues have a direct effect on your daily life and health. – You are worried about the impact of everyday products made of plastic / chemicals on your health. – You are worried about the impact of everyday products made of plastic / chemicals on the environment.	
Responsibility	Type of answer (Totally agree - Totally disagree)
– The big polluters should be mainly responsible for making good the environmental damage they cause. – In your opinion, is each of the following currently doing too much, about the right amount, or not enough to protect the environment? a) Big companies and industry b) Citizens themselves c) Your region /area d) Your country / National government e) The EU	
EU Policy Obligations	Type of answer (Totally agree - Totally disagree)
– EU environmental legislation is necessary for protecting the environment in your country – The EU should be able to check that EU environmental laws are being applied correctly in your country – The EU should assist non-EU countries to improve their environmental standards	

Table A.2 Information to define inconsistencies

Concerns & Interests
(1) Decline or extinction of species and habitats, and of natural ecosystems (forests, fertile soils) (2) Shortage of drinking water (3) Frequent droughts or floods (4) Pollution of rivers, lakes and groundwater – Marine pollution (5) Air pollution (6) Noise pollution (7) Climate change (8) Agricultural pollution (use of pesticides, fertilisers, etc.) and soil degradation (9) Plastic production (10) Use of chemicals
Actions performed
– Choose a more environmentally-friendly way of travelling (walking, cycling, public transport, electric car) (Related to 5,6,7,3) – Avoid buying over-packaged products (9,4) – Avoid single-use plastic goods other than plastic bags (e.g. plastic cutlery, cups, plates, etc.) or buy reusable plastic products (9,4,1) – Separate most of your waste for recycling (1,4) – Cut down your water consumption (2) – Cut down your energy consumption (e.g. by turning down air conditioning or heating, not leaving appliances on stand-by, buying energy-efficient appliances) (7,5,3) – Buy products marked with an environmental label (10,8) – Buy local products (Related to 7,1,3) – Use your car less by avoiding unnecessary trips, working from home (teleworking), etc. (7,3)
Level of activity Many - Some - Few - One - None (actions)
Information channels
– National newspapers – Regional or local newspapers – Magazines – TV News – Radio – Films and documentaries on television – Family, friends, neighbours or colleagues – Books or scientific publications – Brochures or information materials – Events (conferences, fairs, exhibitions, festivals, etc.) – Museums, national or regional parks – Online social networks – The Internet (other websites, blogs, forums, etc.)

Table A.3: Social categories with socio-demographic groups selected from the Eurobarometer

Socio-demographic groups
Gender: Men / Women
Age: 15-24 / 25-39 / 40-55 / 55+
Socio-professional category: Self-employed / Managers / White collar workers / Manual workers / House persons / Unemployed / Retired / Student
Marital Status: Married / Single living with a partner / Single / Divorced / Widowed
Difficulty paying bills: Always / Frome time to time / Never
Considered as belonging to: Working class / Lower-middle class / Middle class / Upper-middle class / Upper class
Living in: Rural village / Small town / Large town
Politics: Left / Centre / Right
Internet use: Everyday / Often / Never

B – Complete statements of the test tasks

Table B.1: User testing tasks (Evaluation Phase II with the broader public).

Task 1: Identify two countries on the map which are less committed to the environment.
Task 2: Which are the three least committed socio-demographic groups for Romania with regards to the environment?
Task 3: Imagine that we want to compare the level of commitment for the 'Students' socio-demographic group for the countries around Romania. Which country has the least committed students? And the most committed?
Task 4: Can you explain what the dots around the glyphs are? Can you explain the information available for the socio-demographic group 'Widowed' in Austria?
Task 5: Find the way to access the Summary View? Which are the least committed socio-demographic groups-countries in Europe?

C – NASA-TLX description

NASA-TLX questionnaire categories (Cao et al., 2009a), includes the following aspects related to perceived difficulty and cognitive resource use:

- Mental Demand or level of concentration required.
- Physical Demand *(not applicable for the type of tasks we are dealing with in this study).*
- Temporal Demand *(NASA-TLX).*
- Perceived Effort *(NASA-TLX) or perceived difficulty while performing the task.*
- Frustration Level *(NASA-TLX).*

C.1 – NASA-TLX based questionnaire results

Below, we present the opinion results for the categories of 'Mental demand', 'Temporal demand', 'Perceived effort' and 'Frustration level'.

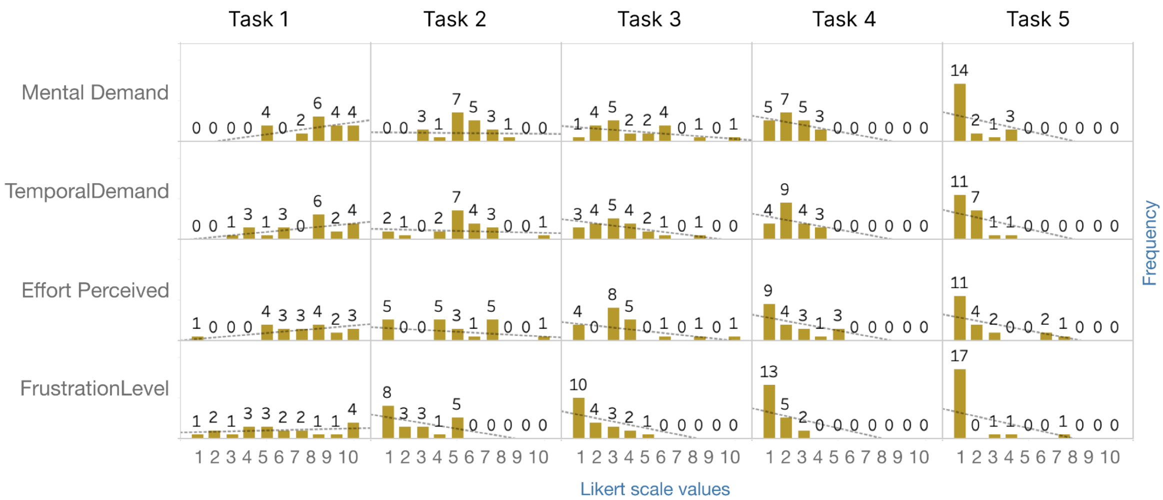


Fig. C.1: Frequency answers to NASA-TLX questionnaire

Using histograms we can analyse the frequencies through which users evaluated the different aspects of each task (see Fig. C.1). We have ignored category 2 ('Physical demand'), as it was not applicable due to the nature of tasks which did not involve any significant physical effort.

For Task 1, 'Mental demand' and 'Temporal demand' were the categories with the highest scores with most being valued between 6 and 10. The 'Perceived effort' was universally considered to be the most taxing of all the tasks (receiving scores between 5 and 8). For 'Frustration level', opinions were spread out over the evaluation spectrum with 10 (very frustrating) being the most voted value with 4 votes.

Task 2 obtained middle-of-the-road values for 'Mental Demand' and 'Temporal Demand' with 7 votes for level 5 in both categories. However, 'Perceived effort' received moderate values, where levels from 1 to 6, were the most voted. Lastly, 'Frustration level' was considered low, 1 being the most voted level (with 8 votes) (See Fig. C.1 column 2).

Task 3 presented more defined 'shapes' in the distribution, having low-medium values for the categories of 'Mental demand' and 'Temporal demand', but high values for 'Perceived effort' and 'Frustration Level '(see Fig. C.1 column 3) received low values which indicates that both were considered easy tasks to perform and with a low perceived effort.

Tasks 4 and 5 repeated the pattern of the previous tasks, but with even lower values for the categories of 'Mental demand', 'Time demand', 'Perceived effort' and 'Frustration level'.
Task 1 obtained high values in 'Mental Demand' and 'Temporal Demand' and moderate values in 'Perceived Effort' and 'Frustration level'. Tasks 2 and 3 presented moderate values in 'Mental Effort', 'Temporal Demand' and 'Perceived effort'. However, the values (level of confidence) were high and 'Frustration level' presented low values.

In line with expectations, tasks 4 and 5 presented low values for 'Mental Demand', 'Temporal Demand', 'Perceived Effort' and 'Frustration Level' (See Fig. C.1).

D – Bipolar Laddering results

Table D.1: Positive and negative comments gathered during the Bipolar Laddering test. For each aspect we show the number of times an aspect was mentioned (frequency) and the average value it received.

Positive aspects		
Description	Freq.	Avg.
Usefulness of visualisation customization for the different views and depending on the task or interest	12	9.7
Presentation of inconsistencies is very useful information, saves time and helps to reach conclusions	11	9.1
Comparison of socio-demographic groups is useful and intuitive	8	9.4
Interactions / Exploration are intuitive and clear: the user knows how to interact at each moment	7	9.4
Information density and amount of insights gathered are presented in a clear and familiar way	6	7.3
Contrast with dark background is very clear	5	8.4
Summary view is very useful as a complement to the general overview	5	8.4
Simplicity of the representations	5	8.2
Utility of charts, step-by-step information and level of deep analysis	2	8
Totals by region/category useful for checking some general insights of the area/region under analysis	1	7
Global view represented as concentric circles gives a lot of information and works well for problems of overlapping	1	8
TOTAL	63	8.4
Negative aspects		
Description	Freq.	Avg.
Complexity of the global visualisation using concentric circles summarising all opinions for all socio-demographic groups	7	8.2
Inverted scale (darker colours for lower values) seems strange until you understand the user requirements.	4	5.2
Vague legend descriptions in the initial representations: Information about the diameter size of the concentric circles is needed.	6	4.4
Showing all labels for all socio-demographic groups simultaneously	3	4.6
The sorting by negative aspects seems strange until you understand the importance of highlighting the less committed groups	2	4.5
The dots to indicate inconsistencies in one socio-demographic group are too visible	2	2.5

Some labels in the summary view must be relocated to a different place.	1	4
Radial plots are complicated at first sight	1	4
Comparison of the status of socio-demographic groups is not intuitive	1	6
Help or more detailed information is needed at some points/sections	1	3
Not enough contrast for some figures	1	4
TOTAL	29	4.6

10.Publications

BAMS Article

Users' Cognitive Load

A Key Aspect to Successfully Communicate Visual Climate Information

Luz Calvo, Isadora Christel, Marta Terrado, Fernando Cucchietti,
and Mario Pérez-Montoro

ABSTRACT: The visual communication of climate information is one of the cornerstones of climate services. It often requires the translation of multidimensional data to visual channels by combining colors, distances, angles, and glyph sizes. However, visualizations including too many layers of complexity can hinder decision-making processes by limiting the cognitive capacity of users, therefore affecting their attention, recognition, and working memory. Methodologies grounded on the fields of user-centered design, user interaction, and cognitive psychology, which are based on the needs of the users, have a lot to contribute to the climate data visualization field. Here, we apply these methodologies to the redesign of an existing climate service tool tailored to the wind energy sector. We quantify the effect of the redesign on the users' experience performing typical daily tasks, using both quantitative and qualitative indicators that include response time, success ratios, eye-tracking measures, user perceived effort, and comments, among others. Changes in the visual encoding of uncertainty and the use of interactive elements in the redesigned tool reduced the users' response time by half, significantly improved success ratios, and eased decision-making by filtering nonrelevant information. Our results show that the application of user-centered design, interaction, and cognitive aspects to the design of climate information visualizations reduces the cognitive load of users during tasks performance, thus improving user experience. These aspects are key to successfully communicating climate information in a clearer and more accessible way, making it more understandable for both technical and nontechnical audiences.

KEYWORDS: Forecasting; Seasonal forecasting; Decision support; Software; Model interpretation and visualization

https://doi.org/10.1175/BAMS-D-20-0166.1
Corresponding author: Luz Calvo, luz.calvo@bsc.es
In final form 8 July 2021

AFFILIATIONS: Calvo, Christel, Terrado, and Cucchietti—Barcelona Supercomputing Center, Barcelona, Spain; **Pérez-Montoro**—Faculty of Information and Audiovisual Media, Universitat de Barcelona, Barcelona, Spain

The accessibility to climate information has implications for how society makes the best use of scientific knowledge to adapt to climate change (Harold et al. 2016). For years, climate service providers have faced the challenge of how to best communicate climate-related data and information, together with their inherent uncertainty, in an easy and understandable way for both expert and nonexpert users (Kaye et al. 2012; Lorenz et al. 2015). This has resulted in service providers often using visual representations to supply climate knowledge tailored for users' decision-making (Gerst et al. 2020; Taylor et al. 2015).

As a standardized mapping approach to represent climate uncertainty is lacking, different techniques have been applied for this purpose, often resulting in users spending more time trying to understand the mapping approach than focusing on the interpretation of the presented information itself (Kaye et al. 2012). In general, users' familiarity with a type of data visualization has been found to play a significant role in the process of reading and making sense of maps and graphs (Lorenz et al. 2015). Commonly used representations of climate data include choropleths, heat maps, and line charts (Taylor et al. 2015). Although familiar to users, these elements have limitations when used to communicate climate data to decision-makers in a way that is transparent, understandable, and that does not lead to a false sense of certainty. An example of such limitations is seen when communicating climate predictions from the next two weeks up to a few decades into the future. A characteristic of such predictions is that they are probabilistic, meaning they provide information on the probability of a certain climate outcome to occur (e.g., winds below or above a threshold, not optimal for the energy production). In addition, climate predictions are often given in the form of large amounts of data covering the whole globe, and their quality (i.e., level of success of a prediction against observationally based information) depends on the specific location and time (Kaye et al. 2012). Both aspects, probabilities and forecasts quality (referred to as skill by the climate science community), add complexity to the visual communication and can eventually compromise the understanding of climate predictions by users (Bonneau et al. 2014; Terrado et al. 2019).

Visualizations of complex climate data tend to prioritize solutions that take into account the greatest combination of variables and dimensions, e.g., color, size, distance or brightness of the glyphs (the graphic symbols used to represent a value). This complexity calls for formats paying special attention to visual encoding, which encompasses the translation of multidimensional data into visual elements on a chart or map representation. Visual encoding is useful in the sense that it allows to convey a higher amount of information in a single visualization (Grainger et al. 2016; Lloyd 1997). However, it rarely considers if the information needs to be displayed all at a time, with a certain visual aesthetic, or if it will be too complex for its proper interpretation (Cleveland and McGill 1985; Sager et al. 2007). Indeed, aesthetics might be worth considering if striving to create something memorable, that helps to raise awareness about a specific scientific challenge (Borkin et al. 2013). However, an attractive image cannot qualify as effective unless it accurately conveys something meaningful or credible (Holmes 1984; Kosara 2013).

According to Stephens et al. (2012), for a visualization to be effective, it is important to consider a balance between density (the amount of data represented), robustness (the representation of scientific confidence and consensus), and saliency (the relevance of the information to user needs). Although visualizing climate forecast uncertainties and associated probabilities has been thought to increase users' trust (Joslyn and LeClerc 2011;

Roulston et al. 2006), it does not automatically lead to better decisions (Greis et al. 2015). This is especially critical when the visual elements used to represent uncertainty compete with the limited cognitive resources of the users (i.e., their attention, recognition, and working memory) (Antifakos et al. 2004; Riveiro et al. 2014; Davis et al. 2020). Visualizations with a heavy cognitive load, that is, involving a high amount of working memory resources, can have negative effects on users' understanding and learning and can impact their ability to complete a task or make an informed decision (Cairo 2012; Few 2009; McInerny et al. 2014). In such cases, the use of interactive elements can offer mechanisms for progressively dosing the information that is to be shown (Veras Guimarães 2019; Bertin 2010; Ware et al. 2002; Yoghourdjian et al. 2018).

Methodologies grounded on the fields of user-centered design (UCD), user interaction, and cognitive psychology have a lot to contribute to the climate data visualization field (Christel et al. 2018; Bevington et al. 2019). UCD techniques involve users throughout the design process in order to create highly usable visualization tools based on their needs (Davis et al. 2020; Dong et al. 2008; Yucong et al. 2019). User interaction makes use of interactive elements to let users decide what to see, when, or to show only those values that are relevant for a given task. This facilitates decision-making, returns control to the users, and allows them to discard the nonrelevant information at each moment (Lau and Vande Moere 2007; Gerharz and Pebesma 2009; Ware 2012). Within the framework of UCD methodologies, technologies such as the eye-tracker have been used to quantify and analyze visual patterns, attention, and cognitive aspects. Adding cognition and perception (i.e., processes about how humans acquire knowledge, understanding and interpretation) can help detect and solve initial design problems in UCD. Rather than just favoring visual exploration, these disciplines offer increasingly effective methods to develop and evaluate visualization systems that explicitly consider real-world user requirements (Block 2013).

Tools and learnings from the UCD field have already been applied to the visualization of climate information and climate services. For instance, Argyle et al. (2017) showed how incorporating usability evaluation into the design of decision support tools can improve the efficiency, effectiveness, and user experience of a weather forecasting application. Other studies have similarly applied UCD to inform the design of climate information websites, apps, and prototypes (Ling et al. 2015; Oakley and Daudert 2016; Khamaj et al. 2019). Design elements have also been introduced in the development of climate services to increase their usability, in particular for the renewable energy sector (Christel et al. 2018). On the other hand, cognitive and psychological sciences have been applied to the visualization of climate data. Some examples are the use of cognitive psychology methods to help make information provided by IPCC graphs more accessible to expert and nonexpert audiences (Harold et al. 2016) and improve users' task performance (Hegarty et al. 2010). Differences in the interpretation of climate graphs between experienced and nonexperienced users have been explored elsewhere (Atkins and Mcneal 2018; Gerst et al. 2020), both for climate change variables and for temperature and precipitation outlooks.

The aim of this paper is to provide quantitative and qualitative evidence of how the use of user-centered design, visualization techniques, and interaction elements can reduce the cognitive load of nonexpert users during tasks' performance. We compare two different map visualizations of uncertainty. First, the one used in Project Ukko, a climate service tool prototype that provided climate predictions tailored to the energy sector (http://project-ukko.net/). The second visualization was a redesigned version of Project Ukko, using a simplified visual encoding of uncertainty and the use of interactive elements. We quantified the effects of the redesign on the users' experience performing typical daily tasks, using indicators that include response time, success ratios, eye-tracking measures, and user perceived effort and comments, among others. We show that involving experts from different disciplines [climate experts, user experience (UX) and visualization experts, communicators] in the co-production process and simplifying the visual encoding used in the visualization, has an impact on the users' cognitive load, favoring response times and confidence in decision-making.

Study context: Previous work with experts

Project Ukko introduced design to explore new forms of representation for wind predictions, which was considered a groundbreaking step in the visualization of climate services (Christel et al. 2016). The representation of uncertainty in Project Ukko was achieved through multiple visual resources, namely, glyph thickness representing the intensity (speed) of predicted wind, glyph inclination representing the predicted change in wind conditions (i.e., higher-than-normal, normal, or lower-than-normal wind), and opacity representing the quality of the prediction (skill).

After the development of Project Ukko, a user test was carried out in order to detect functional or usability problems. The test identified a series of conflicting aspects in the tool, mainly related to visual encoding (Makri 2015) and flexibility of use, that needed improvement. These aspects included the use of too many categories for each represented variable, the difficulties to detect color hue associated with narrow glyphs, and the excess of information in nonrelevant areas of the map.

These findings indicated that the development of a new visualization was needed in order to solve the problems detected. For informing this new redesign of Project Ukko, we started by organizing a workshop to gather expert user needs and limitations experienced while using the tool. We also conducted four interviews with target users, including climate experts and operators and managers of wind power plants, to obtain specific behavioral information related to their daily work activities. As a result, some important requirements were identified, e.g., that users often based their decisions on simple metrics or threshold values predefined by their companies. This allowed the redesign of a new version of the tool: the S2S4E Decision Support Tool (S2S4E 2020), an operational climate service which integrates subseasonal to seasonal climate predictions for renewable energy production. In S2S4E, a series of changes were applied to the initial Project Ukko visualization following well-established visual recommendations (Tufte 2001; Few 2009; Yoghourdjian et al. 2018; Veras Guimarães 2019). Some interactive elements were also incorporated to allow users to filter the nonrelevant information, thus reinforcing their cognitive abilities during daily work activities (Gerharz and Pebesma 2009). Specifically, the modifications include (see comparison in Fig. 1): (i) hiding values that do not meet the minimum guarantees of quality (prediction skill); (ii) simplifying the visual encoding for some variables (intensity and predicted change); and (iii) reducing the visual noise by using colors that are similar to the background color for nonrelevant predicted changes, i.e., middle category, showing conditions that can be considered "normal" or close to the average historical observations for a particular region (Chun 2017).

Methodology

To make a comparison and validation from the point of view of perception and cognitive load, we assessed if the changes applied to the original tool (Project Ukko) to create the redesigned one (S2S4E) were fully effective for a general public and not only for an expert audience. To avoid interference in the comparison of cognitive load measurements between both visual representations, not all the improvements included in the S2S4E tools were shown in this study (e.g., color blindness palette, labeling improvement and customizable elements).

The experiment. We conducted an experiment with nonexperts to assess if users' tasks performance when using Project Ukko was improved after redesigning the tool taking into account user-centered design, visualization techniques, and interaction elements. A sample

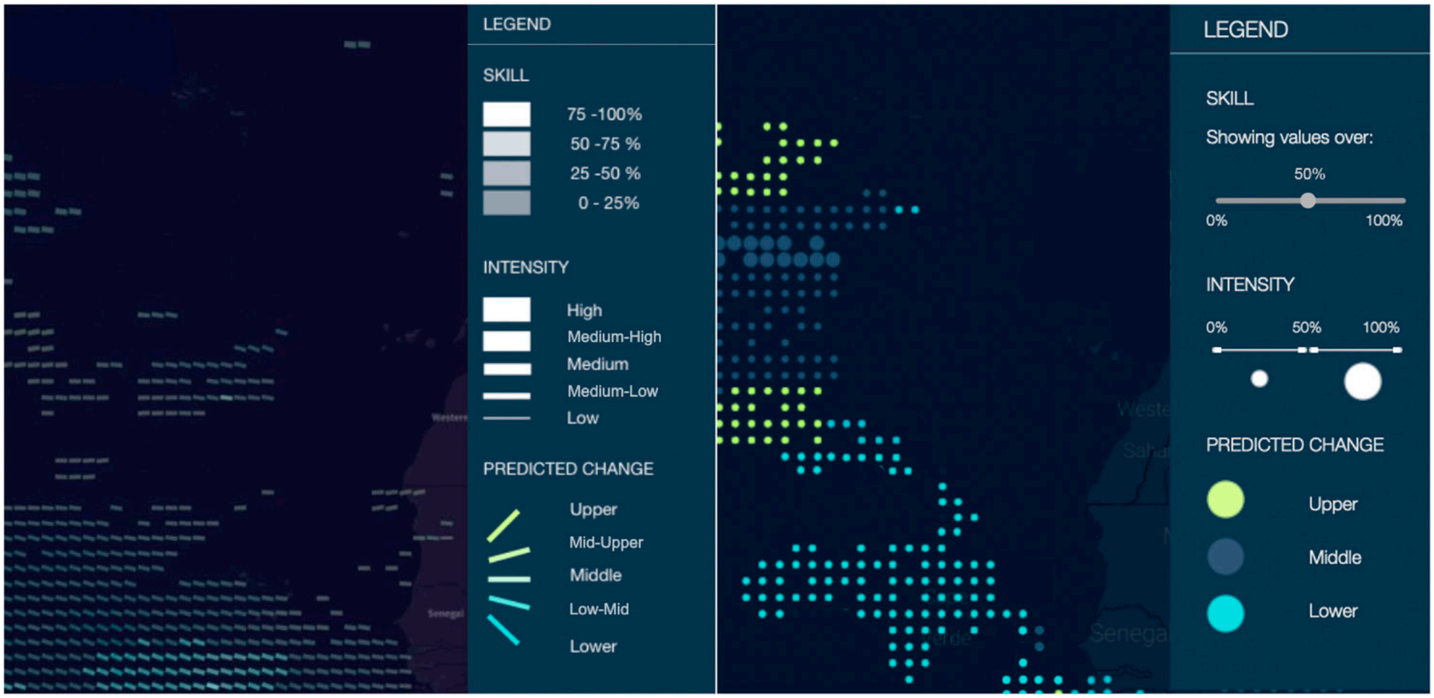


Fig. 1. (left) Project Ukko and (right) the redesigned tool based on user requirements research and visual encoding techniques. Applied changes: filtering out glyphs under a certain prediction skill threshold; simplification of the intensity representation from five to only two size; and simplification of the predicted change color scale from five to three colors, where a similar color to the background one is used for values close to the historical average and more distinguishable colors are used for values higher and lower than the historical average.

of 20 people was tested in order to guarantee 95% of detection of problems (Virzi 1992; Faulkner 2003). The sample was composed by 50% of men and 50% of women aged between 22 and 50, with a low-to-medium knowledge in data visualization, which entailed being familiar with common charts or basic maps but without expertise in visualization of climate or uncertainty information. Recruited participants were students and administrative staff in academia from the human resources, communication, and finance departments. Participants were asked not to have consumed exciting substances that could affect the test results.

The study consisted in two main analyses: (i) a quantitative analysis through the application of a user testing session with two tasks to be completed with both tools (Project Ukko and redesign), and (ii) a qualitative analysis of the positive versus the negative aspects for both tools through the application of a bipolar laddering pocket methodology (Pifarré and Tomico 2007). We also implemented a brief two-question quiz (Schrepp et al. 2017) to determine which visualization needed more mental effort, and which was the tool preferred by participants for decision-making processes.

The development of user testing sessions followed well-established recommendations of planning, moderation, and analysis (Nielsen 1993; Faulkner 2003). User testing sessions were moderated by an expert and recorded to measure time and responses afterward (Holtzblatt et al. 2004). Participants were asked to perform two tasks (task 1 and task 2 below) with each of the tools, based on two typical daily work activities of the intended users (Anderson et al. 2011; Block 2013).

Participants were provided with some context about the real-world conditions in which the tasks would take place. In addition, all the information and conditions necessary to carry out the tasks were presented in the statements, visualization, and captions shared with participants, for them to be able to perform the tasks without having an expert or climate science background. It was simply necessary to identify the requested areas or properties visually (Trivedi 2012). Task 1 was aimed to test if the tool allows good detection of areas with particular conditions while task 2 was aimed to test the differentiability of glyph representation at specific locations.

Task 1 statement: Locate or identify an area on the map that is appropriate to the location of a wind power plant. The area must meet a series of conditions: For Project Ukko, the suitable area should have a prediction skill over 50%, high or medium-high intensity, and an upper or mid-upper predicted change in wind speed (Fig. 1, left panel). For the redesigned version, the area should have a skill over 50%, intensity over 50%, and upper predicted change (Fig. 1, right panel). *[This task required users to be able to identify at least one of the areas that met the conditions specified.]*

Task 2 statement: Identify aloud the conditions that occur in the points included in the highlighted area on the map in terms of skill, intensity, and predicted change. *[This task required users to be able to identify the characteristics of two kinds of glyphs contained in a certain map area]*.

For both task 1 and task 2, we measured the cognitive load and task performance from a quantitative perspective using the following indicators:

1) Success rates when completing a task, including total or partial success (Freitas et al. 2002; Winckler et al. 2004; Ellis and Dix 2006).
2) Response time when completing a task.
3) Number of fixations (i.e., number of gaze points located very close in space, when the eyes are locked toward an object).
4) Fixation duration (i.e., period of time allocated to a fixation (Wang et al. 2014; Majooni et al. 2018).
5) Number of accesses to legend for the completeness of tasks (Pretorius et al. 2005; Klingner et al. 2008).

We also compared qualitative aspects in the use of both tools taking into account the level of user satisfaction and the decision-making facilitation using the following indicators:

6) Subjective effort perception (i.e., user perception on which tool was easier to use while performing the tasks (two-question quiz).
7) Preference between both tools regarding the decision-making process (two-question quiz).
8) Positive and negative aspects (bipolar laddering pocket methodology).

The number of fixations, the fixation duration, and the number of visual paths to legend (quantitative indicators 3, 4, and 5) were monitored using an eye-tracker (GP3 Eye tracker of gazepoint), a sensor-based device that measures eye positions (i.e., point of regard) or eye movement.

To investigate whether there were differences between tools or tasks, a chi-squared test was performed for indicator 1 and a two-way analysis of variance (ANOVA) was performed for indicators 2 to 5 using R software (R Core Team 2018).

During the performance of tasks 1 and 2, we promoted the thinking-aloud technique to identify and detect usability problems (Nielsen 1993; Olmsted-Hawala et al. 2010). Also, the sequence of tasks was randomly presented to the users to avoid biases derived from the learning acquired during the completion of the tasks (Li et al. 2013). We used the bipolar laddering pocket methodology (qualitative indicator 8), which is a reduced version of the bipolar laddering technique. It consists of asking the user about three positive and three negative aspects of both tools (instead of the 10 aspects requested in the extended modality). Then the users rate (from 1 to 10) the aspects mentioned regarding their importance (for the positive ones) or their severity (for the negative ones). By evaluating users' subjective opinions using a scale from 1 to 10 we are able to better quantify input results that are qualitative, which helps identify the key aspects to work with or prioritize the problems that need to be solved (Pifarré and Tomico 2007).

Results

Quantitative analysis of Project Ukko vs redesign. The assessment of success rates (indicator 1) indicated that the number of successfully completed tasks was significantly better when using the redesigned tool than when using Project Ukko, X^2 (9, N = 20) = 60.6, $p < 0.001$ (see Figs. 2a,b). In the case of Project Ukko, only 15.8% of the participants successfully completed task 1 and 21.1% task 2. In contrast, with the redesigned tool, task success was much higher, reaching 97.4% and 68.4% for task 1 and task 2, respectively. Although a higher proportion of failures and abandonments occurred during task 1, especially with Project Ukko, completion of task 2 showed various cases of partial success (i.e., identified just one of the two types of glyphs presented) for both tools, reaching 31.6% with the redesigned tool.

The average time to solve a task (indicator 2) was significantly lower for task 1 compared to task 2 ($p = 0.001$) and for the redesigned tool when compared to Project Ukko ($p < 0.001$) (see Figs. 2c,d).

The eye-tracker measurements of the number of fixations (indicator 3) was significantly lower for task 1 compared to task 2 ($p = 0.024$) but showed no significant differences between Project Ukko and the redesigned tool ($p = 0.234$) (see Figs. 2e,f). The fixation duration (indicator 4) did not show a significant difference between tasks ($p = 0.061$) or between both tools ($p = 0.651$) (see Figs. 2g,h). The number of accesses to legend (indicator 5) followed a similar trend as observed for the indicator of response time, with Project Ukko showing higher numbers than the redesigned tool ($p = 0.017$) and significantly lower values for task 1 when compared with task 2 ($p = 0.021$) (see Figs. 2i,j).

Qualitative analysis of project Ukko vs redesign. In terms of perceived effort (indicator 6), a much bigger effort was perceived for Project Ukko (89.9% of participants) than the redesigned

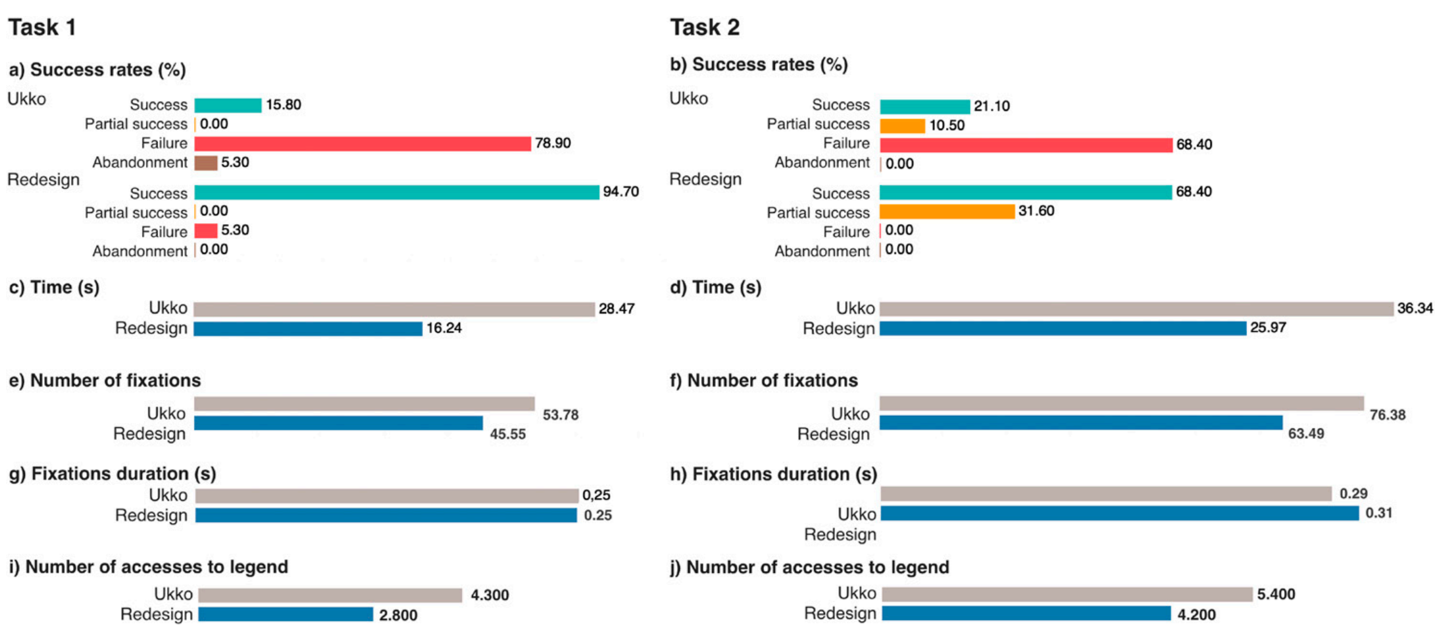


Fig. 2. (a),(b) Comparison of participants' success rates, (c),(d) average time to solve the tasks, (e),(f) number of fixations, (g),(h) fixations duration, and (i),(j) number of accesses to legend when using Project Ukko and the redesigned tool for tasks 1 and 2, respectively.

version (10.5% of participants). Likewise, regarding the tool preferred for decision-making (indicator 7), 89.50% of the participants stated that they would choose the redesigned tool as their working tool for daily tasks decision-making.

By analyzing the results of the bipolar laddering pocket technique (indicator 8) we found that the number of positive aspects mentioned by participants was just 15 for Project Ukko against 42 mentioned for the redesigned tool (see Fig. 3). Regarding the negative aspects, 32 and 12 aspects were pointed out for Project Ukko and the redesigned tool, respectively. On average, positive aspects were rated with an average score of 7.3 for Project Ukko and 8.75 for the redesign. Conversely, negative aspects were rated with a greater severity for Project Ukko, with an average score of 7.4, than for the redesigned tool, with an average score of 5.8.

Discussion

Considering user requirements when developing climate data visualizations is key to improve decision-making. Moreover, the simplification of a complex visualization through changes in visual encoding and interactivity often increases efficiency. Here we use quantitative and qualitative indicators to assess whether the redesign of the Project Ukko tool, taking into account user requirements, visual encoding and interactivity, enhances communication, users' cognitive capacity, and translates into a better task performance.

When comparing the experience of participants with both Project Ukko and the redesigned tool, the quantitative indicators of success rate and response time when completing a task (indicators 1 and 2) demonstrate that the changes made to the redesigned tool increased the rate of success or partial success and allowed participants to perform the tasks faster. It is necessary to highlight that, in the case of Ukko, the success rate was extremely low for task 1. This was due to an incorrect identification of the area, which was done based on its brightness (high skill values), but that did not meet the minimum requirements of thickness (high intensity values). From a usability point of view, this is commonly defined as a false

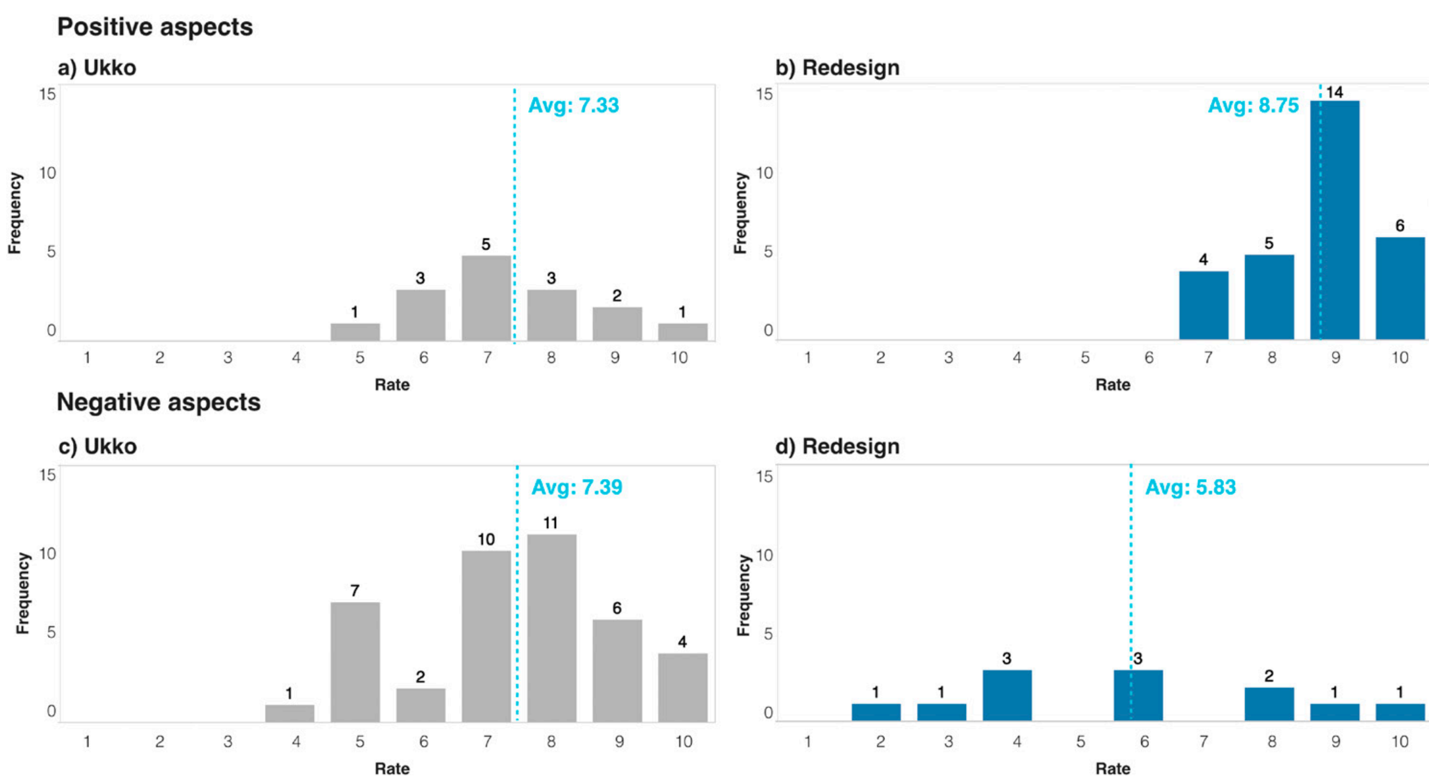


Fig. 3. Frequency histograms of participants' ratings of positive and negative aspects mentioned for (a),(c) Project Ukko and(b),(d) the redesigned tool. All the mentioned aspects (*n* = 97) receive a score from 1 to 10. Positive aspects are rated from 10 (very positive) to 1 (less positive) and negative aspects are rated from 10 (very serious) to 1 (slightly serious).

success, since users believe they have performed the task correctly, when in fact the answer was wrong (Brinck et al. 2002).

Qualitative methods such as the bipolar laddering and two-question quiz applied in this study are based on subjective user assessment (e.g., experience, feelings, intuition, opinions), which is often perceived as less reliable than other quantitative methods (Szafir and Szafir 2016). However, they are well-established practices that support and complement quantitative analysis and have been widely applied (Lim et al. 2019; Navarro et al. 2020). These methods provide useful insights to understand user's preferences and the positive and negative aspects that intervene in the performance and efficiency of users to perform day-to-day tasks.

In general, the changes in shape and size (i.e., the visual encoding of the information) as well as the reduction of categories were highly rated by the participants, enhancing the clarity and ease of use of the redesigned tool. In the case of Ukko, a large number of participants were not able to identify the categories associated with a glyph (skill, intensity, predicted change) nor the exact category, even when they tried to compare them to nearby glyphs. This was because thin lines, which were combined with opacity (a visual encoding that affects visibility), made color detection more difficult. Sometimes the participants were unable to identify the exact thickness of a group of glyphs despite the many accesses to the legend, which led them to randomly select a category to avoid abandoning the task. Overall good practices in data visualization indicate that opacity, combined with color and reduced thickness, can make graphical interpretation worse (Dastani 2002; Jenny and Kelso 2007; Ware 2012). Also, the use of the slope of the glyphs was perceived as a negative aspect by participants. This is confirmed by visual encoding good practices, indicating that changes in slope are more difficult to be interpreted, especially for nominal data (Alexandre and Tavares 2010; Munzner 2014). In addition, in the context of wind data visualization, the use of slopes tends to be related to wind direction (Powers et al. 2017), as was also pointed out by some of the participants during the test. Therefore, using slope to display the wind predicted change can be counterintuitive for users.

Only two sizes and three colors were combined in the redesigned tool. This favors the detection of the areas of interest because the visual encoding does not create competition between variables (in this case, wind intensity and predicted change) (Iliinsky and Steele 2011; Riveiro et al. 2008). Additionally, glyphs below a specific skill threshold (which would be discarded in a decision-making process) can be hidden from the display, allowing the user to focus on feasible options. This has been recognized as an effective means to reduce user memory workload and enhance task performance (Hegarty 2011).

The most frequently mentioned negative aspect of the redesigned tool referred to the color chosen for the middle prediction category ("Small dark glyphs, showing average values, have a similar color to the background"). Users found the color too similar to the background, therefore lacking sufficient contrast. However, this was decided on purpose, since target users are more interested in situations that depart from normal (i.e., upper and lower than average values), since these are the ones in which they need to take action (Kohlhammer and Zeltzer 2004). Therefore, by choosing a color similar to the background for nonrelevant values, we reduced visual noise in the representation. On the other hand, values of upper and lower predicted change (indicating wind conditions above and below normal) use a green-yellow color hue, which stimulates more photoreceptors in the human eye and hence are easily detected by users (UNSW 2015).

Despite color being a crucial element of visual encoding, we did not include further changes in the color scale used in the test to compare Project Ukko and its redesigned visualization. This was decided in order to focus the analysis in the visual encoding and interactivity aspects and to avoid a major change between both tools that could bias the results of the test.

However, in order to improve the accessibility of any climate service, color-blindness should be taken into account in color choices for visual representations (Light and Bartlein 2004). In the case of Project Ukko, even using color-blind-friendly scales, the combination of color, opacity, and certain widths reduced the effective perception by color-blind people. In this sense, the S2S4E Decision Support tool (S2S4E 2020) already included changes in the color palette to improve its accessibility by taking into account color-blindness aspects.

Some changes applied in the redesign of Project Ukko were also related to the use of interactive filters to dose or personalize information. The skill filter in the redesigned version of the tool allows one to explore the uncertainty associated with the predicted change focusing just on the relevant data for the user (i.e., values below a preferred skill threshold are hidden). In the same way, the intensity slider allows the user to establish a preferred threshold, giving more visual presence (larger glyph size) to the values above this threshold. This capacity of filtering and personalizing was highlighted as a strong positive aspect of the redesigned tool. One of the users even suggested adding interactivity to the color legend, to be able to filter by the predicted change. Indeed, interactivity allows users to consume information step by step, explore particular aspects of complex datasets, and display relevant information in their own world view (Beddington 2011; McInerny et al. 2014). This is thus a highly recommended feature for online climate services taking into account that there may be limits to how useful interactive visualizations are if the viewers do not have the required skills to interact with the presented information (diSessa 2004).

The second task proposed to participants was more challenging than the first task regardless of the tool used. The classification of two types of glyphs proposed in task 2 took more time to complete than the identification of an area of interest in task 1. This was especially remarkable for Ukko, with a higher number of categories competing at the same time for visual attention (Alhadad 2018; Munzner 2014). Also, the number of times that participants needed to check the legend (indicator 5) was larger for Project Ukko. The difference between Project Ukko and the redesigned tool is probably related to Project Ukko's negative aspects linked to problems for understanding the legend, which had a more complex visual encoding. The combination of different categories in Project Ukko was also considered as a negative aspect (e.g., "overwhelming representation," "mixing too many categories increases complexity") in contrast to positive aspects of the redesigned tool linked with the simplicity of the representation (e.g., "the representation is very clear," "easy to distinguish between glyphs"). This would also explain the low success rate of task 1 for Ukko where, despite having more accesses to the legend, the area selected by participants did not meet the requirements of the statement in terms of skill, intensity and predicted change.

The difference in the purpose of task 1 and task 2 (identification versus classification, respectively), might also be the reason why participants needed to check the legend more often at the second task. Overall, the obtained number of accesses to the legend (combined with a simpler visualization) suggests a reduction in the cognitive load of the participants during the completion of the tasks with the redesigned tool as they could retain the legend better and therefore reduce the number of times they had to check it.

Regarding the fixation duration, the quantitative indicator behaved almost equally between tasks and between tools, with durations ranging between 0.25 and 0.31 s. The number of fixations are the number of times a user pays attention to a certain point or area of interest on the screen. According to available bibliography, a longer fixation duration may indicate a bigger cognitive load during task performance (Duchowski 2007; Ooms et al. 2014; Andrzejewska and Skawińska 2020; Klingner et al. 2008; Krejtz et al. 2018) or that users have found more interesting elements to fix their attention for a longer time, without necessarily implying a greater difficulty or cognitive load (Henderson and Ferreira 2004; Klingner et al. 2008; Ooms et al. 2012; Krejtz et al. 2018; Andrzejewska and Skawińska 2020).

Advanced brain monitoring tools, such as electroencephalograms and eye-tracker measures of pupillometry, can be useful to further study cognitive load (Anderson et al. 2011; Jiang et al. 2014; Keskin et al. 2020). It would be interesting to test them with other eye-tracker models to explore if this could be due to the accuracy of the model used.

The perceived effort during the task performance (indicator 6) identifies Project Ukko as being more complex to use than the redesigned tool, which was indicated as the tool preferred by 90% of the participants tested in this work (indicator 7). This contributes to the hypothesis that by eliminating or simplifying visual encodings nonrelevant to a target action or task and increasing interactivity, we favor decision-making.

The results and opinions of the bipolar laddering (indicator 8) clearly supported the previous indicators. A total of 29 positive comments were obtained for the redesigned tool, compared to 15 for the original Ukko tool (Table 1). This was also confirmed by the higher average score obtained for the redesign (8.75) when compared to Project Ukko (7.33). In the same way, the redesigned tool received fewer negative comments (12 compared to 41 for Ukko) and they were less serious (obtained scores of 5.83 for the redesign against 7.33 for Ukko). The most serious aspects associated with Ukko referred to the difficulty to differentiate the categorization of glyphs due to the combination of encoding through color, intensity, and thickness, often making participants unable to identify the corresponding category. These aspects had a very high frequency, 25 comments with notable severity and an obtained score of 8.

Table 1. Participants' positive and negative comments for Project Ukko and the redesigned tools with the number of participants that mentioned a particular aspect (freq.) and average rate of its importance/severity (avg.).

Project Ukko					
Positive	**Freq.**	**Avg.**	**Negative**	**Freq.**	**Avg.**
Easy to distinguish extreme areas	7	7.8	Difficult to distinguish opacity	9	8
A very detailed version with lots of information	3	6.6	Mixing width and brightness is too complex	9	8.22
Static legends are more traditional	3	6.6	Thin glyphs with low visibility are impossible to distinguish	7	7.57
Visually attractive	2	7.5	Using the combination of two variables (color and slope) for prediction change is too complex	4	6.25
			Slopes are confusing, they usually are used to show wind direction	3	7.33
			Mixing too many categories increases complexity	3	6.33
			Overwhelming representation	2	7
			Terrain not visible enough	1	6
			Slope value (of glyphs) is difficult to measure	1	5
			Legend is confusing	1	5
Total	15	7.33		41	7.39
Redesign					
Positive	**Freq.**	**Avg.**	**Negative**	**Freq.**	**Avg.**
Shapes and sizes are easier to identify	8	8.25	Small dark glyphs (showing average values) have a color similar to the background	5	5.6
The representation is very clear	6	9.3	The descriptive labels of the skill slider (using mathematical terms) may not be clear to all audiences.	3	6.66
Easy to distinguish between glyphs	5	8.4	Too basic to represent predicted change	1	6
Skill filtering is very useful	4	9.25	By simplifying some of the categories, we lose information	1	6
High contrast	2	9.5	Terrain not visible enough	1	6
Easy location of an area	2	9	All the three filters or descriptive labels could have been interactive (such as color category)	1	4
Simple categorization	2	8			
Total	29	8.75		12	5.83

Glossary

Our work is based on the collaboration of multiple disciplines that, when combined, improve the user experience and favor decision-making. **User-centered design (UCD)** allows us to meet the needs of the users. **Interaction design** allows us to dose and customize the way the data are displayed. **Design** and **data visualization** rules improve the way of visually encoding information and highlight what is really important. **Cognitive psychology** helps us measure the impact of complex visualizations on users and, with the help of other disciplines, favors their understanding.

These disciplines have specific terminology used in this paper:

Accessibility: Discipline and rules that guarantee that websites and technologies are designed and developed so that people with disabilities can use them independently from their capability limitations: auditory, visual, cognitive, physical, or neurological.

Cognition: Mental process of acquiring knowledge and understanding through senses, thought, and experience.

Cognitive load: Refers to the used amount of human working memory resources, which is limited in both capacity and duration.

Cognitive psychology: The scientific study of mental processes such as attention, memory, perception, problem solving, and understanding.

Eye-tracker: Sensor-based device which measures where the participant is looking at (the point of gaze) and the motion of the eyes.

Fixation: A period of time during which the eyes are locked toward an object or visual element.

Color hue: The attribute of color defined by wavelength (red, blue, etc.).

Glyph: A hieroglyphic character or symbol used in visualization as a part of a chart or graph.

Interaction design: Design of interactive products focusing on the way users interact with them, including visual representation, terminology, devices, and behavior.

Multidimensional visualization: Graph or visualization showing more than one variable through visual encoding (color, size, etc.).

Opacity: Property of a visual element that determines how transparent (or less visible) it will be. The lower the opacity value, the more transparent the element is.

Perception: The way in which something is regarded, understood, or interpreted.

Usability testing: Evaluation of a product or a service in order to detect problems or evaluate how easy it is to use it.

User-centered design: Iterative design process in which designers focus on the user needs and involve them in each phase of the design process.

Visual encoding: Translating the data into a visual element on a chart/map or graph using visual properties as length, position, size, color, slope, opacity, etc.

Working memory: The cognitive system with a limited capacity that can hold information temporarily.

Regarding the negative comments obtained for the redesigned tool, five participants referred to the color similarity of some glyphs to the background of the tool (values close to the mean in terms of predictive change). However, after explaining the reasons for this design decision (reduce visual noise by attenuating non relevant points), all users found the change appropriate. Another negative aspect was related to the label terminology of the skill slider, which included mathematical terms that may not be clear to all audiences. When delving into the reasons for the negative assessment, the participant clarified that the control seems useful, but that the texts used in the labels could be clearer or more intuitive.

Paradoxically, some positive comments received for Ukko, referred to its higher density of information, greater detail in the representation of predicted change (five categories instead of the three categories in the redesigned tool) and visual appeal. However, although users who mentioned these aspects believed that these characteristics could be valuable in a context where exploration was the objective of visualization, they did not favor clarity or decision-making.

Conclusions

The behavioral decision-making literature shows how people often struggle to understand particular climate terminology. There is often a mismatch between the understanding of concepts such as probabilities or uncertainty between experts and nonexperts. Although visualizing forecast uncertainties and associated probabilities is thought to increase users' trust, it does not automatically lead to better decisions.

Our results identify relevant aspects that can improve user experience and reduce cognitive load and that are worth considering when designing climate data visualizations. These include choosing representations and categories tailored to specific decisions, avoiding visual encoding that interferes with users' perception of the represented forms, and offering interactive elements that allow users to filter nonrelevant information or highlight relevant information for the decision at hand. In the redesign of Project Ukko we included all these changes at the same time. Hence, we cannot empirically establish the relative influence of each of these individual aspects in the overall reduction of the users' cognitive load. Nevertheless, we demonstrate that all these aspects can help reduce cognitive load, favor decision-making, and thus improve the overall user experience with a climate service.

In future works, it would be interesting to delve into the weight of each of the implemented actions (simplifying the number of categories, avoiding redundant visual encoding, customizing the visualizations based on user needs through interactive controls) in the total reduction of the cognitive load.

Likewise, analyzing the implemented changes in the context of the final tool, in combination with other improvements not assessed in the framework of this study (redesign of the navigation, color-blind aspects, customization, and levels of detail available), can highlight additional benefits. This would delve further into the visual communication of climate information.

Our study highlights that, when combining techniques and knowledge from different disciplines [climate science, design, user-centered design (UCD), user interaction, and cognitive psychology], we are able to find better solutions for the visualization of climate data, especially when aimed at supporting decision-making. In addition, we identify a clear need for co-design and increased empirical testing of the resulting products. We recommend information providers and tool designers in the field of climate services to collaborate more with end users throughout the whole design process to identify what is effective and to leverage the knowledge and well-established techniques from nonclimate related disciplines that have a lot to offer.

Acknowledgments. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreements 776787 (S2S4E), 776613 (EUCP), and (ClimatEurope). This work was also supported by the MEDSCOPE project. MEDSCOPE is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by AEMET (ES), ANR (FR), BSC (ES), CMCC (IT), CNR (IT), IMR (BE), and Météo-France (FR), with co-funding by the European Union (Grant 690462). The research team would like to thank the participants of the test who generously shared their time and opinions for the purposes of this research. This study is a part of the PhD of the corresponding author, Luz Calvo.

References

Alexandre, D., and J. Tavares, 2010: Introduction of human perception in visualization. *Int. J. Imaging*, **4** (10), 60–70.

Alhadad, S. S. J., 2018: Visualizing data to support judgement, inference, and decision making in learning analytics: Insights from cognitive psychology and visualization science. *J. Learn. Anal.*, **5**, 60–85, https://doi.org/10.18608/jla.2018.52.5.

Anderson, E. W., K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva, 2011: A user study of visualization effectiveness using EEG and cognitive load. *Comput. Graph. Forum*, **30**, 791–800, https://doi.org/10.1111/j.1467-8659.2011.01928.x.

Andrzejewska, M., and A. Skawińska, 2020: Examining students' intrinsic cognitive load during program comprehension – An eye tracking approach. *AIED 2020: Artificial Intelligence in Education*, I. I. Bittencourt et al., Eds., Springer, 25–30, https://doi.org/10.1007/978-3-030-52240-7_5.

Antifakos, S., A. Schwaninger, and B. Schiele, 2004: Evaluating the effects of displaying uncertainty in context-aware applications. *UbiComp* 2004*: Ubiquitous Computing*, N. Davies, E. D. Mynatt, and I. Siio, Eds., Springer, 54–69, https://doi.org/10.1007/978-3-540-30119-6_4.

Argyle, E. M., J. J. Gourley, Z. L. Flamig, T. Hansen, and K. Manross, 2017: Toward a user-centered design of a weather forecasting decision-support tool. *Bull. Amer. Meteor. Soc.*, **98**, 373–382, https://doi.org/10.1175/BAMS-D-16-0031.1.

Atkins, R., and K. Mcneal, 2018: Exploring differences among student populations during climate graph reading tasks: An eye tracking study. *J. Astron. Earth Sci. Educ.*, **5**, 85–114, https://doi.org/10.19030/jaese.v5i2.10219.

Beddington, J., 2011: Blackett Review of high impact low probability events. Government Office for Science, 46 pp.

Bertin, J., 2010: *Semiology of Graphics: Diagrams, Networks, Maps*. 1st ed. Esri Press, 456 pp.

Bevington, K., S. Ansari, R. Bilotta, A. B. Courtright, and A. Lang, 2019: Improving climate data visualization and information dissemination on the U.S. drought portal through user experience (UX) design. *2019 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract PA33C-1103.

Block, G., 2013: Reducing cognitive load using adaptive uncertainty visualization*.* Ph.D. thesis, Nova Southeastern University, 132 pp. https://nsuworks.nova.edu/gscis_etd/95/.

Bonneau, G.-P., H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz, 2014: Overview and state-of-the-art of uncertainty visualization. *Scientific Visualization*, C. D. Hansen et al., Eds., Springer, 3–27, https://doi.org/10.1007/978-1-4471-6497-5_1.

Borkin, M. A., A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister, 2013: What makes a visualization memorable? *IEEE Trans. Vis. Comput. Graph.*, **19**, 2306–2315, https://doi.org/10.1109/TVCG.2013.234.

Brinck, T., D. Gergle, and S. D. Wood, 2002: *Usability for the Web*. Elsevier, 497 pp.

Cairo, A., 2012: *The Functional Art: An Introduction to Information Graphics and Visualization*. New Riders Publishing, 384 pp.

Christel, I., and Coauthors, 2016: Project Ukko: Seasonal wind speed predictions at a glance. *WindEurope Summit 2016*, Hamburg, Germany, Wind Europe, https://discovery.dundee.ac.uk/en/publications/project-ukko-seasonal-wind-speed-predicitions-at-a-glance*.*

——, D. Hemment, D. Bojovic, F. Cucchietti, L. Calvo, M. Stefaner, and C. Buontempo, 2018: Introducing design in the development of effective climate services. *Climate Serv.*, **9**, 111–121, https://doi.org/10.1016/j.cliser.2017.06.002.

Chun, R., 2017: Redundant encoding in data visualizations: Assessing perceptual accuracy and speed. *Vis. Commun. Quart.*, **24**, 135–148, https://doi.org/10.1080/15551393.2017.1343153.

Cleveland, W. S., and R. McGill, 1985: Graphical perception and graphical methods for analyzing scientific data. *Science*, **229**, 828–833, https://doi.org/10.1126/science.229.4716.828.

Dastani, M., 2002: The role of visual perception in data visualization. *J. Vis. Lang. Comput.*, **13**, 601–622, https://doi.org/10.1006/jvlc.2002.0235.

Davis, B. M., K. Dickerson, and S. C. Gillmore, 2020: User centered design strategies for improving visualization of sensor data in rotorcraft cockpit displays for degraded visual environment operations. *Advances in Usability and User Experience*, T. Ahram and C. Falcão, Eds., Springer, 131–141, https://doi.org/10.1007/978-3-030-19135-1_13.

diSessa, A. A., 2004: Metarepresentation: Native competence and targets for instruction. *Cognit. Instr.*, **22**, 293–331, https://doi.org/10.1207/s1532690xci2203_2.

Dong, F., G. Ghinea, and S. Y. Chen, 2008: *User Centered Design for Medical Visualization*. Medical Information Science Reference Global, 436 pp.

Duchowski, A., 2007: *Eye Tracking Methodology*. Springer, 334 pp.

Ellis, G., and A. Dix, 2006: An explorative analysis of user evaluation studies in information visualisation. *Proc. of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, Venice, Italy, ACM, 1–7, https://doi.org/10.1145/1168149.1168152.

Faulkner, L., 2003: Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behav. Res. Methods*, **35**, 379–383, https://doi.org/10.3758/BF03195514.

Few, S., 2009: *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. 1st ed. Analytics Press, 329 pp.

Freitas, C. D. S., P. R. G. Luzzardi, R. Cava, M. Winckler, M. Pimenta, and L. Nedel, 2002: On evaluating information visualization techniques. *Proc. of the Working Conf. on Advanced Visual Interfaces*, Trento, Italy, ACM, 373–374, https://doi.org/10.1145/1556262.1556326.

Gerharz, L. E., and E. J. Pebesma, 2009: Usability of interactive and non-interactive visualisation of uncertain geospatial. *Geoinformatik 2009 Konferenzband*, W. Reinhardt, A. Krüger, and M. Ehlers, Eds., Institut für Geoinformatik, 223–230.

Gerst, M. D., and Coauthors, 2020: Using visualization science to improve expert and public understanding of probabilistic temperature and precipitation outlooks. *Wea. Climate Soc.*, **12**, 117–133, https://doi.org/10.1175/WCAS-D-18-0094.1.

Grainger, S., F. Mao, and W. Buytaert, 2016: Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environ. Modell. Software*, **85**, 299–318, https://doi.org/10.1016/j.envsoft.2016.09.004.

Greis, M., T. Ohler, N. Henze, and A. Schmidt, 2015: Investigating representation alternatives for communicating uncertainty to non-experts. *INTERACT 2015: Human-Computer Interaction*, J. Abascal et al., Eds., Springer, 256–263, https://doi.org/10.1007/978-3-319-22723-8_21.

Harold, J., I. Lorenzoni, T. F. Shipley, and K. R. Coventry, 2016: Cognitive and psychological science insights to improve climate change data visualization. *Nat. Climate Change*, **6**, 1080–1089, https://doi.org/10.1038/nclimate3162.

Hegarty, M., 2011: The cognitive science of visual-spatial displays: Implications for design. *Top. Cognit. Sci.*, **3**, 446–474, https://doi.org/10.1111/j.1756-8765.2011.01150.x.

——, M. Canham, and S. Fabrikant, 2010: Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *J. Exp. Psychol. Learn. Mem. Cognit.*, **36**, 37–53, https://doi.org/10.1037/a0017683.

Henderson, J. M., and F. Ferreira, Eds., 2004: *The Interface of language, Vision, and action: Eye Movements and the Visual World*. Psychology Press, 414 pp.

Holmes, N., 1984: *Designer's Guide to Creating Charts and Diagrams.* 2nd ed. Watson-Guptill, 192 pp.

Holtzblatt, K., J. B. Wendell, and S. Wood, 2004: *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. Elsevier, 320 pp.

Iliinsky, N., and J. Steele, 2011: *Designing Data Visualizations*. O'Reilly, 114 pp.

Jenny, B., and N. V. Kelso, 2007: Color design for the color vision impaired. *Cartogr. Perspect.*, **58**, 61–67, https://doi.org/10.14714/CP58.270.

Jiang, X., M. Atkins, G. Tien, R. Bednarik, and B. Zheng, 2014: Pupil responses during discrete goal-directed movements. *Proc. SIGCHI Conf. on Human Factors in Computing System*, Toronto, ON, ACM, 2075–2084, https://doi.org/10.1145/2556288.2557086.

Joslyn, S., and J. LeClerc, 2011: Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *J. Exp. Psychol. Appl.*, **18**, 126–140, https://doi.org/10.1037/a0025185.

Kaye, N. R., A. Hartley, and D. Hemming, 2012: Mapping the climate: Guidance on appropriate techniques to map climate variables and their uncertainty. *Geosci. Model Dev.*, **5**, 243–256, https://doi.org/10.5194/gmd-5-245-2012.

Keskin, M., K. Ooms, A. O. Dogru, and P. De Maeyer, 2020: Exploring the cognitive load of expert and novice map users using EEG and eye tracking. ISPRS Int. Geo-Inf., **9**, 429, https://doi.org/10.3390/ijgi9070429.

Khamaj, A., Z. Kang, and E. Argyle, 2019: Users' perceptions of smartphone weather applications' usability. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **63**, 2216–2220, https://doi.org/10.1177/1071181319631098.

Klingner, J., R. Kumar, and P. Hanrahan, 2008: Measuring the task-evoked pupillary response with a remote eye tracker. *ETRA'08: Proc. of the 2008 Symp. on Eye Tracking Research & Applications*, Savannah, GA, ACM, 69–72, https://doi.org/10.1145/1344471.1344489.

Kohlhammer, J., and D. Zeltzer, 2004: Towards a visualization architecture for time-critical applications. *Proc. of the 9th Int. Conf. on Intelligent User Interfaces*, Funchal, Madeira, Portugal, ACM, 271–273, https://doi.org/10.1145/964442.964502.

Kosara, R., 2013: InfoVis is so much more: A comment on Gelman and Unwin and an invitation to consider the opportunities. *J. Comput. Graph. Stat.*, **22**, 29–32, https://doi.org/10.1080/10618600.2012.755465.

Krejtz, K., A. T. Duchowski, A. Niedzielska, C. Biele, and I. Krejtz, 2018: Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE*, **13**, e0203629, https://doi.org/10.1371/journal.pone.0203629.

Lau, A., and A. Vande Moere, 2007: Towards a model of information aesthetics in information visualization. *2007 11th Int. Conf. Information Visualization (IV'07)*, Zurich, Switzerland, Institute of Electrical and Electronics Engineers, 87–92, https://doi.org/10.1109/IV.2007.114.

Li, Y., N. Hollender, and T. Held, 2013: Task sequence effects in usability tests. *IADIS Int. Conf. Interfaces and Human Computer Interaction and Game and Entertainment Technologies 2013*, Prague, Czech Republic, University of Economics in Prague, 141–148.

Light, A., and P. Bartlein, 2004: The end of the rainbow? Color schemes for improved data graphics. *Eos, Trans. Amer. Geophys. Union*, **85**, 385, https://doi.org/10.1029/2004EO400002.

Lim, K., A. Selamat, R. Alias, O. Krejcar, and H. Fujita, 2019: Usability measures in mobile-based augmented reality learning applications: A systematic review. *Appl. Sci.*, **9**, 2718, https://doi.org/10.3390/app9132718.

Ling, C., L. Hua, C. D. Karstens, G. J. Stumpf, T. M. Smith, K. M. Kuhlman, and L. Rothfusz, 2015: A comparison between WarnGen system and probabilistic hazard information system for severe weather forecasting. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **59**, 1791–1795, https://doi.org/10.1177/1541931215591387.

Lloyd, R., 1997: *Spatial Cognition: Geographic Environments*. Springer, 118 pp.

Lorenz, S., S. Dessai, P. M. Forster, and J. Paavola, 2015: Tailoring the visual communication of climate projections for local adaptation practitioners in Germany and the UK. *Philos. Trans. Roy. Soc.*, **373A**, 20140457, https://doi.org/10.1098/rsta.2014.0457.

Majooni, A., M. Masood, and A. Akhavan, 2018: An eye-tracking study on the effect of infographic structures on viewer's comprehension and cognitive load. *Inf. Vis.*, **17**, 257–266, https://doi.org/10.1177/1473871617701971.

Makri, S., 2015: EUPORIAS seasonal wind prediction prototype: User evaluation. City University London, 36 pp., https://earth.bsc.es/wiki/lib/exe/fetch.php?media=library:external:services:euporias_seasonal_wind_prediction_prototype_-_user_evaluation_report_nov_2015.pdf.

McInerny, G. J., and Coauthors, 2014: Information visualisation for science and policy: Engaging users and avoiding bias. *Trends Ecol. Evol.*, **29**, 148–157, https://doi.org/10.1016/j.tree.2014.01.003.

Munzner, T., 2014: *Visualization Analysis and Design*. 1st ed. CRC Press, 428 pp.

Navarro, I., O. de Reina, D. Fonseca, M. Gomez, and Á. Ferrer, 2020: Virtual reality using smart-devices in educational frameworks: Case study Museum Casa Batlló. *Int. J. Technol. Hum. Interact.*, **13**, 50–61, https://doi.org/10.4018/IJTHI.2017100104.

Nielsen, J., 1993: Evaluating the thinking-aloud technique for use by computer scientists. *Advances in Human-Computer Interaction*, H. R. Hartson and D. Hix, Eds., Ablex Publishing, 69–82.

Oakley, N. S., and B. Daudert, 2016: Establishing best practices to improve usefulness and usability of web interfaces providing atmospheric data. *Bull. Amer. Meteor. Soc.*, **97**, 263–274, https://doi.org/10.1175/BAMS-D-14-00121.1.

Olmsted-Hawala, E. L., E. D. Murphy, S. Hawala, and K. T. Ashenfelter, 2010: Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-dissemination websites for usability. *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, Atlanta, GA, ACM, 2381–2390, https://doi.org/10.1145/1753326.1753685.

Ooms, K., P. De Maeyer, V. Fack, E. Van Assche, and F. Witlox, 2012: Interpreting maps through the eyes of expert and novice users. *Int. J. Geogr. Inf. Sci.*, **26**, 1773–1788, https://doi.org/10.1080/13658816.2011.642801.

——, A. Coltekin, P. De Maeyer, L. Dupont, S. Fabrikant, A. Incoul, and L. Van der Haegen, 2014: Combining user logging with eye tracking for interactive and dynamic applications. *Behav. Res. Methods*, **47**, 977–993, https://doi.org/10.3758/s13428-014-0542-3.

Pifarré, M., and O. Tomico, 2007: Bipolar laddering (BLA): A participatory subjective exploration method on user experience. *Proc. of the 2007 Conf. on Designing for User EXperiences, DUX 2007*, Chicago, IL, ACM, 2–13, https://doi.org/10.1145/1389908.1389911.

Powers, J. G., and Coauthors, 2017: The weather research and forecasting model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Pretorius, M., A. P. Calitz, and D. van Greunen, 2005: The added value of eye tracking in the usability evaluation of a network management tool. *SAICSIT'05: Proc. of the 2005 Annual Research Conf. of the South African Institute of Computer Scientists and Information Technologists*, Cape Town, South Africa, South African Institute for Computer Scientists and Information Technologists, 1–10, https://doi.org /10.5555/1145675.1145676.

R Core Team, 2018: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org.

Riveiro, M., G. Falkman, and T. Ziemke, 2008: Improving maritime anomaly detection and situation awareness through interactive visualization. *Proc. of the 11th Int. Conf. on Information Fusion*, Cologne, Germany, Institute of Electrical and Electronics Engineers, 1–8.

——, T. Helldin, G. Falkman, and M. Lebram, 2014: Effects of visualizing uncertainty on decision-making in a target identification scenario. *Comput. Graph.*, **41**, 84–98, https://doi.org/10.1016/j.cag.2014.02.006.

Roulston, M. S., G. E. Bolton, A. N. Kleit, and A. L. Sears-Collins, 2006: A laboratory study of the benefits of including uncertainty information in weather forecasts. *Wea. Forecasting*, **21**, 116–122, https://doi.org/10.1175/WAF887.1.

S2S4E, 2020: Decision support tool. Accessed 16 December 2020, https://s2s4e.eu/dst.

Sager, L., R. Grier, C. Jackson, Y. Levchuk, and E. Stelzer, 2007: Visualization techniques for revealing uncertainty to decision makers. *Human Systems Integration Symp. 2007*, Annapolis, MD, American Society of Naval Engineers.

Schrepp, M., A. Hinderks, and J. Thomaschewski, 2017: Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *IJIMAI*, **4**, 103–108, https://doi.org/10.9781/ijimai.2017.09.001.

Stephens, E. M., T. L. Edwards, and D. Demeritt, 2012: Communicating probabilistic information from climate model ensembles—Lessons from numerical weather prediction. *Wiley Interdiscip. Rev.: Climate Change*, **3**, 409–426, https://doi.org/10.1002/wcc.187.

Szafir, D. A., and D. Szafir, 2016: Cognitive load in visualization: Myths & misconceptions. *Workshop on Creation, Curation, Critique and Conditioning of Principles and Guidelines in Visualization (C4PGV)*, Baltimore, MD, IEEE, https://cmci.colorado.edu/visualab/papers/16-c4pgvcogload.pdf.

Taylor, A. L., S. Dessai, and W. B. de Bruin, 2015: Communicating uncertainty in seasonal and interannual climate forecasts in Europe. *Philos. Trans. Roy. Soc.*, **373A**, 20140454, https://doi.org/10.1098/rsta.2014.0454.

Terrado, M., and Coauthors, 2019: The weather roulette: A game to communicate the usefulness of probabilistic climate predictions. *Bull. Amer. Meteor. Soc.*, **100**, 1909–1921, https://doi.org/10.1175/BAMS-D-18-0214.1.

Trivedi, M. C., 2012: Role of context in usability evaluations: A review. *J. Adv. Comput. Sci. Appl.*, **3**, 69–78, https://doi.org/10.5121/acij.2012.3208.

Tufte, E. R., 2001: *The Visual Display of Quantitative Information*. 2nd ed. Graphics Press, 200 pp.

UNSW, 2015: Colour mixing and colour vision. Accessed 16 December 2020, www.animations.physics.unsw.edu.au/jw/light/colour-mixing.htm.

Veras Guimarães, R., 2019: Visual encoding quality and scalability in information visualization. Ph.D. thesis, Faculty of Science, University of Ontario Institute of Technology, 151 pp., http://ir.library.dc-uoit.ca/handle/10155/1024.

Virzi, R. A., 1992: Refining the test phase of usability evaluation: How many subjects is enough? *Hum. Factors*, **34**, 457–468, https://doi.org/10.1177/001872089203400407.

Wang, Q., S. Yang, M. Liu, Z. Cao, and Q. Ma, 2014: An eye-tracking study of website complexity from cognitive load perspective. *Decis. Support Syst.*, **62**, 1–10, https://doi.org/10.1016/j.dss.2014.02.007.

Ware, C., 2012: *Information Visualization: Perception for Design*. Elsevier, 536 pp.

——, H. Purchase, L. Colpoys, and M. McGill, 2002: Cognitive measurements of graph aesthetics. *Inf. Vis.*, **1**, 103–110, https://doi.org/10.1057/palgrave.ivs.9500013.

Winckler, M. A., P. Palanque, and C. M. D. S. Freitas, 2004: Tasks and scenario-based evaluation of information visualization techniques. *Proc. of the 3rd Annual Conf. on Task Models and Diagrams*, Prague, Czech Republic, ACM, 165–172, https://doi.org/10.1145/1045446.1045475.

Yoghourdjian, V., D. Archambault, S. Diehl, T. Dwyer, K. Klein, H. C. Purchase, and H.-Y. Wu, 2018: Exploring the limits of complexity: A survey of empirical studies on graph visualisation. *Vis. Inf.*, **2**, 264–282, https://doi.org/10.1016/j.visinf.2018.12.006.

Yucong, Y., F. Sauer, K.-L. Ma, K. Aditya, and J. Chen, 2019: A user-centered design study in scientific visualization targeting domain experts. *IEEE Trans. Vis. Comput. Graph.*, **26**, 2192–2203, https://doi.org/10.1109/TVCG.2020.2970525.

Climatic Change (2022) 172: 18
https://doi.org/10.1007/s10584-022-03365-4

REVIEW ARTICLE

Towards more effective visualisations in climate services: good practices and recommendations

Marta Terrado[1] · Luz Calvo[1] · Isadora Christel[1]

Received: 10 November 2021 / Accepted: 30 April 2022 / Published online: 28 May 2022

Abstract

Visualisations are often the entry point to information that supports stakeholders' decision- and policy-making processes. Visual displays can employ either static, dynamic or interactive formats as well as various types of representations and visual encodings, which differently affect the attention, recognition and working memory of users. Despite being well-suited for expert audiences, current climate data visualisations need to be further improved to make communication of climate information more inclusive for broader audiences, including people with disabilities. However, the lack of evidence-based guidelines and tools makes the creation of accessible visualisations challenging, potentially leading to misunderstanding and misuse of climate information by users. Taking stock of visualisation challenges identified in a workshop by climate service providers, we review good practices commonly applied by other visualisation-related disciplines strongly based on users' needs that could be applied to the climate services context. We show how lessons learned in the fields of user experience, data visualisation, graphic design and psychology make useful recommendations for the development of more effective climate service visualisations. This includes applying a user-centred design approach, using interaction in a suitable way in visualisations, paying attention to information architecture or selecting the right type of representation and visual encoding. The recommendations proposed here can help climate service providers reduce users' cognitive load and improve their overall experience when using a service. These recommendations can be useful for the development of the next generation of climate services, increasing their usability while ensuring that their visual components are inclusive and do not leave anyone behind.

Keywords Transdisciplinarity · User experience · Data visualisation · Graphic design · Psychology · Usability

✉ Marta Terrado
marta.terrado@bsc.es

Luz Calvo
luz.calvo@bsc.es

Isadora Christel
isadora.jimenez@bsc.es

[1] Department of Earth Sciences, Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain

1 Introduction

1.1 Need for effective visualisations in climate services

The visual communication of climate information is one of the cornerstones of climate services (Calvo et al. 2021a). Visualisations are the entry point to information that supports stakeholders' decision- and policy-making processes by increasing the efficiency of tools and bringing knowledge closer to relevant actors (Kamateri et al. 2015). To serve this purpose, visualisations should contain the right balance among information density (amount of represented data), robustness (scientific confidence and consensus) and saliency (relevance of the information to user needs) (Stephens et al. 2012). However, achieving this balance is not trivial and will depend on the stakeholder group and the visualisation is aimed for (e.g. academia, business, government, NGOs, citizens), the type of data communicated (e.g. physical, economic, political) and the purpose of communicating such information to the target audience (e.g. increase understanding, call to action, decision-making, etc.) (Raaphorst et al. 2020).

Data visualisations transform abstract information into two- or three-dimensional geometric representations (e.g. points, lines, surfaces) to facilitate analysis, comprehension and communication of models, concepts and data (Nocke 2014). Thus, visualisations use different visual modes (e.g. interactive tools and static representations) and different visual channels (e.g. combination of shape, colour, size, opacity and other attributes) to translate data. However, the use of these particular visual modes and channels can add further levels of complexity to the communication of climate information, which can affect users' attention, recognition and working memory (Quiroga et al. 2004; Calvo et al. 2021).

The scientific community has traditionally used standard visualisation techniques depending on the nature of the climate data to be displayed, including time and bar charts, box plots, scatter plots, probability distribution functions and maps (e.g. flood maps, heat maps, choropleth maps) (Haase et al. 2000; Stephens et al. 2012; Ash et al. 2014; Nocke 2014). Despite being widely used formats in the scientific domain, such representations are biased towards expert audiences with the appropriate background knowledge and give less consideration to broader audiences increasingly involved in climate adaptation processes but less familiar with the traditional ways of presenting climate data.

Climate service providers have largely faced the challenge of how to best communicate climate-related data in an easy and understandable way for both experts and non-experts. Aspects to be considered when delivering climate services include data encoding (how to best visualise data), user decoding (how to enhance user understanding of these data), service usability (how to ensure ease-of-use) and service efficacy (how to ensure credibility, relevance and usefulness) (Grainger et al. 2016). This suggests that a climate service that is partially or wrongly used (or not used at all) requires evaluation and potential redesign taking these aspects into account (Raaphorst et al. 2020). In this sense, the last years have seen an increase in the number of studies comparing different visualisation formats (Kaye et al. 2012; Daron et al. 2015, 2021; Lorenz et al. 2015; Taylor et al. 2015; Kause et al. 2020) and testing such formats with target audiences to assess their usability (McMahon et al. 2015; Christel et al. 2018; Laudien et al. 2019; Calvo et al. 2021). However, despite the increasing number of tools and platforms providing climate data visualisations, there is still limited guidance on how to develop usable, accessible and effective visualisations for climate services.

1.2 Lack of standards in climate service visualisations

The need for standards in climate services is a topic that has recently gained momentum. Guidelines for the quality management of climate services have been proposed by the World Meteorological Organisation (WMO 2018) and broad standards for climate change adaptation and mitigation also exist (ISO 9000 and 14,000 families of standards). In addition, some basic recommendations on communicating forecast uncertainty have been published (WMO 2008; Davis et al. 2015), but there is no standardised approach guiding the development of the visual component of climate services. This lack of guidance has favoured the proliferation of different practices, often resulting in stakeholders spending more time trying to understand the approach used for representing data than focusing on the interpretation of the information itself. This lack of standards has also been reported to lead to misunderstanding or misuse of climate information (Hewitt et al. 2020).

Due to the diversity of content and stakeholders, determining a one-size-fits-all visualisation practice for climate services would be a daunting task, if not impossible. However, a number of highlights, including both common practices and challenges for the development of visualisations in climate services, were identified in a workshop involving climate service providers (Terrado et al. 2022) and constitute the basis of the present review (see Section 2.1). The overarching take home message from the workshop was that climate service visualisations need to be developed by transdisciplinary teams with clear collaboration frameworks among academia and other stakeholders, recognising the essential role of social sciences, humanities, visualisation, behavioural sciences and other complementary research fields. Various studies claim that transdisciplinary aspects have received little attention in the climate services field, which still has low awareness of the lessons that can be learned from other disciplines (McInerny et al. 2014; Estrada and Davis 2015). Methodologies grounded in the fields of user experience, data visualisation, graphic design or psychology, which are strongly based on the needs of the users, can be an asset for the development of visualisations in the climate services field (Christel et al. 2018; Calvo et al. 2021a). In this work, we review good practices commonly applied in these disciplines and provide a list of recommendations that can serve as guidance for climate service providers aiming to co-develop more effective climate service visualisations.

2 Methods

2.1 Taking the pulse of the climate services community

This study takes as departing point the challenges identified in an interactive visualisation workshop organised in the framework of a Coordination and Support Action on climate services (Terrado et al. 2022). The aim of the workshop was to monitor current practices applied by climate services providers when developing visualisations and identify issues that need to be tackled for moving towards the development of more effective visualisations. The workshop consisted in a 2.5-h interactive session in breakout groups. A total of 25 participants attended the workshop, representing 22 projects working in climate services, including EU H2020 and ERA4CS projects and a few national projects and private contracts. Participants were from different European countries, including Spain, Belgium, Germany, Italy, the Netherlands, Ireland, Norway, Portugal and the UK. The feedback gathered during the workshop was analysed using affinity maps, a qualitative analysis method using a coding technique that helped organise the information gathered into groups or themes of similar items and topics to categorise and extract relevant knowledge (Lucero 2015). Common challenges that emerged from the workshop discussions included the following: (1) moving towards transdisciplinary co-production approaches, (2) testing visualisations with potential users, (3) the existence of a plethora of approaches to represent uncertainty in climate data, (4) the appropriate use of interactive elements in visualisations, (5) differences in terminology used by scientific and stakeholder communities and (6) taking the vernacular language of target audiences into account. These six challenges provided the basis for the present review, which investigates which are the good practices applied in other disciplines that can guide the development of enhanced climate service visualisations. Since the workshop included projects developing climate services at local, regional, European and global scales, the insights obtained from the analysis are considered valid for climate services providers worldwide.

2.2 What can the climate services field learn from other disciplines?

We performed a literature review to identify well-established practices in the disciplines of user experience, data visualisation, graphic design and psychology with the objective to compile a list of recommendations that can benefit the climate services field. A first set of recommendations was identified through desk review and the list was subsequently validated by 4 experts from the mentioned disciplines and adjusted according to the feedback received. A brief introduction of each discipline and the aspects related to the visual components that have been considered in this study are described in Sections 2.2.1–2.2.4.

2.2.1 User experience

User experience (UX) is a discipline that considers the user's feelings when using a product, application, system or service, the quality of the user's perception and how easy or pleasing to use the product or service is. UX encompasses various aspects, and some of them are considered in this study, such as:

- *User-centred design (UCD)*: iterative process that focuses on the user needs and involves users in each step of the development (Abras et al. 2004; Yucong et al. 2019).
- *Interaction design*: process in the area of human–computer interaction that focuses on creating digital products, visualisations and interfaces with logical behaviours and actions that allow users to manipulate and analyse information in an intuitive way (Yi et al. 2007; Dimara and Perin 2020).
- *Personalisation*: aspect that offers the users the possibility to customise how information is presented to them (Wiens et al. 2020).
- *Error prevention*: aspect that improves user experience by preventing users from making mistakes (Senders and Moray 1991).
- *Information architecture*: way in which users organise and structure the information available with the aim to favour the design of intuitive interfaces and minimise the user errors (Plaisant et al. 1998).

UX techniques have been broadly applied to various fields, and some applications in climate science include the design of apps, prototypes and decision support tools (Oakley and Daudert 2016; Argyle et al. 2017; Khamaj et al. 2019).

2.2.2 Data visualisation design

Data visualisation is the graphical representation of data or the translation of information into a visual context. It also refers to the techniques applied to represent this data by encoding it as visual objects (e.g., points, lines or bars) in graphics as well as the display of labels and legends (Wilke 2019). Some data visualisation design aspects considered in this study are as follows:

- *Visual information display*: relative positioning and sizing of the different components of a visualisation (Ware 2012).
- *Type of representation*: involves the choice of an optimal representation for data (e.g. heatmap, time series, time comparison, ranking, region map) according to the purpose of the visualisation and the type of information displayed (Kaye et al. 2012; McInerny et al. 2014).
- *Visual encoding*: use of visual channels to represent values, including colour, size, shape, orientation, brightness, texture or location (McInerny et al. 2014).
- *Labels and legends*: design of well-thought and clearly displayed labels and legends that contain the key information for the user to be able to correctly interpret the visualisation (no less and no more).

Fundamentals of data visualisation design have been applied to the visualisation of climate information and services with the aim to deliver complex information in a simplified way, e.g. through the reduction of dimensionality (Tsai 2010) and the representation of information through graphical features (Grainger et al. 2016; Christel et al. 2018).

2.2.3 Graphic design

Graphic design encompasses the use of a coherent 'look and feel' throughout the visualisation, which includes but is not limited to aspects considered for the definition of an appropriate layout, the use of a suitable background colour, contrast and alignment, as well as the correct presentation of information without copywriting or other errors (e.g. broken hyperlinks) (Moere and Purchase 2011; Zallio 2021). In the field of climate services, graphic design techniques have been applied to the visualisation of probabilistic information and improvement of its usability (Dasgupta et al. 2015; Christel et al. 2018; Terrado et al. 2018).

2.2.4 Psychology

The field of psychology applied to data visualisation studies human thought processes including attention, memory, perception, decision-making, problem-solving, language acquisition as well as how we understand and interpret visual information. Taking psychological aspects into account can help detect and solve initial design problems (Wagemans et al. 2012; Schiewe

2019), develop and evaluate visualisations that explicitly consider real-world user requirements (Block 2013). Psychology aspects considered in this study are as follows:

- *Perception and cognition*: analysis of human mental mechanisms that describe the way in which something is understood or interpreted as well as the acquisition of knowledge and understanding through senses, thought and experience (i.e. Gestalt psychology) (Rosli and Cabrera 2014).
- *Pre-attentive processes and attention*: study of how humans process visual stimuli from an array of given information and the speed with which users can process this information. It also involves users' selective attention, which helps to focus on the most relevant information while reducing distractions (Janes et al. 2013).
- *Memorability*: use of certain resources or visual coding that can favour users' recall of the information presented (Borkin et al. 2013).

In climate science, psychology methods have been applied to make information more accessible to expert and non-expert audiences (Harold et al. 2016) and to improve the efficiency of users during the performance of particular tasks (Hegarty et al. 2010; Calvo et al. 2021). The memorability aspect has been commonly used to communicate climate information to the general public (e.g. warming stripes, climate spiral).

3 Results and discussion

A number of heuristics (i.e. broad rules of thumb) and good practices from the fields of user experience, data visualisation design, graphic design and psychology are presented in the form of recommendations in Table 1. These recommendations not only refer to the visualisation itself but also apply to the entire climate service interface, which encompasses other elements of the climate service product. The recommendations try to cover a range of possible situations faced by climate service providers when developing visualisations to be used by stakeholders. However, this list must be understood in a flexible way, and climate service providers should reflect on the adequacy of the different recommendations in their contexts and taking their possibilities and resources into account. Also, depending on the type of climate service provided, its purpose, the target user and the decision at hand, some recommendations will simply not be relevant or applicable. The analysis of the effectiveness of these recommendations in particular cases is available in the literature and is out of the scope of the present review. Although recommendations are divided according to the different disciplines and aspects considered in this work, the boundaries among such divisions are porous and allow for some overlapping. Sections 3.1–3.4 expand on each of the recommendations and provide examples in the context of climate services, both described throughout the text and displayed in Fig. 1 (interactive tool with sub-seasonal and seasonal forecasts for the energy sector) and Fig. 2 (static representation with observations of sea ice extent and volume).

3.1 Recommendations from the field of user experience

3.1.1 User-centred design (UCD)

UCD is an iterative design process that starts with the identification and gathering of user needs. This can be done by involving stakeholders in knowledge exchange and co-learning

Table 1 Recommendations from other disciplines and aspects that can improve the efficiency of climate services visualisations

Discipline	Aspects	Recommendation
User experience	User-centred design (UCD)	- Gather user requirements using participatory approaches - Keep information simple and digestible - Involve users from the initial analysis to the final evaluation stage - Explore multiple visualisation options with users - Assess user performance in quantitative and qualitative terms
	Interaction design	- Include interactive elements and visual features appropriate for users' skills in an intuitive way (e.g. select, zoom, sort and filtering options) - Ensure that interactions have a visible response for users - Use progressive disclosure of information when appropriate - Include hyperlinks and tooltips for contextualisation - Consider multi-device visualisation options
	Personalisation	- Allow users to customise the display reflecting individual preferences
	Error prevention and system status	- Avoid ambiguous visual encoding of key information - Use selectors rather than open fields to minimise mistakes - Use indicators of system status - Use confirmation questions
	Information architecture	- Have a clear hierarchy of the information (e.g. take into account fields' dependencies, navigation) - Adapt terminology and language to the user's context - Use terminology consistently (e.g. avoid using different terms for the same concept) - Reduce jargon and ambiguous terms - Add tooltips and links to glossaries and extended descriptions - Ensure that help documentation is easy to find

Table 1 (continued)

Discipline	Aspects	Recommendation
Data visualisation design	Visual information display	- Use chart size and position to convey hierarchy and relevance (e.g. in dashboards or multiple simultaneous displays) - Use patterns and conventions that users are accustomed to - Embed the visualisation in a narrative context
	Type of representation	- Choose the optimal and familiar representations (e.g. bars, heatmaps) according to the objective of the visualisation - Favour representations with references to the real world
	Visual encoding	- Choose visual codification carefully, since some options might be difficult to interpret - Use visual codification that allows discrimination between relevant and non-relevant values - Make data visualisation inclusive (e.g. use contrast, colour-blind friendly palettes, clear fonts) - Select a colour palette that matches the nature of the data to be represented (e.g. qualitative, sequential, divergent) - Take into account the meaning associated to colour (e.g. use blue for wet or cold conditions)
	Labels/legends	- Include in the legend all relevant information required to interpret the visualisation - Locate legend in a visible area and, when possible, integrate it in the visualisation - Facilitate labels' reading
Graphic design	Design aspects	- Use appropriate design to support and improve user experience (e.g. colours, typography, display of information) - Be consistent in the visual language used in the visualisation, graphic elements and styles
Psychology	Perception and cognition	- Provide no more than the necessary information, but sufficient to ease user's understanding - Reduce the complexity of the information presented - Reduce spatial distance between similar elements and between these elements and the legend or caption - Use predictability, let the user successfully foresee the result of an interaction
	Pre-attentive processes and attention	- Use pre-attentive processing elements (i.e. indicating where to look first) - Reduce the use of elements that distract attention - Use visual metaphors coherently
	Memorability	- Use visualisation elements to reinforce memorability

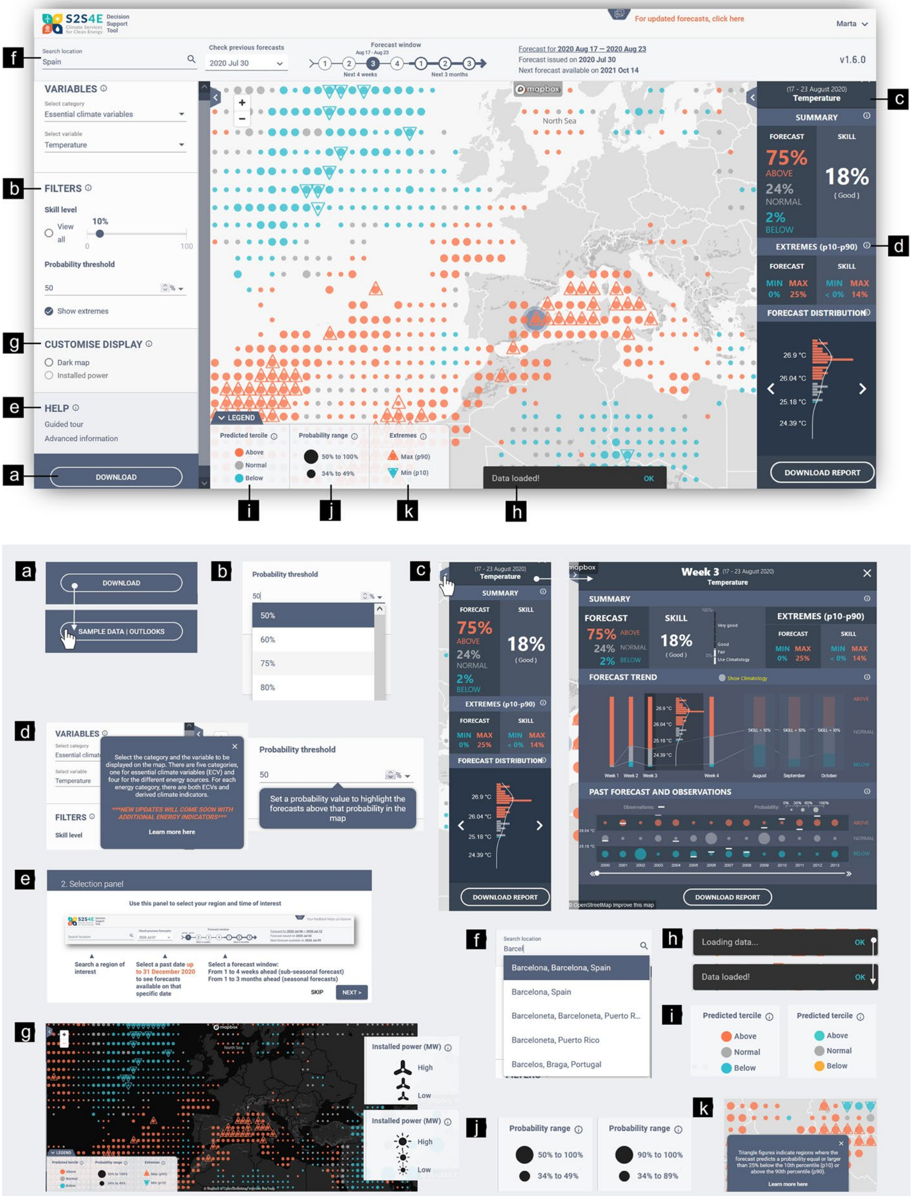


Fig. 1 Interactive decision support tool for the energy sector that considers aspects from different disciplines (source: S2S4E 2020). In the upper panel, highlighted features that are further expanded in the lower panel, exemplify particular aspects explained in the paper. *User experience aspects*: **a** button reflecting available actions; **b** possibility to filter information for skill, probabilities and extremes; **c** basic panel that can be expanded into advanced panel; **d** tooltips and hyperlinks; **e** help documentation section; **f** search location; **g** customisation options; **h** feedback of system status. *Visualisation design aspects*: **g** use of intuitive patterns and conventions, **i** and **j** glyph map variables' representation (e.g. temperature, precipitation) with circles of changing size and colour and use of contrasting colour hues in a dynamic legend. *Graphic design aspects*: **i** typography with good readability and use of colour blind-friendly palettes. *Psychology aspects*: **k** use of triangle symbols for extremes to enhance attention, **i** simple visual encoding

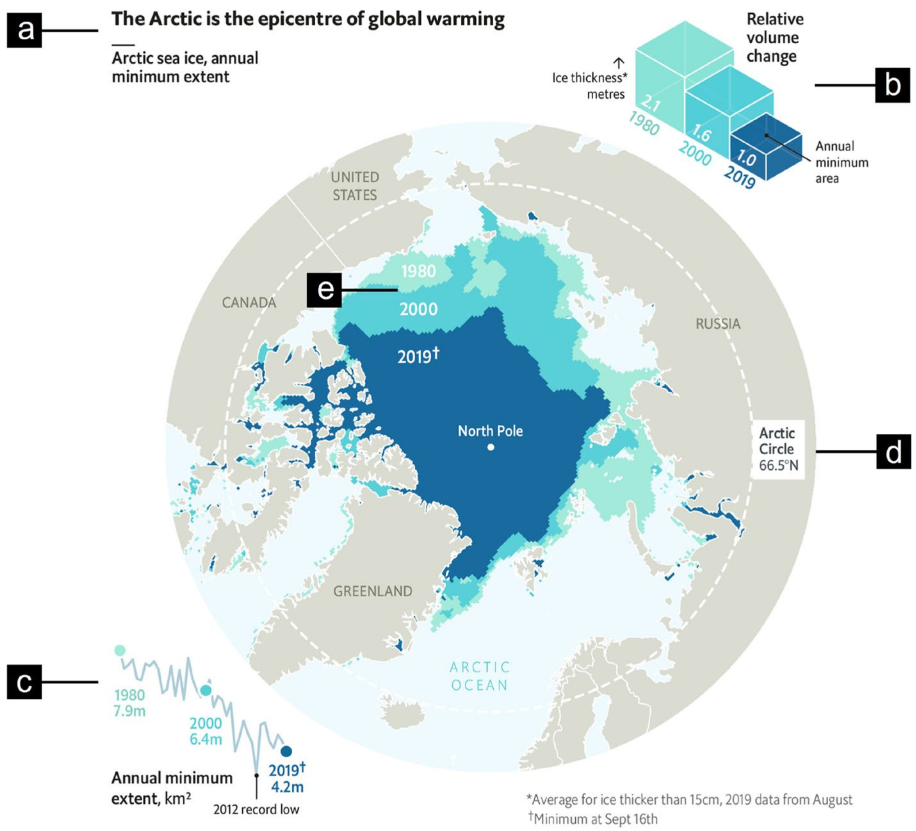


Fig. 2 Static representation of observed sea ice annual minimum extent and relative volume change that considers aspects from different disciplines (source: The Economist 2019). Highlighted features exemplify particular aspects explained in the paper. *User experience aspects*: **a**, **b** and **c** clear positioning of titles and legends. *Data visualisation design aspects*: **d** use of size and position to convey hierarchy, **e** visualisation embedded into a narrative context, **b**, **c** and **e** integration of different types of representations such as infographic, time series and region map, **d** use of soft hues in the base map to emphasise colour scale values, **e** use of single-colour inverted sequential scheme to represent ice, i.e. the darker the colours, the lower the ice values. *Graphic design aspects*: **b**, **c** and **e** consistent use of colours across different visualisation types. *Psychology aspects*: **b**, **c** and **e** use of redundancy to enhance understanding, **e** use of ice cubes metaphor to enhance memorability

using various participatory approaches, such as interviews, workshops, learning labs, surveys, user forums, etc. (Bojovic et al. 2021). Learning about user needs is key to ensure that climate service visualisations include the information required by users and that it is provided in a format that is simple and digestible. Looking at current practices in the climate services field, stakeholders are rarely involved in the initial stages of the visualisation development (Terrado et al. 2022). Rather, they are often reached at a later stage for the testing of the final visualisation. Although this is still useful to identify issues and ensure that the information is interpreted as intended by climate service providers, a truly transdisciplinary co-design approach goes beyond final testing and entails involving stakeholders from the very beginning and throughout all the stages of the visualisation co-development (McInerny et al. 2014; Lorenz et al. 2015; Grainger et al. 2020). This means co-exploring multiple visualisation options while keeping an 'agile' approach that allows to adapt outputs as user needs may change. The Double Diamond is a model that has been commonly used in the UX design process, and that puts special emphasis on the so-called divergent and convergent thinking, where first many different ideas are created before refining and narrowing them down to select the best one (Nguyen and Dupuis 2019).

Once the best option has been selected, it is important to co-evaluate the visualisation through quantitative and qualitative methods that go beyond merely asking users about their views on the service (Sharp et al. 2007; Calvo et al. 2021). Examples of quantitative metrics for user testing include the time used and the number of success, failure and abandonment rates when completing a particular task. These metrics can be complemented with the use of passive monitoring devices, such as eye-trackers, which allow to obtain more accurate measurements, e.g. pupillometry measurements, number of eye fixations, duration of the fixation and number of accesses to legend (Wang et al. 2014; William and Murugesh 2020). Other more advanced techniques even allow to measure the brain activity through electroencephalography (EEG) during the performance of tasks in order to evaluate the mental resources (i.e. cognitive load) that users use to carry them out (Anderson et al. 2011).

While some usability and effectiveness aspects are relatively easy to quantify, such as the increase in users' response speed or the decrease in error rates, other aspects are more challenging. For instance, the assessments of improvements in understanding and insights are highly subjective and need to rely on verbal feedback. Examples of qualitative methods to explore the user experience include the Think Aloud technique, which allows gathering users' comments and impressions during the tests (Charters 2003). The Bipolar Laddering method is also useful to quantify users' insights and their impact, by evaluating the positive and negative aspects found in the visualisation and rating them according to their importance (for positive aspects) and severity (for negative aspects) (Pifarré and Tomico 2007). Additionally, questionnaires such as the NASA-TLX also provide subjective but valuable opinions on cognitive load, perceived effort and frustration (Cao et al. 2009).

3.1.2 Interaction design

Recent developments in climate science are moving beyond the simple presentation of data and provide interactive visual tools for data exploration and analysis (Nocke 2014; Giuliani et al. 2017). This trend was also identified by climate service providers during the visualisation workshop (Terrado et al. 2022) and goes in line with the assumption that understanding of information can be improved through greater interaction (Yi et al. 2007). From the perspective of data visualisation, interaction design includes options related to the selection (i.e. mark something as interesting), exploration (i.e. show additional information), reconfiguration (i.e. show a different arrangement), encoding (i.e. show a different representation), abstraction or elaboration (i.e. show more or less detail), filtering (i.e. show information conditionally) and connection (i.e. show related items) of data and information. These options need to be used in an intuitive way and select those that are appropriate for the skills of potential users. In all cases, feedback matters, meaning that when users hover or click on elements of the visualisation, it is important that they receive a response to understand their command has been well-received (e.g. a button changing colour or size when hovering the mouse over it). An example of a button changing the text when hovering over it can be seen in the tool presented in Fig. 1a. In this case, the text changes from a general 'download' to a more specific 'sample data/outlooks'. These feedback mechanisms can guide users while using the climate service, not only by giving them better control but also strengthening trust in the dynamics of the service.

Functions such as zooming, through which users can change the scale of the representation, and filtering techniques, which allow users to change the set of information displayed based on some specific conditions, also fall in the interaction design category. When applying filters, only data items meeting the specified criteria are presented whereas those not satisfying the condition are hidden from the display or shown differently. Filters can be implemented giving the option to select ranges or particular values, either by moving sliders or clicking on checkboxes. Various filters are available in the interactive tool displayed in Fig. 1b. One of the filters allows the selection of the skill level, by either viewing all predictions regardless of their level of skill or by selecting the skill value to filter the predictions to be displayed through a slider. Another filter allows the selection of a probability value to highlight the forecasts above that probability in the map, either by selecting the value from a dropbox list or directly typing it. A third filter provides the option to show regions with extremely above and below normal probabilities by selecting a checkbox. Using such a variety of interaction options (i.e. for exploration, abstraction, filtering, etc.) empowers the users and facilitates their interpretation of the data displayed.

Other interaction techniques are used to highlight associations and relationships between data items, for instance when multiple views are used (e.g. in dashboards) or to reveal items that are initially not shown in the visualisation. A frequently used interaction mechanism in climate service visualisations is the progressive disclosure of information, which allows users to explore complex datasets and consume information step by step (McInerny et al. 2014). This mechanism can help prevent overwhelming effects that can derive from the high density and levels of information frequently related to climate data. In the interactive tool in Fig. 1, when the user clicks on a particular location in the map, a basic panel opens on the right with a summary of the main information of the prediction (Fig. 1c). However, the tool provides different levels of detail for different users' profiles and needs, since by expanding the panel, advanced users can see additional information on the forecast trend and past forecasts and observations. In general, including interactive elements helps climate service visualisation to feel much more responsive. However, it is important to consider that there may be limitations to how useful interactive visualisations are if the viewers do not have the required skills to interact with the presented information (diSessa 2004). Besides, a certain mistrust in interactivity has been identified in some interactions with scientists, mainly related to the arbitrariness of visual representations generated by interactive adjustments of user-defined thresholds or visualisation parameters (Nocke 2014). This brings us back to the need of applying a UCD and testing the visualisations with users.

Hyperlinks and tooltips are other types of interactive elements that can be included in climate service visualisations to provide some context (e.g. tooltips with terminology definitions, links to glossaries, guided tours to expand on the possibilities of the tool and other support systems). This information should be easy to access and, whenever possible, presented in the right order and at the right moment when the user may require it. The tool in Fig. 1 shows various examples of the use of tooltips and hyperlinks. Tooltips with descriptions and definitions can be accessed by clicking on the information buttons (Fig. 1d) next to key elements such as the titles of the navigation controls on the left bar, the map legend and the basic and advanced panels that open after selection of a particular map location. Some of the descriptions provided also contain a 'Learn more' hyperlink that brings the user to a webpage with advanced information. Other tooltips appear when the user hovers the mouse over some of the elements of the interface (Fig. 1d, lower panel), which provides additional information for the user to take action. The tool also contains a help section with access to a guided tour and advanced information (Fig. 1e).

Finally, interaction design also considers the new information consumption habits. Multi-device visualisation options (i.e. computers, mobile phones, tablets) need to be considered for climate service visualisations including options for data display optimised for different screen sizes.

3.1.3 Personalisation

When the visualisation of climate data is aimed at various audiences, basic visualisation options with interaction mechanisms may not suffice. Allowing visualisations to be tailored to reflect individual preferences, such as variables' preferred display thresholds, modification of visual properties like the dark/light mode or selection of a particular location, provide a high level of flexibility to the visualisation (Meijie et al. 2020; Calvo et al. 2021). Figure 1 shows some examples of customisation options, including the user selection of thresholds (e.g. making map icons visible or not in the map according to the selected skill threshold or assigning more relevance to specific probabilities through size) (Fig. 1b), selection of a location, either directly on the map or by searching it (Fig. 1f) and the customisation of the map with a different background colour and the addition of a layer of information about installed power, showing the total installed power of wind or solar energy in each region (Fig. 1g). Changing the actual display (e.g. providing more or less information depending on the user's background knowledge) is a very powerful type of personalisation and allows different users to see the visualisation in different ways, thus adapting it to their needs and enhancing its usability. For instance, the dark mode is often used to enhance contrast (e.g. when the visualisation is projected for an exhibition) whereas the light mode is commonly used for printing purposes. However, personal preferences also play a role, meaning that some users prefer or are more used to using one mode or the other. Some climate services even offer the possibility to save personalisation aspects, so the selected options are kept for future use (as in the case of the tool presented in Fig. 1).

3.1.4 Error prevention

Exit and recovery options allow users to remain in control of the system and avoid getting stuck and feeling frustrated. Moreover, when it is easy for users to back out of a process or undo an action, it fosters a sense of freedom and confidence in the service. However, a good user experience goes beyond allowing recovery once a mistake has been made, but it also prevents users from making certain errors, which are often entrained by mismatches between the users' mental model and the visualisation design. Error prevention in visualisations can be implemented through subtle design solutions, such as avoiding the use of ambiguous visual encoding of key information (e.g. the counter-intuitive use of green colour for warnings). Likewise, it is important to keep in mind that icons, images, terms and concepts that seem clear to particular communities or disciplines may be unfamiliar or confusing for others.

The use of selectors rather than open fields can also help minimise potential mistakes (e.g. typing). Hence the widespread use of selectors in climate services visualisations, with a number of variables, time scales, geographical regions, climate models or climate change scenarios to choose from (see selectors for the different navigation controls on the left and top bars in Fig. 1). However, particular attention should be paid to the programming of selectors and other interactional elements to make them clear, intuitive and fluid for everyone, including people with disabilities.

Error prevention can also be addressed through other more obvious design solutions, such as messages and indicators of the system status (e.g. loading messages to indicate that the user needs to wait). This is especially relevant in the case of climate data, which uses a huge amount of computational resources, affecting the reaction time of climate services platforms (e.g. see feedback provided in the decision support tool in Fig. 1h during and at the end of the process completion). A good visualisation design should always keep users informed through appropriate feedback within a reasonable amount of time, since continuous communication helps create trust in the climate service product and its provider. Confirmation questions warning users against certain actions are also useful to prevent errors. Overall, error prevention has been proved to lead to more successful task completion and reduction of the time required, therefore contributing to a better user experience (Maxion and Reeder 2005).

3.1.5 Information architecture

Information architecture involves the consideration of different strategies for the design of intuitive data visualisations that enhance navigation (Plaisant et al. 1998). This includes having a clear hierarchy of the information that takes into account dependencies among the presented fields and navigation aspects when placing different settings and filters. This hierarchy can also be used to convey the relative relevance of the elements of the visualisation and the key aspects highlighted by each of them (Janes et al. 2013; Shahar 2019). Although this may seem like the natural way to organise information, it is not always applied in practice, sometimes resulting in confusion and frustration on the side of the users. In Fig. 1, navigation controls in the left and top bars are placed from the most general to the most specific to enhance navigation, taking dependencies between fields into account (e.g. through dynamic options presented in the selectors), which are reflected in the order these fields are presented. Attention was also paid to the appropriate placement of tooltips throughout the interface and to the placement of the legend, which is integrated within the visualisation itself. Information architecture is also present in the static representation displayed in Fig. 2a, b and c, which has a clear positioning of titles and legends.

Information architecture also involves adapting the language and terminology used in the visualisation to the users' context (i.e. taking into account the user's vernacular language, expertise and background). Consideration of the users' vernacular language has often been an overlooked usability factor that causes users to abandon further interaction with the service because they are unable to easily select a language they understand. In this regard, Miraz et al. (2016) show that combining the IP address with the language of the users' browser can eliminate this problem. However, Miraz and co-authors are critical about the use of automated non-supervised translation, which often fails to address abbreviations, metaphors and cultural terms. Using terminology consistently throughout the visualisations (i.e. avoiding different terms for the same concept), and reducing jargon and ambiguous terms (e.g. specific to some research disciplines but disconnected from the user's context or that hold different meanings in different contexts or fields) is recommended.

Including tooltips and links to glossaries, thesauruses or extended descriptions as well as ensuring that the help documentation is easy to find is a visualisation aspect related to the architecture of the information. An example of a clear indication of help documentation is given in Fig. 1e, where the left bar contains a help section with access to a guided tour of the tool and to advanced information. Special attention to these elements has been increasingly paid by climate services providers looking for bridging the terminology gap with stakeholders (Terrado et al. 2022).

3.2 Recommendations from the field of data visualisation design

3.2.1 Visual information display

The display of visual information refers to the relative positioning and sizing of different elements in a visualisation. Thus, chart and legend size and position, as well as labels, can be used in a visualisation to convey hierarchy and relevance (Dykes et al. 2010). This is an important aspect, since different visualisation displays of the same information have been found to lead to dramatically different task performances by users (Hegarty 2011). Size and position have been used in Fig. 2d to convey hierarchy by placing the main map showing the changes in Arctic sea ice extent with a bigger size at the centre of the visualisation, with additional smaller representations around it to complement the main message. The display of visual information is especially relevant in the case of dashboards or multiple simultaneous displays, in which there is a need to compare different types of information to make an informed decision. In this regard, various climate services targeting users in the agriculture sector have been developed following a dashboard format, since users need to compare different types of agro-climatic information to make a particular decision (Marcos-Matamoros et al. 2020a, b).

The use of patterns and conventions that the users are accustomed to (e.g. because they are common to other similar interactions) is highly recommended. Using known patterns creates an experience that feels more intuitive for users, making visualisations faster to learn and easier to remember and helping them act with more confidence (Plaisant et al. 1998). Similarly, using a design in which visualisation elements represent the desired outcome following real-world conventions is called *natural mapping* (e.g. use certain reference icons to represent certain variables) and can also enhance user experience. See for example the icons used to represent installed solar and wind power in the climate service tool presented in Fig. 1g (lower panel) or the orientation of the triangular icons used to highlight extreme high and low values (Fig. 1k).

Finally, embedding the visualisation into a narrative context is important because it can help audiences make sense of complex information (Segel and Heer 2010; Figueiras 2014). This can be done by combining graphs to easily compare them or spot relevant changes or trends in the data (Krzywinski and Cairo 2013; Grainger et al. 2016; IPCC 2021). An example of narrative is shown in Fig. 2e, where the different representations embedded in the visualisation tell the story of sea ice melting due to global warming, allowing readers to follow the course of the recession of sea ice both in extent and volume through time (from 1980 to 2019).

3.2.2 Type of representation

Graphs and maps help to understand scientific data, find patterns, identify trends and tell a story. The choice of the optimal representation for a dataset will depend on the visualisation purpose as well as the level of background knowledge of the intended audience and their familiarity with different visualisation formats (McInerny et al. 2014; Raaphorst et al. 2020). In terms of purpose, certain visualisations favour comparisons whereas others focus on the evolution over time or the composition of a set. Maps incorporate a geographical component, which is crucial for spatially explicit data, and therefore, have been extensively used in climate services to represent climate variability and risks, e.g. climate anomalies (Lledó et al. 2018) or tornado threats (Klockow-McClain et al. 2019). Glyph maps are another technique used to represent information, in which single data points are encoded individually, by using different visual channels (shape, size, colour, etc.) in a variety of combinations. A glyph map was used to represent information in Fig. 1, representing values with circles of changing size and colour (Fig. 1i and j). On the other hand, the visualisation in Fig. 2 includes other types of representations, according to their purpose, either to highlight changes over time or space. Thus, while the central map Fig. 2d mostly focuses on showing the spatial reduction of sea ice extent (although it also contains the temporal component by including the extent for three different years), the line chart in Fig. 2c works better to show the decrease of sea ice with time.

The lack of a standard statistical mapping approach has resulted in a vast body of literature on the representation of climate data and its inherent uncertainty (Kaye et al. 2012; Daron et al. 2015; Grainger et al. 2016). In general, users tend to prefer familiar representation formats to novel ones, even if they may not necessarily be the optimal or the best understood (Lorenz et al. 2015; Taylor et al. 2015). Specificities associated to the different types of maps have been identified, e.g. perceptual area distortions affecting choropleth and heatmaps (Speckmann and Verbeek 2010) or overlapping problems of glyph maps (Fuchs et al. 2017), which need to be taken into account when choosing a particular type of representation. Therefore, it is important to keep enough distance between different glyphs, as in the case of Fig. 1, and to adjust these distances to the different levels of zoom (i.e. levels of detail) applied in interactive tools.

The representation of geospatial data using realistic displays has been found to convey more confidence to users (Fabrikant and Lobben 2009). Hence, the success of Google maps and its virtual representation of the surroundings: people want to find themselves. This preference has also been observed in the climate services field, where users prefer detailed map representations (smaller grid cells) that look closer to reality, even if the resolution of the information provided, and therefore their usefulness, is not enhanced. This has been described as *naïve realism* in the literature, reflecting the dichotomy between visualisation preference and effectiveness and showing that users' preferences for types of displays are not necessarily a good indicator of how effective and understandable these representations are (Fabrikant and Lobben 2009; Taylor et al. 2015).

3.2.3 Visual encoding

Visual encoding refers to the use of visual channels to represent values, including colour, size, shape, orientation, brightness, texture and location. Various solutions for visual encoding have been explored in the climate and environmental sciences literature (Light and Bartlein 2004; Kaye et al. 2012), resulting in the recommendation to carefully choose visual codification, since some options might be difficult to interpret. For instance, the use of paleness or colour saturation as a method to illustrate uncertainty has been questioned (MacEachren et al. 2005), since when combined with colour and reduced glyph thickness, has been found to compromise graphical interpretation (Ware 2012).

Visual encoding can be used to favour discrimination among relevant and non-relevant values in a visualisation. For instance, glyph colours similar to the visualisation background colour are recommended to display non-relevant values, making users pay less attention to them. Conversely, colours considerably different from the background, especially those that stimulate more photoreceptors in the human eye (e.g. green or yellow colour hues), are more easily detected by users (Ramamurthy and Lakshminarayanan 2015), and therefore, are recommended for relevant values. This strategy was applied in Fig. 1i. In that case, as target users were interested in situations departing from normal conditions, normal variable values (for temperature, precipitation, wind, etc.) were represented in grey colour whereas upper and lower than average values used a blue-orange colour hue, indicating conditions in which users needed to take action. Glyph size (i.e. small and big circles) was also used to discriminate between low- and high-probability situations. Additionally, the possibility of seeing the glyph or not was provided according to the user's preferences selected in Fig. 1b. Another example of the use of colour is displayed in Fig. 2d, where soft hues are used for the land and the ocean to support colour scales, allowing readers to focus their attention on the sea ice representation.

A commonly overlooked consideration in scientific graphs is the perception by users with disabilities. For instance, colour-blind friendly palettes are recommended to ensure that maps are usable by individuals with colour-deficient vision. This recommendation was explicitly considered for the selection of the colour palettes in the climate service presented in Fig. 1. Specific software that allows to simulate how the visualisation is seen by different types of colour-blind users is available and can be used for making data visualisations more inclusive using colour-blind friendly palettes. Other good practices, such as the use of contrast, clear fonts, white spaces or complementing the use of colour with pattern or texture, have been found useful to favour people with visual impairment (Dong et al. 2010; Marriott et al 2021).

Some broad rules of thumb include the following: avoiding spectral schemes (e.g. rainbow colour schemes) to represent sequential data (Brewer 1997), using yellow with care and using colour intensity to reinforce hue as a visual indicator of magnitude. Colour intensity, also referred to as lightness, brightness or luminosity, provides perceptual ordering for all readers (Kaye et al. 2012), whereas hue is what we typically refer to as colour (e.g. red, blue, green, etc.). At the same time, it is important to select a colour palette that matches the nature of the data to be represented. Qualitative colour palettes work best for the display of categorical variables (e.g. land use categories). Colours used in these palettes should be distinct (i.e. use different intensity, have enough contrast) to ensure accessibility. When mapping data that can be ordered (e.g. probabilities, anomalies), either sequential or divergent colour palettes can be used. Sequential palettes use hue or lightness or a combination of both to create a continuous set and work best for numeric variables that need to be placed in a specific order. On the other hand, divergent colour palettes are a combination of two sequential palettes with a central value in the middle (usually zero) and are often used to communicate positive and negative values. Figure 1i shows an example of contrasting colour hues for the representation of the predicted terciles, while Fig. 2e uses colour hues with increasing intensity to represent the sea ice extent at three different moments in time.

Another aspect that has received large attention is the use of appropriate colour symbolism. For readers of weather and climate maps, a useful association exists with blue and red colours used in global future temperature maps and scenarios graphs (Kaye et al. 2012). Red colours are used to mark maximum values, temperature increases, great risk, anomalies and worst-case scenarios like the RCP8.5 scenario, whereas blue colours denote cold temperatures and precipitation increases, but also illustrate best case scenarios (Schneider and Nocke 2018). Even though this symbolism might not exist for all climate-related variables, ignoring it when it exists could hinder users' interpretation. Thus, following these associations, reversed colour scales were defined in the service shown in Fig. 1i for precipitation and temperature, with red hues associated with above normal temperature and below normal precipitation and blue hues associated with above normal precipitation and below normal temperature. An example of non-conventional colour scale is provided in Fig. 2b, c and e, which uses an inverted scale where the more intense colours are not associated with higher values but rather to the lower ones. Traditionally, for the representation of sea ice, lighter colours closer to white have been used to represent sea ice whereas darker colours have been used for the ocean. However, in this case, the scale has been inverted to emphasise the reduction of sea ice. Despite having a clear rationale, this colour choice can be perceived as counterintuitive for some readers. On the other hand, colour also influences the credibility and understanding of climate change visualisations, since it can impact the emotional and associative reaction of viewers. Avoiding too dark colours at the highest temperature differences has been suggested in some studies to prevent disillusioning associations and feelings of powerlessness that do not call for action (Schneider and Nocke 2018).

3.2.4 Labels/legends

Labels and legends are important elements of the visualisation that need to be well-thought and clearly displayed, since they contain the key information for users to interpret the visualisation correctly. Thus, all the relevant information needs to be included in the legend while additional or non-relevant information should be avoided. This is especially important when using interaction techniques, like zooming in and out. In this case, optimal legends should be dynamic, allowing the legend to simultaneously reflect changes in the visualisation. For instance, the number of categories shown in the legend should reflect the user's selections. An example of dynamic legend was used in the interactive tool shown in Fig. 1j, where the probability range to define the legend categories changed according to the threshold defined by the user in the left bar. It is also important that symbols in the legend relate directly to those used in the map, avoiding additional symbols that cannot be found in the map to prevent confusion.

In addition, the legend should be located in a visible area within the visualisation and, when possible, integrated in the chart to reduce user's cognitive load and facilitate reading (Dykes et al. 2010; Hogräfer et al. 2020). Apart from reducing the distance between the legend and the visualisation, reading can also be enhanced by using horizontal text, appropriate font size and by adding a descriptive title and variable units to the legend. A descriptive title is one that synthesises the main message of the graphic rather than just focusing on a technical description containing, for instance, information on the model, the variable and time scale of the simulation, as it is normally provided as a direct model output. An example of descriptive title is given in Fig. 2a ('The Arctic is the epicentre of global warming').

3.3 Recommendations from the field of graphic design

3.3.1 Design aspects

Graphic design is applied to improve the readability and understandability of information, allowing a wider audience to benefit from scientific results. This fosters the democratisation of knowledge across disciplines and stakeholders (Zallio 2021). Apart from having aesthetic applications, graphic design also provides credibility, consistency and can improve the user experience through the appropriate use of colour, legible typography and a coherent display of information (Moere and Purchase 2011; Zallio 2021). For instance, Fig. 1 uses a typography with a good readability and the elements of the interface are grouped in different sections in the left and right bars, with titles consistently highlighting the most important aspects.

Colour preference varies among different socio-cultural groups and therefore, the choice of a particular colour can have a direct impact on the speed of interactivity and overall user satisfaction (Miraz et al. 2016). This is because colour has an attached cultural significance (e.g. red is often associated with love in the Western World, to misery or mourning in South Africa and to happiness or luck in China). It is also important to be consistent in the visual language, graphic elements and styles used in the context of the visualisation (e.g. if blue is used to indicate cold conditions, it is important to stick to this meaning throughout the visualisation). Consistency aspects influence users' learning when interacting with the visualisation, and therefore, the correct manipulation of the information provided. Overall, design aspects contribute to provide a sense of readiness to the climate service, supporting its trustworthy perception by users. Figure 2 makes a consistent use of colour, for instance, applying the same colours, related to each of the years, throughout the three representations included in the visualisation.

3.4 Recommendations from the field of psychology

3.4.1 Perception and cognition

To favour users' decision-making, visualisations should be designed taking into account the limitations in the users' working memory. A recommendation to reduce users' cognitive load and enhance task performance consists in presenting no more than the necessary information but sufficient to ease user's understanding (Hegarty 2011). In general, presenting less information than needed in visualisations obliges users to keep a detailed representation of the information in their working memory, whereas presenting too much information can lead to visual cluttering or distraction (Rosenholtz et al. 2007; Wickens & Carswell 1995). Another recommendation consists in reducing as much as possible the complexity of the information presented, using a less is more approach, which can foster faster understanding. In the same line, using redundancy in visualisations (through the use of colours and sizes) can also favour a quick understanding, as shown in Fig. 2, where various types of representations are brought together to provide evidence that the sea ice is melting. Reducing the visual path between the visualisation and other auxiliary elements of consultation, such as labels, captions and legends also enhances user cognitive processes (Kause et al. 2020). This aspect is illustrated in both representations shown in Figs. 1 and 2, although in Fig. 2 could be improved by moving the footnote with additional information closer to the corresponding charts.

Apart from being appropriately presented, to be effective, a visual display needs to be accurately perceived. This is due to the principle of discriminability (Kosslyn 2006), which indicates that two properties must differ by a large enough proportion to be distinguished, since our visual systems register relative proportions rather than absolute amounts. Conforming to this principle involves a careful selection of the visual encoding, i.e. appropriate number of categories, glyph size and colour. Other aspects, such as predictability, can also influence visual perception. A service is predictable when a user can foresee the result of an interaction and therefore, it matches the user's expectations. When users do not know what to expect or they expect something different to what they experience, the usability of the climate service is impaired. Assessing how these different aspects of a visualisation are perceived cannot only rely on intuition but demands objective measures (see Section 4.1 on testing of visualisations).

3.4.2 Pre-attentive processes and attention

Climate services need to be designed in such a way that salient information is pushed to the users, capturing their attention and informing them of either unexpected or unforeseen situations or anomalies. There are different mechanisms that can be used to draw users' attention, although care should be paid not to overuse them. Pre-attentive elements are graphical properties, such as form, colour or motion that cause people to process information before they can pay conscious attention to it. Using pre-attentive elements has been proved an effective way to direct the user's attention and boost faster understanding (Janes et al. 2013). Pre-attentive elements commonly used in climate services are eye-stimulating colours for user-relevant values (see Fig. 1i) and different glyph shapes (see Fig. 1j). In Fig. 1j, triangle figures indicate regions where the forecast predicts a high probability of having extreme values. In this case, a fast interpretation is expedited by both colour and the direction of the triangle vertex. The use of non-relevant information that may distract users from their main focus (e.g. animations or flashy components) as well as superfluous elements, incoherent visual metaphors and unfamiliar acronyms and terminology should be avoided.

3.4.3 Memorability

Memorability refers to the use of visual resources that favour user's recall of the information presented. This can be reinforced by the appropriate use of pictograms or visualisation elements that catch users' attention (Borkin et al. 2013; Borkin 2016). Quantifying the memorability of a visualisation is a general metric of the utility of information and an essential step towards determining how to design effective visualisations. It has been shown that attributes like colour enhance memorability and that common graphs are less memorable than unique visualisation types (Borkin et al. 2013). Through the collaboration with designers, data visualisation experts and journalists, memorability has been a broadly considered aspect in the climate services field applied, for instance, to wind data visualisations such as the Windy and Project Ukko (Windyty 2014; Christel et al. 2018). The use of ice cubes in Fig. 2b is also an aspect that enhances memorability through the use of a visual metaphor.

4 Conclusions

The visualisation taxonomy covers a large variety of static and dynamic visualisation types (e.g. charts, graphs, maps) used across social and scientific domains. In the field of climate services, visualisations are used to facilitate readers' identification and understanding of patterns in climate data. However, when not appropriately co-produced, visualisations can exclude readers with visual or cognitive disabilities or those that lack appropriate background knowledge to correctly interpret climate information. This review paper aims to answer the need raised by climate service providers involved in a previous visualisation workshop, of having a list of recommendations to improve visualisations in climate services. For that, in this work, we review good practices commonly applied in other disciplines like user experience, data visualisation design, graphic design and psychology that can be a useful asset for the development of more effective visualisations, given the current lack of guidelines. In addition, as literature published by visualisation-related disciplines is rarely accessed by climate services providers, this list of recommendations and associated publications facilitates the access to this domain knowledge. The separate presentation of user experience, data visualisation design, graphic design and psychology disciplines in the present work obeys only to criteria aimed at facilitating the information flow for readers. However, practices from these fields are rarely applied in isolation, as reflected in the two examples included in this study. Thus, apart from being heavily grounded on the user needs, these disciplines complement each other and are all required for the development of visualisations of climate data and services. This aligns with the feedback provided by four experts from the mentioned disciplines, who highlighted that many of the aspects and recommendations provided in Table 1 could not be only mentioned under one particular discipline, and clearly advocates for the need of transdisciplinary approaches for the co-production of visualisations in climate services.

The recommendations provided in this work highlight useful considerations overlooked by the climate services community, often unaware of the good practices in fields beyond climate science. The list of recommendations is comprehensive yet generic, spanning different target groups (including experts and non-experts in climate science), needs and situations. However, not all recommendations are applicable or relevant to every case and, when relevant, it will be ultimately the task of the climate service providers to balance their possibilities and costs of implementing a recommendation against the potential benefits on the visualisation's usability, accessibility and effectiveness. Moreover, the application of some recommendations can be limited by a lack of the appropriate expertise in the research team or the need for a timely delivery of the results. Nevertheless, we argue that these recommendations can be the departure point for climate service providers willing to increase the impact of visualisations or interested in delving into more specialised literature on the topic.

The lack of guidance for the visualisation of climate information resonates with the discussions on the need for standards for climate services, especially in a time of unprecedented climate impacts affecting society. There is now more than ever an increasing demand for quality-assured climate services that are fit to support mitigation and adaptation strategies to climate change and variability. At this point, intuition is not enough guarantee that a visualisation of climate data works as intended by the climate science community; therefore, testing visualisations through application of quantitative or qualitative assessment methods is required. We show that only by including the expertise from disciplines beyond climate science will the climate services field be able to move towards the co-production of more effective and inclusive visualisations. This will contribute to build trust in science for society, facilitate the appropriate use of climate information and finally boost the uptake of climate services by decision-making and policy actors.

Acknowledgements The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements no. 689029 (Climateurope), 776787 (S2S4E), 776467 (MED-GOLD) and 869565 (VitiGEOSS). The authors would also like to thank the experts that provided feedback to the recommendations delivered in this paper and that ensured their completeness and broader clarity: Diana Urquiza and Diana F. Vélez from BSC as well as Leonor Solis from UNAM.

Author contribution MT: conceptualisation, literature search, writing-original draft preparation, writing-review and editing. LC: conceptualisation, literature search, writing-review and editing. IC: conceptualisation, writing-review and editing.

Funding The research leading to these results received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements no. 689029 (Climateurope), 776787 (S2S4E), 776467 (MED-GOLD) and 869565 (VitiGEOSS).

Data availability Not applicable.

Declarations

Ethics approval This article does not contain any studies with human or animal participants performed by any of the authors.

Consent to participate Not applicable.

Consent to publish Not applicable.

Conflict of interest The author declares no competing interests.

References

Abras C, Maloney-Krichmar D, Preece J (2004) User-centered design. In: Bainbridge W (ed) Encyclopedia of Human-Computer Interaction. Sage Publications, Thousand Oaks

Anderson EW, Potter KC, Matzen LE, Shepherd JF, Preston GA, Silva CT (2011) User study of visualization effectiveness using EEG and cognitive load. Comput Graph Forum 30(3):791–800. https://doi.org/10.1111/j.1467-8659.2011.01928.x

Argyle EM, Gourley JJ, Flamig ZL, Hansen T, Manross K (2017) Toward a user-centered design of a weather forecasting decision-support tool. Bull Am Meteorol Soc 98(2):373–382. https://doi.org/10.1175/BAMS-D-16-0031.1

Ash KD, Schumann RL, Bowser GC (2014) Tornado warning trade-offs: evaluating choices for visually communicating risk. Weather Clim Soc 6:104–118. https://doi.org/10.1175/WCAS-D-13-00021.1

Block G (2013) Reducing cognitive load using adaptive uncertainty visualization. Dissertation, Nova Southeastern University

Bojovic D, St. Clair AL, Christel I, Terrado M, Stanzel P, Gonzalez P, Palin EJ (2021) Engagement, involvement and empowerment: three realms of a co-production framework for climate services. Global Environ Change 68:102271. https://doi.org/10.1016/j.gloenvcha.2021.102271
Borkin MA, Vo AA, Bylinskii Z, Isola P, Sunkavalli S, Oliva A, Pfister H (2013) What makes a visualization memorable? IEEE Trans Visual Comput Graph 19(12):2306–2315. https://doi.org/10.1109/TVCG.2013.234
Borkin MA (2016) Beyond memorability: visualization recognition and recall. IEEE Trans Visual Comput Graph 22 (1). https://doi.org/10.1109/TVCG.2015.2467732
Brewer CA (1997) Spectral schemes: controversial color use on maps. Cartogr Geogr Inf Syst 24(4):203–220. https://doi.org/10.1559/152304097782439231
Calvo L, Christel I, Terrado M, Cucchietti FM, Pérez-Montoro M (2021) Users cognitive load: a key aspect to successfully communicate visual climate information. Bull Am Meteorol Soc. https://doi.org/10.1175/BAMS-D-20-0166.1
Cao A, Chintamani KK, Pandya AK, Ellis RD (2009) NASA TLX: software for assessing subjective mental workload. Behav Res Methods 41(1):113–117. https://doi.org/10.3758/BRM.41.1.113
Charters E (2003) The use of think-aloud methods in qualitative research: an introduction to think-aloud methods. Brock Educ J 12(2):68–82. https://doi.org/10.26522/brocked.v12i2.38
Christel I, Hemment D, Bojovic D, Cucchietti F, Calvo L, Stefaner M, Buontempo C (2018) Introducing design in the development of effective climate services. Clim Serv 9:111–121. https://doi.org/10.1016/j.cliser.2017.06.002
Daron JD, Lorenz S, Wolski P, Blamey RC, Jack C (2015) Interpreting climate data visualisations to inform adaptation decisions. Clim Risk Manage 10:17–26. https://doi.org/10.1016/j.crm.2015.06.007
Daron J, Lorenz S, Taylor A, Dessai S (2021) Communicating future climate projections of precipitation change. Clim Chang 166:23. https://doi.org/10.1007/s10584-021-03118-9
Dasgupta A, Poco J, Wei Y, Cook R, Bertini E, Silva CT (2015) Bridging theory with practice: an exploratory study of visualization use and design for climate model comparison. IEEE Trans Visual Comput Graph 21(9):996–1014. https://doi.org/10.1109/TVCG.2015.2413774
Davis M, Lowe R, Steffen S, Doblas-Reyes FJ, Rodó X (2015) Barriers to using climate information: challenges in communicating probabilistic forecasts to decision makers. In: Drake JL et al (eds) Communicating climate change and natural hazard risk and cultivating resilience: case studies for a multidisciplinary approach. Springer International Publishing, Cham, Switzerland, pp 95–113. https://doi.org/10.1007/978-3-319-20161-0_7
Dimara E, Perin C (2020) What is interaction for data visualization? IEEE Trans Visual Comput Graph 26(1):119–129. https://doi.org/10.1109/TVCG.2019.2934283
diSessa AA (2004) Metarepresentation: native competence and targets for instruction. Cogn Instr 22(3):293–331. https://doi.org/10.1207/s1532690xci2203_2
Dong H, Barr G, Blackburn E, Grant M, Piwek P, Shepherd P, Collins N (2010) Inclusive data visualisation: a multidisciplinary approach. Proceedings of the 3rd International Conference for Universal Design, University of Bath, UK, pp 10–30
Dykes J, Wood J, Slingsgy A (2010) Rethinking map legends with visualization. IEEE Trans Visual Comput Graph 16(6):890–899. https://doi.org/10.1109/TVCG.2010.191
Estrada FCR, Davis LS (2015) Improving visual communication of science through the incorporation of graphic design theories and practices into science communication. Sci Commun 37(1):140–148. https://doi.org/10.1177/1075547014562914
Fabrikant SI, Lobben A (2009) Guest editor's introduction: cognitive issues in geographic information visualization. Cartographica 44:129–144. https://doi.org/10.3138/carto.44.3.139
Figueiras A (2014) Narrative visualization: a case study of how to incorporate narrative elements in existing visualizations. IEEE 18th International Conference on Information Visualisation, Paris, France. https://doi.org/10.1109/IV.2014.79
Fuchs J, Isenberg P, Bezerianos A, Keim D (2017) A systematic review of experimental studies on data glyphs. IEEE Trans Visual Comput Graph Inst Electr Electron Eng 23(7):1863–1879. https://doi.org/10.1109/TVCG.2016.2549018
Giuliani G, Nativi S, Obregon A, Beniston M, Lehmann A (2017) Spatially enabling the global framework for climate services: reviewing geospatial solutions to efficiently share and integrate climate data & information. Clim Serv 8:44–58. https://doi.org/10.1016/j.cliser.2017.08.003
Grainger S, Mao F, Buytaert W (2016) Environmental data visualisation for non-scientific contexts: literature review and design framework. Environ Modell Softw 85:299–318. https://doi.org/10.1016/j.envsoft.2016.09.004

Grainger S, Ochoa-Tocachi BF, Antiporta J, Dewulf A, Buytaert W (2020) Tailoring infographics on water resources through iterative, user-centered design: a case study in the Peruvian Andes. Water Resour Res 56:e2019WR026694. https://doi.org/10.1029/2019WR026694

Haase H, Bock M, Hergenröther E, Knöpfle C, Koppert HJ, Schröder F, Trembilski A, Weidenhausen J (2000) Meteorology meets computer graphics – a look at a wide range of weather visualisations for diverse audiences. Comput Graphics 24(3):391–397. https://doi.org/10.1016/S0097-8493(00)00035-2

Harold J, Lorenzoni I, Shipley TF, Coventry KR (2016) Cognitive and psychological science insights to improve climate change data visualization. Nat Clim Chang 6(12):1080–1089. https://doi.org/10.1038/nclimate3162

Hegarty M, Canham M, Fabrikant S (2010) Thinking about the weather: how display salience and knowledge affect performance in a graphic inference task. J Exp Psychol Learn Mem Cognit 36(1):37–53. https://doi.org/10.1037/a0017683

Hegarty M (2011) The cognitive science of visual-spatial displays: implications for design. Top Cognit Sci 3:446–474. https://doi.org/10.1111/j.1756-8765.2011.01150.x

Hewitt CD, Guglielmo F, Joussaume S, Bessembinder J, Christel I, Doblas-Reyes FJ, Djurdjevic V, Garrett N, Kjellström E, Krzic A, Máñez-Costa M, St. Clair AL (2020) Recommendations for future research priorities for climate modelling and climate services. Bull Am Meteorol Soc 102(3):578–588. https://doi.org/10.1175/BAMS-D-20-0103.1

Hogräfer M, Heitzler M, Schulz HJ (2020) The state of the art in map-like visualization. Comput Graph Forum 39:647–674. https://doi.org/10.1111/cgf.14031

IPCC (2021) IPCC WGI Interactive Atlas: REGIONAl information. Intergovernmental Panel on Climate Change. https://interactive-atlas.ipcc.ch. Accessed 25 Sept 2021

Janes A, Sillitti A, Succi G (2013) Effective dashboard design. Cut IT J 26(1):17–24

Kamateri E, Panopoulou E, Tambouris E, Tarabanis K, Ojo A, Lee D, Price D (2015) A comparative analysis of tools and technologies for policy making. In: Janssen M, Wimmer MA, Deljoo A (eds) Policy practice and digital science. Springer International Publishing, Switzerland, pp 125–156

Kause A, Bruine de Bruin W, Fung F, Taylor A, Lowe J (2020) Visualizations of projected rainfall change in the United Kingdom: an interview study about user perceptions. Sustainability 12(7):2955. https://doi.org/10.3390/su12072955

Kaye NR, Hartley A, Hemming D (2012) Mapping the climate: guidance on appropriate techniques to map climate variables and their uncertainty. Geosci Model Dev Discuss 4(3):1875–1906. https://doi.org/10.5194/gmdd-4-1875-2011

Khamaj A, Kang Z, Argyle E (2019) Users' perceptions of smartphone weather applications' usability. Proc Hum Factors Ergon Soc Annu Meet 63(1):2216–2220. https://doi.org/10.1177/1071181319631098

Klockow-McClain KE, McPherson RA, Thomas RP (2019) Cartographic design for improved decision making: trade-offs in uncertainty visualization for tornado threats. Ann Am Assoc Geogr 110(1):314–333. https://doi.org/10.1080/24694452.2019.1602467

Kosslyn SM (2006) Graph design for the eye and mind. Oxford University Press, New York

Krzywinski M, Cairo A (2013) Storytelling. Nat Methods 10:687. https://doi.org/10.1038/nmeth.2571

Laudien R, Boon E, Goosen H, van Nieuwaal K (2019) The Dutch adaptation web portal: seven lessons learnt from a co-production point of view. Clim Chang 153:509–521. https://doi.org/10.1007/s10584-018-2179-1

Light A, Bartlein P (2004) The end of the rainbow? Color schemes for improved data graphics. Eos 85(40):385–391. https://doi.org/10.1029/2004EO400002

Lledó L, Bellprat O, Doblas-Reyes FJ, Soret A (2018) Investigating the effects of Pacific sea surface temperatures on the wind drought of 2015 over the United States. J Geophys Res Atmos 123:4837–4849. https://doi.org/10.1029/2017JD028019

Lorenz S, Dessai S, Forster PM, Paavola J (2015) Tailoring the visual communication of climate projections for local adaptation practitioners in Germany and the UK. Phil Trans R Soc A 373:20140457. https://doi.org/10.1098/rsta.2014.0457

Lucero A (2015) Using affinity diagrams to evaluate interactive prototypes. In: Abascal J, Barbosa S, Fetter M, Gross T, Palanque P, Winckler M (eds) Human-Computer Interaction - INTERACT 2015, 9297. Springer, Cham, pp 231–248. https://doi.org/10.1007/978-3-319-22668-2_19

MacEachren AM, Robinson A, Hopper S, Gardner S, Murray R, Gahegan M, Hetzler E (2005) Visualizing geospatial information uncertainty: what we know and what we need to know. Cartogr Geogr Inf Sci 32(3):139–160. https://doi.org/10.1559/1523040054738936

Marcos-Matamoros R, González-Reviriego N, Torralba V, Soret A (2020a) Report on the coordinated forecast-phenological-irrigation requirement models for real-time applications. Deliverable 2.6, VISCA project, pp 39

Marcos-Matamoros R, González-Reviriego N, Graça A, Dell'Aquila A, Vigo I, Silva S, Varotsos KV, Sanderson M (2020b) Report on the methodology followed to implement the wine pilot services. Deliverable 3.2, MED-GOLD project, pp 78. https://www.med-gold.eu/wp-content/uploads/docs/776467_MED-GOLD_DEL3.2_Report-on-the-methodology-followed-to-implement-the-wine-pilot.pdf. Accessed 19 Sept 2021

Marriott K, Lee B, Butler M, Cutrell E, Ellis K, Goncu C, Hearst M, McCoy K, Szafir DA (2021) Inclusive data visualisation for people with disabilities: a call to action. Interactions 28(3):47–51. https://doi.org/10.1145/3457875

Maxion RA, Reeder RW (2005) Improving user-interface dependability through mitigation of human error. Int J Hum Comput Stud 63(1–2):25–50. https://doi.org/10.1016/j.ijhcs.2005.04.009

McInerny GJ, Chen M, Freeman R, Gavaghan D, Meyer M, Rowland F, Spiegelhalter DJ, Stefaner M, Tessarolo G, Hortal J (2014) Information visualisation for science and policy: engaging users and avoiding bias. Trends Ecol Evol 29(3):148–157. https://doi.org/10.1016/j.tree.2014.01.003

McMahon R, Stauffacher M, Knutti R (2015) The unseen uncertainties in climate change: reviewing comprehension of an IPCC scenario graph. Clim Chang 133:141–154. https://doi.org/10.1007/s10584-015-1473-4

Miraz MH, Excell PS, Ali M (2016) User interface (UI) design issues for multilingual users: a case study. Univers Access Inf Soc 15:431–444. https://doi.org/10.1007/s10209-014-0397-5

Moere AV, Purchase H (2011) On the role of design in information visualization. Inf Vis 10(4):356–371. https://doi.org/10.1177/1473871611415996

Nguyen J, Dupuis M (2019) Closing the feedback loop between UX design, software development, security engineering, and operations. Proceedings of the 20th Annual SIG Conference on Information Technology Education, pp 93- 98. https://doi.org/10.1145/3349266.3351420

Nocke T (2014) Imagens for data analysis: the role of visualization in climate research processes. In: Schneider B, Nocke T (eds). Image politics of climate change: visualizations, imaginations, documentations. Bielefeld: transcript Verlag. https://doi.org/10.14361/transcript.9783839426104.55

Oakley NS, Daudert B (2016) Establishing best practices to improve usefulness and usability of web interfaces providing atmospheric data. Bull Am Meteorol Soc 97(2):263–274. https://doi.org/10.1175/BAMS-D-14-00121.1

Pifarré M, Tomico O (2007) Bipolar laddering (BLA): a participatory subjective exploration method on user experience. Proceedings of the 2007 conference on Designing for User eXperiences, article 2: 2–13. https://doi.org/10.1145/1389908.1389911

Plaisant C, Shneiderman B, Mushlin R (1998) An information architecture to support the visualization of personal histories. Inf Process Manage 34(5):581–597. https://doi.org/10.1016/S0306-4573(98)00024-7

Quiroga LM, Crosby ME, Iding MK (2004) Reducing cognitive load. Proceedings of the 37th Annual Hawaii International Conference on System Sciences. https://doi.org/10.1109/HICSS.2004.1265328

Raaphorst K, Koers G, Ellen GJ, Oen A, Kalsnes B, van Well L, Koerth J, van der Brugge R (2020) Mind the gap: towards a typology of climate service usability gaps. Sustainability 12(4):1512. https://doi.org/10.3390/su12041512

Ramamurthy M, Lakshminarayanan V (2015) Human vision and perception. In: Karlicek R, Sun CC, Zissis G, Ma R (eds) Handbook of advanced lighting technology. Springer, Cham. https://doi.org/10.1007/978-3-319-00295-8_46-1

Rosenholtz R, Li Y, Nakano L (2007) Measuring visual clutter. J Vision 7:1–22. https://doi.org/10.1167/7.2.17

Rosli MHW, Cabrera A (2014) Application of Gestalt principles to multimodal data representation. Proceedings of the IEEE VIS 2014 Arts Program, VISAP'14: Art+Interpretation, Paris, France

S2S4E (2020) S2S4E Decision Support Tool website. https://www.s2s4e.eu/dst. Accessed 5 Nov 2021

Schiewe J (2019) Empirical studies on the visual perception of spatial patterns in choropleth maps. KN J Cartogr Geogr Inf 69:217–228. https://doi.org/10.1007/s42489-019-00026-y

Schneider B, Nocke T (2018) The feeling of red and blue - a constructive critique of color mapping in visual climate change communication. Practice of climate change communication. In: Filho WL, Manolas E, Azul AM, Azeiteiro UM, McGhie H. The handbook of climate change communication 2. Springer International Publishing

Segel E, Heer J (2010) Narrative visualization: telling stories with data. IEEE Trans Visual Comput Graph 16(6):1139–1148. https://doi.org/10.1109/TVCG.2010.179

Senders JW, Moray NP (1991) Human error: cause, prediction, and reduction. Lawrence Erlbaum Associates, Hillsdale

Shahar E (2019) Dashboard design. In: Shahar E (ed) Project Reliability Engineering. Apress, Berkeley. https://doi.org/10.1007/978-1-4842-5019-8_4
Sharp H, Rogers Y, Preece J (2007) Interaction design 2. Wiley
Speckmann B, Verbeek K (2010) Necklace maps. IEEE Trans Visual Comput Graph 16(6):881–889. https://doi.org/10.1109/tvcg.2010.180
Stephens EM, Edwards TL, Demeritt D (2012) Communicating probabilistic information from climate model ensembles - lessons from numerical weather prediction. WIRES Clim Chang 3(5):409–426. https://doi.org/10.1002/wcc.187
Meijie S, Haihong E, Zhonghong O, Yuanxing C (2020) More flexible: a free-customization technique based on log visualization. Proceedings of the 12th International Conference on Measuring Technology and Mechatronics Automation, 375–381. https://doi.org/10.1109/ICMTMA50254.2020.00089
Taylor AL, Dessai S, de Bruin WB (2015) Communicating uncertainty in seasonal and interannual climate forecasts in Europe. Phil Trans R Soc A 373(2055):20140454. https://doi.org/10.1098/rsta.2014.0454
Terrado M, Christel I, Bojovic D, Soret A, Doblas-Reyes FJ (2018) Climate change communication and user engagement: a tool to anticipate climate change. Case studies in climate change communication. In: Filho WL, Manolas E, Azul AM, Azeiteiro UM, McGhie H (eds) The handbook of climate change communication 3. Springer International Publishing. https://doi.org/10.1007/978-3-319-70479-1
Terrado M, Calvo L, Bojovic D, Christel I (2022) Current practice in climate service visualisation: taking the pulse of the providers' community. Bull Am Meteorol Soc 103(3):E828–E837. https://doi.org/10.1175/BAMS-D-21-0194.1
The Economist (2019) The consequences of a rapidly warming Arctic will be felt far afield (Graphic detail section). The Economist website. https://www.economist.com/graphic-detail/2019/09/21/the-consequences-of-a-rapidly-warming-arctic-will-be-felt-far-afield. Accessed 5 Nov 2021
Tsai FS (2010) Comparative study of dimensionality reduction techniques for data visualization. J Artif Intell 3(3):119–134. https://doi.org/10.3923/jai.2010.119.134
Wagemans J, Elder JH, Kubovy M, Palmer SE, Peterson MA, Singh M, von der Heydt R (2012) A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. Psychol Bull 138(6):1172–1217. https://doi.org/10.1037/a0029333
Wang Q, Yang S, Liu M, Cao Z, Ma Q (2014) An eye-tracking study of website complexity from cognitive load perspective. Decis Support Syst 62:1–10. https://doi.org/10.1016/j.dss.2014.02.007
Ware C (2012) Information visualization: perception for design. Elsevier, pp 536
Wickens CD, Carswell M (1995) The proximity compatibility principle: its psychological foundation and relevance to display design. Hum Factors 37:473–494. https://doi.org/10.1518/001872095779049408
Wiens V, Stocker M, Auer S (2020) Towards customizable chart visualizations of tabular data using knowledge graphs. In Ishita E, Pang NLS, Zhou L (eds) Digital libraries at times of massive societal transition. ICADL 2020 Lecture Notes in Computer Science, 12504. Springer, Cham, pp 71–80. https://doi.org/10.1007/978-3-030-64452-9_6
Wilke CO (2019) Fundamentals of data visualization: a primer on making informative and compelling figures. O'Reilly Media, Inc, Sebastopol, pp 371
William JA, Murugesh R (2020) Potential eye tracking metrics and indicators to measure cognitive load in human computer interaction research. J Sci Res 64(1):268–275. https://doi.org/10.37398/JSR.2020.640137
Windyty (2014) Windy. https://www.windy.com. Accessed 25 Sept 2021
WMO (2008) Guidelines on communicating forecast uncertainty. WMO/TD No. 1422. https://library.wmo.int/doc_num.php?explnum_id=4687. Accessed 19 Sept 2021
WMO (2018) Guidelines on quality management in climate services. WMO-No. 1221, ISBN 978–92–63–11221–7. https://library.wmo.int/doc_num.php?explnum_id=5174. Accessed 19 Sept 2021
Yi JS, Kang YA, Stasko JT, Jacko JA (2007) Toward a deeper understanding of the role of interaction in information visualization. IEEE Trans Visual Comput Graph 13(6):1224–1231. https://doi.org/10.1109/TVCG.2007.70515
Yucong Y, Sauer F, Ma KL, Aditya K, Chen J (2019) A user-centered design study in scientific visualization targeting domain experts. IEEE Trans Visual Comput Graph 26(6):2192–2203. https://doi.org/10.1109/TVCG.2020.2970525
Zallio M (2021) Democratizing information visualization. A study to map the value of graphic design to easier knowledge transfer of scientific research. In: Soares M, Rosenzweig E, Marcus A (eds) Design, user experience, and usability: UX research and design. 10th International Conference, DUXU 2021

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

BAMS
Meeting Summary

Current Practice in Climate Service Visualization

Taking the Pulse of the Providers' Community

Marta Terrado, Luz Calvo, Dragana Bojovic, and Isadora Christel

Climateurope Workshop on the Visualization of Climate Services

What: Barcelona Supercomputing Center's (BSC) Earth Sciences Department organized the workshop in the framework of the Horizon 2020–funded Coordination and Support Action Climateurope. The workshop aimed to discuss different aspects of the state-of-the-art of visualizations used in climate services and produce a publication on the synthesis and recommendations. We invited participants from different projects linked to the Climateurope network, including EU Horizon 2020 (H2020) and European Research Area for Climate Services (ERA4CS) projects as well as a few national projects and private contracts. The workshop was attended by representatives of 22 projects.

When: 20 November 2020

Where: Online

KEYWORDS: Climate services; Communications/decision making; Adaptation; Uncertainty

https://doi.org/10.1175/BAMS-D-21-0194.1
Corresponding author: Marta Terrado, marta.terrado@bsc.es
In final form 3 August 2021

AFFILIATIONS: Terrado, Calvo, Bojovic, and Christel—Department of Earth Sciences, Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain

Effective visualizations of climate information that can be easily understood by non–climate experts are strongly needed. At present, the absence of a common or standardized visualization approach for climate services results in the application of different practices by climate service providers, sometimes leading to users' misinterpretation or misuse of climate information. In this report, we analyze the outputs from a workshop with climate service providers that had the aim to identify current practices and challenges faced when developing visualizations in the field of climate services. The analysis of the results obtained depicted the current status of the climate services visualization field, identified the main lessons learned by different projects, and highlighted challenges that required further research efforts. Insights obtained from the analysis are valid for climate service providers worldwide.

The visualization workshop was organized in the framework of the EU-funded Horizon 2020 (H2020) Climateurope Coordination and Support Action on climate services. The workshop was attended by representatives from 22 projects working in climate services, including EU H2020 and European Research Area for Climate Services (ERA4CS) projects and a few national projects and private contracts. During the 2.5-h interactive session, participants were divided in breakout groups to discuss their experiences in the development of climate service visualizations and to share the lessons learned throughout the process. Breakout group discussions included topics on different visualization aspects, such as the added value of inter- and transdisciplinary approaches, different ways to visually communicate uncertainty, the use of interactive elements, and the importance of an appropriate terminology and language in visualizations. The feedback gathered during the discussions was analyzed using affinity maps, a qualitative analysis method using a coding technique that helped organize the information gathered in the discussion session into groups or themes of similar items and topics to categorize and extract relevant knowledge (Lucero 2015). The role of visualization in the field of climate services was also discussed (Fig. SB1 in the sidebar). Further details regarding the workshop can be found in Terrado et al. (2021).

Current practice in climate service visualization

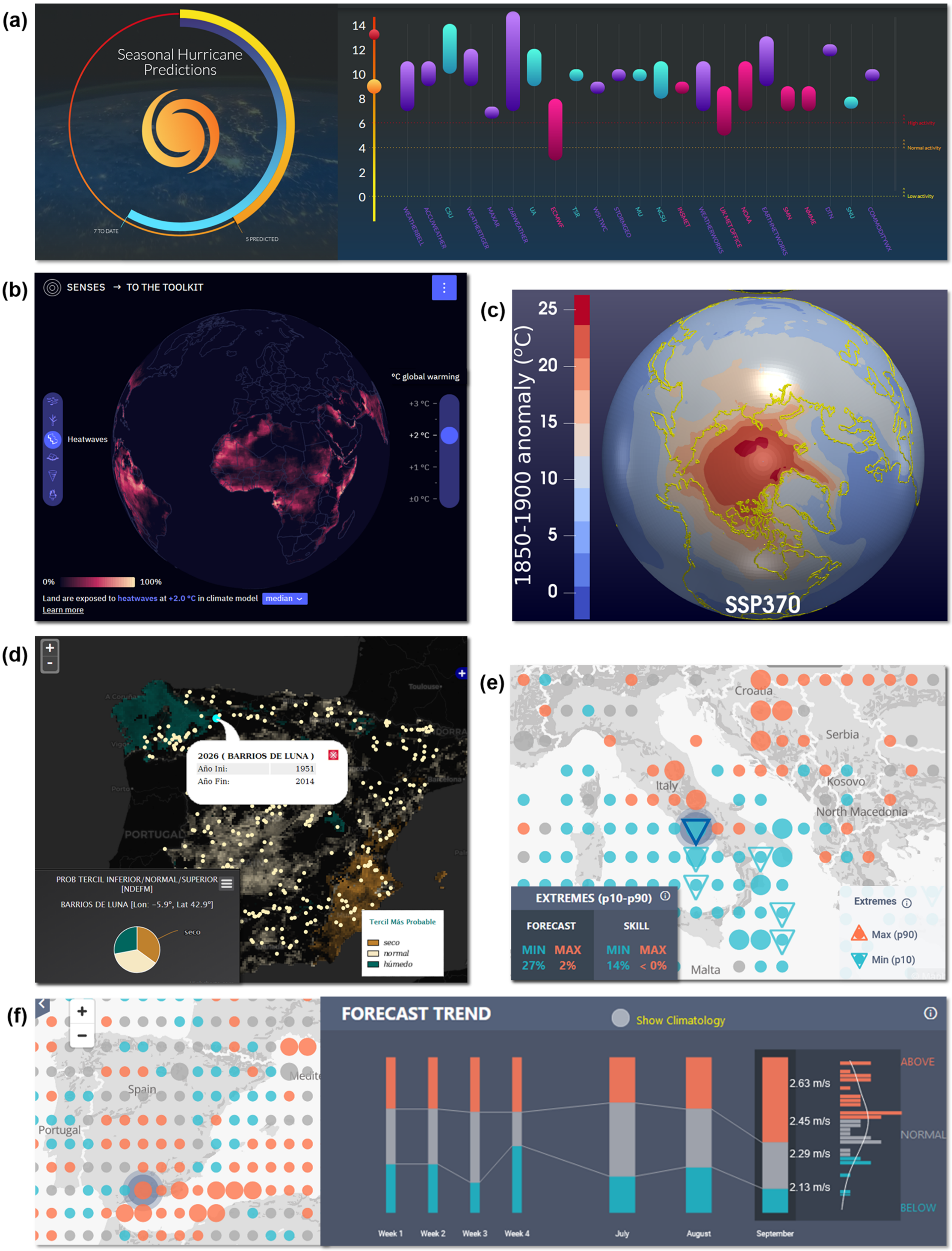
The lack of clear guidelines for the development of climate service visualizations has resulted in climate service providers displaying information in different ways, depending on factors such as the level of visualization expertise of the working team, the level of user engagement applied, or the type of feedback received during the codesign process. Despite the specificities of the different climate services projects, some common patterns emerged from the discussions and are presented in the sections below.

Moving toward transdisciplinary coproduction approaches. Discussions during the workshop revealed that the development of climate service visualizations often applied interdisciplinary approaches, which involve coordinated collaboration among different disciplines (Max-Neef 2005). In addition to climate scientists, some project teams counted with experts in science communication, social sciences and humanities, and sometimes experts in user experience, which made it possible to define an interdisciplinary collaboration space. However, many projects acknowledged a lack of representation of particular profiles (e.g., data visualization experts, interface designers, specific expertise gathered under the umbrella of social sciences and humanities), making particular requirements for the codevelopment of the visualization more challenging to fulfill. In general, projects were aware of the importance of bringing stakeholders' perspectives and experience into the design of climate service visualizations, which is a necessary step to move from interdisciplinary to transdisciplinary approaches (Max-Neef 2005; Bojovic et al. 2021). However, not all projects attained full transdisciplinarity and, in various instances, stakeholders were only reached out for final visualization testing.

Testing visualizations with potential users. Testing visualizations with potential users emerged as a common practice in the climate services field. However, not all the projects acknowledged an active involvement of stakeholders during the whole visualization design process, with some projects only performing the evaluation with the stakeholders at the end. Although this is still useful to identify some issues and ensure that the information is interpreted as intended by climate service providers, a truly collaborative and transdisciplinary codesign goes beyond final testing and entails involving stakeholders from the very beginning and throughout all the stages of the climate service codevelopment (McInerny et al. 2014; Lorenz et al. 2015; Grainger et al. 2020).

A plethora of approaches to represent uncertainty in climate data. Different ways of visualizing uncertainty were identified according to the visualization purpose, the target audience, and the time scale of the information provided. This applied to both first- and second-order uncertainty (Spiegelhalter et al. 2011; Taylor et al. 2015). Whereas first-order uncertainty refers to information on the likelihood of an event happening according to a particular forecast (i.e., probabilities or risk), second-order uncertainty refers to "uncertainty about the uncertainty" or ignorance (i.e., skill or spread), and exists because forecasts are not able to capture all the factors influencing the climate. Information on first- and second-order uncertainty can either be integrated in the same visual representation or presented separately using two different visualizations. Regarding first-order uncertainty, while some projects opted for showing the mean or median value of the modeled results without an indication of its probability of occurrence (Figs. 1a,b), other projects decided to display the information as anomalies (Fig. 1c), that is, the variation of a variable relative to the climatological normal or long-term average. Additional options to represent first-order uncertainty used by climate services projects included showing information through a number of categories, such as terciles or quintiles, either indicating the most likely category (Fig. 1d) or reporting the probability of the different categories to occur (Fig. 1e). An alternative option consisted in providing the probability distribution function (Fig. 1f), which gives an overview of the different amounts of change in a climate hazard and their respective likelihoods for a single point in time and a specific geographical area.

Even though showing second-order uncertainty was generally considered an exercise of transparency, some workshop participants considered that this information could be overwhelming or confusing for some users. For this reason, this information was not generally provided by climate services projects. When second-order uncertainty was not presented, some projects used scenario approaches displaying average values for different pathways, capturing the various plausible descriptions of how the system and/or its driving forces may develop in the future (Fig. 2a). In other cases, projects decided to keep particular climate information hidden from the visualization when the level of uncertainty was too high to use this information meaningfully in decision-making (Fig. 2b). In such cases, applied practices ranged from not providing any information about uncertainty, to replacing uncertain forecasts by a reference value (e.g., climatology). Further alternatives consisted in giving users the



(a)
Seasonal Hurricane Predictions
7 TO DATE
5 PREDICTED
14
12
10
8
6
4
2
0
High activity
Normal activity
Low activity
WEATHERBELL
ACCUWEATHER
CSU
WEATHERTIGER
MAXAR
2WEATHER
UA
ECMWF
TSR
WSI-TWC
STORMGEO
MU
NCSU
INSMET
WEATHERWORKS
UK MET OFFICE
NOAA
EARTHNETWORKS
SMN
NMME
DTN
SNU
COMMODITYWX
(b)
SENSES → TO THE TOOLKIT
Heatwaves
°C global warming
+3 °C
+2 °C
+1 °C
±0 °C
0%
100%
Land are exposed to heatwaves at +2.0 °C in climate model median
Learn more
(c)
1850-1900 anomaly (°C)
25
20
15
10
5
0
SSP370
(d)
A Coruña
Vigo
Porto
PORTUGAL
Toulouse
2026 (BARRIOS DE LUNA)
Año Ini: 1951
Año Fin: 2014
PROB TERCIL INFERIOR/NORMAL/SUPERIOR [NDEFM]
BARRIOS DE LUNA [Lon: -5.9°, Lat 42.9°]
seco
Tercil Más Probable
seco
normal
húmedo
(e)
Croatia
Serbia
Italy
Kosovo
North Macedonia
Malta
EXTREMES (p10-p90)
FORECAST
SKILL
MIN MAX
27% 2%
MIN MAX
14% < 0%
Extremes
Max (p90)
Min (p10)
(f)
Spain
Portugal
FORECAST TREND
Show Climatology
2.63 m/s
2.45 m/s
2.29 m/s
2.13 m/s
ABOVE
NORMAL
BELOW
Week 1
Week 2
Week 3
Week 4
July
August
September

Fig. 1. Representation of first-order uncertainty in different climate service visualizations: (a) average forecast value, (b) average forecast value for different scenarios, (c) anomalies, (d) most likely category (e.g., terciles), (e) probability of different categories (e.g., extreme events probabilities), and (f) probability distribution function. Source projects: (a) Seasonal hurricane predictions platform (www.seasonalhurricanepredictions.org), (b) SENSES (www.senses-project.org; Auer et al. 2021), (c) eClimViz (Walton et al. 2021), (d) MEDSCOPE, and (e),(f) S2S4E (www.s2s4e-dst.bsc.es).

possibility to hide or show high-uncertainty information (Fig. 2c) or allowing them to select a specific uncertainty threshold should they have an idea of the level of uncertainty they were ready to bear (Fig. 2d). Some projects also integrated second-order uncertainty through visual encoding (e.g., through transparency) (Fig. 2e) or showed it as a range in the plot, be it the full ensemble range, the standard deviation, confidence intervals, or the signal-to-noise ratio (Fig. 2f). Figures 1 and 2 show a nonexhaustive sample of ways to represent first- and second-order uncertainties in some climate service visualizations.

Use of interactive elements. A clear trend toward developing visualizations that allow the user to interact with the different elements was observed, in line with the assumption that understanding of information can be improved through greater interaction (Yi et al. 2007). The progressive disclosure of information, which aims at the initial simplification of information followed by the possibility to reveal additional options and content, was identified as a commonly applied technique in the field of climate service visualization, which also grants users a more active role (Bostrom et al. 2008; Spiegelhalter et al. 2011). However, in the case of particular types of services (e.g., dashboards), participants pointed out that users had explicitly indicated their need to access all resources simultaneously. For particular formats that allow low or no interactivity (e.g., factsheets, newsletters or bulletins, direct advice), it was also indicated that they could be effectively used. In the end, understanding when and how to integrate interactivity requires careful considerations of both users' requirements and tool's functionality.

Differences in terminology used by scientific and stakeholder communities. The climate services community involved in the workshop identified more than 25 technical terms commonly used in the field of climate science that are confusing or not well understood by stakeholders outside academia. The more frequently repeated terms were "skill," "anomaly," "reliability," "uncertainty," "percentile," "ensemble," and "model" (Fig. 3). Project representatives also mentioned that the use of conventions such as "likely" or "unlikely" or the distinction between temporal forecasting scales (e.g., hindcasts, climate predictions, climate projections) can make sense in the context of climate science, but that stakeholders are not aware of such distinctions. Participants agreed that more resources should be put in place to overcome the terminology barrier, both among different academic disciplines and between academia and stakeholders, since the same term can be differently understood by these groups. Discussions indicated that terminology should be adapted when possible, even if it involves compromising scientific precision. Otherwise, explanations in lay language should be offered. Although this may not be straightforward and can induce some tensions during the coproduction process, overall, it will prevent wasting time discussing complex terminology concepts and will allow stakeholders to focus on the interpretation of the information. The use of glossaries and thesauruses emerged from the discussions as a good practice to try to find a common ground between the different communities. These tools can also include use case examples to illustrate the term in the context of the target users. Using elements such as tooltip hints, defining technical terms, and hyperlinks to additional explanation were also mentioned to help build a greater understanding.

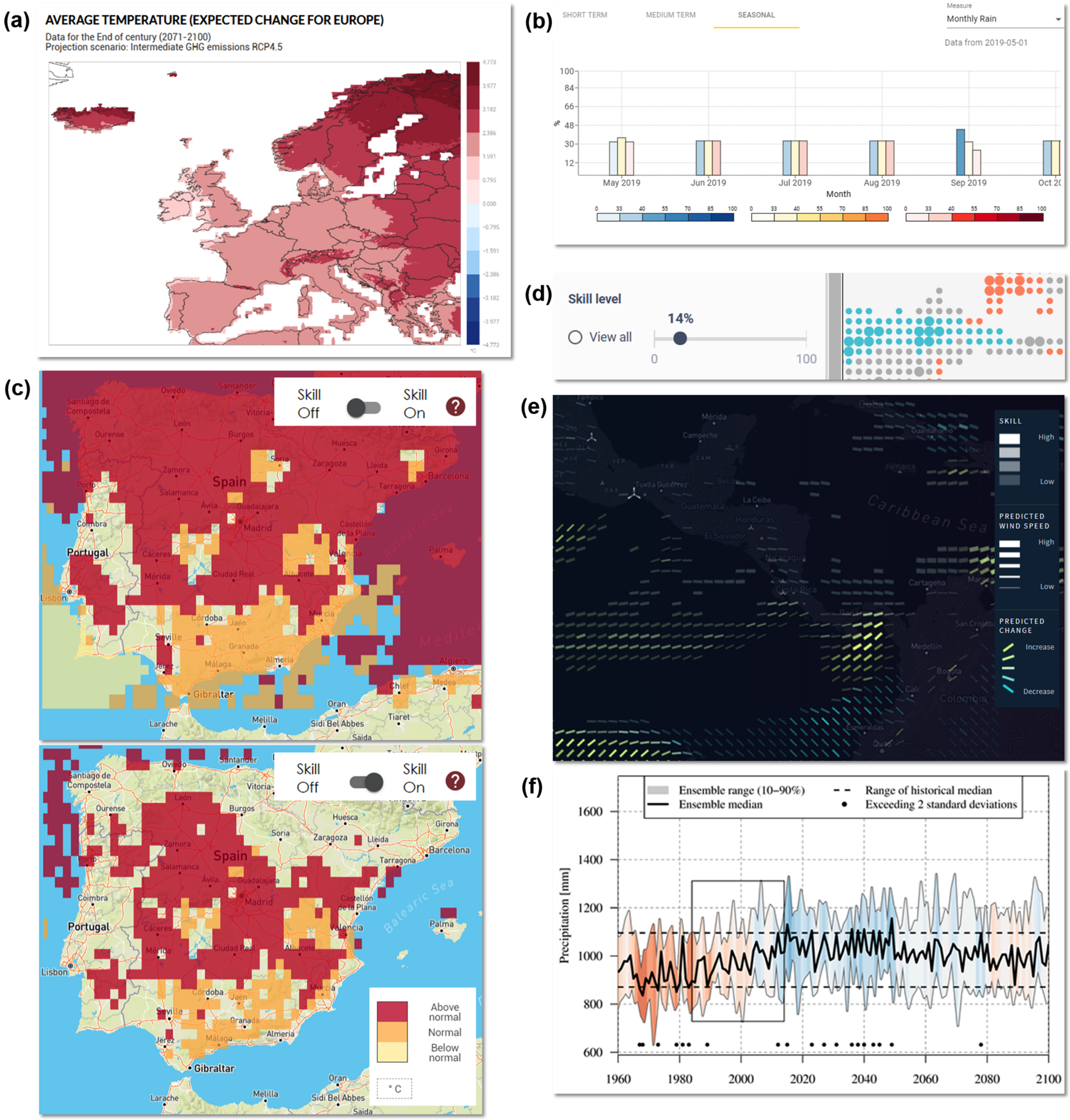


Fig. 2. Representation of second-order uncertainty in different climate service visualizations: (a) forecast information shown, no indication of uncertainty, (b) high-uncertainty information not shown (e.g., replaced by a reference), no indication of uncertainty, (c) possibility to hide/show high-uncertainty information (e.g., activation of a mask), (d) possibility to hide/show information with different levels of uncertainty (e.g., filter by threshold), (e) uncertainty integrated in the visualization through visual encoding, and (f) uncertainty represented as an additional parameter (e.g., uncertainty range). Source projects: (a) C3S Press Data Portal (https://climate.copernicus.eu/press-data-portal), (b) VISCA (Marcos-Matamoros et al. 2020), (c) MED-GOLD (https://dashboard.med-gold.eu), (d) S2S4E (www.s2s4e-dst.bsc.es), (e) Project Ukko (www.project-ukko.net), and (f) CIREG.

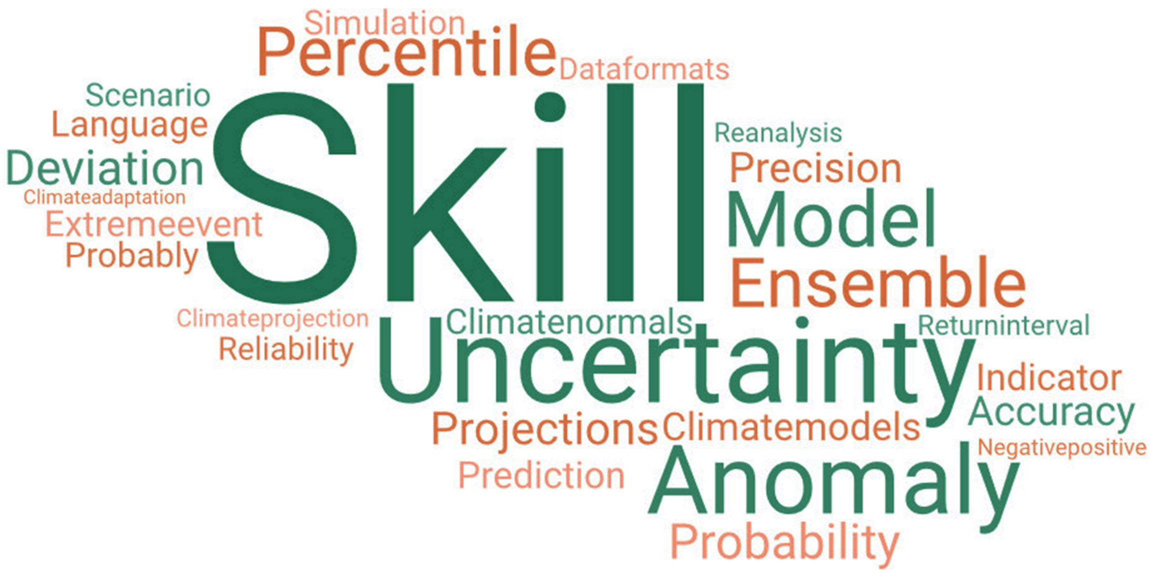


Fig. 3. Word cloud with the technical terms used in the climate services field identified as a challenge by non–climate experts. The most repeated words are displayed in larger size (*n* = 22).

Taking the vernacular language of target audiences into account. An additional factor for misinterpretation arises when the native language of stakeholders is different from that of the producer of the climate information (WMO 2008). Many of the projects involved in the workshop were run at the European or multicountry scale. To adhere to users' preferences, such scales require a multilanguage approach that can become a challenge for the development of visualizations. Therefore, English was a dominant language used in the visualizations developed by the different projects. However, the proportion of climate services in local languages may be substantially higher when moving to the national, regional, and local scales. Despite many climate service visualizations being available in English, project representatives mentioned that, when needed, engagement activities with stakeholders were conducted in local languages to ensure understanding. For that, summary documents, including user guides and illustrative figures, were developed in stakeholders' languages. Related to the use of figures, one of the participants mentioned that the translation of text labels in visuals tends to be more time consuming than translating text explanations. Participants also mentioned that having visualizations in local languages was particularly needed for specific terminology lying in the traditional knowledge domain of local and indigenous communities (e.g., Arctic regions, Pacific Islands). Such terms, that for instance can refer to the characteristics of the local climate, often lack a translation in other languages (e.g., snow types in polar regions; see Eira et al. 2013). Discussions suggested that the definition of an appropriate language should be considered as part of the coproduction of a climate service. This involves tailoring information to match the language in which intended users are accustomed to working as well as the consideration of other elements that enhance usability (Miraz et al. 2016).

Lessons learned

Data visualizations, like charts, graphs, and maps, make it easy for many audiences to identify and understand patterns in climate data. But when not done properly, they can exclude audiences with visual or cognitive disabilities or those that lack appropriate background knowledge. This work analyses the current status of the climate service visualization field and identifies challenges to be tackled for the development of more effective visualizations. The main challenges include the advancement of the climate services field toward a real

The role of visualization in climate services

The climate services community highlighted nine main purposes (Fig. SB1):

1) *Targeted communication and outreach*: Convey messages in a more direct and illustrative way and give the advantage of reaching wider audiences beyond the specialized ones.
2) *Storytelling from data*: Support data exploration and help to extract the underlying information and patterns in data, which makes it possible to tell a story and develop climate change narratives.
3) *Ease decision-making*: Increase understanding, allow users to reach conclusions, and ultimately, enhance their ability to complete a task or make an informed decision.
4) *Simplify complexity*: Deliver complex information in a simplified way, which may require transforming data into a smaller and more manageable dataset or paying special attention to visual encoding, encompassing the representation of information with graphical features.
5) *Transfer knowledge*: Reach various audiences with the aim to facilitate understanding and boost knowledge uptake.
6) *Raise awareness/call for action*: Raise awareness and elicit affective responses, improving the likelihood for action.
7) *Attractiveness*: Aspect worth considering if striving to create something memorable that helps to catch people's attention or increase trust.
8) *Engagement*: Stimulate public willingness to engage with a particular issue and visualization used as a conversation starter and to support the communication process.
9) *Add layers of information*: Combine information coming from various sources and find new relationships among different datasets, which can be interpreted more easily in a visual format.

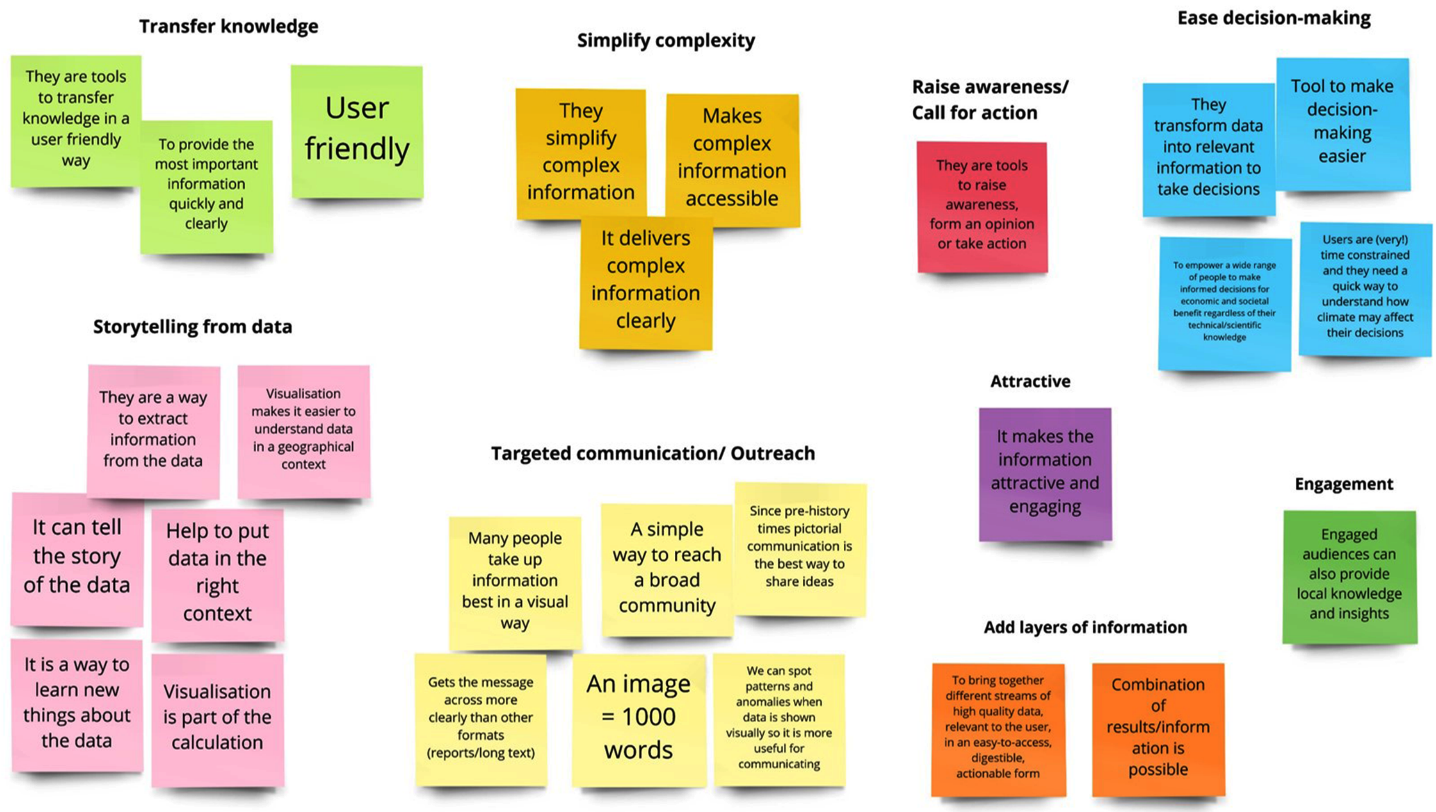


Fig. SB1. The identified purposes of visualization in climate services.

transdisciplinary approach by effectively involving other disciplines and stakeholders in the visualization coproduction process, a better coevaluation of visualizations, a more effective representation of uncertainty in climate data, and bringing the terminology and language closer to those used by target audiences. For the development of more effective climate service visualizations, the climate science field may benefit from advances in other disciplines with a well-founded tradition, such as user experience, data visualization, graphic design, or psychology, which are strongly based on stakeholders' needs. Only by including the expertise from other disciplines will climate service visualizations be able to build trust, prevent misuse of climate knowledge, and boost the uptake of climate information by society. This is a necessary step to move toward the codevelopment of common and agreed guidance and best practices necessary to achieve coherent and effective visualizations in climate services.

Acknowledgments. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreements 689029 (Climateurope) and 776467 (MED-GOLD). The authors want to acknowledge the different climate services projects that participated in the workshop, including Climate-fit.city (73004), Digital-Water.city (20954), IMPREX (641811), PRIMAVERA (641727), RECEIPT (820712), S2S4E (776787), SECLI-FIRM (776868), VISCA (730253), CIREG, Clim2power, CoCliME, EVOKED, ISIpedia, MEDSCOPE, SENSES, WATExR, C3S_429g_BSC, and the inDust COST Action (16202). We would also like to thank the participation of the projects Seasonal Hurricane Predictions, KNMI climate scenarios, and eClimViz. We also acknowledge BSC colleagues Diana Urquiza, Andria Nicodemou, and Sara Octenjak for helping to run the workshop.

References

Auer, C., E. Kriegler, H. Carlsen, K. Kok, S. Pedde, V. Krey, and B. Müller, 2021: Climate change scenario services: From science to facilitating action. *One Earth*, **4**, 1074–1082, https://doi.org/10.1016/j.oneear.2021.07.015.

Bojovic, D., A. L. St. Clair, I. Christel, M. Terrado, P. Stanzel, P. Gonzalez, and E. Palin, 2021: Engagement, involvement and empowerment: Three realms of a co-production framework for climate services. *Global Environ. Change*, **68**, 102271, https://doi.org/10.1016/j.gloenvcha.2021.102271.

Bostrom, A., L. Anselin, and J. Farris, 2008: Visualizing seismic risk and uncertainty: A review of related research. *Ann. N. Y. Acad. Sci.*, **1128**, 29–40, https://doi.org/10.1196/annals.1399.005.

Eira, I. M. G., C. Jaedicke, O. H. Magga, N. G. Maynard, D. Vikhamar-Schuler, and S. D. Mathiesen, 2013: Traditional Sámi snow terminology and physical snow classification—Two ways of knowing. *Cold Reg. Sci. Technol.*, **85**, 117–130, https://doi.org/10.1016/j.coldregions.2012.09.004.

Grainger, S., B. F. Ochoa-Tocachi, J. Antiporta, A. Dewulf, and W. Buytaert. 2020: Tailoring infographics on water resources through iterative, user-centered design: A case study in the Peruvian Andes. *Water Resour. Res.*, **56**, e2019WR026694, https://doi.org/10.1029/2019WR026694.

Lorenz, S., S. Dessai, P. M. Forster, and J. Paavola, 2015: Tailoring the visual communication of climate projections for local adaptation practitioners in Germany and the UK. *Philos. Trans. Roy. Soc.*, **A373**, 20140457, https://doi.org/10.1098/rsta.2014.0457.

Lucero, A., 2015: Using affinity diagrams to evaluate interactive prototypes. *15th IFIP Int. Conf. on Human–Computer Interaction*, Bamberg, Germany, IFIP, 231–248, https://doi.org/10.1007/978-3-319-22668-2_19.

Marcos-Matamoros, R., N. González-Reviriego, V. Torralba, and A. Soret, 2020: Report on the coordinated forecast-phenological-irrigation requirement models for real-time applications. VISCA Project Deliverable 2.6, 39 pp.

Max-Neef, M. A., 2005: Foundations of transdisciplinarity. *Ecol. Econ.*, **53**, 5–16, https://doi.org/10.1016/j.ecolecon.2005.01.014.

McInerny, G. J., and Coauthors, 2014: Information visualisation for science and policy: Engaging users and avoiding bias. *Trends Ecol. Evol.*, **29**, 148–157, https://doi.org/10.1016/j.tree.2014.01.003.

Miraz, M. H., P. S. Excell, and M. Ali, 2016: User interface (UI) design issues for multilingual users: A case study. *Univ. Access Inf. Soc.*, **15**, 431–444, https://doi.org/10.1007/s10209-014-0397-5.

Spiegelhalter, D., M. Pearson, and I. Short, 2011: Visualizing uncertainty about the future. *Science*, **333**, 1393–1400, https://doi.org/10.1126/science.1191181.

Taylor, A. L., S. Dessai, and W. B. de Bruin, 2015: Communicating uncertainty in seasonal and interannual climate forecasts in Europe. *Philos. Trans. Roy. Soc.*, **A373**, 20140454, https://doi.org/10.1098/rsta.2014.0454.

Terrado, M., I. Christel, D. Urquiza, S. Octenjak, A. Nicodemou, D. Bojovic, J. Bessembinder, and J. Walton, 2021: Lessons learned from current practices in climate service visualisation and recommendations. Climateurope Project Rep., 16 pp., www.climateurope.eu/10020-2/.

Walton, J., S. Adams, W. Hayek, P. Florek, and H. Dyson, 2021: Dynamic 3-D visualization of climate model development and results. *IEEE Comput. Graphics Appl.*, **41**, 17–25, https://doi.org/10.1109/MCG.2020.3042587.

WMO, 2008: Guidelines on communicating forecast uncertainty. WMO Tech. Doc. WMO/TD 1422, 22 pp., https://library.wmo.int/doc_num.php?explnum_id=4687.

Yi, J. S., Y. A. Kang, J. T. Stasko, and J. A. Jacko, 2007: Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Visualization Comput. Graphics*, **13**, 1224–1231, https://doi.org/10.1109/TVCG.2007.70515.

Measuring the Effectiveness of Static Maps to Communicate Changes over Time

Luz Calvo, Fernando Cucchietti, and Mario Pérez-Montoro

Abstract—Both in digital and print media, it is common to use static maps to show the evolution of values in various regions over time. The ability to communicate local or global trends, while reducing the cognitive load on readers, is of vital importance for an audience that is not always well versed in map interpretation. This study aims to measure the efficiency of four static maps (choropleth, tile grid map and their banded versions) to test their usefulness in presenting changes over time from a user experience perspective. We first evaluate the effectiveness of these map types by quantitative performance analysis (time and success rates). In a second phase, we gather qualitative data to detect which type of map favors decision-making. On a quantitative level, our results show that certain types of maps work better to show global trends, while other types are more useful when analyzing regional trends or detecting the regions that fit a specific pattern. On a qualitative level, those representations which are already familiar to the user are often better valued despite having lower measured success rates.

Index Terms—Information visualization, Cognition, Static maps, User interfaces, Perception

1 Introduction

Since ancient times maps have been used as visual representations linked to the expression of temporal evolution. As far back as Roman times, they were used for the administrative management of the regions of the empire, as well as for the planning of military campaigns [1], [2].

It is from the seventeenth century and the rise of social cartography when maps began to be used to link geospatial components to certain periods of time and the changes which occurred during these periods [3], [4].

If we explore the diverse and numerous representations in temporal evolution through the use of maps, we find solutions that range from the most creative to the most practical, with applications in national, regional and even urban areas [3], [5], [6]. The use of maps to express temporal evolution contemplates multiple representations (point maps, bubble maps, line and flow maps, heat maps, etc.) [7], but most of these representations (due to the rise of digital technology) lean towards the use of animation mechanisms to favor the sensation of continuity between the different time steps represented on the maps [8].

However, the use of static maps is still present today in the information we consume. For example, scientific or technical reports used by public or private organizations which seek to make information more accessible to their readers [9], [10]. Furthermore, divulgative material and reports used in media such as newspapers, magazines and their digital counterparts, intended for the general public, pay special attention to using visual and written resources which are more familiar and comprehensible to their readers [11], [12], [13], [14], [15].

- *Luz Calvo and Fernando Cucchietti are with the CASE Department of the Barcelona Supercomputing Center (BSC). E-mail: luz.calvo,fernando.cucchietti@bsc.es*
- *Mario Pérez-Montoro is with the Department of Library and Information Science and Audiovisual Communication, University of Barcelona. E-mail: perez-montoro@ub.edu*

The most familiar geospatial representation used to show changes over time is the choropleth map. It is popular both with expert and non-expert audiences in the field of visualization [16], [17], [18]. However, due to an inherent area-size bias problem, choropleth maps make communicating trends in data more problematic. A clear alternative could be the use of tile grid maps, but they in turn, present their own identification problems [19], [20], [21], [22] and are not so familiar to a general audience. Recent studies propose new versions (banded maps) of traditional choropleths, employing visual and intuitive metaphors that seek to favor the communication of changing social phenomena [23]. However, measuring how effectively these maps communicate temporal changes in informative scenarios (press, reports, online news, etc.), as well as their limitations, is essential when using them as a visual reinforcement mechanism.

In this paper, we evaluate four types of maps (choropleth, tile grid maps and their banded versions). We first study their effectiveness by quantitative performance analysis (time and success rates) with a controlled experiment. Secondly, we gather qualitative data to identify which type of map better favors decision-making. With this double evaluation, we intend to analyze whether there is any bias that may interfere between the real effectiveness of each map and the users' perception of effectiveness [24], [25].

2 Previous work

2.1 Choropleth and tile grid map considerations

Choropleth maps are a widely known representation and are common in media communications because they are popular with the target audiences. The use of these representations to illustrate articles, helps readers to understand the news items which they accompany, as well as to improve the perception of reliability [26], [27], [28]. However, despite their familiarity for many audiences, choropleth maps present some associated problems such as "dark-is-more bias" (i.e ranking of color lightness perception), the "area-size bias" (i.e small areas are less dominant than larger ones) and the "data-classification effect" (the established classification intervals used to detect patterns) [29], [30]. Tile grid maps, a kind of cartogram (see Fig. 1b), are another widely used representation that communicates, and in an effective manner, broader trends and patterns summarizing data [31]. They represent the different areas of the map using uniform size and shapes (often squares) and are arranged close to their real-world positions. The key to cartogram design is to use distortion or metaphors to change the size, shape, boundaries and/or geographical representation of regions, depending on statistical parameters or following a conceptual convention, but in a way that keeps the map recognizable to the reader [32], [33]. While less common, tile grid maps and their variants [34] offer an improvement over choropleth maps: for example, in solving the problem of distortion resulting from "area-size bias" [35], [36], [37]. However, we must take into account some of their known limitations such as lookalike aspects (they must bear a resemblance to the original map), topology and inaccuracy (neighboring areas must be maintained), and misdirection (relative positions between neighboring areas must be close to reality) in order to avoid misinterpretation [38], [39]. In addition, not all geographies can be easily represented in a conceptual manner, while at the same time maintaining their capabilities to be identified without resorting to the use of legends, labels or geographical visual clues [40], [41]. In the same way, another drawback is that they are often used as a creative solution instead of as a cartographic projection used to analyze a particular problem [21].

To illustrate this, we can analyse the representation of the map of London using blocks of the same size to represent each neighborhood. This visualization takes advantage of The River Thames, using its meanders to provide a visual clue within the structure of the map as a metaphor that helps the audience to identify the city [42]. Nevertheless, we must bear in mind that these representations may not be easily identified by people who are not familiar with the geography of the area. Therefore, it is necessary to use text references (see Fig. 1) to ensure the correct identification of the regions on the map [43]. Depending on the map and its topography, we must also take into account human perception rules such as figure-ground, depth, readability, identification, and balance [44]. The specific characteristics of the data (i.e. number of categories to show, magnitudes of the values of the different regions or areas, etc.) also force and limit us to using specific visualizations linked to a certain use-context [45]. Apart from the map representation itself, other factors such as visual search patterns and cognition must be considered in order to measure the efficiency of the alternatives discussed above. The analysis of visual search patterns on maps is a complicated process as it involves complex cognitive operations, which include the detection of relationships and the identification of different visual encodings (location, size, color, shape, etc.). The perception of maps requires the identification of a greater number of symbols in a greater range of variability, in which short-term memory is insufficient to cover the global analysis of a task [46], [47].

2.2 Communicating changes over time through the use of static maps

If we focus our attention on the problems related to the representation of change over time, the multiple static map strategy juxtaposes two or more maps to provide a simultaneous visual comparison in specific time units or snapshots. A possible solution could be to employ dynamic strategies to present the maps in a temporal sequence or through an evolution of a geographical pattern using various sequences to favor memory processes [48], [49], [50].

However, these techniques also have limitations: the constraints of the human visual system (change blindness, foveal and peripheral attention) suggest that humans often do not easily perceive changes within dynamic graphics [51], [52]. Hence the importance of taking into account these barriers of limited visual and cognitive user processing capabilities when designing time-related map representations. Studies on the cognitive effects and differences between the use of static and animated maps, lead us to solutions such as providing user instructions and the division of maps into fragments to help the users focus their attention on the relevant areas [53].

Du et al. (2018), proposed a novel visualization technique which attempts to solve the time dynamic sequencing issue: The banded choropleth map (see Fig. 1d), which divides subregions of a map into partitions of the same area to represent different time steps [23]. This technique makes better use of space than representations of small multiples, and performs better than animations, where it would have been necessary to consume both more memory and cognitive resources. However, it inherits the aforementioned problems associated with its precursor: the choropleth map.

3 Experiment

We assessed the effectiveness of these static maps by performance (in terms of completion time and success rate) for the proposed visualization tasks. We also studied the perceived effectiveness (level of usefulness of each map to perform each task) and the comments shared by participants.

3.1 Maps under study

We focused our study on measuring the effectiveness of the most widely used representation in the media to communicate social data over time: the choropleth [54], [55]. We also included one of its simplest alternatives, the tile grid map, that solves its area associated bias [19], [29], [30]. For both maps, we added two other new alternatives, based on a recent study: the banded maps, in order to compare their effectiveness in showing temporal changes in static scenarios [23]. There are other cartographic alternatives to choropleth maps that we did not consider such as graduated symbols or isoline maps [34] as when comparing their performance, the choropleth map is the most familiar, and the easiest alternative for solving several tasks [17].

Throughout this paper we will evaluate two different kinds of representations:

1) Representations which use a different map for each time step and which we refer to as **small multiple**

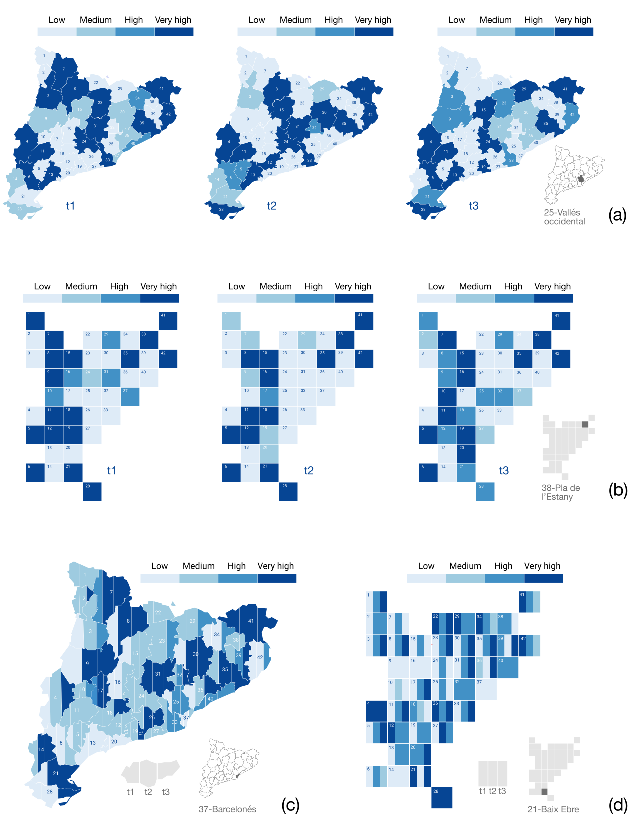


Fig. 1. Visualisations evaluated in the study: First, small multiple maps using Choropleth (a) and Tile Grid map (b). Second, their banded versions, unifying in the same map three different time steps: Banded Choropleth [23] (c) and Banded Tile Grid map (d).

representations. These include choropleth and tile grid maps (see Fig. 1a and Fig. 1b).

2) Representations which combine the time steps within the same map and which we refer to as **banded representations**. These include banded choropleth and banded tile grid maps (see Fig. 1c and Fig. 1d).

As aforementioned, we left out other popular representations such as isoline maps or heat maps as they frequently present multiple simultaneous points of interest that change in a clear spatial direction [56], [57], [58]. An example of this is found in contour maps and uncertainty cones, which are commonly used in the communication of weather or atmospheric data.

3.2 Task selection and design

Our objective was to evaluate the effectiveness of choropleth and tile grid maps, and their banded versions, to show changes over time. We chose the tile grid map as an alternative to the choropleth as it overcomes some of its limitations (as mentioned in the previous sections) [30]. We aimed to conduct an experiment that covered the most common and possible tasks related to changes over time. The final tasks set out in the test were taken from a list of all the combinations for the three variables included in the study (region, time step, and value) and then the tasks which did not involve any changes over time were discarded (see Table 1 in the Suppl. material) [59]. All the tasks included a change of value in one or more areas (or regions) of the map for one or more time steps. The statements of the tasks also described everyday questions that users frequently have to answer while analysing and exploring these time related scenarios, and were based on the insights compiled from the initial user research phase.

Once these questions had been set, we reviewed them with potential users in order to define which questions and tasks they wanted to answer and perform while working with the maps. We also discussed the intended use of each selected type of task and a representative statement for each one. As a result, the final selected tasks were the following:

1) **Task 1.** Detect trends within a specific region. Task statement: Could you describe the evolution of COVID-19 new cases for region number n (region name) over time (t1, t2, t3)?
2) **Task 2.** Identify the regions that meet certain conditions (evolution patterns) over time. Task statement: Could you name five regions that have remained stable in terms of number of new cases throughout these three periods of time?
3) **Task 3.** Determine the global trend (all regions) over time. Statement: Could you determine what the global trend for all of Catalonia was over time?

We classify these tasks based on Bertin's taxonomy based about the level of analysis [60] and in turn on the attributes presented in each representation [61]. Based on this, we can differentiate between 1.Elementary tasks, referring to individual elements (Task 1 and Task 2) or 2.Synoptic tasks, referring to the whole reference set (Task 3). In turn, these three tasks can be classified according to their objective and the relationships established in each representation. Task 1, for example, is oriented towards a specific target (or region), as is Task 3, which aims to identify the overall trend of a given territory. Task 2, however, is based on a number of conditions that can be translated into constraints: Identify a number of regions that meet a specific pattern [61].

3.3 Hypotheses formulation

Our hypotheses are derived from prior cartogram evaluations, perception studies, popular critiques of cartograms and tile grid maps and previous research on geo-temporal representations [50], [62], [63]. The most common observations refer to area-size bias in the case of choropleths and shape distortions in the case of tile grid maps. Although these conceptual representations may hinder the recognition of familiar geographic regions [21], they do solve the area-size bias problem associated with choropleths [30]. It is suggested that recognition can be improved through the use of labeling elements [19]. With these limitations in mind, we added numeric labels to the representations to identify the regions of interest in all maps, although they may have been redundant in some cases (such as the choropleth), where the geography (shape) is more recognizable.

Prior to the development of the study, we held co-creative sessions and ethnographic studies combined with short interviews with five participants. These participants were researchers of the Barcelona city council in the areas of economy, environment and health, who were familiar with static map representations. During this initial phase, we observed their behaviour while performing daily tasks and gathered information on the main obstacles encountered while using static maps to analyze changes over time and to identify trends [64]. This also helped us identify which common user questions arise when exploring changes using maps, and to establish the hypotheses of our study [48], [65].

Before the formulation of our hypotheses it must be noted that when we say that one map or one representation type "is more effective" or "works better" than another for a particular task, it is because there are significant quantitative differences present (participants commit fewer errors and take less time to perform the tasks). With regards to the qualitative assessment, the participants may find one representation easier or more useful than another while solving a task, or express more positive comments about a certain representation.

H1: Banded choropleth and banded tile grid maps will work better for detecting pattern changes in a specific region over time [Task 1].

H2: Banded maps will work better for detecting regions that meet a certain pattern or condition over time [Task 2].

H3: Tile grid maps will work better than both banded versions to show global trends over time [Task 3]. Tile grid maps will also work significantly better than choropleths due to area-size bias distortion when showing global trends over time.

3.4 Participants

For this study we used a convenience sample of 32 people, composed of 15 men and 17 women between 22-50 years of age, with low-medium map visualization knowledge. All the participants were familiar with choropleth maps, but none of them had any knowledge about the other representation types. The recruited participants were students, researchers and administrative staff from the fields of human resources, communication and accounting. Before the test sessions, users were asked not to have consumed stimulant substances that could have affected their performance and the test results. The participants were residents of Catalonia, 25% of them of foreign origin but long term residents (over 10 years) and familiar with the geographical distribution of Catalonia. The participants took part in the experiment one at a time in a private room, in order to ensure a disturbance free environment and at the start of business hours (from 9 to 12 h), complying with the protocols required for COVID-19 prevention. The duration of each test was about 15-20 minutes. The test sessions were carried out over the course of three weeks. All participants completed the experiment and no data was eliminated. The participants did not receive any remuneration or compensation for their participation.

3.5 Context introduction and pilot test

Prior to the experiment, we carried out a short context introduction session [66], [67] where each of the participants filled out a standard bioethics form confirming and accepting their free participation in the experiment. They were also briefed on the purpose of the research and the type of representations under study: For each type of map, we showed an example, together with a short description on how to interpret it (time steps, representation types, legends, etc.). The participants also completed a short questionnaire with details on their age group, level of visualization expertise, employment background and confirming they did not suffer from color blindness.

Moreover, a pilot testing session was carried out with two users under the same conditions presented to the rest of the participants. This pilot test was conducted in order to reveal undetected problems related to the design of the test [68], [69]. The results of these two participants were not included in the final study results.

3.6 Test environment

We developed a simple online tool to guide the participants through the experiment, presenting each of the tasks, and which was controlled by a moderator who assisted the participants during the test. This assistance consisted of answering any queries related to the task statements but not in helping to solve the tasks themselves or giving any clues on the solutions [70]. The sessions were also recorded in order to review the results in terms of quantitative data (time and success rates) and qualitative data (opinions and valuations). Participants were informed that times and results, together with their comments, would be collected. [71]. The study was performed using a laptop (MacBook Pro 13-inch, 2017, 3.1 GHz and 16 GB RAM) and connected to a 27" 4K BenQ display (3840x2160). No interaction (mouse or keyboard input) was required as the participants could solve the task by analyzing the visualizations presented on the screen and responding aloud to the moderator. After having read the statement, and confirming to the moderator that they were ready, the time to perform the task was measured from the moment the visualization appeared on the screen and until the participant gave the answer (for further and specific details of each task time measurements see Table 6 of the Suppl. material).

3.6.1 *Task-based questions (Quantitative evaluation)*

For the task-based part of the study, the participants answered three multiple-choice questions about different visualizations, using the four types of representations under analysis. More details on the task performance are presented in section 3.7.2 (see also Fig. 2 and Table 6 of the Suppl. material for specific details of each task).

The results of each task ('success', 'error', 'abandonment') as well as the completion times needed to perform each task were gathered.

3.6.2 *Decision-making & perceived difficulty (Qualitative evaluation)*

After finishing the three tasks, we required the users to review the four representations (one after another) involved in each task. Users were asked about which representation favoured decision making and facilitated the resolution of the task [72]. They were also encouraged to share their thoughts on the positive or negative aspects of each map, as well as any other observations while performing the task.

This helped us to understand some answers, mistakes or why users valued some maps more positively than others [73], [74].

This approach of analyzing each task individually as they are completed, favors review quality by using short-term memory, which allows users to provide comments with the tasks still fresh in their minds, instead of having a questionnaire at the end of the test [75]. At the same time, this approach has no effect on the quantitative metrics measured (completion times and success rates).

3.7 Values shown, maps and other considerations

3.7.1 Geography

The tasks for the study were carried out on a map of the different regions of Catalonia (Spanish autonomous region) where a hypothetical evolution of new COVID-19 cases was presented over a period of three time steps (one week per time step). The choice of the Catalan autonomous region was due to the fact that it is not a region that is usually represented with conceptual map types (for example, tile grid maps) as would typically happen with countries such as the United States, or cities like London [23], [76]. However, the characteristics associated with the geography (shape) of Catalonia and the dimensions of its regions allow us to represent it easily using a conceptual map. The shape of the regions within Catalonia using choropleths is highly recognizable by users who know its geography. However, few people (including natives) are able to recognize the exact position of all of its regions, which is why we offered auxiliary legends to help locate the region or regions involved in the tasks.

3.7.2 Map design and values presented

All the maps included in the study were represented under the same conditions: All the maps were presented on white backgrounds. The regions were displayed with white contours and numerical labelling. This labelling was needed for the identification of the regions. For Task 1, an auxiliary legend was added which included the location of the region together with its name, in order to facilitate the location of the regions in all the representations or map types (see Fig. 1). For the banded versions, an additional legend was also included to help identify the different time steps in the maps (see Fig. 1c and Fig.1d). Four categories were established in relation to the maximum number of infections, and were classified as follows: 'Low', 'Medium', 'High' and 'Very High', based on a colorblind-safe palette from the ColorBrewer tool [77]. The color palette used for values, texts and legends is presented in Fig. 1 of the Suppl. material.

It should be noted that the information represented does not correspond to any real data. For all the tasks the values of the regions are random and change from task to task (and from map to map), this is done to avoid learning-bias between representations within the same task [78].

Task 1: The region to locate and the possible patterns presented are random. The user has to identify the trend among several alternatives. Depending on the answer, it will be necessary to contrast the answer with the exact values at each time step (See Table 6 and Fig. 2 of the Suppl. material).

Task 2: The regions to be located were selected randomly as well as the values presented, but not the pattern required, which was always stable. That is to say, the regions with the same colour for all time steps ('Low', 'Medium', 'High' or 'Very high'). The user has to identify 5 regions of the 9 available in each map (See Table 6 of the Suppl. material).

Task 3: The global trend is always stable (the same number of regions belonging to a certain category ('Low','Medium','High','Very high') in the three time steps, but randomly combining changes between the bigger regions and the smallest regions of Catalonia (See Table 6, Table 7 and Fig. 3 of the Supplemental Material). The user has to identify the correct trend between three options (given in the statement): increase, decrease or stable.

For each task, the maps were presented using a pool composed of eight variants of each representation. Each participant answered three tasks for four representations (the maps under study). Therefore 96 possible representations (3x4x8) were prepared in this study to be shown randomly in the user testing performance.

3.7.3 Number of time steps

This study only takes into consideration static representations (print press, online reports or websites) which do not use any type of interactive element to favor comparison (sliders, video player controls, carousels, etc.) [79], [80]. To establish the most common number of time steps to be represented, we reviewed several reports published by the EU and US governments and official institutions. We also discussed which was the most common number of time steps with specialized users, in this case, municipality communication staff. Based on this previous analysis, the most frequent options were: two time steps (to compare the situation between t1 and t2), three time steps (normally to present four-month periods in a year) and four time steps, to show data evolution over the quarters of a year (three-month periods). To show more than four time steps, alternatives such as traditional line charts are preferable [81], [82]. We settled on an intermediate number of time steps, in this case, three.

3.7.4 Other particularities of the tasks

In the tasks for which more than one solution was possible, the amount of correct options available was always the same. For example, to identify five regions which met a specific condition (Task 2), there were nine possible regions in each representation that met the stated criteria. For Tasks 1 and 3 only one correct answer is allowed (For further details, see Table 6 and Fig.2 of the Suppl.material).

The statements of the tasks focused on the COVID-19 pandemic, and referred to the number of new cases detected by region. This was the background for the development of the test. However, the data and values presented were made up, purely circumstantial and could be applied to any other social data scenario (marketing, finance, administration, ethnographics, etc.) [83]. Users were informed that the values presented were not real and varied from map to map and from task to task in order to avoid 'automatic' answers for the different map representations [84]. Although the data shown in each map differed, the types of maps were presented in a Latin square random order to avoid possible

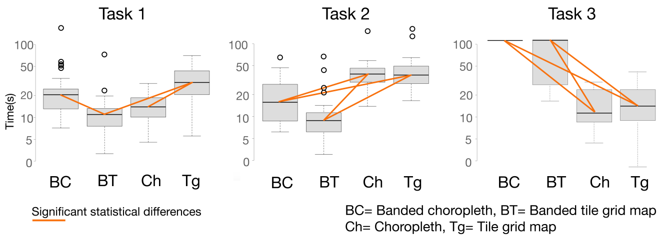


Fig. 2. Boxplots showing the distribution for the results related to completion time for all representations in the three tasks presented. Orange lines show the significant statistical differences between maps.

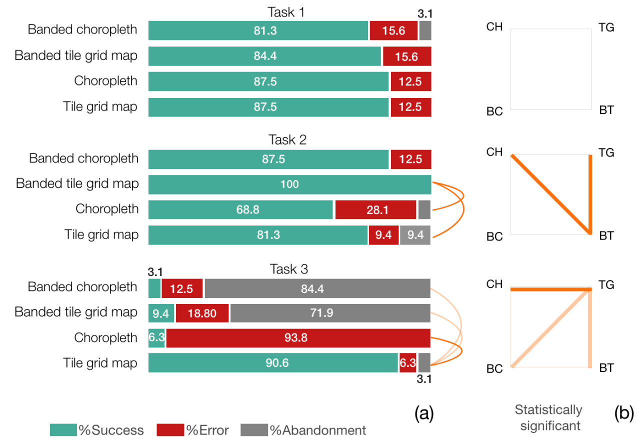


Fig. 3. (a) Success, failure and abandonment rates by task and type of map. (b) Relationships of significant statistical differences between representations. Opacity indicates a higher or lower level of significance.

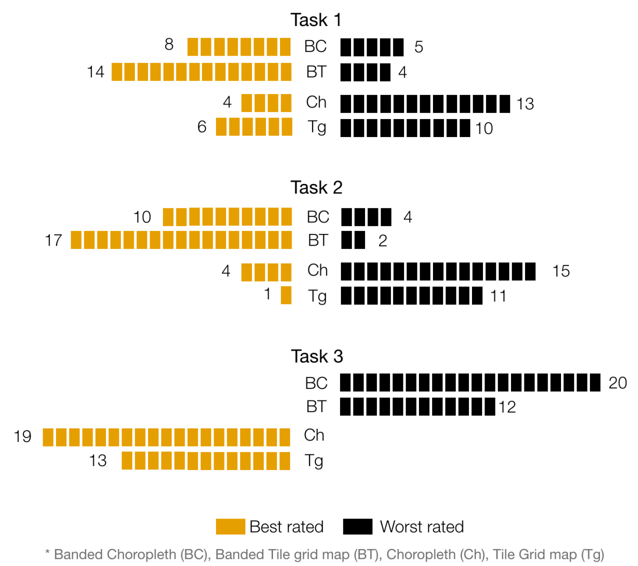


Fig. 4. Most and least favourable maps (number of positive and negative votes) to solve the task and which favour decision making.

additional bias [85], [86], [87]. The users were encouraged to abandon a task at any time if they considered it to be too complex or tedious, or if the answer given would be purely the result of guesswork. The abandonment rate is also a good indicator of if the user would not be able to solve the task or understand the trend in a real world scenario (away from a test environment). In turn, it offered us valuable information on the perceived effort required for the task [88]. For Task3, moreover, given that that the probabilities of a success were 33.3% (choosing between 'increasing', 'decreasing' or 'stable'), voluntary abandonment reduces this risk of getting false correct answers.

When users abandoned a task, the time assigned to the result was the maximum permitted to solve the task [89], [90], [91]. The maximum time allotted was two minutes (120 seconds). This time was calculated based on the average time needed by users to perform the task during the pilot test [89]. In order to classify the results, it is important to determine beforehand what is to be considered as a success, a failure (error), or an abandonment, even when some mistakes may be due to the interpretation of the statement, legends or colors, and not necessarily a consequence of the map representation itself. The conditions for checking the results of each task are defined in the script of the test (see Table 6. of the Suppl. material).

4 Results

4.1 Quantitative metrics. Task-based study

We used ANOVA F-tests with significance level =.05 to carry out the statistical analysis for the time results together with a Tukey post-hoc test to reveal the most significant pairwises. We used a Chi-square test for independence with α=.05 to evaluate categorical variables in the success rate results. The within-subject independent variables were the four map representations. The dependent measurements are the completion time and success rate.

The null hypothesis was that the representation type does not affect completion times or success rates. When the probability of the null hypothesis (p-value) is less than 0.05 (or, equivalently the F-value is greater than the critical F-value, Fcr), the null hypothesis is rejected.

For further information, the values corresponding to the confidence intervals obtained for each task are included in Table 4 and Fig. 5 of the Suppl. material.

4.1.1 *Detecting trends within a specific region (Task 1)*

Regarding the time needed to perform the task, we found statistically significant differences depending on the map type (f(3)=10.63, p=0.000). A Tukey post-hoc test revealed that the most significant pairwise differences were between the banded tile grid map and tile grid map, and choropleth and tile grid map (for more details see Fig. 2 in the main text and Table 2 of the Suppl. material). Values on success rates (X2(1,N=32)=3.3231, p-value=0.7673) indicated that the results obtained were not statistically significant (see Fig. 3 in the main text and Table 3 of the Suppl. material). In this case, the H1 hypothesis was ruled out, as the banded versions were expected to work more efficiently (completion times and success rates) than the small multiple maps due to the visual distance between maps. In the case of success rates, we observed no significant differences between the four representations. In terms of completion time, the banded tile grid map and choropleth performed significantly better.

4.1.2 *Identifying the regions that meet certain conditions (Task 2)*

Regarding the time needed to perform the task, we found statistically significant differences depending on the map type (f(3)=21.6, p=0.000). A Tukey post-hoc test revealed significant pairwise differences between the banded versions and the small multiple versions (see Fig. 2 in the main text and Table 2 of the Suppl. material). The same happened with the success rate (see Fig. 3 in the main text and Table 3 of the Suppl. material). In this case, the H2 hypothesis is fulfilled since the differences in terms of completion times and success rates are significantly better for the banded versions.

4.1.3 *Determining the global trend (Task 3)*

Regarding the time needed to perform the task, we found statistically significant differences depending on the map type (f(3)=73.86, p=0.000). A Tukey post-hoc test revealed significant pairwise differences between the banded versions and the small multiple versions (see Fig. 2 in the main text and Table 2 of the Suppl. material). The same happened with the success rates, which were also significantly better for the tile grid map than for the choropleth (see Fig. 3 in the main text and Table 3 of the Suppl. material). The H3 hypothesis is also confirmed. Applying time correction in the case of abandonment results, we observed significant differences between the small multiple maps and the banded versions. Moreover, and regarding the success rates, there were also significant differences observed between the choropleth and the tile grid map, the latter obtaining better success rates. The banded versions presented the highest abandonment rates (84.4% and 71.9%).

4.2 Qualitative evaluation. Decision-making. Perceived difficulty

After carrying out each of the tasks, users cast their vote for the representation which best favored the detection of the trend (helping in decision-making, simplicity, level of comfort) and the representation which least favored the analysis. They also shared with us the problems encountered together with general feedback while performing the tasks for the different maps.

For Task 1, the responses were more balanced, with the best valued representations being the banded maps (banded tile grid map, followed by banded choropleth) compared to the small multiples (see Fig. 4). In Task 2, the results were much more polarized, and the differences between the banded (the best valued) and the small multiple representations (once again, the worst valued) increased substantially (see Fig. 4). The banded representations not only received the highest evaluation but also received hardly any negative votes. For Task 3, the "usefulness" of small multiple maps was considered unanimous: users valued the banded representations very negatively as they did not favor the resolution of the task at all. Users valued the choropleth map as the most favorable for carrying out the task (see Fig. 4), due to its familiarity (despite having a high error rate). Table 1 shows the most significant aspects, classified by topic, task and map type, as well as the number of times they were mentioned by users (see also Table 5 of the Suppl. material to review the most frequent comments collected during the study).

Fig. 5. In this graph we have combined four of the most significant variables, using a scatter plot representation. For the three tasks, we compare: success rate (x axis), time spent to accomplish the task (y axis), number of times a graph was voted as the most effective (size), and the type of map (color and glyph type or shape). The larger glyphs indicate the best valued map representations. The grey area highlights the most favourable results: highest success rates and lowest completion times.

TABLE 1
Summary of the most frequent comments organized by category, map type and task.

Task	Map	Comment	Freq.
1	BC-BT	Banded representations favour the task because of the proximity between maps	13
1	Ch–Tg	Difficulty to locate a region depending on its position, boundary regions and their color	9
2	BC-BT	Banded representations favour the task due to the proximity between maps	9
2	BC-BT	Small multiple maps require a random search pattern	9
2	BC-BT	Small multiple maps require more cognitive load and concentration	8
3	BC-BT	Banded versions are too complex to establish a pattern	17
3	Ch	Choropleth is perceived as the better representation due to its familiarity	9
3	Tg	Tile grid map favours the task	5

4.3 Summarizing the results

In this section, and with the help of Fig.5, we attempt to draw conclusions arising from the combination of quantitative and qualitative results. We compare the two quantitative measurements: success rate in carrying out the task (x-axis) and the average time (in seconds) needed to solve the task (y-axis) for each of the tasks and each of the representations. Through the use of size, we also represent those maps that received better subjective evaluations (the larger the size, the more votes received as the map type that favors decision-making or that is more useful when solving the task). We used 'color' to differentiate choropleths from tile grid maps and 'shape' to differentiate small multiple representations from their banded versions. We have also highlighted in grey, the most efficient area on the grid (delimiting higher success rates and shorter completion times). Therefore, the representations contained in these areas are the most recommended when carrying out the task from a quantitative point of view. Theoretically, the representations contained within this area should have received the highest number of votes from the audience in terms of their effectiveness (larger icons).

In the case of Task 1 (detecting changes within a specific region), the best valued representations turned out to be the banded tile grid map and choropleth (highest success rates, and least time to perform the task), the former also being the best valued. However, the choropleth, despite obtaining better quantitative results (time and success rates) was valued worse than the banded choropleth. We will take a closer look at these results in the discussion section (see Fig. 5). For Task 2, the banded representations are clearly the most effective, both from a quantitative point of view (time and success rates), and a qualitative point of view (number of votes and positive comments received) (see Fig. 5). For Task 3, the only effective map is the tile grid map (based on its success rates). In this case, the banded alternatives did not receive any votes and saw the highest abandonment rates and consequently the lowest success rates. Paradoxically, the choropleth obtained a higher number of positive votes (being perceived as the most favorable map to solve the task) and had lower completion times [92], [93] (see Fig. 5).

4.4 Limitations

When trying to cover the spectrum of possible tasks and possible statements, we limited ourselves to a particular subset of common tasks and simple statements for task setup. There are also numerous limitations when choosing the possible geographic areas, when representing data through cartograms, such as size, distance between regions and maps, level of detail (continent, country, region), and the number of time steps. As mentioned previously, another limitation is that the scenarios of the proposed tasks were not based on real data. This was because we had to present multiple different values in order to avoid learning bias between the different maps for the same task.

With regards to the number of time steps selected, we assumed that the results obtained would be negatively impacted by increasing the time steps. That is, the difficulty for n + 1 would be equal to, or greater than for n time steps due to the increase in visual stimuli [94]. Additionally, in the case of banded representations, the number of time steps must permit sufficient width for the bands to guarantee the correct color interpretation (especially for color-blind users) [95], [96].

Similarly, the context-related aspects detailed in sections 3.4, 3.5 and 3.6 such as the presence of a moderator, the fact that the tasks were being timed, etc. could have a positive or negative impact on the test results: Depending on the characteristics/personality of the participant this could result in more motivation to complete the tasks, but also more stress when solving them [71].

It is relevant to mention that the original version of this study included quantitative results obtained through an eye-tracker device (model Gazepoint GP2) [97]. However, the metrics obtained in terms of fixation duration and pupillometry were not significant, so it was decided to simplify the study in this aspect. Nevertheless, the use of the eye-tracker is a recommended practice that helped us to understand and confirm visual and search pattern behaviour of our users [98]. (see Fig. 4 of the Suppl. material).

5 Discussion

In the analysis of temporal changes in geographical representations, the optimization of cognitive resources is essential for favoring comparison, the detection of changes and the identification of patterns [30]. The effectiveness of the representation used is also closely related to the task to be performed [50]. We used quantitative and qualitative indicators to assess the differences between small multiple representations and banded representations.

5.1 Detecting trends within a specific region

From user feedback, Task 1 was the easiest to solve. All four maps or representations obtained very similar results in terms of success rates. However, the banded tile grid map and the choropleth showed significant differences in completion times when compared to the other two maps (see Fig. 5). Contrary to what we expected, we did not observe a marked advantage for both banded representations although they were the best valued by users (see Fig. 4). The comments regarding the small multiple representations were based on the difficulty in comparing the values in each time step for a certain region due to the distance between maps: "With the small multiple maps the process is the following, you locate the region in each of the three maps and then identify (or compare) the changes in the three regions, whereas in the combined (banded) maps everything is in the same place". "It requires a little more effort because you have to compare the regions in all three maps. The comparison is more direct in the banded versions". When evaluating these comments we had to pose the question: Why did choropleths obtain better times than banded choropleths despite not being the best rated or receiving the most favorable comments? This can be explained by analyzing the users' observations on the banded choropleth: "Depending on the color of the bands in surrounding regions, sometimes it is more difficult to establish the boundaries. This does not happen, however, with the bands in the tile grid map" [99]. Similar comments were made for the simple choropleth:

"The characteristic shape often helps to identify a region on the three maps, but other times, when the nearby regions share the same color, it is no longer as simple" [100], [101], [102]. Two users further pointed out that there is not much difference in the results due to the simplicity of the task but this could be changed by increasing the number of regions to compare: "We are only looking at the trend of one region. If we were asked to observe the changes of two or three regions simultaneously, the banded versions would behave much more efficiently than the traditional ones (referring to choropleth and tile grid map)".

5.2 Identifying the regions that meet certain conditions

Task 2 presented significant differences (completion times and success rates) between the banded versions and the small multiple versions (the former obtaining shorter completion times and higher success rates). Furthermore, the subjective evaluations were more favorable for the banded versions (see Fig. 4). "It is much easier to reach a conclusion with the banded version, all the information is in the same place, and you don't waste time comparing between maps" [103], [104].

The comments shared by users helped us to understand the way in which they solved this task and processed the visual information: "In the case of the small multiple maps (choropleth and tile grid map) the search is a bit random, you choose a specific region in t1 and check if it has not changed for t2 and t3". "You start with the most easily locatable regions (clearer positions) on the map. When you find one that meets the requirements, you pick another one at random, and so on". "Sometimes the comparison can be bidirectional: you pick a region randomly at t3 and check that it has the same color for t2 and t1".

Some users highlighted aspects such as the shape and location of the regions (repeating comments they made on Task 1): "It is easier to visually locate regions with a characteristic shape (with choropleths) or in an easy-to-locate position (with tile grid maps): Positions at the boundaries are easier to find than the interior ones, in which you have to consult the labels for each square (region) to ensure that it is the correct one."

A user also commented on the high cognitive load associated with Task 2 and small multiple representations, including factors such as attention: "You are so focused on reviewing regions and jumping from map to map, that you no longer remember if you are repeating any of those regions mentioned". "Have I already repeated this? The task is much more comfortable and straightforward to solve in the case of banded maps".

To finish the analysis of Task 2, it would be interesting to discuss the simplicity of the condition proposed for the search, which is 'regions that remain stable over time'. This requires locating monochrome regions. However, this task seems to require more cognitive load for the small multiple representations: "It requires a lot of concentration, and being aware of changes, jumping from map to map. With a more complex pattern (a particular combination of t1, t2, t3) it would be slightly more difficult in the case of the banded versions but much more difficult in the case of the small multiple ones" [105], [106].

5.3 Determining the global trend

In this task, the banded maps showed a high abandonment rate. Users abandoned the task when they realized the answer was going to be guesswork: "I have no idea. I would be unable to identify the trend. The only solution would be to count the sub-regions for the time steps, and I could spend the whole morning doing that". "If I had to answer, it would be random". "The graph has too much information, it is impossible to determine the trend". These comments lead us to rule out the banded maps as a suitable option for communicating global trends.

Users were able to finish the task using both small multiple maps, but the choropleth obtained just 6.3% success when compared to the tile grid map which obtained 90.6% success. In spite of this, users valued the choropleth more positively (as the most helpful and effective map for solving the task) due to its familiarity: "Without doubt, the representation that best favors decision-making is the choropleth. It is a very familiar representation to me". This is what we refer to as a 'false success', which is when the user thinks that they have solved the task correctly but in fact, they are mistaken [107], [108]. Only in some cases were users able to identify the area-size bias problem related to choropleth maps with comments such as: "The choropleth is the clearest one to show the global trend. Wait! No! If it is related to the number of regions, I think that the trend is not increasing".

Despite the good results in the case of tile grid maps, users were not as confident in their answers when compared to using choropleths: "I think the pattern remains stable, but I am not sure".

It is noteworthy that for Task 3 the choropleth was the best valued map due to its familiarity, but this was not the case for Task 1 or Task 2. This may be explained if we suppose that when the perceived effectiveness-difficulty is similar for each map, we tend to value the one which is more familiar to us more favorably. When the perceived effectiveness-difficulty between the maps is very clear, we tend to choose more objectively [25], [109], [110].

6 CONCLUSION

We describe the evaluation of two static representations together with their corresponding banded versions for communicating changes over time. We measure completion times and success rates while the participants resolve habitual tasks common to their speciality fields. In addition, we analyze the participants' subjective opinions for each representation in terms of effectiveness. We also examine the general comments gathered in terms of the strong and weak points for each representation, as well as the problems encountered while carrying out the tasks.

The quantitative and qualitative metrics showed significant differences between the maps and their effectiveness depending on the task being performed. The **banded tile grid map** stood out as the best option in tasks associated with detecting region trends as well as when identifying regions that fit a certain pattern or condition (Task 1 and Task 2). When communicating global changes over time (Task 3), we saw that both banded versions would need to be discarded because of the difficulty in interpreting them and the high abandonment rates recorded. The choropleth map, while familiar to the audience, is not a suitable option due to the high error rates caused by the area-size bias problem (93.8%). The only viable candidate, the tile grid map, is also open to debate. It obtained the highest success rates (90.6%) and optimal completion times, but the lack of confidence shown by the participants, leads us to advise against its use. Due to the doubts expressed while using tile grid maps, and the preference for a less effective representation (choropleth), we conclude that no static map included in this study is sufficiently effective for the communication of global trends over time. The use of new representations in the media (such as banded maps or even tile grid maps) would be a high-risk option as the information presented might not be understood by a general audience. However, in more specialized scenarios with well versed users, the **tile grid map** is preferable to the traditional choropleth.

Given that the familiarity of certain representations affects the evaluation of their perceived effectiveness, and that choropleth maps are the most used in non-interactive solutions, we believe that the alternatives (tile grid map and both banded versions) should be studied further, as well as the learning curve required by the audience to guarantee their comprehension. While it is unlikely that a single evaluation study can be conducted to analyze all possible scenarios, we believe that this study can be a useful starting point, while providing insights for future map variants to be used with the broader public: Banded versions facilitate the task of comparison in some contexts and tile grid maps show promise against traditional choropleths to communicate global trends. Further study is required to measure the possible effectiveness in media communication scenarios or in aiding with decision making.

7 Next steps

With regards to the results obtained, we believe that it would be advisable to perform a new study employing real, large datasets, which would allow us to play with different snapshots (of time and values), while simultaneously preventing learning bias in the user testing. The use of multiple geographies in the same test and the exploration of other possible tasks or scenarios is also advisable. In the introduction, as well as in the section dealing with time steps and their limitations, we mention dynamic or interactive mechanisms to make up for the limitations of static maps. Although these alternatives do have advantages and are often used for representations of multiple time steps, they can present the same problems as static maps from a cognitive load perspective. For this reason, their effectiveness should be questioned when seeking to communicate trends and changes over time or when aiding decision making processes. Another aspect of interest for further analysis is how familiarity and a similar level of difficulty can influence the perception of effectiveness when interpreting maps.

Acknowledgments

The authors would like to thank the users and experts that participated in the early design stages of the experiment, the participants who kindly provided their input during the user testing sessions and to the journalist and teacher, Steve Watts for his inestimable contribution to the revision of this manuscript.

References

[1] M. Friendly, D. Denis, and H. Truman, "Milestones in the history of thematic cartography, statistica l graphics, and data visualization," 01 2001, 00001.

[2] C. Nicolet, *Space, Geography, and Politics in the Early Roman Empire*. University of Michigan Press, 1991, 00008 Google-Books-ID: 7IQpCaCpQ1MC.

[3] M. Kernik and S. Nelson, "9. Social Maps," in *Mapping, Society, and Technology*. University of Minnesota Libraries Publishing, Sep. 2017. [Online]. Available: https://open.lib.umn.edu/mapping/chapter/9-social-maps/

[4] L. Vaughan, *Mapping Society:The Spatial dimensions of social cartography*. London: UCL Press, 2018. [Online]. Available: https://www.jstor.org/stable/j.ctv550dcj

[5] C. Brunsdon, J. Corcoran, and G. Higgs, "Visualising space and time in crime patterns: A comparison of methods," *Computers, Environment and Urban Systems*, vol. 31, no. 1, pp. 52–75, Jan. 2007, 00146. [Online]. Available: http://shorturl.at/ikruw

[6] T. Sobral, T. Galvão, and J. Borges, "Visualization of Urban Mobility Data from Intelligent Transportation Systems," *Sensors*, vol. 19, no. 2, p. 332, Jan. 2019, 00003. [Online]. Available: https://www.mdpi.com/1424-8220/19/2/332

[7] L. Chou, "Top 10 Map Types in Data Visualization," Aug. 2019. [Online]. Available: https://towardsdatascience.com/top-10-map-types-in-data-visualization-b3a80898ea70

[8] C. A. Blok, "Dynamic visualization variables in animation to support monitoring of spatial phenomena," Jan. 2005, 00072 Accepted: 2005-05-20T10:44:19Z ISBN: 9789068093674 ISSN: 0169-4839 Publisher: Utrecht University. [Online]. Available: http://localhost/handle/1874/2895

[9] G. Melloni, A. Caglio, and P. Perego, "Saying more with less? Disclosure conciseness, completeness and balance in Integrated Reports," *Journal of Accounting and Public Policy*, vol. 36, no. 3, pp. 220–238, May 2017, 00132. [Online]. Available: http://shorturl.at/opuzP

[10] P. Noriega, T. Pinto, J. Assunção, F. Lourenço, C. Ungureanu, and E. Ribeiro, "Web Press Carousel and His Effects Over News Memorization," in *Advances in Ergonomics in Design*, ser. Advances in Intelligent Systems and Computing, F. Rebelo and M. Soares, Eds. Cham: Springer International Publishing, 2016, pp. 53–61.

[11] S. Bogacz and J. G. Trafton, "Understanding Static and Dynamic Visualizations," in *Diagrammatic Representation and Inference*, ser. Lecture Notes in Computer Science, M. Hegarty, B. Meyer, and N. H. Narayanan, Eds. Berlin, Heidelberg: Springer, 2002, pp. 347–349, 00008.

[12] E. PAIS, "La crisis del coronavirus en EL PAÍS," 2021. [Online]. Available: https://elpais.com/sociedad/crisis-del-coronavirus/

[13] A. Kevorkian, T. Grenet, and H. Gallee, "Tracking the Covid-19 pandemic : Simple visualization of the epidemic states and trajectories of select European countries & assessing the effects of delays in official response." Public and Global Health, preprint, Mar. 2020. [Online]. Available: http://medrxiv.org/lookup/doi/10.1101/2020.03.14.20035964

[14] T. N. Y. Times, "Coronavirus in the U.S.: Latest Map and Case Count," *The New York Times*, Mar. 2020, 00045. [Online]. Available: https://www.nytimes.com/interactive/2021/us/covid-cases.html

[15] W. Post, "U.S. coronavirus cases: Tracking deaths, confirmed cases by state," 2021. [Online]. Available: https://shorturl.at/psJLT

[16] M. Dodge, R. Kitchin, and C. Perkins, Eds., *Rethinking Maps: New Frontiers in Cartographic Theory*. London: Routledge, Jun. 2011, 00290.

[17] K. Slomska-Przech and I. M. Golebiowska, "Do Different Map Types Support Map Reading Equally? Comparing Choropleth, Graduated Symbols, and Isoline Maps for Map Use Tasks," *ISPRS International Journal of Geo-Information*, vol. 10, no. 2, p. 69, Feb. 2021, 00002 Number: 2 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2220-9964/10/2/69

[18] N. J. W. Thrower, *Maps & man: an examination of cartography in relation to culture and civilization*. Englewood Cliffs, N.J: Prentice-Hall, 1972, open Library ID: OL5707142M.

[19] B. D. Dent, "Communication Aspects of Value-by-Area Cartograms," *The American Cartographer*, vol. 2, no. 2, pp. 154–168, Jan. 1975, publisher: Taylor & Francis _eprint: https://doi.org/10.1559/152304075784313278. [Online]. Available: https://doi.org/10.1559/152304075784313278

[20] D. Dorling, "Area Cartograms: Their Use and Creation," in *The Map Reader*, M. Dodge, R. Kitchin, and C. Perkins, Eds. Chichester, UK: John Wiley & Sons, Ltd, Apr. 2011, pp. 252–260. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/9780470979587.ch33

[21] W. Tobler, "Thirty-Five Years of Computer Cartograms," *Annals of the Association of American Geographers*, vol. 94, no. 1, pp. 58–73, 2004. [Online]. Available: https://www.jstor.org/stable/3694068

[22] P. Vardhan, "Tile Narrative: Scrollytelling with Grid Maps," OSF Preprints, Tech. Rep., Aug. 2020. [Online]. Available: https://osf.io/xr64m/

[23] Y. Du, L. Ren, Y. Zhou, J. Li, F. Tian, and G. Dai, "Banded choropleth map," *Personal and Ubiquitous Computing*, vol. 22, no. 3, pp. 503–510, Jun. 2018, 00003. [Online]. Available: https://doi.org/10.1007/s00779-018-1120-y

[24] A. Jacobs, J. Pinto, and M. Shiffrar, "Experience, Context, and the Visual Perception of Human Movement," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 30, no. 5, pp. 822–835, 2004, place: US Publisher: American Psychological Association.

[25] M. H. Martens and M. R. J. Fox, "Do familiarity and expectations change perception? Drivers' glances and response to changes," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 10, no. 6, pp. 476–492, Nov. 2007. [Online]. Available: http://shorturl.at/eoARZ

[26] O. Dovbysh, "Automating the News: How Algorithms are Rewriting the Media," *Digital Journalism*, vol. 0, no. 0, pp. 1–3, Jul. 2020, publisher: Routledge _eprint: https://doi.org/10.1080/21670811.2020.1777883. [Online]. Available: https://doi.org/10.1080/21670811.2020.1777883

[27] S. Esteves and M. Neves, "Are Charts Going Digital? The Case of Data Visualization on Portuguese Media," in *Human Interaction, Emerging Technologies and Future Applications III*, ser. Advances in Intelligent Systems and Computing, T. Ahram, R. Taiar, K. Langlois, and A. Choplin, Eds. Cham: Springer International Publishing, 2021, pp. 404–409.

[28] T. Spinde, F. Hamborg, K. Donnay, A. Becerra, and B. Gipp, "Enabling News Consumers to View and Understand Biased News Coverage: A Study on the Perception and Visualization of Media Bias," in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, ser. JCDL '20. New York, NY, USA: Association for Computing Machinery, Aug. 2020, pp. 389–392, 00001. [Online]. Available: http://doi.org/10.1145/3383583.3398619

[29] M. P. Armstrong and N. Xiao, "Retrospective Deconstruction of Statistical Maps: A Choropleth Case Study," *Annals of the American Association of Geographers*, vol. 108, no. 1, pp. 179–203, Jan. 2018. [Online]. Available: https://doi.org/10.1080/24694452.2017.1356698

[30] J. Schiewe, "Empirical Studies on the Visual Perception of Spatial Patterns in Choropleth Maps," *KN - Journal of Cartography and Geographic Information*, vol. 69, no. 3, pp. 217–228, Sep. 2019. [Online]. Available: https://doi.org/10.1007/s42489-019-00026-y

[31] L. Besançon, M. Cooper, A. Ynnerman, and F. Vernier, "An Evaluation of Visualization Methods for Population Statistics Based on Choropleth Maps," *arXiv:2005.00324 [cs]*, May 2020, 00001 arXiv: 2005.00324. [Online]. Available: http://arxiv.org/abs/2005.00324

[32] D. Keim, S. North, C. Panse, and J. Schneidewind, "Efficient cartogram generation: a comparison," in *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.*, Oct. 2002, pp. 33–36, iSSN: 1522-404X.

[33] S. Nusrat, M. J. Alam, and S. Kobourov, "Evaluating Cartogram Effectiveness," *IEEE transactions on visualization and computer graphics*, vol. 24, no. 2, pp. 1077–1090, 2018.

[34] Datavizcatalogue, "Chart Combinations: Tile Grid Maps - Dataviz Catalogue Blog," Jun. 2018. [Online]. Available: http://datavizcatalogue.com/blog/chart-combinations-tile-grid-maps/

[35] S. Nusrat and S. Kobourov, "Visualizing Cartograms: Goals and Task Taxonomy," *CoRR*, Feb. 2015.

[36] S. Nusrat, "The state of the art in cartograms," 2016.

[37] J. S. Ueland and P. J. Vincent, "Mapping Florida: Beyond The Choropleth Map," in *Vol. 35 (2004): The Florida Geographer*, 2004.

[38] S. Langton and R. Solymosi, "Cartograms, hexograms and regular grids: a commentary on minimising misrepresentation in spatial data visualisations," SocArXiv, preprint, Aug. 2019. [Online]. Available: https://osf.io/t6agd

[39] K. Wongsuphasawat, "Quality Metrics for Tile Grid Maps," 00002.

[40] T. L. C. Griffin, "Cartographic transformation of the thematic map base," *Cartography*, vol. 11, no. 3, pp. 163–174, Mar. 1980. [Online]. Available: https://doi.org/10.1080/00690805.1980.10438102

[41] A. Fotheringham, C. Brunsdon, and M. Charlton, *Quantitative Geography: Perspectives on Spatial Data Analysis*, 1st ed. London ; Thousand Oaks, Calif: SAGE Publications Ltd, 2000.

[42] AfterTheFlood, "London Squared," 00238. [Online]. Available: http://tools.aftertheflood.com/londonsquared/

[43] M. Engebretsen and H. Kennedy, Eds., *Data Visualization in Society*. Amsterdam University Press, Apr. 2020, 00003. [Online]. Available: https://www.aup.nl/en/book/9789463722902/data-visualization-in-society

[44] C. Board and R. M. Taylor, "Perception and Maps: Human Factors in Map Design and Interpretation," *Transactions of the Institute of British Geographers*, vol. 2, no. 1, pp. 19–36, 1977. [Online]. Available: https://www.jstor.org/stable/622191

[45] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, *Visualization Aspects. In: Visualization of Time-Oriented Data. Human-Computer Interaction Series*. London: Springer, 2011.

[46] A. Ciołkosz-Styk, "The visual search method in map perception research," *Geoinformation Issues*, vol. 4, no. 1, p. 33—42, 2012.

[47] E. Swoboda and P. Vighi, *Early Geometrical Thinking in the Environment of Patterns, Mosaics and Isometries*. Springer, Sep. 2016.

[48] E. Moon, S. M. Schneider, P. Ferronato, M. Salah, S. Ruecker, and D. McDonagh, "The Role of Empathic Design Research in the Prototyping of Interactive Visualizations," *The Design Journal*, vol. 22, no. sup1, pp. 1515–1527, Apr. 2019. [Online]. Available: https://shorturl.at/estJO

[49] T. A. SLOCUM, S. H. ROBESON, and S. L. EGBERT, "Traditional Versus Sequenced Choropleth Maps/An Experimental Investigation," *Cartographica: The International Journal for Geographic Information and Geovisualization*, Oct. 2006. [Online]. Available: http://www.utpjournals.press/doi/abs/10.3138/CG7N-0158-1537-6177

[50] V. P. Araya, E. Pietriga, and A. Bezerianos, "A Comparison of Visualizations for Identifying Correlation over Space and Time," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, pp. 375–385, 2019.

[51] K. Goldsberry and S. Battersby, "Issues of Change Detection in Animated Choropleth Maps," *Cartographica: The International Journal for Geographic Information and Geovisualization*, Oct. 2009. [Online]. Available: http://www.utpjournals.press/doi/abs/10.3138/carto.44.3.201

[52] C. Fish, K. P. Goldsberry, and S. Battersby, "Change Blindness in Animated Choropleth Maps: An Empirical Study," *Cartography and Geographic Information Science*, vol. 38, no. 4, 2011, 00043. [Online]. Available: https://doi.org/10.1559/15230406384350

[53] M. Harrower, "The Cognitive Limits of Animated Maps," *Cartographica: The International Journal for Geographic Information and Geovisualization*, Dec. 2007. [Online]. Available: http://www.utpjournals.press/doi/abs/10.3138/carto.42.4.349

[54] A. Skowronnek, "Beyond choropleth maps : A review of techniques to visualize quantitative areal geodata," in *INFOVIS READING GROUP WS 2015/16*, Chicago, Illinois, USA, 2016.

[55] M. L. Johnson, "Mapping the news: Case studies in gis and journalism," *Cartographic Perspectives*, no. 52, p. 84–86, Sep. 2005. [Online]. Available: shorturl.at/kmptA

[56] N. R. Kaye, A. Hartley, and D. Hemming, "Mapping the climate: guidance on appropriate techniques to map climate variables and their uncertainty," *Geoscientific Model Development Discussions*, vol. 4, no. 3, pp. 1875–1906, Aug. 2011, 00047. [Online]. Available: https://www.geosci-model-dev-discuss.net/4/1875/2011/

[57] B. Hall, "Reading Between the Contour Lines of the Wind Speed Maps: Giving Tornadoes Their Due | IIBEC," Jan. 2019. [Online]. Available: https://iibec.org/giving-tornadoes-their-due/

[58] A. Cairo, "Uncertainty and graphicacy How should statisticians, journalists, and designers reveal uncertainty in graphics for public consumption?" 00002.

[59] R. Bernhaupt, P. Palanque, F. Manciet, and C. Martinie, "User-test results injection into task-based design process for the assessment and improvement of both usability and user experience," in *Human-Centered and Error-Resilient Systems Development*, C. Bogdan, J. Gulliksen, S. Sauer, P. Forbrig, M. Winckler, C. Johnson, P. Palanque, R. Bernhaupt, and F. Kis, Eds. Cham: Springer International Publishing, 2016, pp. 56–72.

[60] J. Bertin, *Semiology of Graphics*. University of Wisconsin Press, 1983.

[61] N. Andrienko, *Exploratory Analysis of Spatial and Temporal Data*. Springer-Verlag, 2006.

[62] P. Carrard, "Mapped stories: Cartography, history, and the representation of time in space," *Frontiers of Narrative Studies*, vol. 4, no. 2, pp. 263–276, Dec. 2018, 00001 Publisher: De Gruyter. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/fns-2018-0022/html

[63] Y. Zhang and R. Maciejewski, "Quantifying the Visual Impact of Classification Boundaries in Choropleth Maps," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, pp. 1–1, Jan. 2016.

[64] J. Trischler, S. J. Pervan, S. J. Kelly, and D. R. Scott, "The Value of Codesign: The Effect of Customer Involvement in Service Design Teams," *Journal of Service Research*, vol. 21, no. 1, pp. 75–100, Feb. 2018, 00106 Publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/1094670517714060

[65] R. E. Roth, A. Çöltekin, L. Delazari, H. F. Filho, A. Griffin, A. Hall, J. Korpi, I. Lokka, A. Mendonça, K. Ooms, and C. P. J. M. v. Elzakker, "User studies in cartography: opportunities for empirical research on interactive maps and visualizations," *International Journal of Cartography*, vol. 3, no. sup1, pp. 61–89, Oct. 2017, 00038 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/23729333.2017.1288534. [Online]. Available: https://doi.org/10.1080/23729333.2017.1288534

[66] M. Richards, "Data visualization and usability testing," in *Human-Computer Interaction (SIGHCI) 2019*, Cancún,Mexico, 2019, iSSN: 978-0-9966831-8-0. [Online]. Available: https://www.aisel.aisnet.org/amcis2019

[67] W. C. Tomlin, "Putting It All Together: Usability Testing Data Analysis and Recommendations," in *UX Optimization: Combining Behavioral UX and Usability Testing Data to Optimize Websites*, W. C. Tomlin, Ed. Berkeley, CA: Apress, 2018, pp. 147–175.

[68] A. Alharbi and P. Mayhew, "Users' performance in lab and non-lab enviornments through online usability testing: A case of evaluating the usability of digital academic libraries' websites | IEEE Conference Publication | IEEE Xplore," 2015, 00008. [Online]. Available: https://ieeexplore-ieee-org.sire.ub.edu/document/7237139

[69] A. Schade, "Pilot Testing: Getting It Right (Before) the First Time," 2015. [Online]. Available: https://www.nngroup.com/articles/pilot-testing/

[70] N. Babich, "Usability Testing: Moderated vs Unmoderated," Mar. 2020. [Online]. Available: https://medium.com/thinking-design/usability-testing-moderated-vs-unmoderated-adbccc37404b

[71] A. Sonderegger, A. Uebelbacher, and J. Sauer, "The ux construct – does the usage context influence the outcome of user experience evaluations?" in *Human-Computer Interaction – INTERACT 2019*, D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, and P. Zaphiris, Eds. Cham: Springer International Publishing, 2019, pp. 140–157.

[72] M. Hassenzahl and N. Sandweg, "From mental effort to perceived usability: transforming experiences into summary assessments," in *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '04. New York, NY, USA: Association for Computing Machinery, Apr. 2004, pp. 1283–1286. [Online]. Available: https://doi.org/10.1145/985921.986044

[73] J. K. McDonald, P. J. Rich, and N. B. Gubler, "The Perceived Value of Informal, Peer Critique in the Instructional Design Studio," *TechTrends*, vol. 63, no. 2, pp. 149–159, Mar. 2019, 00004. [Online]. Available: https://doi.org/10.1007/s11528-018-0302-9

[74] J. K. McDonald and E. Michela, "'This is my vision': how students depict critiques along with themselves during critiques," *J. of Design Research*, vol. 18, no. 1/2, p. 57, 2020. [Online]. Available: http://www.inderscience.com/link.php?id=112057

[75] A. Kittur, H. Chi, and B. Suh, "Crowdsourcing for Usability : Using Micro-Task Markets for Rapid , Remote , and Low-Cost User Measurements," 2007, 0000.

[76] G. McNeill and S. Hale, "Generating Tile Maps," *Computer Graphics Forum*, vol. 36, pp. 435–445, Jun. 2017, 00025.

[77] "ColorBrewer: Color Advice for Maps," 00041. [Online]. Available: https://colorbrewer2.org

[78] G. S. Erickson, *New Methods of Market Research and Analysis*. Edward Elgar Publishing, Oct. 2017, 00012 Google-Books-ID: WR06DwAAQBAJ.

[79] C. Juergens, "Trustworthy COVID-19 Mapping: Geo-spatial Data Literacy Aspects of Choropleth Maps," *KN - Journal of Cartography and Geographic Information*, vol. 70, no. 4, pp. 155–161, Dec. 2020. [Online]. Available: https://doi.org/10.1007/s42489-020-00057-w

[80] R. Li, "Visualizing COVID-19 information for public: Designs, effectiveness, and preference of thematic maps," *Human Behavior and Emerging Technologies*, vol. 3, no. 1, pp. 97–106, 2021.

[81] N. Bhutta, J. Bricker, A. C. Chang, L. J. Dettling, S. Goodman, K. B. Moore, S. Reber, A. H. Volz, and R. A. Windle, "Changes in U.S. Family Finances from 2016 to 2019: Evidence from the Survey of Consumer FinancesChanges in U.S. Family Finances from 2016 to 2019: Evidence from the Survey of Consumer Finances," Federal Reserve Bulletin, Feb 2020.

[82] U. C. Bureau, "2020 census," U.S. Department of Commerce, Feb. 2020.

[83] B. Gordon, G. Loudon, S. Gill, and J. Baldwin, "Product user testing: The void between laboratory testing and field testing," in *nternational Association of Societies of Design Research Conference 2019*. Manchester, UK: Manchester School of Art, Sep. 2019, accepted: 2019-12-02T16:28:03Z. [Online]. Available: https://repository.cardiffmet.ac.uk/handle/10369/10865

[84] R. E. Cordes, "Task-Selection Bias: A Case for User-Defined Tasks," *International Journal of Human–Computer Interaction*, vol. 13, no. 4, pp. 411–419, Dec. 2001, 00049.

[85] L. Strickland, S. Loft, and A. Heathcote, "Investigating the effects of ongoing-task bias on prospective memory," *Quarterly Journal of Experimental Psychology*, vol. 73, no. 9, pp. 1495–1513, 2020, pMID: 32160817. [Online]. Available: https://doi.org/10.1177/1747021820914915

[86] Y. Li, N. Hollender, and T. Held, "Task sequence effects in usability tests," in *Interfaces and Human Computer Interaction*, Prague, Czech Republic, 07 2013.

[87] K. E. Thomas and C. J. König, "Knowledge of Previous Tasks: Task Similarity Influences Bias in Task Duration Predictions," *Frontiers in Psychology*, vol. 9, p. 760, 2018, 00006. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2018.00760

[88] S. Stigberg, J. Karlsen, H. Holone, and C. Linnes, *Nordic Contributions in IS Research: 8th Scandinavian Conference on Information Systems, SCIS 2017, Halden, Norway, August 6-8, 2017, Proceedings*. Springer, Jul. 2017, google-Books-ID: CpouDwAAQBAJ.

[89] J. Sauro and E. Kindlund, "How long should a task take? identifying specification limits for task times in usability tests," in *Human Computer Interaction International Conference 2005*, Las Vegas, USA, 01 2005.

[90] J. Sauro and J. Lewis, "Quantifying the User Experience - 2nd Edition," p. 350pp, 2016.

[91] J. Rubin, D. Chisnell, and J. Spool, *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wiley, 2011. [Online]. Available: https://books.google.es/books?id=l_e1MmVzMb0C

[92] S. Fabrikant and A. Lobben, "Introduction: Cognitive issues in geographic information visualization," *Cartographica*, vol. 44, pp. 139–143, 09 2009.

[93] Z. Gyozo Torok and A. Torok, "Cognitive Data Visualization—A New Field with a Long History," in *Cognitive Infocommunications, Theory and Applications*, Aug 2019, pp. 49–77.

[94] R. A. Rensink, "Visualization as a stimulus domain for vision science," *Journal of Vision*, vol. 21, no. 8, pp. 3–3, Aug. 2021, publisher: The Association for Research in Vision and Ophthalmology. [Online]. Available: https://jov.arvojournals.org/article.aspx?articleid=2776535

[95] A. Buckley, "Guidelines for the effective design of spatio-temporal maps," in *International Cartographic Conference 2013*, Dresden, Germany, 08 2013.

[96] M. Lupa, S. Szombara, M. Chuchro, and T. Chrobak, "Limits of Colour Perception in the Context of Minimum Dimensions in Digital Cartography," *ISPRS International Journal of Geo-Information*, vol. 6, no. 9, p. 276, Sep. 2017, 00006 Number: 9 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2220-9964/6/9/276
[97] "Gazepoint: Eye tracking system technology for everyone - ux testing — gazepoint," 00000. [Online]. Available: https://www.gazept.com
[98] S.-H. Kim, Z. Dong, H. Xian, B. Upatising, and J. S. Yi, "Does an eye tracker tell the truth about visualizations?: Findings while investigating visualizations for decision making," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, pp. 2421–2430, 12 2012.
[99] S. A. Hillyard and T. F. Münte, "Selective attention to color and location: An analysis with event-related brain potentials," *Perception & Psychophysics*, vol. 36, no. 2, pp. 185–198, Mar. 1984, 00704. [Online]. Available: https://doi.org/10.3758/BF03202679
[100] C. A. Brewer, "Color Use Guidelines for Mapping and Visualization," *Modern Cartography Series*, pp. 123–147, Jan. 1994. [Online]. Available: https://pennstate.pure.elsevier.com/en/publications/color-use-guidelines-for-mapping-and-visualization
[101] C. Healey, "Choosing effective colours for data visualization," in *Proceedings of Seventh Annual IEEE Visualization '96*, Oct. 1996, pp. 263–270, iSSN: null.
[102] J. Liu, T. Tang, W. Wang, B. Xu, X. Kong, and F. Xia, "A Survey of Scholarly Data Visualization," *IEEE Access*, vol. 6, pp. 19 205–19 221, 2018, 00035 Conference Name: IEEE Access.
[103] M.-J. Kraak, B. Köbben, and Y. Tong, "Integrated Time and Distance Line Cartogram: a Schematic Approach to Understand the Narrative of Movements," *Cartographic Perspectives*, no. 77, pp. 7–16, Oct. 2014. [Online]. Available: http://shorturl.at/alFP2
[104] V. Gesù and V. Starovoitov, "Distance-based functions for image comparison," *Pattern Recognition Letters*, vol. 20, pp. 207–214, Feb. 1999.
[105] P. Koenig, F. Zaidi, and D. Archambault, "Interactive searching and visualization of patterns in attributed graphs," in *Proceedings of the Graphics Interface 2010 Conference, May 31 - June 02, 2010, Ottawa, Ontario, Canada*, D. Mould and S. Noël, Eds. ACM / Canadian Human-Computer Communications Society, 2010, pp. 113–120. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/1839214.1839235
[106] S. Holtz, G. Valle, J. Howard, and P. Morreale, "Visualization and pattern identification in large scale time series data," in *2011 IEEE Symposium on Large Data Analysis and Visualization*, 2011, pp. 129–130.
[107] L. Collins, "How to analyse usability testing results." [Online]. Available: https://info.webusability.co.uk/blog/how-to-analyse-usability-testing-results
[108] W. L. i. R.-B. U. Experience, "Prioritizing Web Usability: Book by Jakob Nielsen and Hoa Loranger." [Online]. Available: https://www.nngroup.com/books/prioritizing-web-usability/
[109] J. Hansen and M. Wänke, "Liking What's Familiar: The Importance of Unconscious Familiarity in the Mere-Exposure Effect," *Social Cognition*, vol. 27, no. 2, pp. 161–182, Apr. 2009, 00107 Publisher: Guilford Publications Inc. [Online]. Available: https://guilfordjournals.com/doi/10.1521/soco.2009.27.2.161
[110] A. Zissman and E. Neimark, "The Influence of Familiarity on Evaluations of Liking and Goodness of Several Types of Music," *The Psychological Record*, vol. 40, no. 4, pp. 481–490, Oct. 1990, 00019. [Online]. Available: https://doi.org/10.1007/BF03399535

Luz Calvo is a user experience researcher at the Data Visualization and Analysis Group at the Barcelona Supercomputing Center (BSC). She is a technical engineer in Telecommunications and has a degree in UX, Multimedia management and Data visualization. She also teaches Information architecture at the University of La Rioja (UNIR) and UX strategy at the Centro Universitario de Artes y Diseño de Barcelona (BAU). She has more than seven years of professional experience in several private technology consultancies as UX manager and Interaction designer. She works as a user experience researcher and data visualization designer. She is currently in her fourth year of her PhD in Data Visualization at the University of Barcelona (UB).

Fernando Cucchietti leads the Data Visualization and Analysis Group at the Barcelona Supercomputing Center (BSC). He holds a PhD in quantum computing, and his work focuses on scientific data visualization (data heavy graphical interfaces, images, and videos), and data science applied to industrial problems (machine learning and artificial intelligence). He also works with the team on science dissemination, creating video documentaries and interactive data visualizations that explain the science done at BSC. Fernando is also the Chief Data Officer at Mitiga Solutions, working to extract value and insights from data for the air transport industries.

Mario Pérez-Montoro is a (full) Professor at the University of Barcelona (Spain) and Director of the PhD Program in Information and Communication at this University. He holds a PhD in Philosophy and Education from the University of Barcelona (Spain) and a Master in Information Management and Systems from the Polytechnic University of Catalonia (Spain). He studied at the Istituto di Discipline della Comunicazione at the Università di Bologna (Italy) and has been a Visiting Scholar at the Center for the Study of Language and Information (CSLI) at Stanford University (California, USA) and at the School of Information at UC Berkeley (California, USA). His work has focused on interaction design and information visualization.