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].

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. Second, 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].

- *Luz Calvo and Fernando Cucchietti are with the CASE Department of the Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain. 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, 08007 Barcelona, Spain. E-mail: perez-montoro@ub.edu.*

Manuscript received 17 October 2021; revised 17 June 2022; accepted 19 June 2022. Date of publication 12 July 2022; date of current version 1 September 2023.
This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Bioethics Committee of the Universitat de Barcelona Institutional Review Board under Application No. IRB00003099 and performed in line with the bioethics directives of Universitat de Barcelona.
(Corresponding author: Luz Calvo.)
Recommended for acceptance by R. Metoyer.
This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2022.3188940, provided by the authors.
Digital Object Identifier no. 10.1109/TVCG.2022.3188940

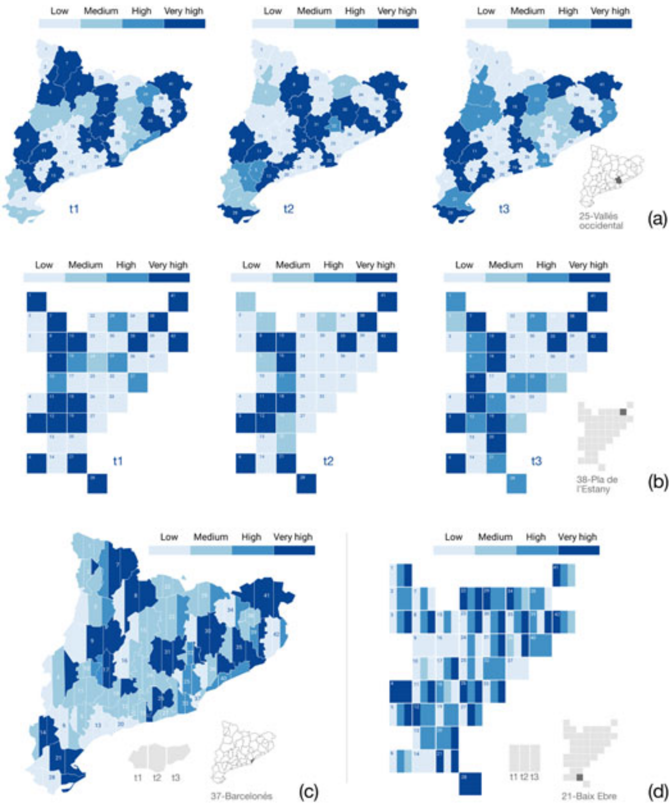

Fig. 1. Visualisations evaluated in the study: First, small multiple maps using Choropleth (a) and Tile Grid map (b). Second, their banded versions, unifying in the same map three different time steps: Banded Choropleth [23] (c) and Banded Tile Grid map (d).

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.

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

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 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 Supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TVCG.2022.3188940) [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

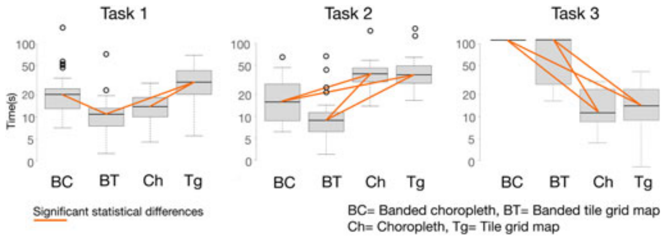

Fig. 2. Boxplots showing the distribution for the results related to completion time for all representations in the three tasks presented. Orange lines show the significant statistical differences between maps.

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, available online).

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, available online 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, available online.

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, available online).

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, available online).

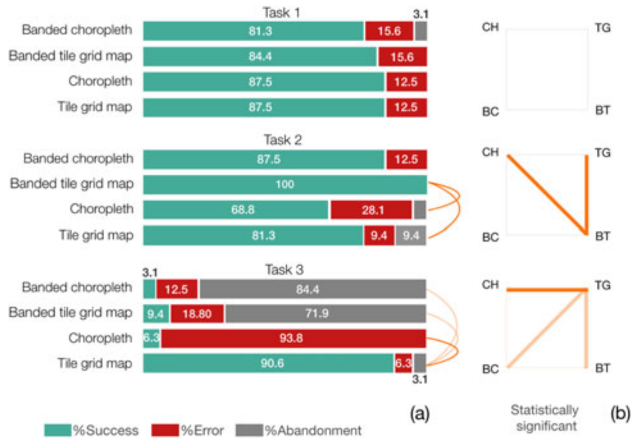

Fig. 3. (a) Success, failure and abandonment rates by task and type of map. (b) Relationships of significant statistical differences between representations. Opacity indicates a higher or lower level of significance.

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, available online). 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, available online).

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 additional bias [84], [85], [86]. 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 [87]. 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 [88], [89], [90]. 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 [88]. 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, available online).

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, available online.

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, available online). 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, available online). 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, available online). The same happened with the success rate (see Fig. 3 in the main text and Table 3 of the Suppl. material, available online). 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, available online). 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, available online). 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

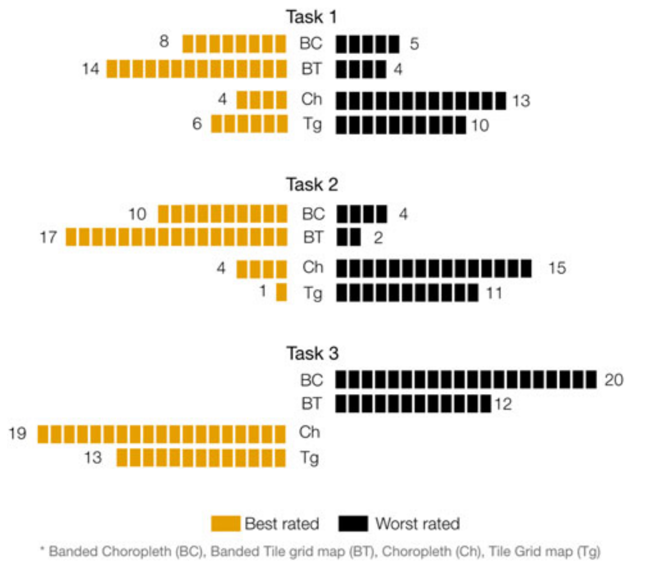

Fig. 4. Most and least favourable maps (number of positive and negative votes) to solve the task and which favour decision making.

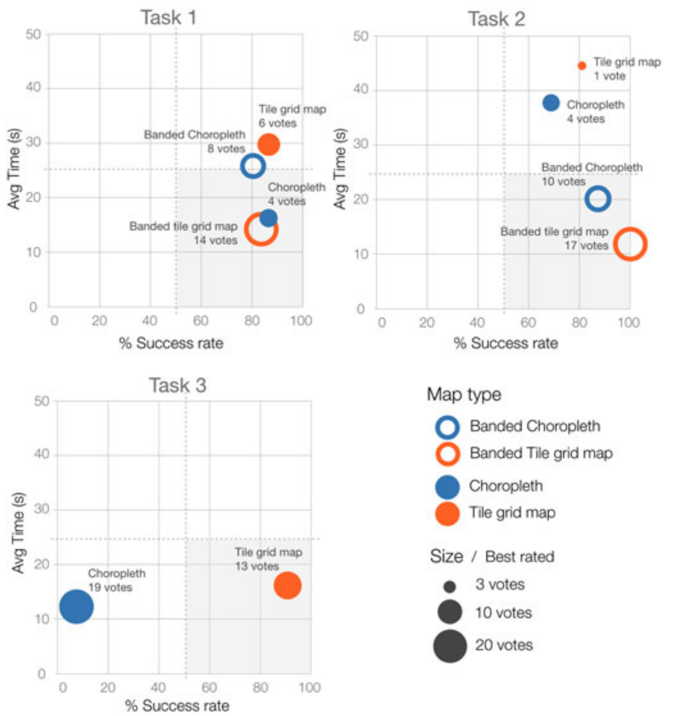

Fig. 5. In this graph we have combined four of the most significant variables, using a scatter plot representation. For the three tasks, we compare: success rate (x axis), time spent to accomplish the task (y axis), number of times a graph was voted as the most effective (size), and the type of map (color 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.

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, available online to review the most frequent comments collected during the study).

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 [91], [92] (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 [93]. 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) [94], [95].

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) [96]. 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 [97]. (see Fig. 4 of the Suppl. material, available online).

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" [98]. 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"[99], [100], [101]. 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" [101], [102].

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"[103], [104].

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 [105], [106]. 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], [107], [108].

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 and D. J. Denis, "Milestones in the history of thematic cartography, statistical graphics, and data visualization," 2001. Accessed: Aug. 10, 2022. [Online]. Available: http://www.datavis.ca/milestones/

[2] C. Nicolet, *Space, Geography, and Politics in the Early Roman Empire*. Ann Arbor, MI, USA: Univ. Michigan Press, 1991.

[3] M. Kernik and S. Nelson, "9. Social Maps," in *Mapping, Society, and Technology*. Minneapolis, MN, USA: Univ. 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, U.K.: 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," *Comput., Environ. Urban Syst.*, vol. 31, no. 1, pp. 52–75, Jan. 2007. [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, Jan. 2019, Art. no. 332. [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. [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," *J. Accounting 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*, F. Rebelo and M. Soares, Eds., Berlin, Germany: Springer, 2016, pp. 53–61.

[11] S. Bogacz and J. G. Trafton, "Understanding static and dynamic visualizations," in *Diagrammatic Representation and Inference*, M. Hegarty, B. Meyer, and N. H. Narayanan, Eds., Berlin, Germnay: Springer, 2002, pp. 347–349.

[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," 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. [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*. Evanston, IL, USA: Routledge, Jun. 2011.
[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 Int. J. Geo-Informat.*, vol. 10, no. 2, Feb. 2021, Art. no. 69. [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, NJ, USA: Prentice-Hall, 1972.
[19] B. D. Dent, "Communication aspects of value-by-area cartograms," *Amer. Cartographer*, vol. 2, no. 2, pp. 154–168, Jan. 1975.
[20] D. Dorling, "Area cartograms: Their use and creation," in *The Map Reader*, M. Dodge, R. Kitchin, and C. Perkins, Eds. Hoboken, NJ, USA: Wiley, Apr. 2011, pp. 252–260. [Online]. Available: https://onlinelibrary.wiley.com
[21] W. Tobler, "Thirty-five years of computer cartograms," *Ann. Assoc. Amer. 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://virtual.ieeevis.org/year/2020/paper_w-comm-1022.html
[23] Y. Du, L. Ren, Y. Zhou, J. Li, F. Tian, and G. Dai, "Banded choropleth map," *Pers. Ubiquitous Comput.*, vol. 22, no. 3, pp. 503–510, Jun. 2018.
[24] A. Jacobs, J. Pinto, and M. Shiffrar, "Experience, context, and the visual perception of human movement," *J. Exp. Psychol.: Hum. Percep. Perform.*, vol. 30, no. 5, pp. 822–835, 2004.
[25] M. H. Martens and M. R. J. Fox, "Do familiarity and expectations change perception? Drivers' glances and response to changes," *Transp. Res. Part F Traffic Psychol. Behav.*, 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," *Digit. Journalism*, vol. 8, no. 7, pp. 972–974, 2020.
[27] S. Esteves and M. Neves, "Are charts going digital? The case of data visualization on portuguese media," in *Hum. Interaction, Emerg. Technol. and Future Appl. III*, T. Ahram, R. Taiar, K. Langlois, and A. Choplin, Eds, Berlin, Germany: Springer, 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 *Proc. ACM/IEEE Joint Conf. Digit. Libraries*, 2020, pp. 389–392.
[29] M. P. Armstrong and N. Xiao, "Retrospective deconstruction of statistical maps: A choropleth case study," *Ann. Amer. Assoc. Geographers*, vol. 108, no. 1, pp. 179–203, Jan. 2018.
[30] J. Schiewe, "Empirical studies on the visual perception of spatial patterns in choropleth maps," *KN - J. Cartography Geographic Informat.*, vol. 69, no. 3, pp. 217–228, Sep. 2019.
[31] L. Besançon, M. Cooper, A. Ynnerman, and F. Vernier, "An evaluation of visualization methods for population statistics based on choropleth maps," 2020, *arXiv:2005.00324*.
[32] D. Keim, S. North, C. Panse, and J. Schneidewind, "Efficient cartogram generation: A comparison," in *Proc. IEEE Symp. Informat. Vis.*, 2002, pp. 33–36.
[33] S. Nusrat, M. J. Alam, and S. Kobourov, "Evaluating cartogram effectiveness," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 2, pp. 1077–1090, Feb. 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. G. Kobourov, "Task taxonomy for cartograms," in *Proc. 17th Eurograph. Conf. Visu.*, Feb. 2015, pp. 61–65.
[36] S. Nusrat, "The state of the art in cartograms," *Comput. Graph. Forum*, vol. 35, no. 3, pp. 619–642, 2016.
[37] J. S. Ueland and P. J. Vincent, "Mapping florida: Beyond the choropleth map," *The Florida Geographer*, vol. 35, p. 11, 2004.
[38] S. Langton and R. Solymosi, "Cartograms, hexograms and regular grids: A commentary on minimising misrepresentation in spatial data visualisations," Aug. 2019. [Online]. Available: https://osf.io/t6agd
[39] K. Wongsuphasawat, "Quality metrics for tile grid maps," in *Proc. IEEE Conf. Inf. Vis.*, 2016, Art. no. 22.
[40] T. L. C. Griffin, "Cartographic transformation of the thematic map base," *Cartography*, vol. 11, no. 3, pp. 163–174, Mar. 1980.
[41] A. Fotheringham, C. Brunsdon, and M. Charlton, *InglésQuantitative Geography: Perspectives on Spatial Data Analysis*, 1st ed. Thousand Oaks, CA, USA: SAGE, 2000.
[42] AfterTheFlood, London squared, 2020. [Online]. Available: http://tools.aftertheflood.com/londonsquared/
[43] M. Engebretsen and H. Kennedy, Eds, *Data Visualization in Society*, Amsterdam, The Netherlands: Amsterdam Univ. Press, Apr. 2020. [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," *Trans. Inst. Brit. 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*, Berlin, Germany: Springer, 2011.
[46] A. Ciołkosz-Styk, "The visual search method in map perception research," *Geoinformation Issues*, vol. 4, no. 1, pp. 33–42, 2012.
[47] E. Swoboda and P. Vighi, *Early Geometrical Thinking in the Environment of Patterns, Mosaics and Isometries*, Berlin, Germany: 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," *Des. J.*, 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 Int. J. Geographic Informat. Geovisualization*, vol. 27, pp. 67–88, Oct. 2006. [Online]. Available: http://www.utpjournals.press
[50] V. P. Araya, E. Pietriga, and A. Bezerianos, "A comparison of visualizations for identifying correlation over space and time," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 1, pp. 375–385, Jan. 2020.
[51] K. Goldsberry and S. Battersby, "Issues of change detection in animated choropleth maps," *Cartographica Int. J. Geographic Informat. Geovisualization*, vol. 44, pp. 201–215, Oct. 2009. [Online]. Available: http://www.utpjournals.press
[52] C. Fish, K. P. Goldsberry, and S. Battersby, "Change blindness in animated choropleth maps: An empirical study," *Cartography Geographic Informat. Sci.*, vol. 38, no. 4, pp. 350–362, 2011.
[53] M. Harrower, "The cognitive limits of animated maps," *Cartographica Int. J. Geographic Informat. Geovisualization*, vol. 42, pp. 349–357, Dec. 2007. [Online]. Available: http://www.utpjournals.press
[54] A. Skowronnek, "Beyond choropleth maps : A review of techniques to visualize quantitative areal geodata," 2016. [Online]. Available: https://alsino.io/static/papers/BeyondChoropleths_AlsinoSkowronnek.pdf
[55] M. L. Johnson, "Mapping the news: Case studies in GIS and journalism," *Cartographic Perspectives*, no. 52, pp. 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 Develop. Discuss.*, vol. 4, no. 3, pp. 1875–1906, Aug. 2011. [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," Oct. 2017. [Online]. Available: https://ec.europa.eu/eurostat/cros/powerfromstatistics/OR/PfS-OutlookReport-Cairo.pdf
[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*, Berlin, Germany: Springer, 2016, pp. 56–72.
[60] J. Bertin, *Semiology of Graphics*. Madison, WI, USA: Univ. Wisconsin Press, 1983.
[61] N. Andrienko, *Exploratory Analysis of Spatial and Temporal Data*, Berlin, Germany: Springer, 2006.
[62] P. Carrard, "Mapped stories: Cartography, history, and the representation of time in space," *Front. Narrative Stud.*, vol. 4, no. 2, pp. 263–276, Dec. 2018. [Online]. Available: https://www.degruyter.com/document

[63] Y. Zhang and R. Maciejewski, "Quantifying the visual impact of classification boundaries in choropleth maps," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 371–380, Jan. 2017.
[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," *J. Serv. Res.*, vol. 21, no. 1, pp. 75–100, Feb. 2018.
[65] R. E. Roth et al., "User studies in cartography: Opportunities for empirical research on interactive maps and visualizations," *Int. J. Cartography*, vol. 3, no. sup1, pp. 61–89, Oct. 2017.
[66] M. Richards, "Data visualization and usability testing," 2019. [Online]. Available: https://shorturl.at/ktBVZ
[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., New York, NY, USA: 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," in *Proc. Sci. Inf. Conf.*, 2015, pp. 151–161. [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 versus unmoderated," Mar. 2020. [Online]. Available: https://medium.com/thinking-design/usability-testing-moderated-versus-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 *Proc. IFIP Conf. Hum.-Comput. Interaction*, 2019, pp. 140–157.
[72] M. Hassenzahl and N. Sandweg, "From mental effort to perceived usability: Transforming experiences into summary assessments," in *Proc. Extended Abstr. Hum. Factors Comput. Syst.*, 2004, pp. 1283–1286.
[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.
[74] J. K. McDonald and E. Michela, "'This is my vision': How students depict critiques along with themselves during critiques," *J. Des. Res.*, vol. 18, no. 1/2, 2020, Art. no. 57. [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," in *Proc. ACM Conf. Hum. Factors Comput. Syst.*, 2007.
[76] G. McNeill and S. Hale, "Generating tile maps," *Comput. Graph. Forum*, vol. 36, pp. 435–445, Jun. 2017, 00025.
[77] ColorBrewer: Color advice for maps, 2022. [Online]. Available: https://colorbrewer2.org
[78] G. S. Erickson, *New Methods of Market Research and Analysis*, Cheltenham, U.K.: Edward Elgar Publishing, Oct. 2017.
[79] C. Juergens, "Trustworthy COVID-19 mapping: Geo-spatial data literacy aspects of choropleth maps," *J. Cartography Geographic Inf.*, vol. 70, no. 4, pp. 155–161, Dec. 2020.
[80] R. Li, "Visualizing COVID-19 information for public: Designs, effectiveness, and preference of thematic maps," *Hum. Behav. Emerg. Technol.*, vol. 3, no. 1, pp. 97–106, 2021.
[81] N. Bhutta et al., "Changes in U.S Family finances from 2016 to 2019: Evidence from the survey of consumer finances," *Federal Reserve Bull.*, vol. 106, no. 5, p. 42, Oct. 2020.
[82] J. Semega, M. Kollar, J. Creamer, and A. Mohanty, "Income ad poverty in the United States: 2018," *United States Census Bureau*, vol. 60, no. 266, pp. 42, Sep. 2019.
[83] B. Gordon, G. Loudon, S. Gill, and J. Baldwin, "Product user testing: The void between laboratory testing and field testing," Sep. 2019. [Online]. Available: https://repository.cardiffmet.ac.uk/handle/10369/10865
[84] R. E. Cordes, "Task-selection bias: A case for user-defined tasks," *Int. J. Hum.–Comput. Interaction*, vol. 13, no. 4, pp. 411–419, Dec. 2001.
[85] L. Strickland, S. Loft, and A. Heathcote, "Investigating the effects of ongoing-task bias on prospective memory," 2020. [Online]. Available: http://shorturl.at/qFGSY
[86] K. E. Thomas and C. J. König, "Knowledge of previous tasks: Task similarity influences bias in task duration predictions," *Front. Psychol.*, vol. 9, 2018, Art. no. 760. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2018.00760
[87] S. Stigberg, J. Karlsen, H. Holone, and C. Linnes, *Nordic Contributions in IS Research: 8th Scandinavian Conference on Information Systems*, Berlin, Germany: Springer, Jul. 2017.
[88] J. Sauro and E. Kindlund, "How long should a task take? Identifying specification limits for task times in usability tests," in *Proc. Human Comput. Interact. Int. Conf.*, 2005.
[89] J. Sauro and J. Lewis, *Quantifying the User Experience* - 2nd Ed. San Mateo, CA, USA: Morgan Kaufmann, 2016, Art. no. 350.
[90] J. Rubin, D. Chisnell, and J. Spool, *en-gbHandbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*, 2nd Ed., Hoboken, NJ, USA: Wiley, 2008. [Online]. Available: https://www.wiley.com/en-gb/Handbook+of+Usability+Testing
[91] S. Fabrikant and A. Lobben, "Introduction: Cognitive issues in geographic information visualization," *Cartographica*, vol. 44, pp. 139–143, 2009.
[92] Z. Gyozo Torok and A. Torok, "Cognitive data visualization—a new field with a long history," *Cognitive Infocommunications, Theory and Applications*, Berlin, Germany: Springer, 2019, pp. 49–77.
[93] R. A. Rensink, "Visualization as a stimulus domain for vision science," *J. Vis.*, vol. 21, no. 8, pp. 3–3, Aug. 2021. [Online]. Available: https://jov.arvojournals.org/article.aspx?articleid=2776535
[94] A. Buckley, "Guidelines for the effective design of spatio-temporal maps," in *Proc. 26th Int. Cartographic Conf.*, 2013, p. 21.
[95] M. Lupa, S. Szombara, M. Chuchro, and T. Chrobak, "Limits of colour perception in the context of minimum dimensions in digital cartography," *ISPRS Int. J. Geo-Informat.*, vol. 6, no. 9, Sep. 2017, Art. no. 276. [Online]. Available: https://www.mdpi.com/2220–9964/6/9/276
[96] Gazepoint: Eye tracking system technology for everyone - UX testing | gazepoint, 2022. [Online]. Available: https://www.gazept.com
[97] 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," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2421–2430, Dec. 2012.
[98] S. A. Hillyard and T. F. Münte, "Selective attention to color and location: An analysis with event-related brain potentials," *Percep. Psychophys.*, vol. 36, no. 2, pp. 185–198, Mar. 1984.
[99] C. A. Brewer, "Color use guidelines for mapping and visualization," in *Proc. Conf. Modern Cartography Ser.*, 1994, pp. 123–147. [Online]. Available: https://pennstate.pure.elsevier.com/en/publications/color-use-guidelines-for-mapping-and-visualization
[100] C. Healey, "Choosing effective colours for data visualization," in *Proc. 7th Annu. IEEE Vis.*, 1996, pp. 263–270.
[101] 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*, vol. 77, pp. 7–16, Oct. 2014. [Online]. Available: http://shorturl.at/alFP2
[102] V. Gesù and V. Starovoitov, "Distance-based functions for image comparison," *Pattern Recognit. Lett.*, vol. 20, pp. 207–214, Feb. 1999.
[103] P.-Y. Koenig, F. Zaidi, and D. Archambault, "Interactive searching and visualization of patterns in attributed graphs," in *Proc. Conf. Graph. Interface*, 2010, pp. 113–120.
[104] S. Holtz, G. Valle, J. Howard, and P. Morreale, "Visualization and pattern identification in large scale time series data," in *Proc. IEEE Symp. Large Data Anal. Vis.*, 2011, pp. 129–130.
[105] L. Collins, How to analyse usability testing results, May 2019. [Online]. Available: https://info.webusability.co.uk/blog/how-to-analyse-usability-testing-results
[106] W. L. i. R.-B. U. Experience, Prioritizing web usability: Book by Jakob Nielsen and Hoa Loranger, 2006. [Online]. Available: https://www.nngroup.com/books/prioritizing-web-usability/
[107] J. Hansen and M. Wänke, "Liking what's familiar: The importance of unconscious familiarity in the mere-exposure effect," *Social Cogn.*, vol. 27, no. 2, pp. 161–182, Apr. 2009. [Online]. Available: https://guilfordjournals.com
[108] A. Zissman and E. Neimark, "The influence of familiarity on evaluations of liking and goodness of several types of music," *Psychol. Rec.*, vol. 40, no. 4, pp. 481–490, Oct. 1990.

Luz Calvo is currently working toward the PhD degree in data visualization with the University of Barcelona (UB). She is a user experience researcher with the Data Visualization and Analysis Group, 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 with the University of La Rioja (UNIR) and UX strategy, 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.

Fernando Cucchietti received the PhD degree in quantum computing. He leads the Data Visualization and Analysis Group with the Barcelona Supercomputing Center (BSC). 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. He is also the chief data officer with Mitiga Solutions, working to extract value and insights from data for the air transport industries.

Mario Pérez-Montoro received the master's degree in information management and systems from the Polytechnic University of Catalonia (Spain) and the PhD in philosophy and education from the University of Barcelona (Spain). He is a (full) professor with the University of Barcelona (Spain) and director of the PhD Program in Information and Communication with the University of Barcelona (Spain). He studied with 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.