



An Exploratory Study of Models of Mobile Map User Experience

Sulaxan Somaskantharajan¹ · Auriol Degbello^{1,2} · Jakub Krukar¹

Received: 18 January 2023 / Accepted: 16 March 2023 / Published online: 29 April 2023
© The Author(s) 2023

Abstract

Several user studies have been conducted to evaluate the User Experience (UX) of thematic mobile maps, but models describing the results beyond point studies are still lacking. This article explored mathematical functions to predict the UX on the visualization types Choropleth Maps and Graduated Symbol Maps. Ten different Choropleth Maps and ten different Graduated Symbol Maps were utilized to conduct a user study, in which 30 participants solved information-gathering tasks on a mobile device. The data from the first 20 participants served as input to build 12 mathematical models on the accuracy, efficiency, perceived mental demand, perceived performance, perceived effort demanded and perceived frustration level for solving the given map tasks. The predictive performance of the models was then evaluated using data from the remaining ten participants and the predictions were within 30% of unseen empirical data. The models obtained are relevant to the design of adaptive and plastic geovisualizations on mobile devices.

Keywords Thematic maps · Mobile ux · Mathematical modelling · Regression models · Colour distance · Size distance · Choropleth map · Graduated symbol map

Eine explorative Studie über Modelle der Nutzererfahrung bei mobilen Karten

Zusammenfassung

Mehrere Nutzerstudien wurden bereits durchgeführt um die User Experience (UX) von thematischen mobilen Karten zu evaluieren. Nichtsdestotrotz fehlen es noch Modelle, die die Ergebnisse über einzelne Studien hinaus beschreiben. In diesem Artikel wurden mathematische Funktionen zur Vorhersage der UX für die Visualisierungstypen „Choropleth Maps“ und „Graduated Symbol Maps“ entwickelt. Zehn Choroplethenkarten und zehn Karten mit abgestuften Symbolen wurden in einer Nutzerstudie untersucht, bei der 30 Teilnehmer Aufgaben zur Informationsbeschaffung auf einem mobilen Gerät lösten. Die Daten der ersten 20 Teilnehmer dienten als Input für die Erstellung von 12 mathematischen Modellen zur Genauigkeit, Effizienz, wahrgenommenen geistigen Anforderung, wahrgenommenen Leistung, wahrgenommenen Anstrengung und wahrgenommenen Frustration bei der Lösung der vorgegebenen Kartenaufgaben. Die Vorhersagekraft der Modelle wurde dann anhand der Daten der verbleibenden 10 Teilnehmer ausgewertet, und die Vorhersagen lagen innerhalb von 30% der ungesehenen empirischen Daten. Die vorgeschlagenen mathematischen Modelle sind relevant für die Gestaltung von adaptiven und plastischen Geovisualisierungen auf mobilen Geräten.

✉ Auriol Degbello
auriol.degbello@tu-dresden.de
Sulaxan Somaskantharajan
s_soma01@uni-muenster.de
Jakub Krukar
krukar@uni-muenster.de

¹ Institute for Geoinformatics, University of Münster, Münster, Germany

² Chair of Geoinformatics, TU Dresden, Dresden, Germany

1 Introduction

Geographical maps serve as a tool to visually present information about the world in an abstract and simple way. These can be ‘topographic maps’ that show the shape of the Earth’s surface to support orientation, or ‘thematic maps’ that focus on a specific topic, also often named ‘single-topic maps’ (Wade et al. 2006). These representations enable map users to gain knowledge and support them in decision-making

processes. Regardless of the type of geovisualization, the perception of the conveyed information should be easy for map users. Hence, geographic visualizations underwent lots of improvements to keep refining map readability and map user experience (Bessadok and Domingues 2011; Harrie 2009). Over time, information technology has developed immensely fast so that ‘personal’ computers became ubiquitous, mostly in the form of mobile phones. For this reason, mobile maps have become essential for decision-making processes and for gathering geographic information, not only for map experts but everyone with different skill levels of map use.

In the field of thematic mapping, different types of maps can be used to visualize the same quantitative data: choropleth map, graduated symbol map, proportional symbol map, dot density map, isoline map, dasymetric map, cartogram and flow map (see Słomska-Przech and Gołębiowska 2021; Gołębiowska et al. 2021). The user experience of these geovisualizations is increasingly studied (Słomska-Przech and Gołębiowska 2021; Brychtová 2015b; Degbelo et al. 2018; Gorte and Degbelo 2022; Brychtová and Coltekin 2015, 2016; Brychtová and Çöltekin 2017b). The goals of these studies are diverse, for example, compare the impact of different visualization types on the user performance during map reading tasks (Degbelo et al. 2018; Gorte and Degbelo 2022; Słomska-Przech and Gołębiowska 2021), or investigate the impact of different design parameters (spatial distance, colour distance, legend position, font size) on the interpretation of maps (Brychtová 2015b; Brychtová and Coltekin 2015, 2016; Brychtová and Çöltekin 2017b).

Although these user studies have provided valuable insights, models systematically describing how the user experience of thematic maps varies as a function of changes in design parameters are still lacking. There are different types of models in HCI (see Oulasvirta 2019 for examples) and ‘model’ denotes here a representation in mathematical terms of the behaviour of a dependent variable in response to a design parameter (i.e. mathematical models). The benefits of these mathematical models are at least threefold. First, and from the theoretical point of view, they enable generalizations beyond the results of empirical ‘point studies’. That is, mathematical models can be seen as the conceptual ‘glue’ (Oulasvirta and Hornbæk 2016) that links different empirical studies examining the impact of a design parameter on a dependent variable. Second, and from the practical viewpoint, they can inform design by allowing numerical derivation of design consequences (see Oulasvirta and Hornbæk 2021). In that sense, and as indicated in (Cockburn and Gutwin 2010; Bailly et al. 2014), they can reduce the dependence on empirical evaluations when testing the performance of a particular prototype. Third, these models are

needed to help computers ‘understand’ UX. That is, they are needed to formulate optimization constraints for computational user interface design (Oulasvirta 2016; Oulasvirta et al. 2020) and intelligent geovisualizations (Degbelo and Kray 2018).

This work systematically investigated the impact of two design parameters on the user experience of map-based products on mobile devices. The focus was on the following research questions:

1. Which mathematical function best describes the relationship between user experience and the colour distance on a choropleth map?
2. Which mathematical function best describes the relationship between user experience and the size distance between symbols on a graduated symbol map?
3. How do the mathematical functions perform on unseen data?

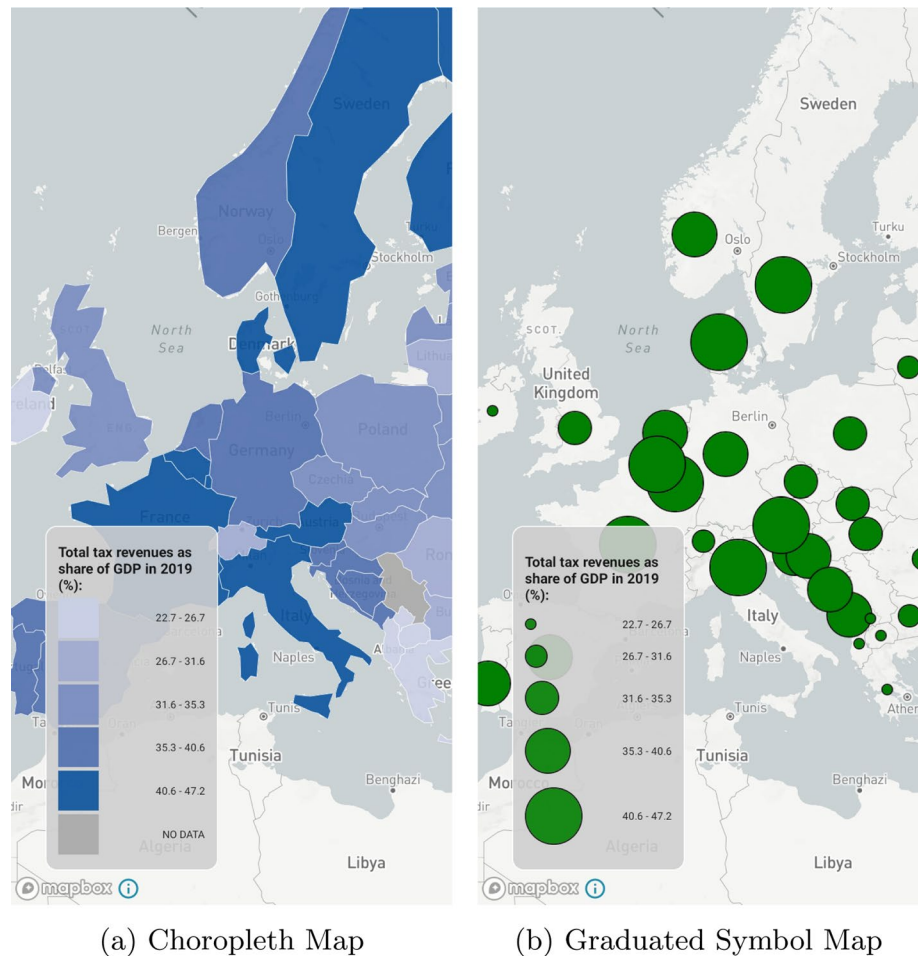
In this exploratory study, we are particularly interested in learning about the form of the mathematical functions. The two types of maps investigated are shown in Fig. 1. We chose choropleth maps (CM) and graduated symbol maps (GSM) as a starting point because they are the most frequently used to visualise open data (see e.g. Degbelo et al. 2020). The contributions of this work are threefold:

- An open-source prototype that generates choropleth maps with different colour distances and graduated symbol maps with different symbol size distances for mobile devices;
- Twelve mathematical functions describing the user experience of choropleth and graduated symbol maps. Their evaluation has shown that they are promising: their predictions remain within 30% of unseen empirical data;
- Hypotheses about the relationships between colour/size distance and pragmatic as well as hedonic aspects of user experience on mobile devices.

2 Related work

The aim of the work is to learn about mathematical functions that can be used to describe the user experience of mobile maps (CM and GSM). Hence, this section briefly reviews previous work on user studies on mobile map UX, user studies on choropleth and graduated symbol maps, and mathematical models of interaction in HCI. In essence, though previous work has contributed several user studies about digital maps on mobile devices, mathematical models of UX are still rare.

Fig. 1 Visualization types investigated in the work: **a** colour distance refers to the average of the distances between the colours used to represent adjacent classes in the map legend; **b** size distance denotes the average of the distances between the circles used to represent adjacent classes in the map legend



2.1 User studies on mobile maps user experience

There is a growing body of work studying the user experience on mobile maps. The goal is to detect issues when interacting with mobile maps and find ways to improve the overall map user experience. For instance, Bartling et al. (2021) created an online survey to test 84 (topographic) map variations to evaluate the relationship between user context and user experience. Participants interacting with polygons on most of the map variations had high task success, comfort, and confidence ratings. Bertel et al. (2017) compared the effect of a visual display condition and a tactile display condition on spatial knowledge acquisition on mobile devices. Their study revealed a distinctive strength for each condition: the visual display condition helped build up survey knowledge more, while the tactile condition helped build up mental route knowledge more. Another work focusing on topographic maps proposed a method using speech interaction for editing maps on mobile devices (Degbelo and Somaskantharajan 2020). Participants used 11 speech commands to add features to the mobile map. This work showed that using speech recognition is feasible and usable, while

the user experience was rated rather average. Einfeldt and Degbelo (2021) studied the impact of the sequence of UI elements and tab/scrolling navigation for forms on mobile devices. They reported that sequence matters while the UI navigation modalities had no impact on the UX. Finally, Horbiński et al. (2020) investigated the placement of buttons for map-related tasks on mobile devices and pointed at a discrepancy between the users' preferences and the current practice in designing mobile maps.

While the works aforementioned have provided insight into users' wishes and mental models from different perspectives, they present two key differences with the current work: (1) their focus was on topographic maps, not thematic maps as addressed here; and (2) they studied factors of positive user experience, not mathematical models of UX.

2.2 User Studies on Choropleth and Graduated Symbol Maps

There are also a couple of studies that attempted to assess the effectiveness of CMs and GSMs for specific tasks. For instance, Schiewe (2019) observed a dark-is-more bias in his

study. The dark-is-more bias refers to the fact that even without a legend, people associated the darkest colour hues with the largest data values. He also reported that including a map legend improves the correctness of the association between colour values (dark/light) and data values (large/small), especially for non-experts. The comparison of graduated symbol maps to data tables was the topic of Degbelo et al. (2018), and the authors reported that graduated symbol maps made ‘space-alone compare’ information more visible, while the tables made ‘space-in-time compare’ information more visible to users. Choropleth, graduated symbol and isoline maps were compared in Słomska-Przech and Gołębiowska (2021), and the results showed that choropleth maps were the most effective to complete map-relevant tasks. In addition, a combination of the visual variables colour and orientation was introduced as ‘Choriented Map’ in Gorte and Degbelo (2022), where participants answered questions on choropleth, graduated symbol and choriented maps on mobile devices. Choriented maps resulted in comparable and sometimes better usability and performance than CMs and GSMs. Change blindness in animated choropleth maps was investigated in Fish et al. (2011). They pointed out that map readers have difficulty detecting changes in animated choropleth maps, and tend to overestimate their own change detection abilities.

The studies by Brychtová and colleagues (Brychtová 2015b; Brychtová and Coltekin 2016; Brychtová and Çöltekin 2017b) are closest to the current work because they investigated the impact of colour distance on map readability on choropleth maps. Nonetheless, there are a few important differences. First, the focus of Brychtová and Coltekin (2016) was on the colour distance between map labels and background colours, not on colour distances between adjacent classes on the legend as in this work. Second, Brychtová and Çöltekin (2017b) investigated colour distances between adjacent classes like done in this work, but used six levels (0, 2, 4, 6, 8, 10) of colour distances on a Desktop device, while we use 10 levels on mobile devices (see Sect. 3). Third, their work focused on six classes whereas we use five. Fourth, they have investigated both sequential and qualitative colour schemes, while the current study is only focused on a sequential scheme. Finally, they focused on a fewer number of dependent variables (i.e. performance-related), while our study includes also subjective metrics (see Sect. 3).

2.3 Mathematical Models of Interaction in HCI

Mathematical models may be divided into analytical models (i.e. models that use a known mathematical formula to describe phenomena) and machine/deep learning models (i.e. created from machine learning algorithms through training using labelled data and/or unlabelled data). Previous work in HCI has contributed both types of models. As to

analytical models, much work has focused on understanding the selection of menu items, and the UX of websites. Cockburn et al. ’s Search-Decision-Pointing (SDP) model (Cockburn et al. 2007) incorporates the Fitts’s law (time taken to move to a target item) and the Hick–Hyman law (time to find an item) to predict menu selection time. It also accommodates users’ increasing expertise. Scrolling hierarchical lists was the topic of Cockburn and Gutwin (2009). They pointed out that when users can anticipate the location of items in a list, the time to acquire them is best modelled by logarithmic functions. When they cannot anticipate the location of items, linear models are more appropriate. Cockburn and Gutwin (2010) proposed the Constrained Input Navigation model to predict human navigation and selection performance in constrained-input scenarios. Bailly et al. (2014) proposed a mathematical model to predict total selection time in linear menus. Their model incorporates three components: serial search (top-to-bottom inspection of items), directed search (direct glance at the target item) and pointing.

In the context of websites, Tuch et al. (2009) reported an inverse-linear relationship between visual complexity and pleasure (i.e. start pages with low visual complexity were rated by users as more pleasurable). A replication of several studies confirmed that inverse-linear relationships (Miniukovich and Marchese 2020). Miniukovich and De Angeli (2014) also observed a negative correlation between visual complexity and aesthetics for mobile apps. They proposed linear regression models for visual complexity, visual aesthetics, and visual complexity+aesthetics on mobile devices. Finally, Reinecke et al. (2013) proposed a model to predict the initial impression of a website’s aesthetics based on its colourfulness and visual complexity.

Examples of work on machine/deep learning models include Li et al. (2018) and Ramakrishnan and Kaur (2020)). With the assumption that UX is affected by page load times, Ramakrishnan and Kaur (2020) compared 18 machine learning models to predict page loading time. They reported that the Radial Basis Function and Random Forest models seem more promising for that task. Li et al. (2018) used recurrent neural networks to predict human performance regarding the execution of a sequence of selection tasks. The type of task considered was the selection of a target item from a vertical menu/list.

2.4 Summary

Cartography and Information Visualization research have contributed insights into users’ preferences regarding map-related tasks and Human-Computer Interaction research has produced useful models for the areas of pointing, menu interaction, and user experience of websites. Nonetheless, there is still a need for models of users’ experience with visualizations in general, and maps in particular. The models

developed during this work attempt to address that gap, focusing on question-answering tasks with maps on mobile devices.

3 Method

As discussed in Nosek et al. (2018), there is the dichotomy between prediction (i.e. have an idea about how the world works and make new observations to test whether that idea is a reasonable explanation) and postdiction (i.e. use existing observations of nature to generate ideas about how the world works). This work does some *postdiction* on the topic of predictive models. That is, it uses observations made in the user study to generate ideas about what predictive models of mobile map UX could look like, for some dependent variables. Hence, the work is at the exploratory end of the exploratory-confirmatory spectrum. We took the following steps:

- Step 1: set a number n_{steps} of colour/size distances to collect UX values for. The value of n_{steps} in this work was 10 because at least ten observations per predictor variable are needed to build a regression model that allows good estimates (see e.g. Babyak 2004).
- Step 2: collect a number n_{values} of user experience values for each colour/size distance. For this exploratory study, n_{values} was set to four, which means that the experiment was designed so that four different participants provide a UX value for a given colour/size distance. We built a prototype to generate mobile maps with different colour/size distances.
- Step 3: find a representative value for each colour/size distance. The representative value was computed using the median of the four values from Step 2. The median is a more robust summary statistic compared to the mean because one single outlier can drastically impact the mean (see e.g. Daszykowski et al. 2007).
- Step 4: find the regression line that fits best the 10 representative UX values across all colour/size distances.
- Step 5: collect some empirical data about the predictive performance of the best regression model. This was done by comparing predicted and observed values for $n_{\text{testvalues}} = 2$. That is, the predicted values were compared to the UX values provided by two different participants. In the absence of baseline models to compare our models to, this step is useful to document the expected performance of the identified regression lines on unseen data.

Since the work is learning the mathematical models from data, the reader may question why this is not done using machine learning models. While machine/deep learning

models produce more accurate predictions, their drawback is that (1) they typically require a lot of training data, and (2) it is challenging to get insights into what the model actually learned. For this reason, we chose to focus on regression models so that we can get mathematical formulas as results, and hence, have outcomes that are good candidates to inform further work in analytical modelling. We provide a definition of key terms next (Sect. 3.1), followed by the description of the prototype (Sect. 3.2), the study design (Sect. 3.3) and the analysis strategy (Sect. 3.4).

3.1 Definition of Terms

On both visualization types (choropleth and graduated symbol maps), geographical data is divided into five separate classes using the data classification technique *natural breaks*. Natural breaks divide the data into “natural” classes, by determining the best arrangement of values (Chen et al. 2013). The transformation of this data into visuals implies the use of colour value (CM) and size (GSM) as visual variables, and hence a choice of a colour distance and size distance.

3.1.1 Colour Distance

Colour distance is defined as a metric between two colours within a colour space.¹ The colour distance used in this work is the CIEDE2000 colour difference formula developed by the International Commission on Illumination (CIE) (Sharma et al. 2005). It is based on the CIELAB colour space applying the coordinates L^* (lightness), a^* (red/green value) and b^* (blue/yellow value) to define a colour. Sequential colour schemes (Brewer 1994) were used on choropleth maps. In this scheme, the darker the colour value, the larger the value it represents. Colour values for choropleth maps were generated by the Sequential Colour Scheme Generator developed by Brychtová (2015a). The Sequential Colour Scheme Generator provides colour values after specifying the origin of the colour scheme, the number of classes and the colour distances between adjacent classes. Ten different choropleth maps were developed. The colour distances used on these maps ranged from 2 to 11. The colour distances between neighbouring classes remained the same on one map: for example, if the colour distance between Class A and B was 2, the colour distance between Class B and C was also 2, and so on.

¹ See for example Brychtová and Çöltekin (2017a) for a discussion about the steps to compute the colour distance between two colour shades.

3.1.2 Size Distance

Size distance indicated the circle diameter difference between neighbouring classes. In this work, ten different size distances were used ranging from 2.5 to 25 in 2.5 steps. Circle sizes were specified in pixels. For all size distances, the class representing the lowest values of the dataset were represented with a circle diameter of 10 pixels. The size distance between adjacent classes stayed the same on a map. This means for instance that a size distance of 2.5 provided circles with diameters of 10, 12.5, 15, 17.5 and 20; the largest size distance of 25 generated circles with diameters of 10, 35, 60, 85 and 110.

3.2 Prototype

The prototype was created by using React-Native, which is an open-source JavaScript framework to develop mobile applications on multiple platforms (e.g. iOS, Android) utilizing the same code base. React-Native uses a component-based approach to build user interfaces fast and responsive. It also comes with a reload feature that allows one to make and see the changes in real-time. Both choropleth and graduated symbol maps are integrated with Mapbox GL JS² in React-Native. Mapbox is an easy-to-implement mapping system and gives developers the control to style and customize maps individually by adding custom markers, polygons or polylines. The geographic focus of this work was Europe. To visualize SDG (Sustainable Development Goal) data for European countries, their geometries were needed first. For better performance purposes, a GeoJSON with low resolution (110 m) was applied, which is sufficient for conducting the user study. The GeoJSON was downloaded freely.³ Datasets displayed on the maps were open-source and were provided as JSON datasets from the articles at Our World in Data.⁴ The generation of the maps for the user study via the prototype is described below. The source code is available on GitHub.⁵

3.3 Study Design

3.3.1 Generating Choropleth Maps

Sequential colour schemes were applied on choropleth maps. There were five different blue colours with a fixed colour distance between adjacent classes. The colour hexes were

calculated by the sequential colour scheme generator developed by Brychtová (2015a). Since no colour can be labelled as *the* standard colour for choropleth maps, the choice of blue as a colour for the experiment stems from the authors' subjective preference. European countries (geometries) are coloured according to the dataset value they represent and the class these values fall into. Countries without data are coloured grey.

3.3.2 Generating Graduated Symbol Maps

Circles are the (de facto) standard symbols for graduated symbol maps, and hence our GSMs used them as symbols. The GSMs were generated by displaying circles of five different sizes in relation to the dataset value and the class it belongs to. The centroid, the arithmetic mean position of a polygon could be external to a country, if the shape is irregular, which is unfavourable for the user study. For example, the arithmetic mean position of Norway could be outside of Norway (for instance in Sweden), which means that this circle could no longer be assigned to Norway. Therefore, circles were placed at the countries' *pole of inaccessibility*, the most distant internal point from the polygon outline (Garcia-Castellanos and Lombardo 2007). No circle was presented for a country without data.

3.3.3 Tasks

We focused on question-answering tasks with mobile maps. These questions are typically useful during the exploratory stage of data analysis. Sarikaya et al. (2018) identified six types of high-level data characteristics: trends, outliers, clusters, frequency, distribution, and correlation. To make the study manageable, the focus was on cluster-questions (i.e. identification of geographic entities that belong to a common class). Since the questions involved the 'attribute-in-space' operand of Roth (2013a), they are attribute-in-space/cluster questions. We originally considered including several types of questions (e.g. cluster, frequency), but it proved challenging to distribute them in a balanced way across map types and participants. Also, it would have been more challenging to extract models relevant only to one question type, if the data was collected by mixing up question types. For that reason, we came to the conclusion that it is best to build models that address one type of question at a time (trends, outliers, clusters, frequency, and so on), and see in the future how these can be combined into models that generalize over question types. We did not control for panning and zooming behaviour, nor did we measure it. Since the map was interactive, the participants were free to pan/zoom as much as they wanted to, in order to find answers to the questions.

The datasets interacted with touched on different topics: proportion of women in managerial positions (SDG 5.5.2);

² <https://github.com/rnmapbox/maps>, last accessed: January 17, 2023.

³ <https://geojson-maps.ash.ms>, last accessed: January 17, 2023.

⁴ <https://ourworldindata.org>, last accessed: January 17, 2023.

⁵ https://github.com/SulaxanSo/Mapbox_UX.

share of children who report being bullied; share of people disagreeing that vaccines are safe; and share of tax revenue. Though the datasets communicated temporal information, that temporal information was only shown in the title (see Fig. 1). Hence, there was no need to provide assistance for temporal navigation across different years. The structure of the study and the information-gathering tasks are available in the supplementary material (Sect. 7, Appendix A).

3.3.4 Study Variables

The independent variables of the study were: different types of map visualization (Choropleth Maps and Graduated Symbol Maps); different colour distances on Choropleth Maps, and different size distances on Graduated Symbol Maps. Previous studies reported that the spatial distance between symbols affects the performance of users during map reading tasks on choropleth maps (Brychtová and Çöltekin 2017b) and graduated symbol maps (Cybulski 2020). Nonetheless, these studies were done using static maps. In the current case, the map is interactive and thus the spatial distance between the enumerations units (countries) was constantly changing because the gap between symbols varies as one zooms in or out. Hence, the implications of these studies are not directly transferable to the current study. As a result, the spatial distance was not considered as a variable during the study.

The main dependent variable is the user experience of the participants. User experience has several facets, notably a pragmatic and a hedonic dimension (see e.g. Hassenzahl 2005). As for the pragmatic dimension, we measured both accuracy (percentage of correct answers) and efficiency (time in seconds needed for solving the tasks). The following formula was used to calculate the accuracy of a participant's answer to a question about the countries fulfilling a constraint (see supplementary material, Sect. 7, Appendix A):

$$\text{accuracy} = \frac{\text{cag}}{\max(\text{ag}, \text{ca})},$$

where cag is the number of correct answers given, ag is the total number of answers given by a participant and ca is the number of correct answers. As for the hedonic dimension, we used a modified dimension of the NASA TLX questionnaire. The NASA Task Load Index (TLX) is a questionnaire that is used to assess workload on six dimensions (Hart 2006). The overall Nasa TLX score derives from the subscales *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort* and *frustration*. Since the user study does not ask for any physical effort, the dimension 'physical demand' was not captured. Participants were not pressured by time, therefore the dimension 'temporal demand' was also omitted. Moreover, the NASA TLX was simplified by

reducing the number of scales to seven. We measured the perceived mental demand (how mentally demanding was the task?); perceived performance (how successful were you in accomplishing what you were asked to do?); perceived effort (how hard did you have to work to accomplish your level of performance?); and perceived frustration (how insecure, discouraged, irritated, stressed and annoyed were you?). All these four hedonic variables were measured on a seven-point Likert scale from 1 to 7, where 1 equals low and 7 equals high. Scale 4 represents the middle of the scale, which means neither low nor high. In sum, user experience was measured through six variables: two touching on the pragmatic dimension and four on the hedonic dimension. The modified version of the NASA TLX questionnaire used in the work is available as supplementary material (Sect. 7, Appendix B).

3.3.5 Procedure and Apparatus

In each study, there was one participant and a moderator. Firstly, the moderator gave a brief explanation of the objective and the procedure of the study. The participant was then asked to sign the consent forms and an additional statement on vision, declaring having normal or corrected-to-normal vision by wearing glasses or contact lenses. After that, the mobile device with the pre-installed prototype application was handed out. A Samsung Galaxy A32 5 G (display size: 720 × 1600 pixels) was used throughout the experiment. On another laptop, an online questionnaire in LimeSurvey was handed over to the participants. They first filled out a questionnaire about background information. They then went on to interact with the prototype application and give their answers to the questions. At the end of the study, the moderator wrapped up the session, by asking the participant if there was some feedback they were willing to share and giving them their rewards (sweets). The study was conducted in a room where sunlight did not obscure the participant's view of the map. For example, unfavourable lighting conditions could affect colour perception on choropleth maps and thus worsen the study results and the time taken to solve the tasks. The study was pilot-tested and approved by the institutional ethics board.

3.3.6 Participants

Thirty users participated in the experiment. They were divided into two different groups. The first group included the first 20 participants whose study results were used to develop the model. In this group, there were eleven male and nine female participants. The majority of the participants were between 23 and 27 years old (11/20). Seven were 28–32 years old and two participants stated to be older than 33 (34 and 58). All participants were frequent mobile map

users, at least once a week. Twelve participants stated to be very familiar with choropleth maps (12/20). The remaining eight participants reported being familiar with choropleth maps. None stated to be unfamiliar with them. Regarding graduated symbol maps, six participants were very familiar with them. Ten participants mentioned being familiar with graduated symbol maps, whereas four participants reported being unfamiliar with them.

The second group included the remaining ten participants, whose data was used to test the predictive performance of the models. It included seven female and three male participants. One participant was between 18 and 22 years old. Four participants stated being 23 to 27 years old. Three participants reported an age between 28 to 32, while two stated being older than 33 (34 and 35). All participants were frequent mobile map users (at least once per week). Most participants stated being very familiar with choropleth maps (7/10), while two participants reported finding them familiar. Only one participant stated to feel unfamiliar with choropleth maps. In connection with graduated symbol maps, three participants were very familiar with them. Five reported being familiar with them, while two stated being unfamiliar with graduated symbol maps.

3.4 Data Analysis

The aim of the modelling process was to find a model that fits best the median values for each dependent variable on both visualization types. The data from the first 20 participants in the experiment was used to learn about the model as said above. Five different polynomial regression functions were tested to fit the data by using the `lm()` linear model function in R, i.e. the tests went from a linear regression model to a fifth-degree polynomial regression model. The best fit was identified by comparing the Bayesian information criteria (BIC) values of these five models. The BIC discounts goodness-of-fit to the extent that it is realized by a model that is overly complex, and hence implements the principle of parsimony (see Vandekerckhove et al. 2015). The BIC leads to models with lower prediction error compared to other criteria (e.g. adjusted R squared, see Sharma et al. 2021). The model with the lowest BIC was selected as the “best” fit for a given set of data points. The predictive performance of the best fit was then evaluated using the data from the remaining ten participants. The units of the different dependent variables (accuracy, efficiency, and so on) are different. To make the results comparable across dependent variables and map types, a delta (Δ) was computed and expressed in percentage of unseen, empirical data. For accuracy and efficiency, $\Delta = \frac{|\text{predicted value} - \text{observed value}|}{\text{observed value}} = \frac{|\text{absolute error}|}{\text{observed value}}$. For Likert-scale ratings, $\Delta = \frac{|\text{predicted value} - \text{observed value}|}{\text{max scale value} - \text{min scale value}} = \frac{|\text{absolute error}|}{6}$. The

values for all deltas are shown in the supplementary material (Sect. 7, Appendix C).

4 Results

As mentioned in Sect. 3.4, the BIC was used to select the best fit for the data among the five polynomial regression models. The BIC is a quantity useful to choose between two or more alternative models (the lower the BIC the better the model), but absolute values of the BIC do not have a practical meaning. On the contrary, the adjusted R-squared (R^2) indicates the percentage of the variation in the response variable that can be explained by the predictor variable in the model, adjusted for the number of predictor variables (Yang and Berdine 2015). The higher the R^2 value, the better the model. Because absolute values of the R^2 are easier to interpret, they will be mentioned while reporting in this section. To reduce the density in the plots, only the best fit is visualized. Table 1 gives an overview of the parameters of the aforementioned predictive models. Each equation for a dependent variable is presented separately in the corresponding section below. As for the error on the test datasets, we report on the average (i.e. median value) of all deltas across all colour/size distances. The median is more robust to outliers.

4.1 Accuracy

Figure 2a, b present the accuracy results on choropleth and graduated symbol maps respectively. The median accuracies ranged from 40.5 to 98.5% on choropleth maps and on graduated symbol maps from 82 to 100%. Regarding choropleth maps, participants performed best at colour distances 10 (98.06%) and 11 (98.53%) and worst at colour distances 2 (40.54%) and 4 (49%). The best performances on graduated symbol maps were at size distances 10, 17.5 and 25 (100% each). On this visualization type, only two size distances performed worse than 95%, namely size distance 2.5 (82%) and 7.5 (91.26%).

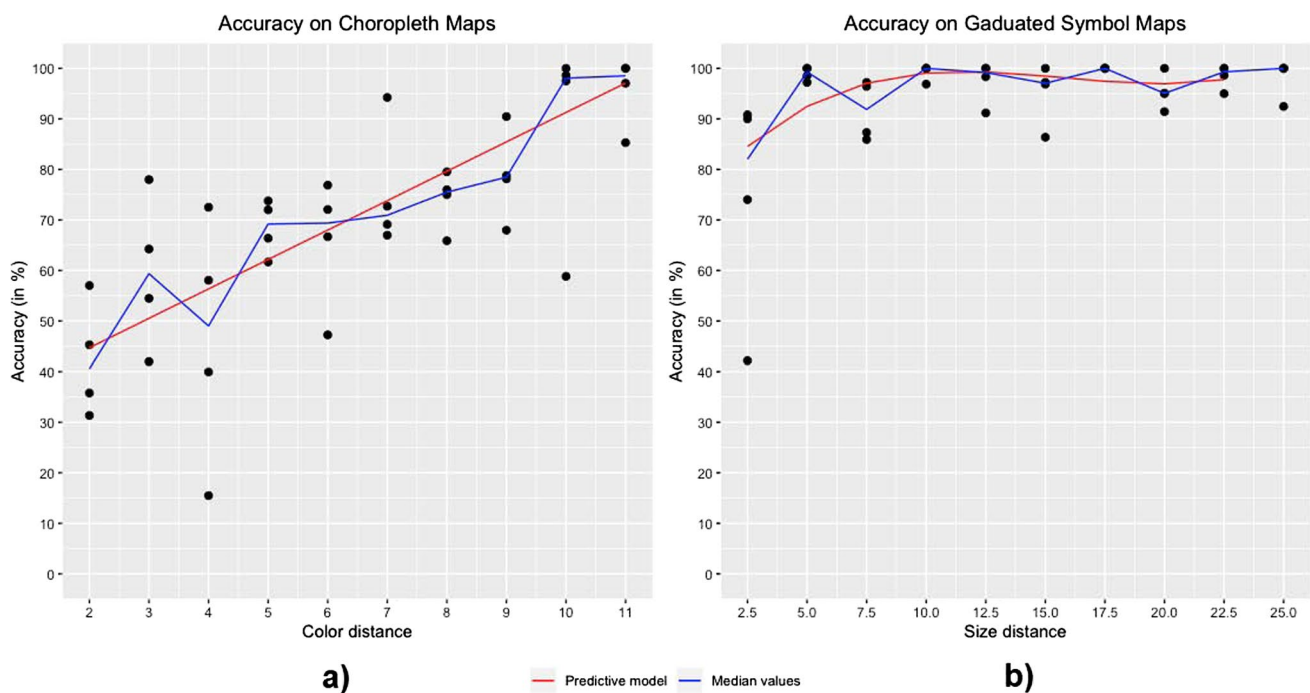
On choropleth maps, the best fit was a first-degree polynomial regression model, which had an adjusted R-squared of 0.8835. The equation for this linear equation is:

$$y = 5.8217x + 33.046.$$

The predictions on the test dataset using this model had an average error (median) of 10.83% of unseen empirical data (Mean: 14.65%, Max: 56.89%, Min: 0.29%, SD: 13.04%). That is, on average, the accuracy values predicted by the model and those of the test data differed by 10.83%.

Table 1 Parameters of the mathematical functions modelling mobile map user experience (*CMs* Choropleth Maps, *GSMs* Graduated Symbol Maps)

Dependent variable	Type of model	Coefficients						Adj. R^2	BIC
		x^5	x^4	x^3	x^2	x	Constant		
Accuracy on CMs	Linear equation					5.8217	33.046	0.8835	70.0344
Accuracy on GSMs	3rd-degree polynomial			0.008322	− 0.393343	5.761621	72.46125	0.5119	62.4492
Efficiency on CMs	Linear equation					− 5.187	170.447	0.477	87.9035
Efficiency on GSMs	Linear equation					− 1.498	157.212	0.1711	93.0881
Mental demand on CMs	Linear equation					− 0.3	6.6	0.4926	30.3436
Mental demand on GSMs	4th-degree polynomial		− 0.0001734	0.0088982	− 0.1381002	0.53885	4.875	0.6815	21.6246
Performance on CMs	Linear equation					0.21515	2.95152	0.4104	26.6259
Performance on GSMs	3rd-degree polynomial			0.001486	− 0.073706	1.11251	0.966667	0.6343	24.9702
Effort on CMs	2nd-degree polynomial				− 0.06439	0.57955	4.63485	0.6504	25.0789
Effort on GSMs	2nd-degree polynomial				0.010606	− 0.315909	6.191667	0.2412	27.8523
Frustration level on CMs	5th-degree polynomial	0.007051	− 0.23048	2.840035	− 16.352855	43.312005	− 36.575758	0.6093	32.1334
Frustration level on GSMs	3rd-degree polynomial			− 0.002959	0.132354	− 1.747591	8.716667	0.4385	35.2081

**Fig. 2** Accuracy on Choropleth (a) and Graduated Symbol Maps (b)

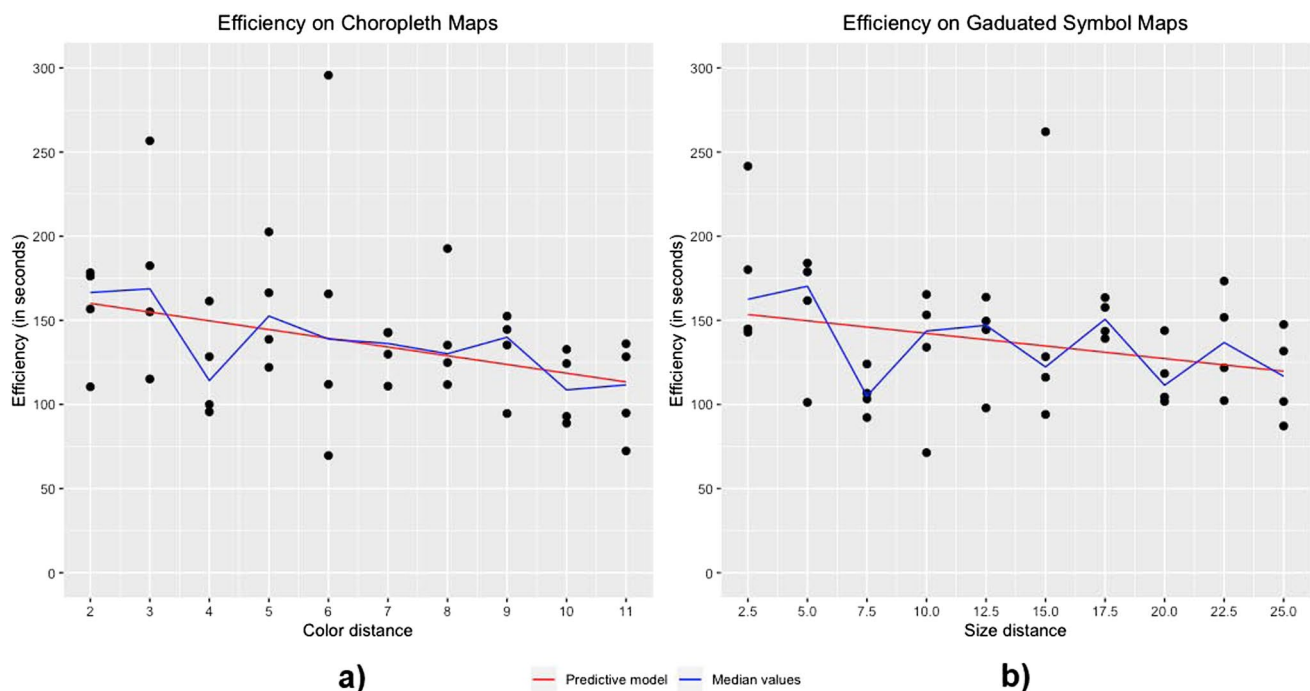


Fig. 3 Efficiency on Choropleth (a) and Graduated Symbol Maps (b)

The best fit on graduated symbol maps was a third-degree polynomial, which has an adjusted R-squared of 0.5119. The equation that describes the model is:

$$y = 0.08322x^3 - 0.393343x^2 + 5.761621x + 72.46125.$$

Here, on average, the accuracy values predicted by the model and those of the test data differed by 2.14% (Mean: 7.33%, Max: 34.51%, Min: 0%, SD: 11.12%).

4.2 Efficiency

The diagrams for efficiency are shown in Fig. 3a, b. Here, each black data point represents the average time taken by participants across all tasks in the four sections (see supplementary material, Sect. 7, Appendix A). The average times needed to answer one question ranged around 50–300 s on both map types. On choropleth maps, participants performed fastest at colour distances 10 (108.64 s) and 11 (111.63 s) and worst at colour distances 2 (166.5 s) and 3 (168.74 s). The results on graduated symbol maps show the fastest performances at size distances 7.5 (104.97 s) and 20 (111.38 s) and the slowest at size distances 2.5 (162.5 s) and 5 (170.27 s) for each question.

A linear equation was the best fit for the efficiency data points on choropleth maps ($R^2 = 0.477$):

$$y = -5.187x + 170.447.$$

On average, the efficiency values predicted by the model and those of the test data differed by 28.80% (Mean: 28.24%, Max: 60.39%, Min: 5.92%, SD: 15.88%).

On graduated symbol maps, the best fit was obtained with a linear equation ($R^2 = 0.1711$):

$$y = -1.498x + 157.212.$$

The efficiency values predicted by the model and those of the test data differed by 22.64% on average (Mean: 22.79%, Max: 51.72%, Min: 5.09%, SD: 13.77%).

4.3 Perceived Mental Demand

Figure 4a shows that the median values ranged between 2.5 (colour distance 10) and 6.5 (colour distance 3) on choropleth maps. On graduated symbol maps (Fig. 4b), the values spanned between 2.5 (size distance 12.5) and 5.5 (size distance 2.5). The mental demand was rated highest at colour distances 3 (6.5) and 5 (6) and lowest at colour distances 10 (2.5) and 11 (3) on choropleth maps. Regarding graduated symbol maps, the mental demand ratings were highest at size distances 2.5 (5.5) and 5 (5) and lowest at 12.5 (2.5) and 15 (3).

On choropleth maps, the best fit was a linear equation regression model ($R^2 = 0.4926$):

$$y = -0.3x + 6.6.$$

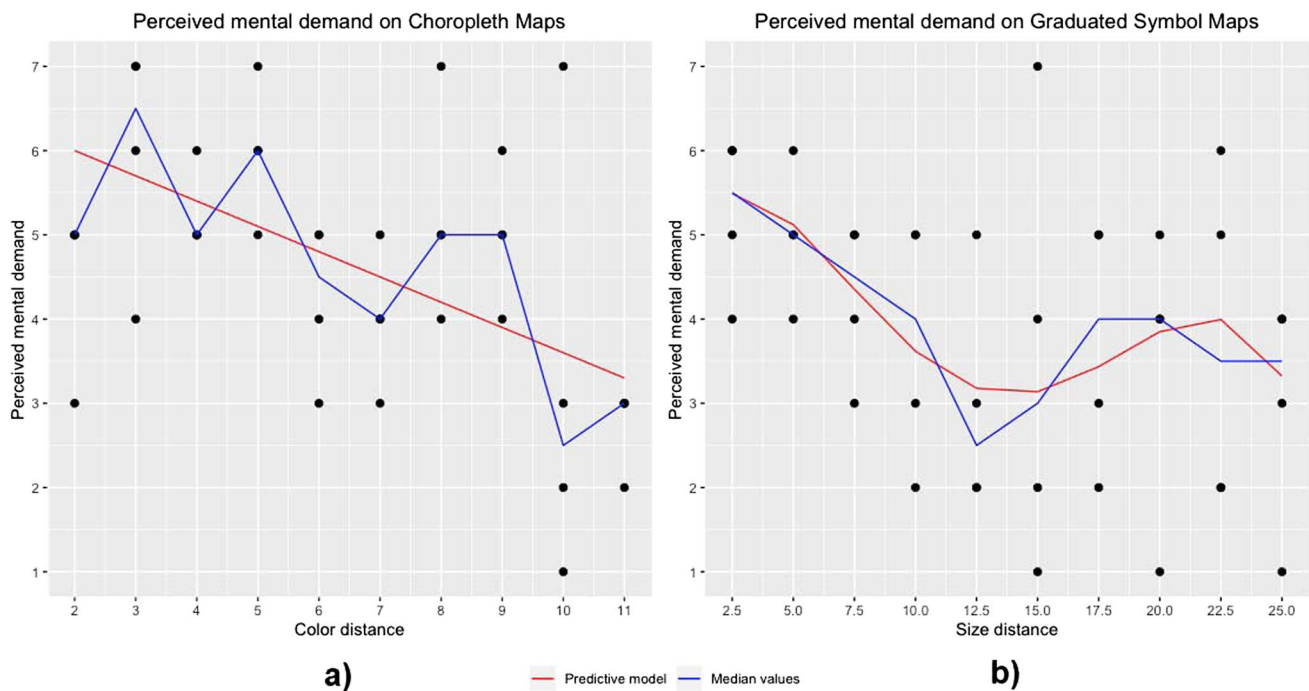


Fig. 4 Perceived mental demand on Choropleth (a) and Graduated Symbol Maps (b)

The mental demand values predicted by the model and those of the test data differed by 14.17% on average (Mean: 18.00%, Max: 56.67%, Min: 1.67%, SD: 13.12%).

The data points regarding graduated symbol maps were best fitted by a fourth-degree polynomial regression model ($R^2 = 0.6815$). The equation of this model is:

$$y = -0.0001734x^4 + 0.0088982x^3 - 0.1381002x^2 + 0.53885x + 4.875.$$

On average, the mental demand values predicted by the model and those of the test data differed by 14.17% (Mean: 18.60%, Max: 52.00%, Min: 0%, SD: 13.21%).

4.4 Perceived Performance

The average performance ratings were 4.35 on choropleth and 5.55 on graduated symbol maps. The values ranged from 2.5 (colour distance 5) and 5.5 (colour distance 10 and 11) with approximately equal distribution over all colour distances (Fig. 5a). On graduated symbol maps (Fig. 5b) the span was between 3 (size distance 2.5) and 6.5 (size distance 10 and 20). Apart from size distance 2.5, all other size distances were rated higher than or equal to 5, which is noticeable. The performance on choropleth maps was rated highest at colour distances 10 and 11 (5.5 each) and lowest at colour distances 5 (2.5) and 3 (3.5). The performance on graduated symbol maps was rated highest at size distances 10 and 20 (6.5 each) and lowest at size distances 2.5 (3) and 22.5 (5).

On choropleth maps, a linear equation fitted the data points best ($R^2 = 0.4104$):

$$y = 0.21515x + 2.95152.$$

The performance values predicted by the model and those of the test data differed by 14.92% on average (Mean: 15.63%, Max: 36.50%, Min: 0.50%, SD: 9.20%).

The best fit on graduated symbol data was a third-degree regression model ($R^2 = 0.6343$):

$$y = 0.001486x^3 - 0.073706x^2 + 1.11251x + 0.966667.$$

The performance values predicted by the model and those of the test data differed by 19.83% on average (Mean: 21.95%, Max: 76.83%, Min: 1.17%, SD: 16.05%).

4.5 Perceived Effort Demanded

The average effort ratings were 5.15 on choropleth and 4.4 on graduated symbol maps. The values ranged from 3 (colour distance 10) to 6.5 (colour distance 5) on choropleth maps. Apart from colour distances 10 and 11, all other colour distances were rated higher than or equal to 5, which is noticeable (Fig. 6a). The effort on choropleth maps was rated highest at colour distances 3 (6) and 5 (6.5) and lowest at colour distances 10 (3) and 11 (3.5). On graduated symbol maps (Fig. 6b), the span was between 3 (size distance 15) and 5.5 (size distance 2.5). The lowest ratings were provided

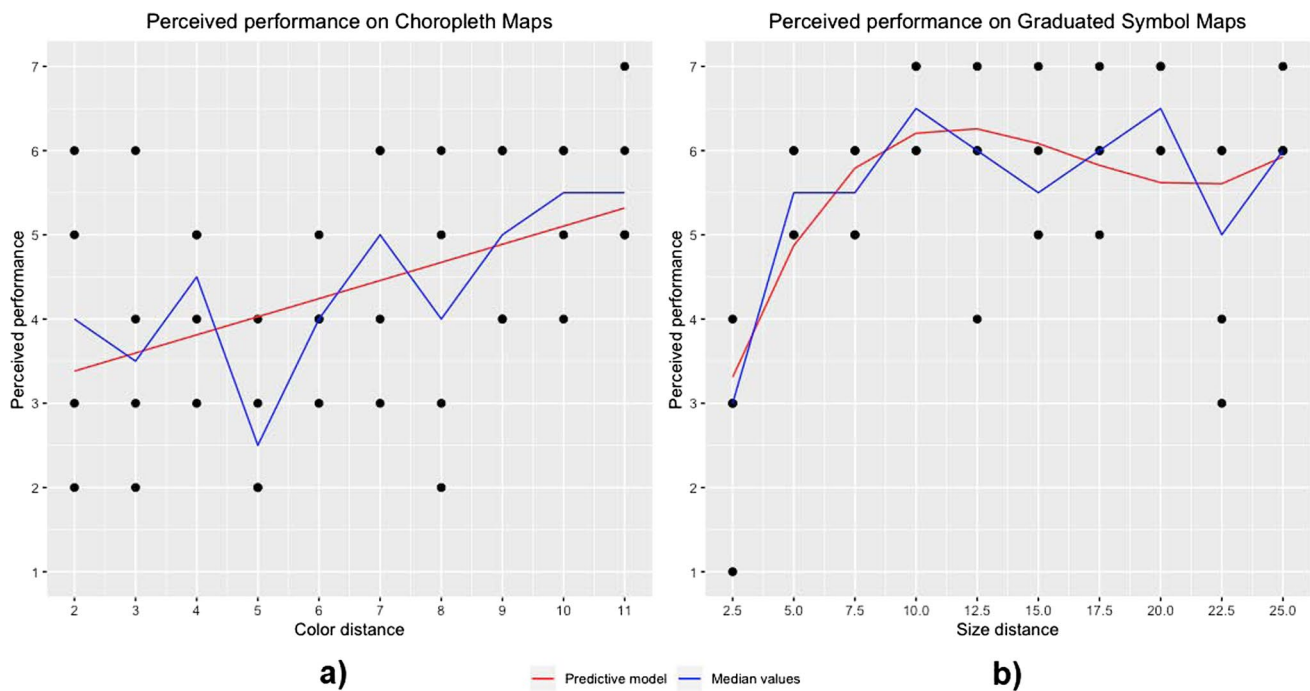


Fig. 5 Perceived Performance on Choropleth (a) and Graduated Symbol Maps (b)

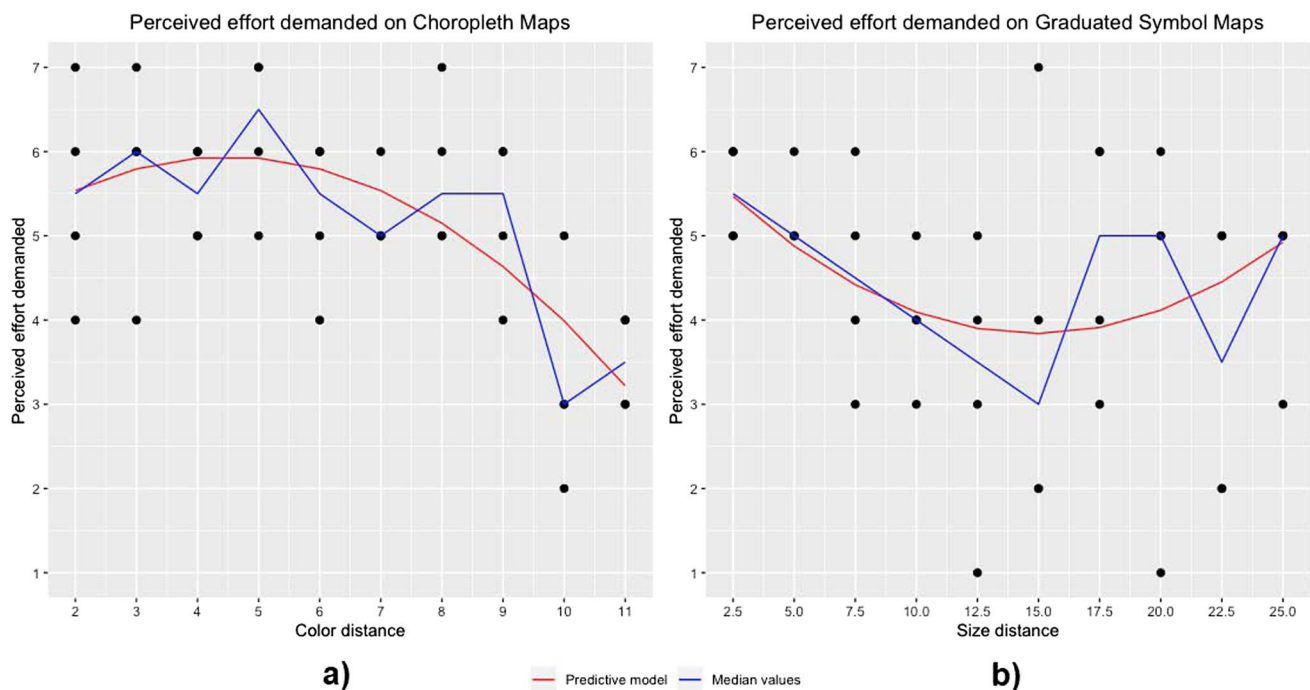


Fig. 6 Perceived effort demanded on Choropleth (a) and Graduated Symbol Maps (b)

at distances 15 (3) and 22.5 (3.5), whereas the highest effort was reported by the participants when working with size distances 2.5 (5.5) as well as 5, 17.5, 20 and 25 with a reported effort of 5.

A second-degree polynomial regression model fitted the data points on choropleth maps best ($R^2 = 0.6504$). The equation of this model is:

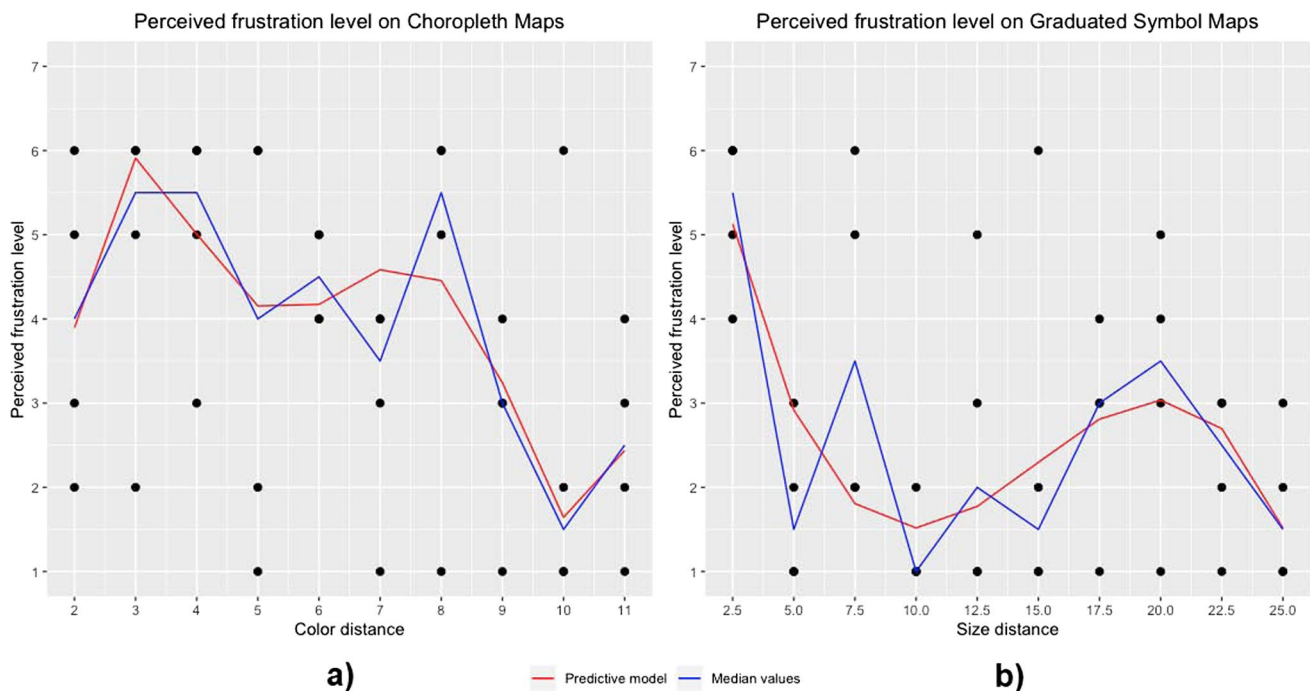


Fig. 7 Perceived frustration level on Choropleth (a) and Graduated Symbol Maps (b)

$$y = -0.06439x^2 + 0.57955x + 4.63485.$$

The effort values predicted by the model and those of the test data differed by 14.83% on average (Mean: 18.22%, Max: 49.83%, Min: 1.33%, SD: 14.14%).

On graduated symbol maps, the best fit was a second-degree regression model ($R^2 = 0.2412$):

$$y = 0.010606x^2 - 0.315909x + 6.191667.$$

The effort values predicted by the model and those of the test data differed by 11.83% on average (Mean: 13.95%, Max: 31.83%, Min: 1.3%, SD: 11.03%).

4.6 Perceived Frustration Level

Choropleth maps were most frustrating at colour distances 3, 4 and 8 (5.5 each) and least frustrating at colour distances 10 (1.5) and 11 (2.5) as seen in Fig. 7a. On graduated symbol maps (Fig. 7b), the highest level of frustration was perceived when working with size distances 2.5 (5.5) and lowest at size distances 10 with 1 (very low frustration) as well as 5, 15 and 25 (1.5 each).

On choropleth maps, a fifth-degree polynomial regression model was the best fit ($R^2 = 0.6093$):

$$y = 0.007051x^5 - 0.23048x^4 + 2.840035x^3 - 16.352855x^2 + 43.312005x - 36.575758.$$

The frustration values predicted by the model and those of the test data differed by 29% on average (Mean: 29.50%, Max: 66.83%, Min: 1.50%, SD: 20.40%).

On graduated symbol maps, the best fit was a third-degree polynomial regression model ($R^2 = 0.4385$). The equation of this model is:

$$y = -0.002959x^3 + 0.132354x^2 - 1.747591x + 8.716667.$$

The frustration values predicted by the model and those of the test data differed by 20.75% on average (Mean: 24.45%, Max: 68.83%, Min: 1.33%, SD: 17.33%).

4.7 Multiverse Analysis

As mentioned in Sect. 3, a key step during the analysis was the choice of a representative value for each colour/size distance. While the median was used as a technique to compute that representative value, there are alternative techniques (e.g. mean or mode) to compute the representative value for each colour/size distance. Since the regression model fits the representative values across all colour/size distances, the form and monotonicity of the

Table 2 Sensitivity analysis: frequencies for the function forms as well as monotonicity trends across $N = 1,048,576$ iterations

	Form = type of nth-degree polynomial					Monotonicity	
	1st (%)	2nd (%)	3rd (%)	4th (%)	5th (%)	Monotonic (%)	Non-Monotonic (%)
Accuracy Choropleth	60.09	5.31	3.69	11.97	18.94	63.55	36.45
Accuracy GSM	35.03	8.74	10.74	7.83	37.66	49.87	50.13
Efficiency Choropleth	62.01	12.08	6.03	5.05	14.84	62.10	37.90
Efficiency GSM	53.41	6.64	12.86	13.33	13.77	64.98	35.02
Mental Demand Choropleth	38.12	20.38	1.86	10.91	28.73	38.12	61.88
Mental Demand GSM	48.16	13.68	5.77	12.33	20.07	66.08	33.92
Performance Choropleth	45.43	19.06	9.03	8.23	18.25	45.62	54.38
Performance GSM	17.09	12.35	24.00	15.81	30.75	48.16	51.84
Effort Choropleth	27.18	32.56	6.95	9.61	23.70	27.62	72.38
Effort GSM	49.38	18.31	6.16	11.15	15.00	70.01	29.99
Frustration Choropleth	54.56	11.41	7.60	5.71	20.72	54.65	45.35
Frustration GSM	52.10	5.88	23.24	11.95	6.83	61.36	38.64

The most frequent type per dependent variable is highlighted in bold

mathematical function may be influenced by *where* that representative value for each colour/size distance exactly lies. Conducting (large-scale) user studies that control several parameters (task, device, user background), so that we can generalize across settings and become more confident about the “true” representative value for each colour/size distance is beyond the scope of this work.

To evaluate the sensitivity of the results to the location of the representative values, we did a sensitivity analysis (a.k.a. multiverse analysis). As discussed in Steegen et al. (2016), a multiverse analysis acknowledges that datasets during an analysis are, to some extent, actively constructed. It involves performing the analysis of interest across the whole set of data sets that arise from different reasonable choices for data processing. It offers an idea of how much the conclusions change because of arbitrary choices in data construction. The key idea was to shed some light on the question “how sensitive are the results to alternative locations of the representative values”?

The analysis iterates over all input data points about a million times ($4^{10} = 1,048,576$ times exactly: we have four values per step and 10 steps). At each iteration, we take a value randomly from the four input values available for a step (= colour/size distance). We then look for the best regression model given those randomly chosen input points and document both its form (1st deg, 2nd deg and so on) and monotonicity (monotonic vs non-monotonic). This is basically equivalent to assuming that any of the values collected has an equal chance to be taken as the “representative value” if more data is collected, and simulating the possible consequences of that scenario. Table 2 summarizes the frequencies of occurrences of the function forms as well as the monotonicity trends across all 1,048,576 iterations. The

following observations can be made about the models for the different aspects of user experience:

- Accuracy Choropleth: likely a 1st-degree polynomial, likely monotonic;
- Accuracy GSM: likely a 5th-degree polynomial, likely non-monotonic;
- Efficiency Choropleth: likely a 1st-degree polynomial, likely monotonic;
- Efficiency GSM: likely a 1st-degree polynomial, likely monotonic;
- Mental Demand Choropleth: the results are inconclusive. Although a 1st-degree polynomial is the most frequent best model, the probability that the model is non-monotonic lies at 62%;
- Mental Demand GSM: likely a 1st-degree polynomial, likely monotonic;
- Performance Choropleth: the results are inconclusive. Although a 1st-degree polynomial is the most frequent best model, the probability that the model is non-monotonic lies at 54%;
- Performance GSM: likely a 5th-degree polynomial, likely non-monotonic;
- Effort Choropleth: likely a 2nd-degree polynomial, likely non-monotonic;
- Effort GSM: likely a 1st-degree polynomial, likely monotonic;
- Frustration Choropleth: likely a 1st-degree polynomial, likely monotonic;
- Frustration GSM: likely a 1st-degree polynomial, likely monotonic.

5 Discussion

The work has extracted 12 mathematical functions to model the UX of mobile maps through the analysis of the data. Across dependent variables, the models for CMs predicted values that were, on average, within [11–29]% of unseen empirical data: 11% (accuracy), 29% (efficiency), 14% (perceived mental demand), 15% (perceived performance), 15% (perceived effort), and 29% (perceived frustration level). The predictions of the models for GSMs were, on average, within [2–23]% of unseen empirical data: 2% (accuracy), 23% (efficiency), 14% (perceived mental demand), 20% (perceived performance), 12% (perceived effort), and 21% (perceived frustration level). This is promising for a first attempt.

5.1 Theoretical Value of the Models

The section addresses four topics: the usefulness of the models for hypothesis generation, their theoretical implications for the study of visual variables, the lessons learned about predictors of mobile map UX, and the relevance of the models for plastic maps.

5.1.1 Hypothesis Generation

As mentioned in Sect. 3, one key aim and outcome of the study is to generate hypotheses about predictive models. A look at the equations (Table 1) suggests some tentative statements about the *form* of the mathematical functions that predict mobile map UX. A linear equation suggests a function of the form *dependent variable* = $a + b * X$, where X denotes either a colour distance or a size distance; a two-degree polynomial function suggests a function of the form *dependent variable* = $a + b * X + c * X^2$; and so on. In the long run, once the form of the model is validated for a class of questions (e.g. clusters), the constants may be learned through calibration work for other settings (e.g. devices). A second direction in which the results can help formulate hypotheses is about the *monotonicity* of the relationship between the colour/size distance and the dependent variable (i.e. whether these are monotonic or non-monotonic relationships). Below, we summarize the hypotheses that can be formulated based on the data. We formulate a hypothesis when the three pieces of evidence available (i.e. function obtained from the analysis using the median value, most frequent function form and monotonicity trends) converge. When this is not the case, we refrain from making a statement. A look at the graphs from Sect. 4, Tables 1 and 2 suggests the following:

- Colour distance and accuracy: the relationship is likely linear (and hence monotonic);
- Size distance and accuracy: the relationship is likely non-linear and non-monotonic;
- Colour distance and efficiency: the relationship is likely linear (and hence monotonic);
- Size distance and efficiency: the relationship is likely linear (and hence monotonic);
- Size distance and perceived performance: the relationship is likely non-linear and non-monotonic;
- Colour distance and perceived effort: the relationship is likely non-linear and non-monotonic;
- No strong hypotheses can be formulated for the following relationships, based on the current data:
 - Colour distance and perceived mental demand
 - Size distance and perceived mental demand
 - Colour distance and perceived performance
 - Size distance and perceived effort
 - Colour distance and perceived frustration
 - Size distance and perceived frustration

In sum, the pieces of evidence (i.e. models based on the median plus sensitivity analysis) have converged for the *pragmatic* aspects of user experience (accuracy and efficiency), but are still conflicting for the majority of *hedonic* aspects of user experience (perceived mental demand, perceived performance, perceived effort, and perceived frustration level). Formulating these hypotheses about the function form and the monotonicity properties would have been challenging using machine/deep learning models. It should also be noted that the hypotheses only apply to mobile devices, cluster-tasks and the range of colour/size distance investigated. Indeed, the UX may exhibit a trend (e.g. linear) on a range of values and another trend (e.g. logarithmic) on another range of values. The trends may also be different if the type of device/task changes, but this needs to be investigated in future work.

5.1.2 Visual Variables

Bertin (1983) suggested seven visual variables, which were subsequently extended into a list of 12 (Roth 2017): location, size, shape, orientation, colour hue, colour value, colour saturation, orientation texture, arrangement, crispness, resolution and transparency. Because the choropleth maps investigated in the work use *colour value* and the graduated symbol maps use *size* as visual variables, the hypotheses made just above on the form of the modelling function and its' monotonicity are implicit hypotheses about the trajectory of the impact of colour value and size on design. Hence, they can inform a theory of

visual variables. For instance, Garlandini and Fabrikant (2009) compared participants' response times (efficiency) for change detection tasks using size and colour value and reported that size was faster. A look at Fig. 3 suggests that the assertion that 'size is faster than colour' cannot hold for all colour and size distances. Though the purpose of the study was not to compare CMs and GSMs, the figure suggests that for the ranges of colour/size distances investigated and the cluster tasks, efficiency data fall in a similar range of values. The colour/size distances were not documented in Garlandini and Fabrikant (2009), but one can guess that these are so that efficiency values in their study could fall in different ranges.

5.1.3 Predictors of Mobile Map UX

The R^2 values ranged from 41 to 88% for colour distances, and from 17 to 68% for size distances (Table 1). These values are relatively high for a *single* regressor and suggest that colour/size distance are good candidate predictors for models of mobile maps UX on CMs and GSMs respectively. Another implication from Table 1 is that authors should discuss the possible impact of colour/size distances on mobile map UX ratings obtained in their experiments. Colour/size distances indeed appear to have a non-negligible contribution to the dependent variable measured.

5.1.4 Plastic Maps

Kray and Degbelo (2019) adapted the idea of user interface plasticity (Thevenin and Coutaz 1999; Coutaz 2010) to the domain of maps and proposed the concept of map plasticity. Like in the original plasticity framework, 'plastic maps' foresee an interactor model. A recommendation made about interactors in Thevenin and Coutaz (1999) was that they should "specify the abstract data types they are able to handle. They should also be able to evaluate their appropriateness as well as their rendering cost" (Page 113). The models of map UX learned in this work could be one way of informing these interactors as they are evaluating their appropriateness in maximizing the user experience of map users.

5.2 Practical Usefulness of the Models

In addition to their usefulness to advance Cartography theory discussed above, the mathematical models have also some practical value. For instance, several authors have mentioned that empirically-derived guidelines for map design are still lacking (Roth 2013b; Kray et al. 2017; Degbelo et al. 2022). Though the R^2 values from 17 to 88% remind us that not every model is a 'good' fit for the data points collected at this point, the empirically-derived models are still useful to

formulate some recommendations (e.g. they give some indications about trends of the dependent variables for both map types). A word of caution: using them for predictions should be done only if the designer considers the error ranges documented above to be acceptable for their task.

The functions modelling accuracy and efficiency (Figs. 2 and 3) show that increasing the colour distance on choropleth maps implies having greater success and needing less time to answer questions. The trend that accuracy improved with increasing colour distances was also observed in Brychtová and Çöltekin (2017b), who recommended the colour distance 10 to cartographers and information visualization designers. Our data suggest that colour distance 11, which was not tested in their work, might yield even better results from the accuracy and efficiency point of view. Taking all six models into consideration, colour distances 10 and 11 are favourites among the colour distances investigated, if designers aim for the 'best' user experience as users answer cluster questions on mobile devices.

On graduated symbol maps, the model anticipates the lowest accuracy at a size distance of 2.5. The remaining size distances might produce accuracies greater than 90%. The efficiency model suggests that a graduated symbol map with a size distance of 25 would be best to answer questions efficiently. Combining these two factors, the size distance of 25 would be best to answer cluster questions, if only pragmatic aspects of UX are important. Map users might perceive the least mental demand at size distances 12.5 and 15 according to our model; perceive the highest performance level at size distances 10 and 12.5; perceive that the least effort is demanded at size distances 12.5, 15 and 17.5; and be less frustrated at size distances 7.5, 10, 12.5 and 25. Taking all these factors into account, the models suggest a graduated symbol map with a size distance of 12.5 as adequate for the 'best' user experience if hedonic aspects of UX should be given more weight.

At last, an area where the models can be of practical value is that of toolkit design. Several toolkits have been proposed in Cartography research to support the creation of choropleth and/or graduated symbol maps. These include for instance SDG Viz (Gong 2019), AdaptiveMaps (Degbelo et al. 2020), the GAV Toolkit (Van Ho et al. 2012) and the Geoviz Toolkit (Hardisty and Robinson 2011). The default colour/size distance used in these toolkits has been rarely if at all discussed. Hence, our work can inform the design of future toolkits for interactive map creation, and more broadly of tools in need of *good defaults* to support the creation of choropleth/graduated symbol maps semi-automatically.

5.3 Limitations

One limitation of the work relates to the number of participants. Only four data points for each colour/size distance were used to build the models and two data points were used to test them. These were helpful to learn what the models *could* look like, but increasing the number of data points is needed to find out what the models *truly* look like. In addition, the number of colour/size distance steps used to construct the model could be extended (i.e. go beyond 10). The evaluation has not yet addressed the question of how well the model performs for colour/distances we have not investigated (i.e. colour/size distances in-between two step values or beyond the range of the ten selected).

In addition, the experiment used real data and this comes with the risk of participants' prior knowledge as a confounding variable. Nonetheless, since the questions were of type *cluster*, answering them necessitated the interaction with the actual datasets given in the experiment, as opposed to general knowledge about the topic. That is, finding all countries that belong to a given cluster is a question that requires interaction steps with the map to find the specific spatial entities within the cluster. Given the nature of the task and the fact that we used four different datasets, the impact of the participants' individual levels of familiarity with each dataset on the overall results is likely minimal. Still, since the level of familiarity with each topic of the four datasets (e.g. share of children who report being bullied, rate of women in senior and middle management positions, and so on) was not explicitly controlled, no statement can be made at this point about the impact of participants' familiarity with the topics on the mathematical functions obtained.

Furthermore, the user group that participated was relatively homogeneous. The age of (29/30) of the participants ranged from 18 to 35, and only one participant was significantly older than the rest of the group (58). Hence, the results apply at best to the age group 18–35 only. From the implementation perspective, choropleth maps used the colour *blue* to visualize geographical data. The extent to which the results generalize to other colours still needs to be tested empirically. Likewise, circles were utilized to represent data on graduated symbol maps. The extent to which the results hold for other symbols such as squares or triangles needs to be investigated as well. In addition, the study investigated questions of the type *cluster* only. Asking different types of questions may lead to different tentative models.

5.4 Future Work

There is still much work to be done to develop mature computational models for the UX of maps and geovisualizations more broadly. These computational models should cover more thematic maps—more tasks—more users—more

devices—more modalities—more design parameters. Once specific computational models for each dimension become available, there will be the question of combining them into a coherent framework and deriving, if possible, a general theory of UX of geovisualizations. Various types of thematic maps were already mentioned in Sect. 1; tasks relevant to thematic maps can be derived from the classifications provided in e.g. Roth (2013a) and Brehmer and Munzner (2013); a typical way of distinguishing users of interactive maps is through their abilities, expertise and/or motivation (see e.g. Roth 2013b; Degbelo 2022); additional devices could include desktop computers, interactive tabletops and surfaces, smartwatches or large displays; additional modalities could go beyond the visual to include the haptic, auditory, olfactory and gustatory modalities (Hogan 2018). Finally additional design parameters, in addition to those already mentioned in Sect. 5.3, could include the number of data classes, the number of alternative visualizations of the geographic dataset offered in a single interface (e.g. as a map, chart, data table), and the use or not of a dark mode.

6 Conclusion

This work has investigated how mathematical functions to describe the user experience of maps on mobile devices could look like, in an exploratory study. The outcomes of the analysis are *hypotheses* about the behaviour of map user experience as one changes colour or size distance. The work also provided a quantification of the expected errors for predictions using the mathematical functions on unseen data. Overall, we have learned that some aspects of user experience are likely linear monotonic (e.g. the relationship between colour distance and accuracy) while some are probably not (e.g. the relationship between size distance and accuracy). A replication of this study in other contexts is needed to confirm these observations.

7 Supplementary Material

All scripts used during the analysis are available on GitHub (https://github.com/SulaxanSo/Mapbox_UX). The supplementary material showing the tasks given to the participants and the error values on the test dataset is available at <https://doi.org/10.6084/m9.figshare.21908079>.

Acknowledgements Auriol Degbelo was funded by the German Research Foundation through the project NFDI4Earth (DFG project no. 460036893, <https://www.nfdi4earth.de/>) within the German National Research Data Infrastructure (NFDI, <https://www.nfdi.de/>). Jakub Kruckar has received funding from the programme “Profilbildung 2020”, an initiative of the Ministry of Culture and Science of the State of North

Rhine-Westphalia. The sole responsibility for the content of this publication lies with the authors.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data collected during the user study is available at https://github.com/SulaxanSo/Mapbox_UX.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Babyak MA (2004) What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med* 66(3):411–421. <https://doi.org/10.1097/01.psy.0000127692.23278.a9>
- Bailly G, Oulasvirta A, Brumby DP, et al (2014) Model of visual search and selection time in linear menus. In: Jones M, Palanque PA, Schmidt A et al (eds) CHI'14 conference on human factors in computing systems. ACM, Toronto, pp 3865–3874. <https://doi.org/10.1145/2556288.2557093>
- Bartling M, Robinson AC, Resch B et al (2021) The role of user context in the design of mobile map applications. *Cartogr Geogr Inf Sci* 48(5):432–448. <https://doi.org/10.1080/15230406.2021.1933595>
- Bertel S, Dressel T, Kohlberg T et al (2017) Spatial knowledge acquired from pedestrian urban navigation systems. In: Proceedings of the 19th international conference on human-computer interaction with mobile devices and services. ACM, New York, pp 32:1–32:6. <https://doi.org/10.1145/3098279.3098543>
- Bertin J (1983) *Semiology of graphics*. University of Wisconsin Press, Wisconsin
- Bessadok F, Dominguès C (2011) Automatic evaluation and improvement of map readability. In: 25th International cartographic conference (ICC2011), pp 2–8
- Brehmer M, Munzner T (2013) A multi-level typology of abstract visualization tasks. *IEEE Trans Vis Comput Graph* 19(12):2376–2385. <https://doi.org/10.1109/TVCG.2013.124>
- Brewer CA (1994) Color use guidelines for mapping. In: Maceachren AM, Taylor DRF (eds) *Visualization in modern cartography*, vol 1994. Pergamon, Oxford, pp 123–148. <https://doi.org/10.1016/B978-0-08-042415-6.50014-4>
- Brychtová A (2015a) Sequential color scheme generator 1.0. <http://eyetracking.upol.cz/color/>. Accessed 17 Jan 2022
- Brychtová A (2015b) Exploring the influence of colour distance and legend position on choropleth maps readability. In: *Modern trends in cartography*. Springer, Berlin, pp 303–314. https://doi.org/10.1007/978-3-319-07926-4_23
- Brychtová A, Coltekin A (2015) Discriminating classes of sequential and qualitative colour schemes. *Int J Cartogr* 1(1):62–78. <https://doi.org/10.1080/23729333.2015.1055643>
- Brychtová A, Coltekin A (2016) An empirical user study for measuring the influence of colour distance and font size in map reading using eye tracking. *Cartogr J* 53(3):202–212. <https://doi.org/10.1179/1743277414Y.0000000103>
- Brychtová A, Çöltekin A (2017a) Calculating colour distance on choropleth maps with sequential colours—a case study with ColorBrewer 2.0. *KN J Cartogr Geogr Inf* 67(2):53–60. <https://doi.org/10.1007/BF03545377>
- Brychtová A, Çöltekin A (2017b) The effect of spatial distance on the discriminability of colors in maps. *Cartogr Geogr Inf Sci* 44(3):229–245. <https://doi.org/10.1080/15230406.2016.1140074>
- Chen J, Yang S, Li H et al (2013) Research on geographical environment unit division based on the method of natural breaks (jenks). *Int Arch Photogramm Remote Sens Spat Inf Sci* 3:47–50. <https://doi.org/10.5194/isprsarchives-XL-4-W3-47-2013>
- Cockburn A, Gutwin C (2009) A predictive model of human performance with scrolling and hierarchical lists. *Hum Comput Interact* 24(3):273–314. <https://doi.org/10.1080/07370020902990402>
- Cockburn A, Gutwin C (2010) A model of novice and expert navigation performance in constrained-input interfaces. *ACM Trans Comput Hum Interact* 17(3):1–38. <https://doi.org/10.1145/1806923.1806927>
- Cockburn A, Gutwin C, Greenberg S (2007) A predictive model of menu performance. In: Rosson MB, Gilmore DJ (eds) *Proceedings of the 2007 conference on human factors in computing systems*. ACM, San Jose, pp 627–636. <https://doi.org/10.1145/1240624.1240723>
- Coutaz J (2010) User interface plasticity: model driven engineering to the limit! In: Sukaviriya N, Vanderdonckt J, Harrison M (eds) *Proceedings of the 2nd ACM SIGCHI symposium on engineering interactive computing systems—EICS '10*. ACM Press, Berlin, pp 1–8. <https://doi.org/10.1145/1822018.1822019>
- Cybulski P (2020) Spatial distance and cartographic background complexity in graduated point symbol map-reading task. *Cartogr Geogr Inf Sci* 47(3):244–260. <https://doi.org/10.1080/15230406.2019.1702102>
- Daszykowski M, Kaczmarek K, Vander Heyden Y et al (2007) Robust statistics in data analysis—a review: basic concepts. *Chemom Intell Lab Syst* 85(2):203–219. <https://doi.org/10.1016/j.chemo.2006.06.016>
- Degbelo A (2022) FAIR geovisualizations: definitions, challenges, and the road ahead. *Int J Geogr Inf Sci* 36(6):1059–1099. <https://doi.org/10.1080/13658816.2021.1983579>
- Degbelo A, Kray C (2018) Intelligent geovisualizations for open government data (vision paper). In: Banaei-Kashani F, Hoel EG, Güting RH et al (eds) 26th ACM SIGSPATIAL international conference on advances in geographic information systems. ACM Press, Seattle, pp 77–80. <https://doi.org/10.1145/3274895.3274940>
- Degbelo A, Somaskantharajan S (2020) Speech-based interaction for map editing on mobile devices: a scenario-based study. In: Alt F, Schneegass S, Hornecker E (eds) *Mensch und computer 2020*. ACM, Magdeburg, pp 343–347. <https://doi.org/10.1145/3404983.3409996>
- Degbelo A, Wissing J, Kauppinen T (2018) A comparison of geovisualizations and data tables for transparency enablement in the open

- government data landscape. *Int J Electron Gov Res* 14(4):39–64. <https://doi.org/10.4018/IJEGR.2018100104>
- Degbelo A, Sarfraz S, Kray C (2020) Data scale as cartography: a semi-automatic approach for thematic web map creation. *Cartogr Geogr Inf Sci* 47(2):153–170. <https://doi.org/10.1080/15230406.2019.1677176>
- Degbelo A, Schmidt B, Henzen C et al (2022) Themen, Trends und Visionen im Spannungsfeld Geo UX. *Kartographische Nachrichten—Fachberichte aus der Praxis* 72(2):4–13. <https://doi.org/10.13140/RG.2.2.32169.67688>
- Einfeldt L, Degbelo A (2021) User interface factors of mobile UX: a study with an incident reporting application. In: Paljic A, Peck T, Braz J et al (eds) *Proceedings of the 16th international joint conference on computer vision, imaging and computer graphics theory and applications (VISIGRAPP 2021)*, vol 2: HUCAPP. SCITEPRESS—Science and Technology Publications, Online, pp 245–254. <https://doi.org/10.5220/0010325302450254>
- Fish C, Goldsberry KP, Battersby S (2011) Change blindness in animated choropleth maps: an empirical study. *Cartogr Geogr Inf Sci* 38(4):350–362. <https://doi.org/10.1559/15230406384350>
- Garcia-Castellanos D, Lombardo U (2007) Poles of inaccessibility: a calculation algorithm for the remotest places on earth. *Scott Geogr J* 123(3):227–233. <https://doi.org/10.1080/14702540801897809>
- Garlandini S, Fabrikant SI (2009) Evaluating the effectiveness and efficiency of visual variables for geographic information visualization. In: Hornsby KS, Claramunt C, Denis M et al (eds) *Spatial information theory—COSIT 2009*. Springer, Berlin, pp 195–211. https://doi.org/10.1007/978-3-642-03832-7_12
- Golebiowska I, Korycka-Skorupa J, Slomska-Przech K (2021) Common thematic map types. In: Wilson JP (ed) *Geographic information science and technology body of knowledge*. <https://doi.org/10.22224/gistbok/2021.2.7>, issue: Q2, ISSN: 25772848
- Gong X (2019) SDG Viz: a web-based system for visualizing sustainable development indicators. *Proc ICA 2*:1–8. <https://doi.org/10.5194/ica-proc-2-39-2019>
- Gorte V, Degbelo A (2022) Choriented maps: visualizing SDG data on mobile devices. *Cartogr J* 59(1):35–54. <https://doi.org/10.1080/00087041.2021.1986616>
- Hardisty F, Robinson AC (2011) The GeoViz toolkit: using component-oriented coordination methods for geographic visualization and analysis. *Int J Geogr Inf Sci* 25(2):191–210. <https://doi.org/10.1080/13658810903214203>
- Harrie L (2009) Methods to Measure Map Readability. In: Mustière S, Sester M, van Harmelen F, et al (eds) *Generalization of spatial information, Dagstuhl Seminar Proceedings (DagSemProc)*, vol 9161. Schloss Dagstuhl—Leibniz-Zentrum für Informatik, Dagstuhl, pp 1–6. <https://doi.org/10.4230/DagSemProc.09161.5>
- Hart SG (2006) Nasa-task load index (nasa-tlx); 20 years later. In: *Proceedings of the human factors and ergonomics society annual meeting*. Sage Publications, Los Angeles, pp 904–908. <https://doi.org/10.1177/1541193120605000909>
- Hassenzahl M (2005) The thing and I: understanding the relationship between user and product. In: Blythe M, Overbeeke K, Monk A et al (eds) *Funology: from usability to enjoyment*. Kluwer Academic Publishers, pp 31–42. https://doi.org/10.1007/1-4020-2967-5_4
- Hogan T (2018) Data sensification: beyond representation modality, toward encoding data in experience. In: Storni C, Leahy K, McMahon M et al (eds) *Design as a catalyst for change—DRS international conference 2018, Limerick*. <https://doi.org/10.21606/drs.2018.238>
- Horbiński T, Cybulski P, Medyńska-Gulij B (2020) Graphic design and button placement for mobile map applications. *Cartogr J* 57(3):196–208. <https://doi.org/10.1080/00087041.2019.1631008>
- Kray C, Degbelo A (2019) Map plasticity. In: *Poster session at conference on spatial information theory: COSIT'19, Regensburg*. <https://doi.org/10.5281/zenodo.3385238>
- Kray C, Schmid F, Fritze H (2017) Guest editorial: map interaction. *GeoInformatica* 21(3):573–576. <https://doi.org/10.1007/s10707-016-0290-x>
- Li Y, Bengio S, Bailly G (2018) Predicting human performance in vertical menu selection using deep learning. In: Mandryk RL, Hancock M, Perry M et al (eds) *Proceedings of the 2018 CHI conference on human factors in computing systems*. ACM, Montreal, p 29. <https://doi.org/10.1145/3173574.3173603>
- Miniukovich A, De Angeli A (2014) Visual impressions of mobile app interfaces. In: Roto V, Häkkinä J, Väänänen-Vainio-Mattila K et al (eds) *Proceedings of NordiCHI '14*. ACM Press, Helsinki, pp 31–40. <https://doi.org/10.1145/2639189.2641219>
- Miniukovich A, Marchese M (2020) Relationship between visual complexity and aesthetics of webpages. In: Bernhaupt R, Mueller FF, Verweij D et al (eds) *Proceedings of CHI 2020*. ACM, Honolulu, pp 1–13. <https://doi.org/10.1145/3313831.3376602>
- Nosek BA, Ebersole CR, DeHaven AC et al (2018) The preregistration revolution. *Proc Natl Acad Sci* 115(11):2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Oulasvirta A (2016) Can computers design interaction? In: Luyten K, Palanque PA (eds) *Proceedings of the 8th ACM SIGCHI symposium on engineering interactive computing systems—EICS '16*. ACM Press, Brussels, pp 1–2. <https://doi.org/10.1145/2933242.2948131>
- Oulasvirta A (2019) It's time to rediscover HCI models. *Interactions* 26(4):52–56. <https://doi.org/10.1145/3330340>
- Oulasvirta A, Hornbæk K (2016) HCI research as problem-solving. In: Kaye J, Druin A, Lampe C et al (eds) *Proceedings of the 2016 CHI conference on human factors in computing systems—CHI '16*. ACM Press, San Jose, pp 4956–4967. <https://doi.org/10.1145/2858036.2858283>
- Oulasvirta A, Hornbæk K (2021) Counterfactual thinking: what theories do in design. *Int J Hum Comput Interact*, 1–15. <https://doi.org/10.1080/10447318.2021.1925436>
- Oulasvirta A, Dayama NR, Shiripour M et al (2020) Combinatorial optimization of graphical user interface designs. *Proc IEEE* 108(3):434–464. <https://doi.org/10.1109/JPROC.2020.2969687>
- Ramakrishnan R, Kaur A (2020) An empirical comparison of predictive models for web page performance. *Inf Softw Technol* 123(106):307. <https://doi.org/10.1016/j.infsof.2020.106307>
- Reinecke K, Yeh T, Miratrix L et al (2013) Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In: *CHI'13: proceedings of the SIGCHI conference on human factors in computing systems*. ACM, Paris, pp 2049–2058. <https://doi.org/10.1145/2470654.2481281>
- Roth RE (2013a) An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization. *IEEE Trans Vis Comput Graph* 19(12):2356–2365. <https://doi.org/10.1109/TVCG.2013.130>
- Roth RE (2013b) Interactive maps: what we know and what we need to know. *J Spat Inf Sci* 6:59–115. <https://doi.org/10.5311/JOSIS.2013.6.105>
- Roth RE (2017) Visual variables. In: Richardson D, Castree N, Goodchild MF et al (eds) *International encyclopedia of geography: people, the earth, environment and technology*. Wiley, Oxford, pp 1–11. <https://doi.org/10.1002/9781118786352.wbieg0761>
- Sarikaya A, Gleicher M, Szafir DA (2018) Design factors for summary visualization in visual analytics. *Comput Graph Forum* 37(3):145–156. <https://doi.org/10.1111/cgf.13408>

- Schiewe J (2019) Empirical studies on the visual perception of spatial patterns in choropleth maps. *KN J Cartogr Geogr Inf* 69(3):217–228. <https://doi.org/10.1007/s42489-019-00026-y>
- Sharma G, Wu W, Dalal EN (2005) The ciede2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color Res Appl* 30(1):21–30. <https://doi.org/10.1002/col.20070>
- Sharma PN, Shmueli G, Sarstedt M et al (2021) Prediction-oriented model selection in partial least squares path modeling. *Decis Sci* 52(3):567–607. <https://doi.org/10.1111/dec.12329>
- Słomska-Przech K, Gołębiowska IM (2021) Do different map types support map reading equally? comparing choropleth, graduated symbols, and isoline maps for map use tasks. *ISPRS Int J Geo-Inf* 10(2):69. <https://doi.org/10.3390/ijgi10020069>
- Steegen S, Tuerlinckx F, Gelman A et al (2016) Increasing transparency through a multiverse analysis. *Perspect Psychol Sci* 11(5):702–712. <https://doi.org/10.1177/1745691616658637>
- Thevenin D, Coutaz J (1999) Plasticity of user interfaces: framework and research agenda. In: Sasse MA, Johnson CW (eds) *Proceedings of INTERACT '99*. IOS Press, Edinburgh, pp 110–117
- Tuch AN, Bargas-Avila JA, Opwis K et al (2009) Visual complexity of websites: effects on users' experience, physiology, performance, and memory. *Int J Hum Comput Stud* 67(9):703–715. <https://doi.org/10.1016/j.ijhcs.2009.04.002>
- Van Ho Q, Lundblad P, Astrom T et al (2012) A web-enabled visualization toolkit for geovisual analytics. *Inf Vis* 11(1):22–42. <https://doi.org/10.1177/1473871611425870>
- Vandekerckhove J, Matzke D, Wagenmakers EJ (2015) Model comparison and the principle of parsimony. In: Busemeyer JR, Wang Z, Townsend JT et al (eds) *The Oxford Handbook of Computational and Mathematical Psychology*, vol 1. Oxford University Press, Oxford. <https://doi.org/10.1093/oxfordhb/9780199957996.013.14>
- Wade T, Sommer S et al (2006) *A to Z GIS. An illustrated dictionary of geographic information systems*. Esri Press, Redlands
- Yang S, Berdine G (2015) Model selection and model over-fitting. *Southwest Respir Crit Care Chron* 3(12):52–55. <https://doi.org/10.12746/swrccc2015.0312.160>