When Size Matters: Towards Evaluating Perceivability of Choropleths

Liam McNabb[1], Robert S. Laramee[1] and Max Wilson[2]

[1]Visual and Interactive Computing Group, Swansea University, Wales
[2] School of Computer Science, University of Nottinghamam, UK

Figure 1: *Presents a subset of administrative areas in London at 1.5% of the scale of the full map.*

Abstract
Choropleth maps are an invaluable visualization type for mapping geo-spatial data. One advantage to a choropleth map over other geospatial visualizations such as cartograms is the familiarity of a non-distorted landmass. However, this causes challenges when an area becomes too small in order to accurately perceive the underlying color. When does size matter in a choropleth map? We conduct an experiment to verify the relationship between choropleth maps, their underlying color map, and a user's perceivability. We do this by testing a user's perception of color relative to an administrative area's size within a choropleth map, as well as user-preference of fixed-locale maps with enforced minimum areas. Based on this initial experiment we can make the first recommendations with respect to a unit area's minimum size in order to be perceivably useful.

CCS Concepts
●Human-centered computing → User studies; Geographic visualization; Information visualization; ●Computing methodologies → Visibility; Perception;

1. Introduction and Motivation

Size is an important facet of user design in visualization. McNabb and Laramee find scalability as the 3rd most cited future work across 40 visualization survey papers, only being preceded by evaluative papers and missing scenarios in classifications [ML17]. In relation to choropleth maps, size is considered a widely recognized problem in the field. Ward et al. state "*A problem of choropleth maps is that the most interesting values are often concentrated in densely populated areas with small and barely visible polygons*" [WGK10]. Lee *et al.* suggest that improving tiny classes would be interesting future work in the field [LSS13].

We conduct a preliminary user-experiment to determine the minimum screen size of a unit area on a choropleth map in order to be accurately perceivable. We also measure the range of unit area sizes where an increase in perceivability error occurs. The experiment focuses on the scalability factors 'human perception' and 'monitor resolution', presented by Eick and Karr [EK02]. We base our user-study on the following hypotheses:

H1 The smaller the size of a unit area, the higher the error rate of perceiving the correct underlying color category.

H2 The smaller unit area, the more time required to perceive its color.

H3 The user will prefer choropleth maps with a larger minimum size over those with smaller sizes. Particularly, users will prefer a trade-off between area resolution in favor of legibility.

The contributions of this paper include:

1. The first user-study of its kind to evaluate perceivability of color on a choropleth map relative to unit area size.
2. The first recommendation on the minimum screen space size a unit area should occupy to maintain perceivability on a choropleth map.
3. The first recommendation on the minimum screen space size a unit area should be to optimize user satisfaction.

The rest of the paper is organized as follows. In Section 2, we explore work in the related fields of perception of choropleths, color, and size. Section 3 breaks down the tasks that we will cover in our user study. Section 4 describes the important variables that are considered in the user study, this is divided into dependent variables and controlled variables. In Section 5, we discuss the procedure of our user study. We provide depth for each task in order to allow the

study to be reproduced. In Section 5.3, we discuss the stored data from the experiments to give an understanding of the results. We then discuss the results of our three tasks in Section 6 in detail, and present our user-study findings. Finally, we give our conclusions.

2. Related Work

We first describe previous research involving user-studies and perception. McNabb and Laramee's Survey of Surveys [ML17], as well as Kijmongkolcahi *et al.* [KARC17] provide papers that aid in the search for related work in the areas of human-centered evaluation of choropleths, color, and perception of size.

Perception of Choropleths: Rittschof and Kulhavy present a user-study which includes a comparison of choropleth maps and cartograms. Cartograms are a different class of geo-visualization because they use distortion to convey data. They study the recall of data using different types of maps, firstly testing maps against the raw data and secondly studying two different map types (cartograms and choropleths) with two variations of exposure time. Their results indicate that choropleth maps are associated with greater recall of information [RK98]. Kasper reviews the effectiveness of Gastner-Newman diffusion cartograms [GN04] for the representation of population data which includes a comparative experiment against thematic maps (choropleth with overlaid circle maps) [KFF11]. The experiment is performed using two informatically equivalent maps using different designs, testing response times and accuracy for varied levels of question difficulty (using Bertin's map reading levels [Ber83]). The results report that the thematic maps are more efficient and effective, particularly with complex tasks. Sun and Li review the effectiveness of cartograms for the representation of spatial data which includes a comparative experiment against thematic maps, including choropleths [SL10]. The test takes place online for 100 subjects, who rank thematic maps (including choropleth maps) against cartograms. The results indicate that the thematic maps are more effective representing quantitative data whilst cartograms are more effective with qualitative data.

Perception of Color: We find two survey papers related to color mapping and perception in visualization. Zhou and Hansen present a survey of color maps and color-map generation techniques in visualization, providing a helpful reference for readers who are faced with color mapping decisions [ZH15]. Silva *et al.* present an overview of color, color scales, and tools to guide expert and non-expert users [SSM11]. The paper examines different domains of visualization individually.

Heer and Stone investigate how a model of color-naming can enable user interfaces to meaningfully mimic a link between visual perception and symbolic cognition. Color saliency is used to define the degree to which a color value is uniquely named. This is tested against a number of popular qualitative and quantitative color palettes to find the average salience which is calculated using the algorithms provided in the paper [HS12b]. Lee *et al.* present a color optimization algorithm for visibility measures over a range of color palettes [LSS13]. They look at classes and the area the class will envelope, and modify the color palette based on this. Smaller classes are given higher saturation and luminance, for example. Fang *et al.* provide an algorithmic approach for maximizing the perceptual distances among a set of colors, with a focus on color re-assignment,

which may occur when values change, and the phenomenon of local maxima, which becomes a problem when new colors need to be added [FWD*17]. Szafir presents quantitative studies measuring color difference perception for points, bars, and lines [Sza18]. The experiment for each of the three marks is conducted using Amazon's Mechanical Turk with items of varying size in terms of pixels, and recorded the distinguishability of the points, lines, and bars. The goal of the paper is a first step in discerning quantitative understanding of color perception in visualization. Our paper is closely related to this, with a focus on the size of perceivable areas in color, and also focuses on maps rather than mark types.

Perception of Size: Borgo *et al.* present a new order of magnitude mark (OOMM) and an empirical study comparing logarithmic, linear, scale-stack, and their own custom marks to test magnitude estimation, target identification, ratio estimation and trend analysis [BDJ14]. The results found the OOMM marks outperform state of the art techniques in quantitative analysis. Gramazio *et al.* present a study looking at the relationship between perceivable size, grouping, and search performance. The user is asked to find a distinct mark amongst a set of distractor marks which was applied to a grid layout as well as a scattered layout [GSL14]. The study finds that grouped, or single color layouts have a much faster response time, and visual marks below $0.508°$ showed a large drop in performance time. Jansen and Hornbæk provide a psychophysical investigation of size as a physical variable, where participants estimated size of objects between solid bars and spheres. Solid bars were found to be estimated based on their length whilst spheres were measures based on their surface area [JH16]. Ronne Jakobsen and Hornæk produce an experiment reviewing the use of overview+detail, focus+context, and zooming over 3 different display sizes. The results indicate that performance time for each technique was worse for smaller screens [RJH11]. Healey and Sawant conduct an experiment to review identified pixel resolution and visual angle needed to distinguish different values of luminance, hue, size, and orientation [HS12a]. Our study differs by testing on a set of non-uniform shapes, as would be found on many choropleth maps.
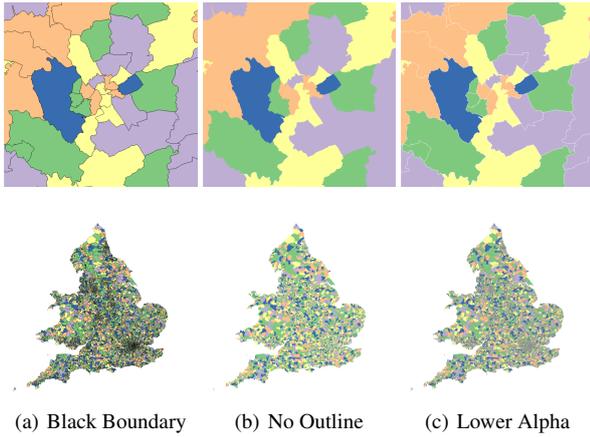
3. Experimental User Tasks

We research the hypotheses presented in Section 1 by testing singular interpretation, comparative interpretation, and comparative preference using the following user tasks:

- **T1** Identify the color of a given unit area on a choropleth map.
- **T2** Compare the relative color of two unit areas.
- **T3** Select the preferred minimum size unit area on a dynamic choropleth map using a slider.

Details about how the tasks are performed can be found in Section 5.2.

4. Experimental Variables

In this section, we discuss the user-study variables, and some of the factors we consider in the design of the experiment. These are split into the dependent variables which we test, and variables we control.

(a) Black Boundary   (b) No Outline   (c) Lower Alpha

Figure 2: *Three examples for boundary design at two different scales. The selected process is to use boundaries with a low alpha level.*

4.1. Dependent Variables

Our primary dependent variable is size of a unit-area on a choropleth map. Unit area size is an important attribute to study due to it's common use and variability in standard choropleth maps. We test variable unit-area sizes by looking at how small a unit area can be in order to perceive the color it encodes based on a given color-map. We measure size in terms of the number of color-mapped pixels in a given unit area. This measure is somewhat screen dependent since different displays feature different pixel densities, thus we use the same display throughout the experiments (see Figure 3 for pixel size demonstration). The hardware specification of the display is given in Section 5.1 . We also measure the size of each unit area as a percentage of screen space.
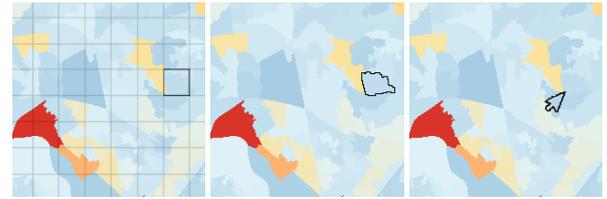
4.2. Control Variables

We discuss visual variables that we control in order to study the perception of unit-area size. Some user-study design decisions have to be made carefully.

Area Uniformity: The first thing we must decide is whether to test uniform shape areas, where all areas are the same shape during a given task, or non-uniform areas where areas have distinct shapes. For the experiment, we focus on non-uniformly shaped areas used in real choropleth maps. This decision is made to study and emulate real-world scenarios, and therefore we use real-world administrative areas for this experiment. We attempt to collect enough experimental test data to reduce the influence of unit area shape on the experimental findings.

Area Boundary Design: We must decide how differentiation

Figure 3: *Area Size approximation demonstration. (left) 1 pixel. (middle) 50 pixels. (right) 200 pixels*
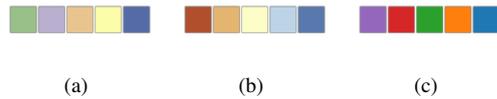

(a) Grid Concept   (b) Contour Highlight   (c) Cursor Concept

Figure 4: *Three concepts for communicating requested area. The selected process was the contour highlight*

between neighboring areas is implemented. A boundary line is useful for discerning where one area begins and another ends. Although this can be considered an important feature in a regular choropleth map, we must also consider the perceptual influence associated with different types of boundaries. First is border thickness. When the user examines smaller unit areas, the border can become relatively thick and obscure a unit area's color, causing unnecessary error increase. Three choropleth design options are (a) to include boundary lines, due to their popular use in choropleth maps, (b) remove the outline of boundaries to avoid obscuring any perceivable color, or (c) add boundary lines at a reduced opacity with the goal of reducing their influence. Refer to Figure 2. We must also consider boundary color an important variable when designing boundaries. We use a black line with an alpha value of 20 (7.5%), to keep a standard profile, using Figure 2(c) as our boundary design. We use a boundary thickness of one pixel.

Communicating the target unit area to the user: One of the most important factors of the experiment is to verify whether the user can accurately interpret the underlying color of an area of screen space. In order to do this, we need to accurately communicate the target area itself without obscuring or enhancing the area's color. This becomes non-trivial when perceiving an area of screen space at less than 1%. We considered the use of a grid to quickly convey the target area. Although the grid could be effective with uniform size areas, the grid would become less accurate with non-uniform size areas. Grid granularity would also need to be taken into account and conditions in which cell labeling would become harder to communicate. Using the mouse cursor to communicate the area to the user may allow the user to intuitively see the target area. The disadvantage of this is it could cause a wider variance in performance time, and the mouse pointer itself could cause occlusion. A super-imposed cursor could be added to the image to control mouse movement, however this could still result in scenarios with multiple small areas close together being difficult to diagnose. Refer to Figure 4 for examples.

We use Robinson's design criteria [Rob11] in order to select an identification criteria that does not alter or accentuate the value itself. We use style reduction highlighting using the lower alpha boundaries discussed in the area boundary design, and give contour highlighting to outline the communicated area using a thicker black outline. We use a pixel thickness of 4 pixels to aid the use for smaller areas in order to reduce the effects of framing areas. On piloting the identification criteria, this alone was deemed to be

(a)   (b)   (c)

Figure 5: *3 options for color mapping. (a) Colorbrewer's 5 color palette (qualitative). (b) Colorbrewer's 5 color palette (diverging). (c) Tableau 10 color palette (reduced to 5 for consistency). We use (a) for this study [Bre17, Tab18].*

insufficient causing users to overlook small areas. On top of this criteria, we added a blinking effect to help the user locate the target area, and render the highlighted color to something that stands out over the existing boundary. In this case we use a black boundary which stands out from our color palette. The black outline blinks at 1Hz for a period of 1 second. In order to further aid the search for the highlighted area, we place a grey circle around the target area, where the target area sits in the center of the circle, and fills 10% of the screen, and blinks with the black contour. See Figure 6 (a).

Choice of Color Map: Selecting the appropriate color map is an important decision when testing perception (refer to Section 2). We use colorbrewer [Bre17] due to it's far-reaching application across geography and cartography, and acceptance with basic and advanced users [Bre03]. Colorbrewer is also created by geospatial visualization experts. We would like to ensure perceptual distances and therefore avoid the use of a sequential color map focused on a single hue. We can incorporate qualitative and diverging color maps for the color palette. Using Heer and Stone's color saliency test, we can see that qualitative color palettes tend to have better saliency, and can therefore use this as a guideline [HS12b]. Although Tablaeu's color palette provides marginally better saliency, we use Colorbrewer's qualitative color map [Bre17, Tab18]. See Figure 5.

Display Screen Resolution: Keeping resolution consistent is essential to experimenting with perception. Both digital resolution (display resolution) and physical resolution (monitor size) need to be consistent. In order to ensure this, we must forgo a mechanical turk testing environment and work in a controlled lab test environment. This is further explained in Section 5.1.

Lighting Conditions of Environment: We must consider the lighting as an important control variable when conducting this experiment. We make sure that the lighting environment stays the same by using a static study location, that uses interior lighting as the only lighting source. We also must consider the monitor's lighting settings, we do this by calibrating the screen using an International Color Consortium (ICC) profile. We provide a sample of presented question designs for T1 and T2 in Figure 6.
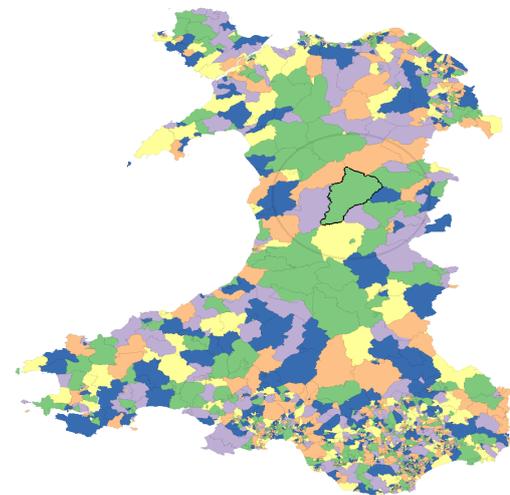
5. Perceivability Experiment

In this section, we discuss the nature of the experiment including our designated equipment, experiment design, and stored experiment data.
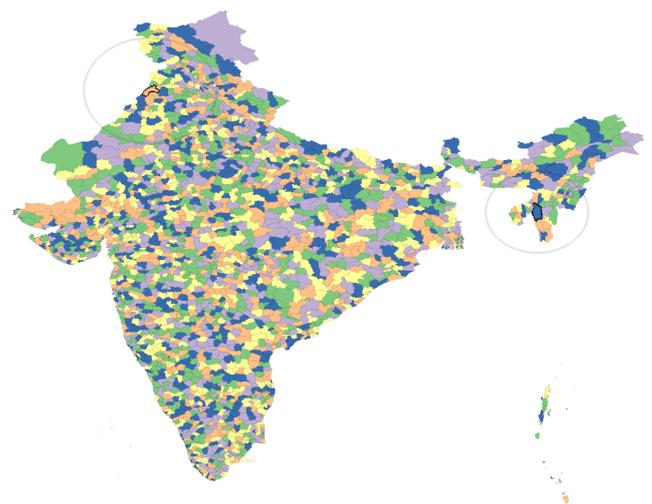
5.1. Equipment

For this process, we use a standard Dell U2412M display resolution of 1920x1080p, on a screen size of 21". We then display the choropleth images to the subject as a static image 1200x960p. The desktop used to test this implementation features an Intel i5-4460 at 3.2GHz with 16GB of RAM and a GeForce GTX960. The test is run using the Qt Framework [The18]. For the ICC color pofile, we use the Dell U2412M (Custom) color profile with suggested settings [TFT18].

5.2. Procedure

The experiment requires approximately 30 minutes per participant. We solicit 10 participants who individually perform the tasks in this study. The subjects were gathered using an advertising campaign.

(a) T1 Sample

(b) T2 Sample

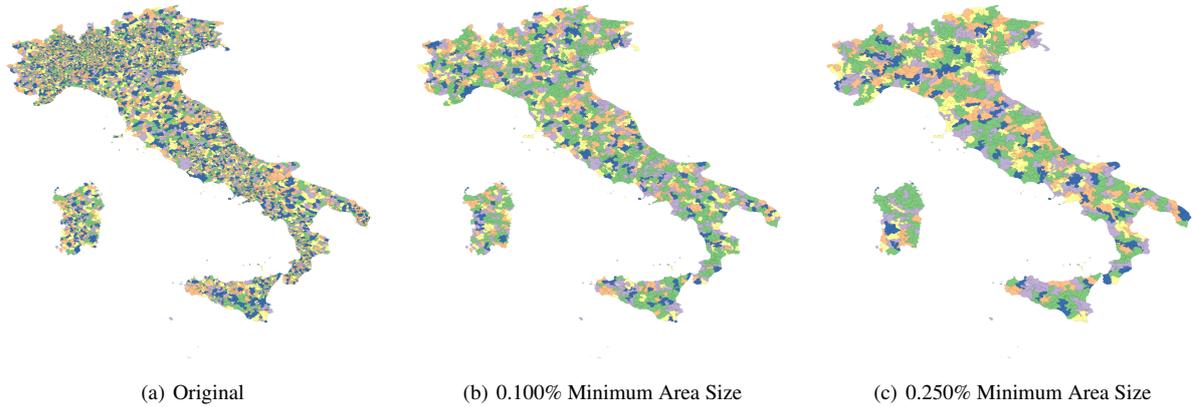**Figure 6:** *Sample questions for T1 and T2.*

(a) Original   (b) 0.100% Minimum Area Size   (c) 0.250% Minimum Area Size

Figure 7: *Sample map designs using McNabb et al's Dynamic Choropleth Map procedure [MLF18] for T3.*

The experiment is designed to test perception of color conveyed by areas with a small screen space. An instructive introduction explains what the test entails, what a choropleth is, how a task is completed, and sample questions. Some sample questions are used to test for basic understanding of the tasks as well as a test for color blindness or eyesight problems. The sample questions provide a task where the largest area is used as the target. Test subjects who fail the training questions do not have their results used. For T1 a standard question presents a choropleth map image mapped by a color palette. There is one highlighted area in which the subject diagnoses the conveyed color. Each section of the color palette is presented as an option for the subject to select. See the Supplementary Material. The subject's color-choice answer is saved to a database for further analysis along with the true color of the area. These are saved as well as the performance time, choropleth identifier, area identifier, and the area's size in terms of pixels and screen space percentage. The experiment is run in task sets of 10, where each set provides a set of random choropleth maps and random unit area targets with a screen space ranging from 1 pixel upwards. When the set is complete we produce a short transitional state where we ask the participant for a confidence rating, and ease of understanding rating. See Figure the Supplementary Material. Each task is associated with a random choropleth, random color from the color map, and randomized choice of unit area on the choropleth in order to prevent learning effects. We run this for 5 sets.

T2 is presented indicating two unit areas on the same choropleth map. The user is asked to diagnose whether the units are the same color, or differ. After selecting their answer, the answer is saved as well as the the same set of information in T1, for each area. This is run in sets of 5, with 10 tests per set. See Figure 6 for sample images.

T3 is presented by showing a dynamic map presenting the same data using the Dynamic Choropleth Map procedure [MLF18] by McNabb *et al*. We let the user dynamically select their preferred minimum screen space. The map for each position on the slider is split to enable options for the default map, moving up in increments of 0.001% up to 0.01%, and an increment of 0.01% min-

imum screen space up to 0.25%. The user will simply select the preferred size using the slider and a new map will replace it. This is done for 5 different comparisons, with the selected options being stored for further analysis. Refer to Figure 7. These maps will be given sequentially as we want to gather data for each of these maps, where the slider of each represents different constrained minimum areas.

5.3. Stored Experimental Data

Refer to Table 1 for a visual breakdown of the stored experimental data. Data stored per task in T1 for analysis includes: map identifier, target area identifier, user-selected area color, target area color, task time, task identifier, and set identifier. T2 saves all of the T1's information, but with the additional data for the second target area. For both T1 and T2, we store the set identifier, confidence level, and difficulty level per set. For T3 we store: task identifier, task time, and preferred screen space identifier. T1 does not store any information for target two as the task only focuses on one target area.
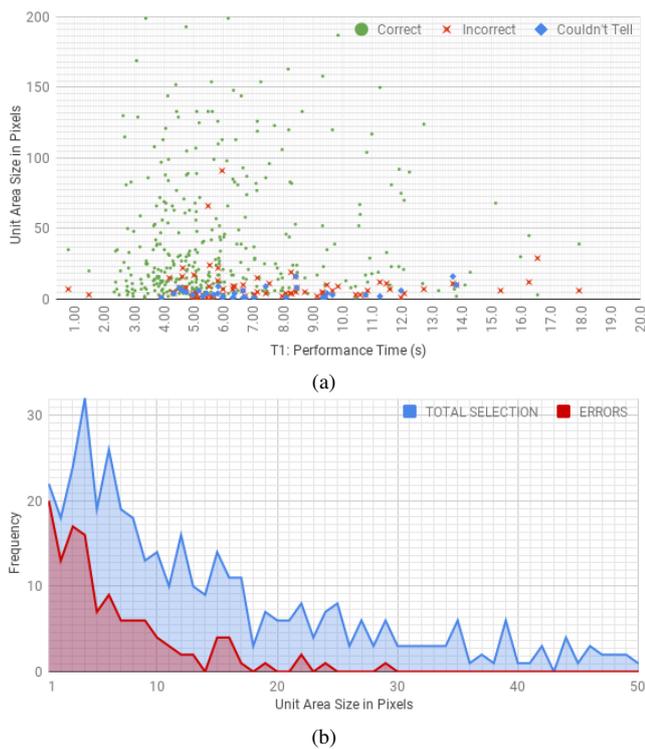
Stored Data:	T1	T2	T3
User ID	✓	✓	✓
Set ID	✓	✓	✗
Test ID	✓	✓	✗
Performance Time	✓	✓	✓
Map ID	✓	✓	✓
Target ID	✓	✓x2	✓
Screen space for Target	✓	✓x2	✗
Pixels for Target	✓	✓x2	✗
User Selected Color	✓	✓	✗
Target Color	✓	✓	✗
Set Confidence	✓	✓	✗
Set Ease	✓	✓	✗
Slider Value	✗	✗	✓

Table 1: *Table indicating the data saved for each task. Refer to Section 5.3.*

T2 includes the most data saved. T3 does not save any information for sets, screen space of targets (as targets already refer to screen space), pixels, or the target answer, as this is used to test preference. Therefore we save the value selected using a the interactive slider. See Table 1.

6. User-Study Results

In this section we discuss the results of three tasks which include diagnosing the color value of areas on a choropleth map, comparing the color of areas on a choropleth map, and determining user preference for choropleth maps with respect to minimum area size. The task included 10 participants with 50 data samples per participant for both T1 and T2, and 5 data samples per participant for T3 (from a selection of 130 potential candidates). Of the participants, we use an unconstrained sample of users from both genders, a range of ages (18-35), at different expertise levels. Candidates were provided with £10 amazon vouchers after the completion of the study's completion.

(a)

(b)

Figure 8: *T1: (a) A scatter plot depicting the correlation between performance time (seconds) and area size (in pixels). Color is mapped to Answer Category (green for correct answer, red for incorrect answer, blue for 'couldn't tell'). (b) Area chart presenting the T1 error rate based on area size measured in pixels, where the blue area represents the total samples per area size, and the red area represents the error per sample.*
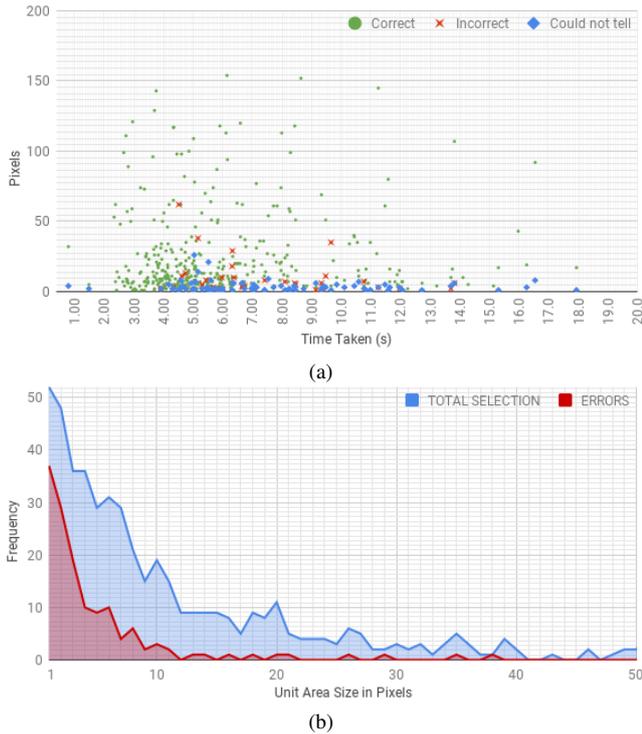
6.1. T1 Analysis

Task 1 involved the user determining the color of highlighted areas. Each user completed the task with 50 data samples, excluding outliers for a total of 500 data samples across all users. Of these 500, we use 465 of these for analysis, filtering out areas that are over 200 pixels in size, or outliers in terms of performance time (see Section 8). For T1 we present our findings using a scatter plot and area chart (Figure 8). The unit plot displays the relationship between the performance time, unit area size, and user results. We can see a large distribution of results where the user correctly interpreted the color mapped to the area. For incorrect results we see that the results primarily fall very close to 0. On top of this, we identify when a user selects the 'couldn't tell' option. This seems to fall along the lower half of the incorrect distribution. In terms of performance time, there is a clear gap between 0-2 seconds. This seems to be related to user search time, and the highlight animation. This is followed by another 2 seconds where there is a noticeable gap in incorrect answers. This might be influenced by pre-attentive processing . The area chart presents the accuracy of T1 error based on unit area. We also provide a histogram to present the frequency of size representation in the study. For T1, we see clearly that accuracy is particularly low below 5 pixels in size. With a range of error just over 90% for 1 pixel. We see from this chart that the error becomes prevalent under 10 pixels.

Statistical Analysis: We use our results of T1 in our analysis of H1 (size vs error) and H2 (size vs time) using the Pearson correlation co-efficient where $\alpha = 0.05$. As our sample size favors smaller pixels due to the random nature, we use normalized error rate to calculate the relationship between area size and error. We conclude using Pearson's correlation co-efficient that there is a relationship between area size and error rate, where $r(50) = -0.77$. This supports our hypothesis H1. For H2, we conclude there is a relationship between area size and performance time, where $r(466) = -0.325$. This supports our hypothesis H2. Due to this we can verify that our first two hypotheses are supported statistically by our results for T1.

6.2. T2 Analysis

Task 2 involves comparing 2 highlighted areas in order to diagnose whether they are mapped to the same color value. Each user completed the task with 50 data samples for a total of 500 data samples across all users. Of these 500 we use 490 of these for analysis, filtering out areas that are over 200 pixels in size, or outliers in terms of performance time (see Section 8). We produce similar graphs to facilitate comparison between T1 and T2 (see Figure 9). Similarly to T1, correct answers exhibit a wide distribution of pixels and performance time, however we find that users were far more liberal with the use of of the 'cannot tell' option. This suggests that the binary comparison was more difficult than the selection of colors, confirming the assumption that color salience may have been an important factor to consider.
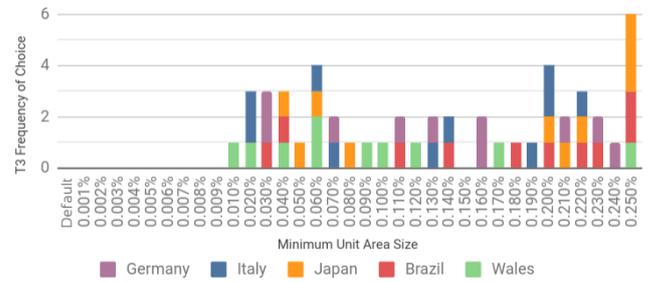
In terms of accuracy, we can see that the error rate for T2 was less extreme for small area sizes, but has a slightly higher distribution which leads to error up to 39 pixels (although 38 pixels has only one test case, 35 pixels has a better sample size). The lower error rate at smaller sizes could be related to a couple of visual features. First, as there are 5 map-able colors, when diagnosing colors

(a)

(b)

Figure 9: *T2: (a) Scatter Plot depicting the correlation between performance time (seconds) and area size (pixels) of the smallest target area. Color is mapped to user performance (green for correct answer, red for incorrect answer, blue for could not tell). (b) Area chart presenting the T2 error rate based on area size measured in pixels, where the blue area represents the total samples per area size, and the red area represents the error per sample.*

in T2, a decision can be made by testing the contrary. A user may not be able to distinguish a color in T1, however if they can distinguish that the two colors are not the same without distinguishing the correct color, they can effectively answer the question. We must also consider that although we provide and encourage the use of the 'couldn't tell' button, some users compared this to a test and therefore ignored the button. For T1, guessing a correct answer has a 1/5 chance of being correct, however, selecting the 'no' option has a 4/5 chance of being a correct answer in T2.

Statistical Analysis: We use our results of T2 in our analysis of H1 (size vs error) and H2 (size vs time) using the Pearson correlation co-efficient where $\alpha = 0.05$, similarly to our analysis of T1. We follow the same convention to analyse error as in T1, by using normalized error. From this analysis we conclude using Pearson's correlation co-efficient that there is a relationship between area size and error rate, where $r(50) = -0.433$. This supports our hypothesis for H1. For H2, we conclude there is a relationship between area size and performance time, where $r(490) = -0.358$. This also supports our hypothesis H2. As T1 and T2 both support the hypotheses of H1 and H2, we can be confident that our hypotheses for each are evident.

Figure 10: *T3: Histogram presenting the frequency of minimu unit area size preference, where the preference is defined as 'your preference of area size against the data relayed, where larger areas provide more uncertain data'.*

6.3. T3 Analysis

For T3, we allow the user to select the most "useful" map from a set of fixed-locale choropleths which constrain areas size, using the Dynamic Choropleth Map procedure by McNabb *et al* [MLF18]. We defined "useful" as being 'your preference of area size against the data relayed, where larger areas provide more uncertain data'. A histogram is used to visualize the frequency of selection for T3 and the results of this task are quite unexpected (see Figure 10). We expected that the preference would be found at the lower end of the spectrum, however over half of the study participants selected the largest possible enforced minimum screen space. On top of this, our pilot study prompted us to increase the number of selectable areas closer to the default screen space which did not reflect any results in the study. This may have been caused by the discrepancy in size variation between the first 10 selections, and the following selections. In order to rectify this, a follow-up study may be necessary that shows screenspace on a logarithmic scale to remove that type of bias, or a larger selection of choices, that are quantized into bins. Unlike H1 and H2, we cannot effectively run statistical analysis on our finding due to the subjective nature of the task. However, our observations do show evidence to support hypothesis H3. A more thorough investigation into this task is a good area for future work.

6.4. Anecdotal Results

During the period of testing, there were a few interesting observations we noticed. We found that users were generally caught off guard by the difficulty of the test, although the random selection of areas was mentioned, some users believed that this was not the case, leading to the conclusion that many are unaware of the frequency of small areas. To continue with this, users tended to feel like they had performed poorly, making statements remarking the difficulty. Some users regarded surprise at the difficulty of T2, suggesting that they believed it would be easier to identify whether two colors would be the same, but found themselves in situations where it was difficult to identify either color. For T3, one user explained that his decision for larger areas was based on the fact that they prefer islands to consist of only one area, which suggest the users' preferred minimum screen space could be related to area contiguity.

7. Supplementary Material

We include supplementary material that provide high resolution images of the scatter plots included. We also provide a video (https://bit.ly/2M9wIvY) previewing a demonstration of the user-study in action.

8. Limitations & Future Work

There are some limitations to the study we review. The first we must consider is some error in performance time. Although we record performance time as variable. We concluded that users would not be tasked to complete questions under a time limit. This was due to the assumption that this may cause unnecessary strain on the user which would lead to higher error. Thus there is a larger distribution in task time than anticipated caused by human factors including participants: re-adjusting posture, asking questions, or making comments. We found boundary design to be a bigger factor for small sizes than anticipated. In the pilot study, we found that users had a much better accuracy overall, and comments were made about the area's color changing. This was actually due to the boundary's framing of the pixel, enhancing the perception of color based on the surrounding black contour. We try to control this as closely as possible, however this could be examined further in a follow-up study. Finally, as we use real-world maps and areas, we do not control the aspect ratio of the areas, which may lead to more error in some situations such as narrow areas. This is an aspect that can be explored in an alternative study. Future work also includes a more in-depth statistical analysis of T1 and T2.

9. Conclusion

We conduct a preliminary study to evaluate perceptual evaluation of performance error, performance time, and preference of area's relative to size in choropleth maps. We hypothesize that the smaller the size of a unit area, the higher the error rate of perceiving the correct underlying color category; the smaller the unit area, the more time required to perceive its color; and the user will prefer choropleth maps with a larger minimum size over those with smaller sizes. Particularly, users will prefer a trade-off between area resolution in favor of legibility. For these three statements, we support the first two using Pearson's r to show a relationship between the variables, with strong observations for the third that give an initial understanding of how users' perceive the relationship between area size and data on a choropleth map. We give a recommendation of at least 10 pixels displayed for each area to be accurately interpreted.

10. Acknowledgements

References

[BDJ14] BORGO R., DEARDEN J., JONES M. W.: Order of magnitude markers: An empirical study on large magnitude number detection. *IEEE transactions on visualization and computer graphics 20*, 12 (2014), 2261–2270. 2

[Ber83] BERTIN J.: *Semiology of graphics: diagrams, networks, maps*. University of Wisconsin press, 1983. 2

[Bre03] BREWER C. A.: A transition in improving maps: The color-brewer example. *Cartography and Geographic Information Science 30*, 2 (2003), 159–162. 4

[Bre17] BREWER C. A.: Colorbrewer, 2017. accessed 2017/08/10. 4

[EK02] EICK S. G., KARR A. F.: Visual scalability. *Journal of Computational and Graphical Statistics 11*, 1 (2002), 22–43. 1

[FWD*17] FANG H., WALTON S., DELAHAYE E., HARRIS J., STORCHAK D., CHEN M.: Categorical colormap optimization with visualization case studies. *IEEE transactions on visualization and computer graphics 23*, 1 (2017), 871–880. 2

[GN04] GASTNER M. T., NEWMAN M. E.: Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences of the United States of America 101*, 20 (2004), 7499–7504. 2

[GSL14] GRAMAZIO C. C., SCHLOSS K. B., LAIDLAW D. H.: The relation between visualization size, grouping, and user performance. *IEEE transactions on visualization and computer graphics 20*, 12 (2014), 1953–1962. 2

[HS12a] HEALEY C. G., SAWANT A. P.: On the limits of resolution and visual angle in visualization. *ACM Transactions on Applied Perception (TAP) 9*, 4 (Oct. 2012), 20–41. 2

[HS12b] HEER J., STONE M.: Color naming models for color selection, image editing and palette design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, ACM, pp. 1007–1016. 2, 4

[JH16] JANSEN Y., HORNBÆK K.: A psychophysical investigation of size as a physical variable. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 479–488. 2

[KARC17] KIJMONGKOLCHAI N., ABDUL-RAHMAN A., CHEN M.: Empirically measuring soft knowledge in visualization. *Computer Graphics Forum 36*, 3 (2017), 73–85. 2

[KFF11] KASPAR S., FABRIKANT S., FRECKMANN P.: Empirical study of cartograms. In *25th International Cartographic Conference* (2011), vol. 3, p. 5. 2

[LSS13] LEE S., SIPS M., SEIDEL H.-P.: Perceptually driven visibility optimization for categorical data visualization. *IEEE Transactions on visualization and computer graphics 19*, 10 (2013), 1746–1757. 1, 2

[ML17] MCNABB L., LARAMEE R. S.: Survey of surveys (sos)-mapping the landscape of survey papers in information visualization. In *Computer Graphics Forum* (2017), vol. 36, Wiley Online Library, pp. 589–617. 1, 2

[MLF18] MCNABB L., LARAMEE R. S., FRY R.: Dynamic choropleth maps - using amalgamation to increase area perceivability. In *The 22nd International Conference on Information Visualization (IV)* (Salerno, Italy, July 2018). 5, 7

[RJH11] RØNNE JAKOBSEN M., HORNBÆK K.: Sizing up visualizations: effects of display size in focus+ context, overview+ detail, and zooming interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 1451–1460. 2

[RK98] RITTSCHOF K. A., KULHAVY R. W.: Learning and remembering from thematic maps of familiar regions. *Educational Technology Research and Development 46*, 1 (Mar 1998), 19–38. 2

[Rob11] ROBINSON A. C.: Highlighting in geovisualization. *Cartography and Geographic Information Science 38*, 4 (2011), 373–383. 3

[SL10]   SUN H., LI Z.: Effectiveness of cartogram for the representation of spatial data. *The Cartographic Journal 47*, 1 (2010), 12–21. 2

[SSM11]   SILVA S., SANTOS B. S., MADEIRA J.: Using color in visualization: A survey. *Computers & Graphics 35*, 2 (2011), 320 – 333. Virtual Reality in Brazil Visual Computing in Biology and Medicine Semantic 3D media and content Cultural Heritage. 2

[Sza18]   SZAFIR D. A.: Modeling color difference for visualization design. *IEEE transactions on visualization and computer graphics 24*, 1 (2018), 392–401. 2

[Tab18]   TABLEAU SOFTWARE: Tableau, 2018. accessed 2018/01/31. 4

[TFT18]   TFT CENTRAL: Icc profiles and monitor settings database, 2018. accessed 2018/03/24. 4

[The18]   THE QT COMPANY: *Qt Framework*. The Qt Company, 2018. 4

[WGK10]   WARD M. O., GRINSTEIN G., KEIM D.: *Interactive data visualization: foundations, techniques, and applications*. CRC Press, 2010. 1

[ZH15]   ZHOU L., HANSEN C.: A survey of colormaps in visualization. *IEEE Transactions on Visualization and Computer Graphics 22*, 8 (2015). 2