Evaluating the Influence of Surprise and Suppression Techniques in Map Visualizations

by

Akim Ndlovu

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Computer Science

by

September 2023

APPROVED:

Professor Lane T. Harrison, Thesis Advisor

Professor Neil Heffernan, Thesis Reader

Abstract

Choropleth maps are commonly used to visualize geospatial data, such as disease outbreaks across various geographic regions. However, well-known biases associated with choropleth maps, such as the effect of area in map exploration tasks and population statistics in the interpretation of event rates, have led to extensive research on how to overcome such biases to avoid misleading users. Two recently developed techniques, Surprise and VSUPs (Value Suppressing Uncertainty Palettes) may be considered as viable solutions for overcoming biases in choropleth maps, but have yet to be empirically tested with users of visualizations. In this thesis, we explore how well people make use of Surprise and VSUPs in map exploration tasks, by conducting a crowdsourced experiment where $n = 300$ participants are assigned to one of Choropleth, Surprise (only), and VSUP conditions (depicting rates and Surprise in a suppressed palette). We improve participants' exploration of each stimulus through the use of various interaction techniques (*e.g.* zooming and panning), and adapt tasks from prior studies to reduce noise from participants' responses. Quantitative analysis shows clear differences in the interpretation of metrics such as rate, surprise and population, with surprise maps leading people to map locations with significantly high population and VSUPs performing similar or better than Choropleths for rate selection. In addition, qualitative analysis suggests that many participants may only consider a subset of the metrics presented to them during map exploration and decision-making. We discuss how these results generally support the use of Surprise and VSUP techniques in practice, and opportunities for further technique development. For replicability and reproducibility, the material for the study (data, study results and code) is publicly available at `https://osf.io/exb95/`.

Acknowledgments

I would like to express my gratitude to my advisor, Prof. Lane Harrison, who has afforded me the opportunity to pursue a dream, continuously believed in me during difficult times, and for his patience and support throughout the process. Many thanks to my co-author Hilson Shrestha for his contribution to the study. I also thank my reader, Prof. Neil Heffernan, for his constructive feedback. Finally, my unwavering gratitude to my friends and family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Choropleth maps are widely used for visualizing trends in geo-spatial data, such as high or low performing regions and regions that show a high degree of correlation or disparity [44, 43]. For example, the vast amount of data gathered during pandemic outbreaks such as Covid-19, has renewed research efforts on the need to visualize and accurately communicate trends for vaccinations, deaths and infections [29, 31].

However, visualizing actual rates limits the trends and conclusions that may be drawn from a choropleth map [19]. Therefore, prior research, has explored techniques that use uncertainty metrics to highlight unexpected values, to counteract biases associated with visualizing data that closely resembles a population distribution. For example, using a choropleth map to visualize the percentage rates of counties or regions with low population and high variance, may result in them being shaded using darker colors, which may be misleading for map readers (see Figure 1.1).

A number of techniques have been proposed to counteract biases in Choropleth maps. These techniques can be broadly classified into two categories, Statistical (*e.g.* Bayesian weighting, weighted regression) and Design (*e.g.* color, legends and scales, map type (morphed, 3D, dot and heat maps)). Statistical techniques lead to the modification or supplementation of a dataset [19], whilst design techniques lead to the implementation of various well researched design considerations that reduce biases in choropleth maps.

The implementation of statistical techniques that result in the generation of a supplemented dataset, may increase the complexity required to understand the information conveyed by the visualization. For example, considering various metrics in a visualization may result in a cognitive overload for some users, whilst providing experts with essential information to make complex queries. Modified datasets on the other hand, have a tendency of losing meaning especially if the metric used to replace the actual value is not well researched or known. For example, smoothed data generated from using a non parametric method such as Kernel Density Estimation, may result in the loss of detail, blurring of boundaries as well as the mis-interpretation of extremes [26].

The use of techniques that may result in a loss of detail about the actual data, has influenced the need to understand the impact of using different metrics to offset biases in map visualizations [40]. For example, the use of Bayesian weighting [19] results in the visualization of an uncertainty metric that measures disbelief about the actual data (see Figure 2.2). Such statistical techniques, may result in misleading visualizations, therefore, it is essential to evaluate their impact on users' perceptions. Hullman [30] suggests that the representation of uncertainty in visualizations, requires people to understand the metric for effective use. However, Correll [10], hypothesizes that metrics similar to "Surprise" are more suited for people with expert statistical knowledge. Whilst the claim by Correll may be generalized across various uncertainty metrics, such claims may limit the intrinsic understanding of these metrics as well as their influence on peoples' takeaways. In addition, such generalizations suggest that uncertainty metrics are only suitable for expert users, which may lead to diverging

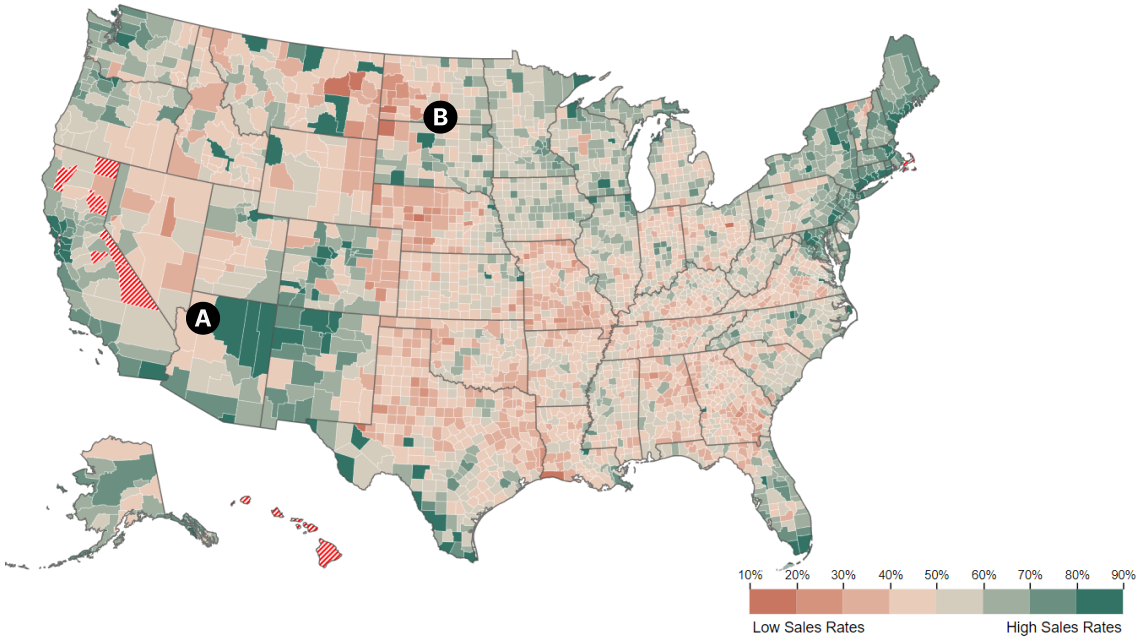

Figure 1.1: Choropleth map showing biases asssociated with population distribution when visualizing Covid-19 vaccination rates in the USA. Both Coconino county (A) and Dewey county (B) are shaded dark green to show high vaccination rates. However, Coconino county (A) has a population of 143476 with a vaccination rate of 80% and Dewey county (B) has a population of 5892 with a vaccination rate of 82%, which may create misleading visual cues for users of the map.

opinions about the information conveyed by a visualization. Therefore, we base our work on design considerations for the work proposed by Correll [19, 20], in order to make such techniques, more accessible for use by non-expert people.

Researchers have examined some model-driven mapping techniques by designing information retrieval, comparison, ranking and aggregation tasks, to understand their impact on pattern recognition and decision making [41, 12, 15, 6]. Other approaches for evaluating such maps include using empirically derived frameworks similar to the one proposed by Roth [48]. Although widely applicable to map evaluation studies, such frameworks may need to be extended when uncertainty is added as a consideration [30]. In addition to task design, MacEachren [40] suggests that various

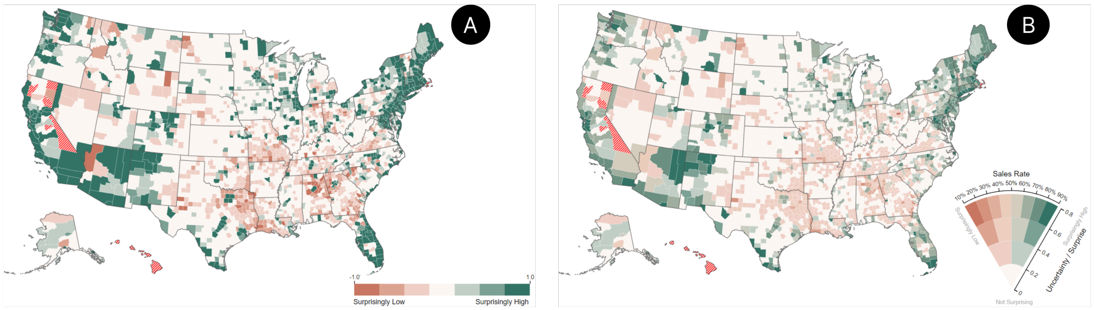

Figure 1.2: Recently developed techniques that may be used to counteract biases in Choropleth maps. The Surprise map (A), encodes surprise values whereas the (B) Value Suppressing Uncertainty Palette (VSUP) map encodes both event rates and surprise values.

visual arrangements such as the use of bi-variate maps, map pairs and sequential representation, may impact users' interpretation of uncertainty in map visualizations.

Two recently developed techniques provide a promising baseline for investigating map debiasing techniques in user studies. Correll and Heer propose the use of "Surprise", a Bayesian weighting technique that offsets biases in map visualizations [19]. Surprise up-weights or down-weights data points that deviate from expected values, by calculating an updated belief about the data based on prior knowledge. Correll [20] also introduce the Value-Suppressing Uncertainty Palette, a map coloring and legend technique which can visualize both uncertainty measures (such as Surprise) and rates on a single map. Although plausible, using statistical techniques such as "Surprise" to overcome biases in choropleth maps, may lead to the development of untrustworthy visualizations [29].

In this thesis, we report a crowdsourced study, where we ask $n = 300$ participants to perform map analysis tasks with one of three visualization conditions: Choropleth, Surprise and Value Suppressing Uncertainty Palettes (VSUPs). Prior to conducting the main study, we conduct multiple pilot studies to refine experiment tasks and design consideration. We improve participants' user experience by using multiple interaction techniques for each stimulus (*e.g.* zooming, panning and tooltips). In addition, we adapt prior map tasks and task taxonomies to study techniques which emphasize different metrics (Surprise, rates, or both). In particular, we leverage previous research by Roth [48] and Besançon et al. [11] to design a universal task for all map conditions, by summarizing map analyses tasks and objectives. Finally, we develop these tasks into a sales and marketing decision making problem to reduce noisy data (see Table 3.2). Quantitative analysis shows clear differences in map analysis outcomes (Figure 4.3), while qualitative analysis suggests that participants in some cases only consider a subset of the metrics available to them. We discuss how these results tentatively support the use of Surprise and VSUP techniques for broader visualization viewing populations, while also highlighting challenges that might be addressed through future design and technique development.

1.1 Research Objective

In this study, we examine the strategies that both expert and non-expert users employ when exploring and generating takeaways from Choropleth, Surprise and VSUP maps.

1.1.1 Research Question

How do Surprise and VSUP maps influence the manner in which people explore and generate takeaways in map reading contexts?

1.2 Contributions

Our main contributions are as follows:

- Evidence of clear differences in map analysis results between three mapping techniques (Choropleth, Surprise and VSUPs).
- Qualitative feedback that suggests that in some cases, participants only consider a subset of the metrics available to them.
- Quantitative findings suggest that Surprise and VSUP maps help people in identifying highly populous counties, which may be beneficial for offsetting issues related to traditional Choropleth maps.

Chapter 2

Background

The design of map visualizations using datasets that are strongly related to the population of a geographic location may result in misleading visual signals or patterns. The limitations of map visualizations that share such characteristics may be overcome through the use of various statistical techniques such as Bayesian weighting, where certain regions on a map are up-weighted or down-weighted based on some prior belief [19]. However, it is necessary to evaluate of such techniques in order to understand their impact on peoples' perception [30].

Research efforts have led to the development of generic frameworks that propose a list of tasks to be considered when evaluating map visualizations. However, such frameworks may be ill-equipped to address the goals of studies that evaluate uncertainty in map visualizations [30]. In addition, prior studies that leverage generic frameworks to evaluate uncertainty in visualizations have shown the presence of cognitive bias in decision making [54].

Whilst some research has been conducted to evaluate statistical techniques for de-biasing map visualizations [35, 30], their findings may not be universally applicable to other techniques.

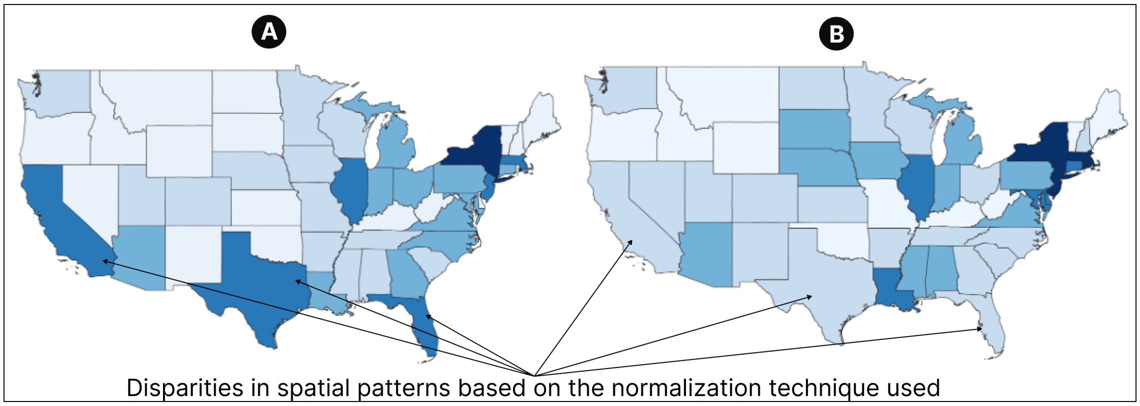

Figure 2.1: Choropeth map of (A) Covid-19 total cases (B) Covid-19 cases per capita, with darker shaded areas showing states with a high number of Covid-19 cases. The use of unnormalized data similar to A (raw counts) results misleading patterns that mimmic the population distribution whilst ignoring the intended visual signal (Covid-19 cases), which are better represented as percentages or per capita rates as shown in B [21].

2.1 Biases in Maps Visualizations

Continous research has sought to understand and address the challenges associated with effectively representing various metrics in map visualizations. However, the representational technique of choice is highly dependant on what information the designer intends to communicate as well as the intended audience. Therefore, it is imperative that designers understand biases associated with map visualizations, techniques to overcome them, as well as their impact on users' perception. For example, metrics such as raw counts, may be closely related to a geographic regions population which may lead to the communication of misleading visual signals or patterns when using only color as a visual encoding (see Figure 2.1) [21]. Such effects may make it challenging for users to effectively rank color intensity on map visualizations, which may result in a reduced consensus of the conveyed information [49].

Furthermore, other design strategies such as the binning technique used to associate values with color encodings (*e.g.* Jenks, Quantile and Manual scales) [39] and the technique used to map the earths surface to a 2-Dimensional plane (*e.g.* Mercator, Robinson Conic, Azimuthal), may influence users' perception by creating patterns that differ from the underlying data [13]. For example, binning techniques influence the intensity of the colors displayed on a map, whilst map projections may influence the size, shape and boundaries that are shown on a map. Such distortions may draw users to large geographic regions on a map whilst neglecting smaller ones, which may result in biased map interpretations [49].

While various considerations may be made by visualization designers to overcome biases in maps, it is important to understand their appropriate use-cases in order to avoid the dissemination of harmful or misleading information.

2.2 Techniques for De-biasing Thematic maps

Various supplemental approaches can be applied to spatial datasets in order to reduce the impact of noise or outliers in map visualizations. For instance, visual encodings such as color, may be weighted based on prior beliefs in order to highlight outliers or regions that deviate from expectations [19] (see Figure 2.2), while spatial smoothing can be used to estimate values for a region of interest by calculating weighted averages of neighbouring locations. In addition, normalization techniques such as decimal scaling, logarithmic transformation and min-max normalization can be used to alter the spatial or statistical properties of a dataset [3]. Other techniques such as VSUPs suppress values at high levels of uncertainty by using a quantization tree to assign values at a certain threshold of uncertainty to a root node [20]. However, when applied to thematic map visualizations, these techniques can offset bias as well as introduce bias. For example, the extent to which values have to be suppressed or biased towards a mean value when using spatial smoothing, results in loss of detail

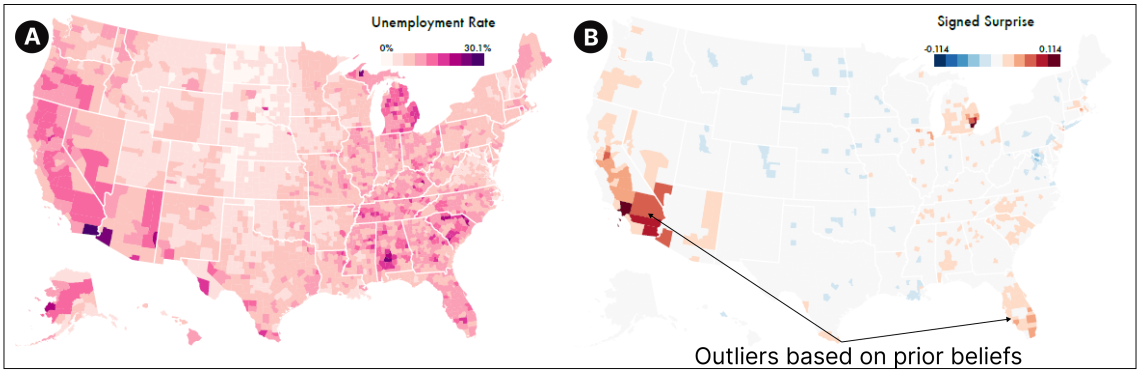

Figure 2.2: (A) Choropleth map of per capita unemployment rates (B) Surprise map showing outlier counties based on prior beliefs. Comparison between a traditional Choropleth map and a Surprise map based on a de Moivres funnel. The Surprise map up-weights or down-weights counties based on their level of variance thereby highlighting outliers on the map. The reduced visual search space allows users to only focus on interesting regions on the map. Counties that have surprisingly high or surprisingly low values are shaded darker [19].

or a distortion of values. Furthermore, altering the statistical properties of a dataset, may change spatial relationships or trends. Therefore, by evaluating such techniques, researchers and designers may gain a better understanding of how to interpret map de-biasing techniques as well as their potential trade-offs.

2.3 Visualizing Uncertainty using Choropleth maps

Using choropleth maps to visualize unexpected events in a dataset, leads to the generation of different visual patterns that may result in a change of belief [19]. A number of approaches and empirical studies exist in literature that visualize uncertainty [18, 33, 38, 57], indicating the importance and positive impact of communicating and applying uncertainty in data visualization [24, 25, 57, 46]. The challenges associated with visualizing uncertainty include, the selection of appropriate representational techniques (visual features, fuzziness, color, layout), as well as ensuring that the information being conveyed by the visualization is not misleading [16]. MacEachren

Objective	Space-Alone	Attributes-in-space	Space-in-time
Identify	find in space (where?)	find attribute value (what? or who?)	find in time (when?)
	identify your house based on an aerial image in Google Earth	*what explosives materials are known to be inside a building that is on fire?*	*how many hotels were in the town in the late 1800s?*
Compare	difference between bearings, distances, extents, or shapes	difference between values (numerical); same or different type (categorical)	difference in time lengths, composites, or resolutions; change
	compare the distribution of patients over 65 years old to the distribution of patients not treated with radiation	*discern between two types of policies on the map*	*compare the historic vegetation to the current vegetation*
Rank	order by nearest/farthest	order by most/least, best/worst	order by first/last
	where are the nearest schools to the toxic chemical release?	*which county has the highest mesothelioma mortality rate?*	*have any apprehensions occurred in the last seven days in this area?*
Associate	strength of connectedness; routing/topology	correlation between variables	trend across time; cause/effect
	this small town community is connected to which major urban systems?	*is socioeconomic status correlated spatially to gonorrhea rates?*	*see if the remediation procedure resulted in reducing the geographic extent of the chemical*
Delineate	division into distinct regions; clustered/disperse	division into distinct types according to attribute values	division into distinct periods; spikes/troughs
	where are the high risk clusters of disease morbidity?	*find clusters of similar attribute values within a set of map features*	*look into a spike of disorderly conduct cases in an area*

Figure 2.3: High-level list of map interaction user goals proposed by Roth [48]. Operands (Space alone, Attributes in space and Space in time), cater for the comparison of both spatial and spatio-temporal maps. Such generic frameworks may be used to develop map specific tasks, to explore their benefits or limitations.

[40] proposes the use of map pairs, bi-variate maps and sequential representation for visualizing the uncertainty of geo-spatial values. A later study by Elmer [23] argues against the use of bi-variate maps to represent uncertainty in favor of map pairs, whilst recent findings by Correll et al. [20] recommend the use of either uni-variate or bi-variate maps when representing uncertainty. The lack of a universal approach for visualizing uncertainty promotes the need for further research.

2.4 Evaluating uncertainty in maps

Prior studies have focused their efforts on information retrieval for uni-variate or bi-variate maps [41, 36, 6], whilst other studies have focused on the ranking, speed and accuracy in decision making [12, 15]. A recent study by Hullman [30] proposes a framework for the evaluation of uncertainty in visualizations. However, studies used to evaluate uncertainty consider different statistical techniques or metrics, restricting the possibility of generalizing the findings. This creates a knowledge gap when we consider the implementation of Bayesian surprise for de-biasing thematic

Name	Description	Break down according to Roth's taxonomy [86]
$Q1_{Pop}$	Read the number of inhabitants in a specific region.	RANK attributes in space for population
$Q2_{Ump}$	Read the statistic of interest (unemployment) in a region.	RANK in attributes in space for unemployment
$Q3_{Comb}$	Combine population and unemployment to find the region with the highest number of unemployed people.	RANK in attributes in space for unemployment $\times$ population
$Q4_{Comb}$	Identify the neighbour of a region that has the highest number of unemployed people.	RANK in attributes in space for unemployment $\times$ population IDENTIFY in space alone
$Q5_{Sum}$	Average the absolute number of unemployed people over multiple regions and compare with another average	SUMMARIZE (from [74]) for unemployment $\times$ population RANK for both summarized regions

Figure 2.4: High-level list of map interaction user goals proposed by Roth [48]. Operands (Space alone, Attributes in space and Space in time), cater for the comparison of both spatial and spatio-temporal maps. Such generic frameworks may be used to develop map specific tasks, to explore their benefits or limitations.

maps. MacEachrean [42], suggests the use of two criteria for evaluating uncertainty in visualizations, which are 1. the usability of uncertainty representations and 2. the manner in which uncertainty information is used and how it affects decision making. Prior studies on the usability of uncertainty visualizations have focused on the encoding of data on a map as well their layout. Bi-variate maps that display uncertainty have been found to be more comprehensible compared to uni-variate maps [40, 27]. However, representing multiple variables on a single map, may result in complex visualizations which make it difficult to identify patterns [9]. Furthermore, the selection of which map representation technique to use (Univariate, bi-variate and side-by-side) should be driven by the purpose of the map (bi-variate maps are more suitable for determining relationships between variables whereas side-by-side univariate maps are more suitable for comparisons), the audience, and the complexity of the data being represented [52]. The use of uncertainty information may lead to the discovery of much more diverse patterns in map visualizations. However, not much research has been conducted on the use of uncertainty information therefore it may be difficult to determine the extent to which such discoveries influence decision making [6].

2.5 Map design considerations

Research efforts have led to continuous development of various map representational techniques (*e.g.* necklace, surprise and VSUP maps) as well as the use of various normalized (see Figure 2.1) or calculated metrics to represent raw geospatial data in map visualizations [19, 20, 51]. Therefore, in order to effectively and accurately communicate information that is not dangerous or misleading, various design considerations have to be taken into account [16, 48]. These include the type of color scheme (*e.g.* color brewer, viridis), the type of scale (*e.g.* jenks, linear, manual, quantile) and interaction (*e.g.* animation, zooming, panning and tooltips)

Work by [16] suggests that various color schemes may play a pivotal role in precise map interpretation by producing distinguishable colors that accommodate for common visual impairments. In addition, they also contribute a concrete framework for the generation of color schemes, suitable for pattern analysis and information retrieval. However, while the selection of an appropriate color scheme may reduce biases associated with the identification of extreme values in maps, it is important for map designers to understand the influence of color schemes in map visualizations.

The correct classification of data by grouping data using a suitable scale, enhances the readability and usefulness of maps. However, maps that do not properly classify data by analyzing its' distribution, may create misleading patterns. For example, when visualizing crime rates for a geographic location using a map, some regions may be grouped as having high crime rates by shading them using dark color, instead of being shaded with slightly less intensity. Such effects may mislead map readers, which may result in biased map interpretation [8, 49].

While static maps offer useful offline capabilitlies (the ability to carry out remote map analysis), they fall short in providing users with dynamic interaction capabilities (*e.g.* zooming, panning, tooltips and dynamic map updates) suitable for modern day communication channels (*e.g.* web based infrastructure and mobile devices). Such map interaction capabilities make the accessibility of map information much easier, which may improve map interpretation, speed and accuracy. For example, zooming and panning interaction primitives, allow map users to explore smaller regions, which may reduce common biases prone to choropleth maps.

2.6 Summary

In this chapter, we narrate how visual perception may be influenced by biases in map visualizations. For example, the use of raw counts instead of percentage rates, results in creation of non-informative map visualizations that highly resemble the underlying population distribution. While visualizations of this nature may be potentially harmful and misleading, various techniques may be implemented to overcome the challenges associated with effectively and accurately communicating information using map visualizations.

Chapter 3

Methodology

We designed three interactive stimuli (Choropleth, Surprise, and VSUP maps) using Covid-19 vaccination and poverty datasets. We conducted two experiments on an online crowdsourcing platform (Prolific), where we collected data from $n = 300$ participants. Pilot studies with vaccine datasets alone, revealed a strong political bias which may skew results. Therefore, we designed a scenario that "masks" the underlying datasets as being about sales rates, using tasks adapted from Roth [48] and Besançon [11].

3.1 Stimuli Design

Our design goal was to minimize notable differences between the stimuli to avoid map interpretation bias, while maximizing on techniques that improve the accessibility of information [37, 47, 7].

3.1.1 Stimuli Design considerations

For each stimulus, we make the following design considerations:

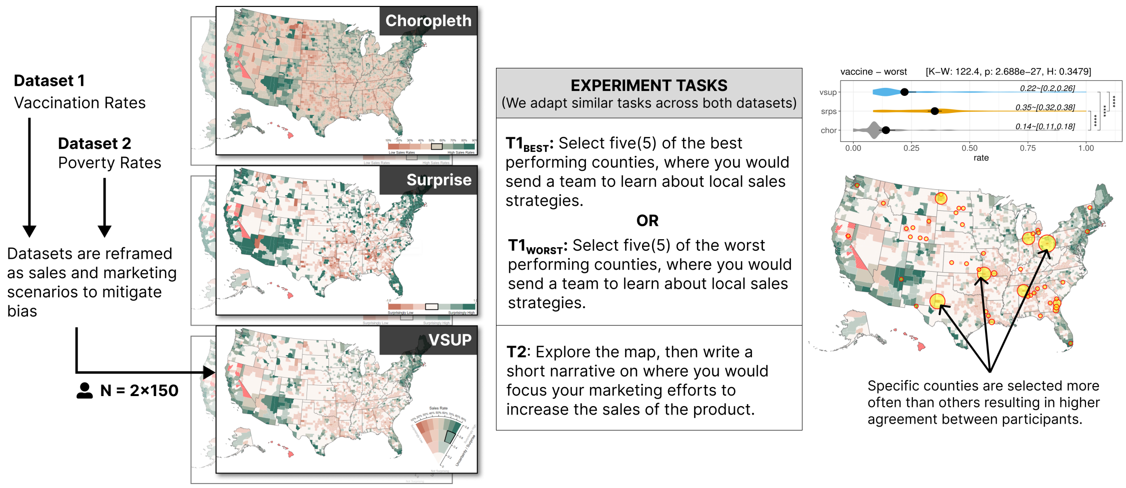

Figure 3.1: How do surprise metrics and suppression encodings influence peoples' takeaways in map visualizations? We use Covid-19 and Poverty datasets to generate the experiment stimulus. We conducted two experiments and randomly assign 300 participants to three map conditions. We collected data using two categories of tasks $T1_{Best}$, $T1_{Worst}$: *Identify* and *T2: Explore*. To mitigate biases for particular dataset contexts, for example vaccination skepticism, we reframed both datasets as a sales and marketing task. We captured participants' exploration meta-data and their feedback regarding the study. We visualized participants' ranking selections from each map to infer their takeaways.

Clicking and Hovering effects

The use of interactive map visualizations, allows users to easily access and engage with the represented data [6]. For example, when users hover or click counties of interest on a map, the underlying data is shown through the use of a tooltip. In addition, the boundary of a selected county may then be highlighted to make it more identifiable, thereby enhancing the user experience [7]. Therefore, we designed the experiment stimulus such that, when a participant hovers over a county, we display a tooltip showing an event rate, a surprise value and the population of the county. In addition, we allow participants to hover over and click on the legend to highlight all counties with a similar color encoding (see Figure 3.9).

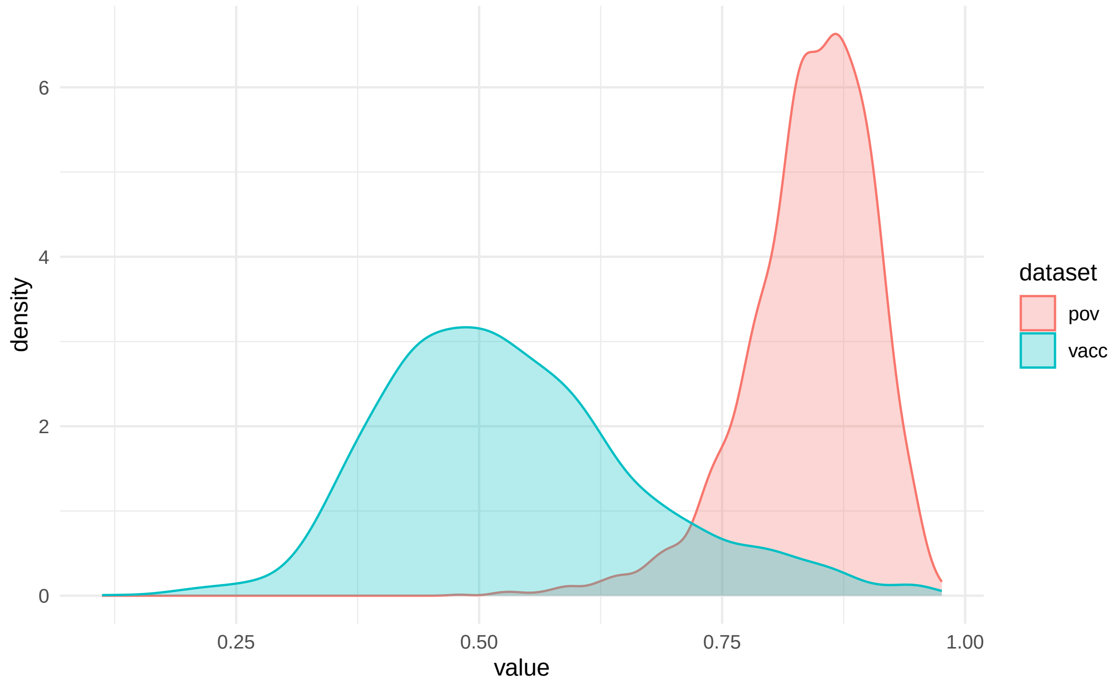

Figure 3.2: Distribution plots for Poverty ad Vaccination datasets. Poverty rates are left skewed whilst vaccination rates are normally distributed with a mean of 0.53.

Zooming and panning

Research has shown that the use of interaction techniques such as zooming and panning allows users to navigate smaller regions on a map visualization [22]. In addition, such techniques may be used to counteract area-size biases associated with choropleth maps. However, whilst such techniques may be beneficial for the accessibility of information related to smaller counties, it is important to consider best design practices for a much more refined user experience [22]. For example, the use of controls compared to pinching or scrolling the mouse wheel may be used to provide better precision and a consistent user experience [50]. In our study, we allow for control based zooming (x2) and panning (up, left, down, right) for each stimulus, to improve participants exploration capability (see Figure 3.3).

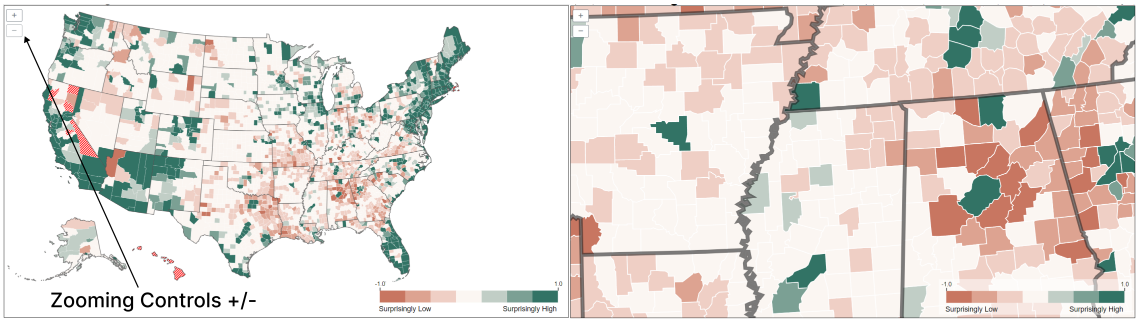

Figure 3.3: Each stimulus is augmented with zoom controls that allow participants to explore smaller counties on the map.

Map color scheme

Color scheme design choices such as, whether to use a diverging, sequential or qualitative scheme, may influence the ability to distinguish between the colors represented (*e.g.* viridis, RColorBrewer). In addition, the number of colors represented in the color palette, has been shown to influence users interpretation and accuracy in extracting information from map visualizations [16]. For example, the use of too many color encodings, may make a map much more difficult to interpret. In addition, some color schemes have been found to be more effective, for representing data related to a specific context. For example, research by [17] found that users prefer the use of spectral color schemes, for health related data. For ecological validity, our map design and color schemes were influenced by The New York Times (NYT) Covid-19 vaccination map [53], and designed to be as consistent as possible between all the stimuli (see Figure 3.1).

Map binning

An intrinsic understanding of the geospatial data to be visualized is important for the selection of a suitable binning strategy for choropleth maps [56]. Various binning techniques (*e.g.* quantile, equal interval and natural breaks) may be used to aggregate

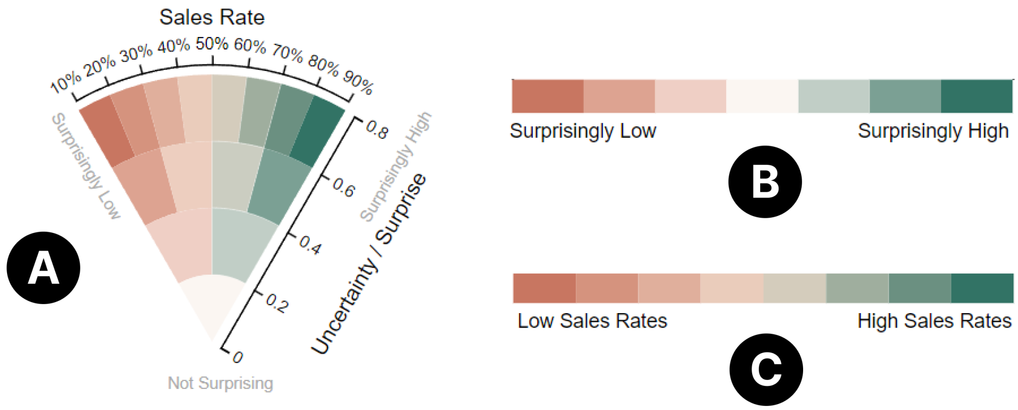

Figure 3.4: A) VSUP legend B) Surprise legend C) Choropleth map legend. We use a diverging color scheme motivated by the NyTimes for ecological validity.

geospatial data. The aggregated data is split into multiple groups and associated with a particular color encoding for visualization purposes. Therefore, it is important for visualization designers to follow suitable guidelines when selecting an appropriate binning technique, to avoid creating maps that are not meaningful or creating maps with obscured visual patterns. For example, equal interval breaks, divide the range of data into equal sized classes and are best used for continous datasets whereas natural breaks use algorithms that specify distinct break points based on the variance of the dataset (*e.g.* Jenks natural breaks). We used a discrete scale from the D3 library (d3.scaleQuantize) to create equal interval breaks for each stimulus and thereafter, mapped the domain values to a corresponding color.

Map domain specification

Manually specifying the domain for each stimulus scales may result in the generation of inconsistent maps. Therefore, we calculate the Mean Absolute Distance (see Equation 3.1), which we use to determine the lower and upper fence of the dataset (see Equation 3.2 and Equation 3.3). The datapoints around the lower and upper

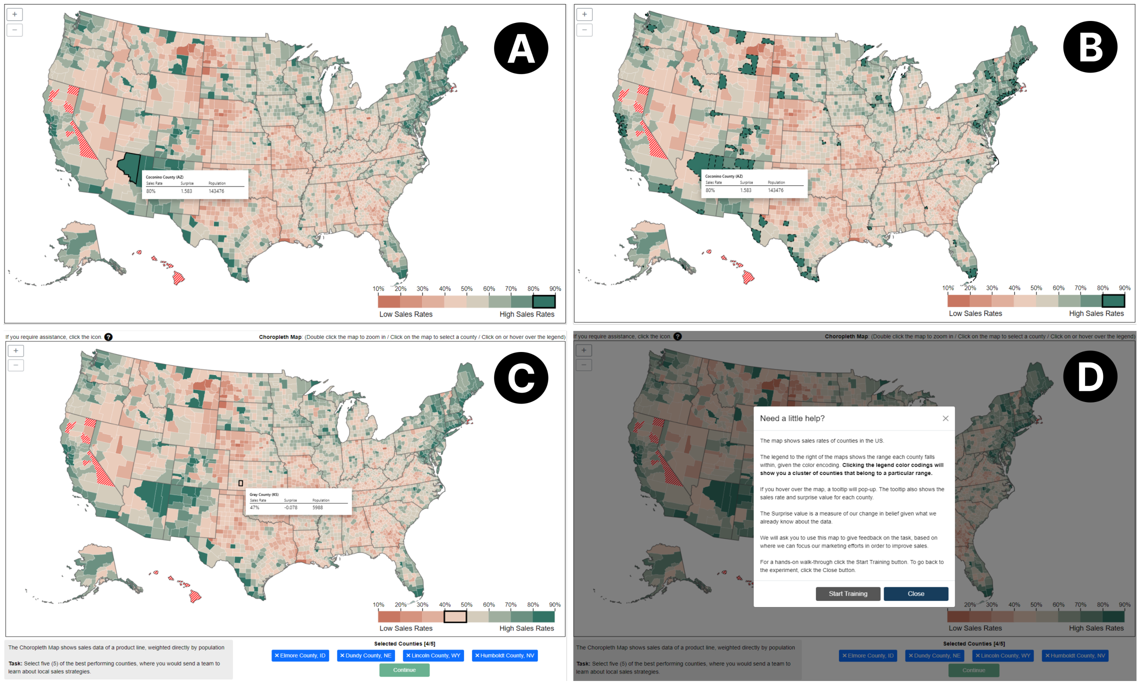

Figure 3.5: We add interaction capability to refine participants' user experience when conducting the experiment. A) When participants hover over a county, a tooltip with rate, surprise and population metrics is shown. B) Participant can click on the legend to highlight all counties associated with a color bin. C) Selected counties are listed alongside the experiment task. D) Whilst conducting the study, participants can access a help section related to the task.

fence are then considered as outliers and represented as extreme values on our maps. For each stimulus, we use the lower fence and upper fence to set the minimum and maximum values for each scale.

$$MAD = median(|X_i - \overline{X}|) \tag{3.1}$$

$$upperFence = q3 + (1.5 * MAD) \tag{3.2}$$

$$lowerFence = q1 - (1.5 * MAD) \tag{3.3}$$

Table 3.1: Preliminary list of tasks used for the pilot study

Qn	Objective	Task Narration
Q_1	Identify attributes in space	Identify a state with the highest surprise in vaccinations.
Q_2	Compare attributes in space	Between state X and state Y, which one has the least surprise in vaccinations?
Q_3	Rank attributes in space	Identify the state with the third most surprise in vaccinations
Q_4	Delineate in space	Which region has the highest surprise in vaccinations?
Q_5	Summarize	Provide a short narrative about the surprise in vaccinations as shown on the map

Map size

Research on choropleth maps has shown that map users tend to overlook smaller areas whilst focusing on larger areas that are easy to identify (Area-size bias) [49]. Therefore, in addition to augmenting each stimulus with zooming and panning capabilities, we set the size of each map to 950×525 pixels to reduce the effect of biases associated with visualizing small scale maps.

Map projection

Map projections allow visualization designers to map the 3-D earths surface onto a 2-D plane (Mercator, Azimuthal and Conic projections). However, the tradeoffs of each map projection should be considered beforehand to avoid introducing biases that may influence users takeaways from map visualizations [28]. For example, the mercator projection preserves angles and is often used for navigation. However, it distorts the shapes and size which may lead to significant inaccuracies in choropleth maps. We use a composite map projection (d3.geoAlbersUSA) [4, 14], to overcome the challenges associated with designing meaningful maps that preserve area.

Table 3.2: To mitigate biases for particular dataset contexts, for example vaccination skepticism, we reframe both datasets as a sales and marketing task. The experiment tasks are split into two categories *T1*$_{Best}$*, T1*$_{Worst}$*: Identify* and *T2: Explore*

	Objective	Task Narration
T1_{Best}	Identify and Rank	Select five (5) of the best performing counties, where you would send a team to learn about local sales strategies.
T1_{Worst}	Identify and Rank	Select five (5) of the worst performing counties, where you would send a team to learn about local sales strategies.
T2	Compare and Delineate (Explore)	Explore the map, then write a short narrative on where you would focus your marketing efforts to increase sales of the product.

Experiment restrictions

Technological advances have led to the development of multiple communication devices which may be used to access web based visualizations (tablets, mobile phones and personal computers). The use of multiple devices to conduct an exploratory study, may lead to inconsistencies in participants responses as well as the visual distortion or misinterpretation of maps [45]. Therefore, prior to conducting the experiment, participants were asked to use either a laptop or desktop device for consistency in map resolution.

3.2 Experiment Datasets

We adapted publicly available county level datasets of Covid-19 vaccinations [1] and poverty rates [2] of the US. Prior to conducting the study, we replicated a Surprise map of per-capita unemployment rates from Correll and Heer [19], that uses a model of the deMoivre's funnel to determine deviations from the average per-capita rate. This method calculates the test statistic (Z_s) from event rates. Bayesian methods are then used to find the likelihood of points being Z_s distant from the center of the

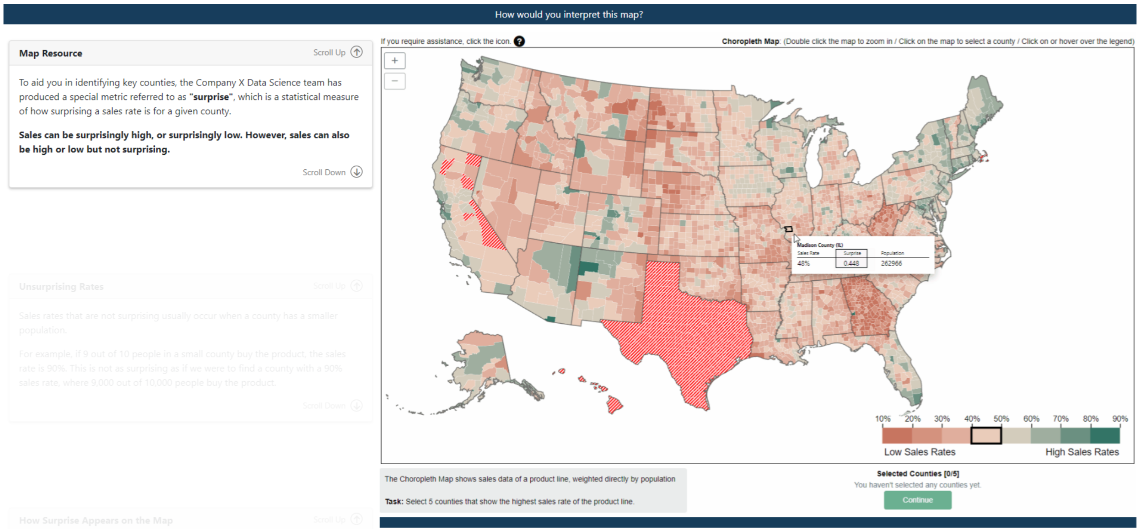

Figure 3.6: A "scrollytelling" training of how to interpret the metrics shown on the map. All trainings' are customized by map condition, and participants are required to go through the training section before conducting the experiment.

funnel, given:

$$Z_{\mathrm{s}} = \frac{O(s) - \overline{x}}{SE_{\mathrm{s}}} \tag{3.4}$$

and:

$$P(s|deMoivre) = 1 - (2 \cdot \int_{0}^{|Z_s|} \phi(x)dx) \tag{3.5}$$

where deMoivre represents the model and $s \in D$ (Dataset). *Surprise* is then calculated by finding the relative distance between the prior and posterior probability distributions through the use of KL-divergence (see Figure 3.7). After replicating the Surprise map of per-capita unemployment rate, we apply the same process to our datasets of interest [1, 2]. To assess the broad relevance of our findings, we use datasets with different distribution characteristics (see Figure 3.2).

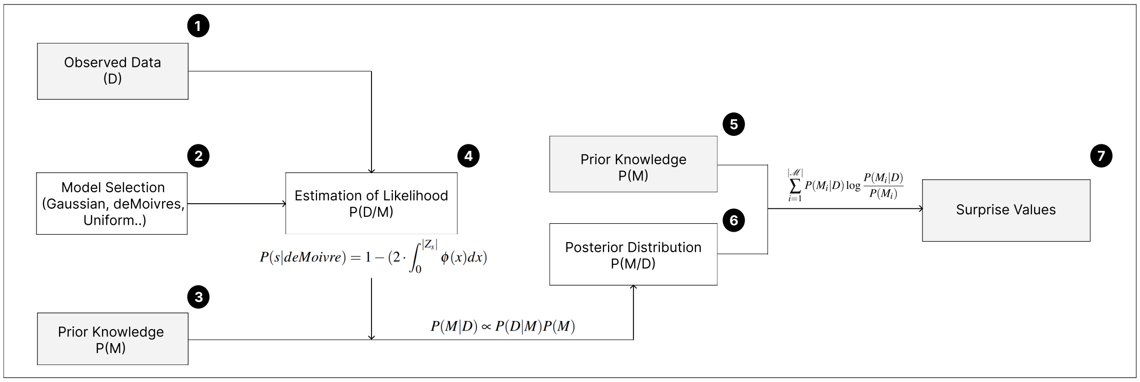

Figure 3.7: Flowchart highlighting how Surprise values are calculated. 1) Some observed data which is denoted by a random variable D. 2) Assumed distribution of the random variable (M) 3) Prior Expectation P(M) of random variable M 4) Conditional likelihood P(D|M) 5) Prior Expectation P(M) of random variable M 6) Posterior distribution P(M|D) 7) Signed Surprise values [19].

3.3 Pilot Study

To refine the user experience for the study, we conducted a pilot study with $n = 30$ participants. We designed our stimuli using a Covid-19 dataset [1] and randomly assigned $n = 10$ participants to each condition. We adapted tasks by Roth [48] to align with the characteristics of our dataset (see Table 3.1). However, our initial spatial analysis of participants' county selections suggest that participants highly consider large counties on the Choropleth map which are shaded dark, compared to the smaller ones (see Figure 3.8. We attribute this to the visual bias created by Choropleth maps, where participants tend to be drawn to larger areas on the map compared to smaller ones [49]. In addition, we observe more consensus on the Choropleth map compared to the Surprise and VSUP maps. We attribute this effect to a lack of comprehension and interpretation of the metrics presented to the participants. Further analysis also shows a higher degree on overlap on the VSUP and Surprise conditions.

Qualitative feedback reflected a high degree of participants' personal beliefs and

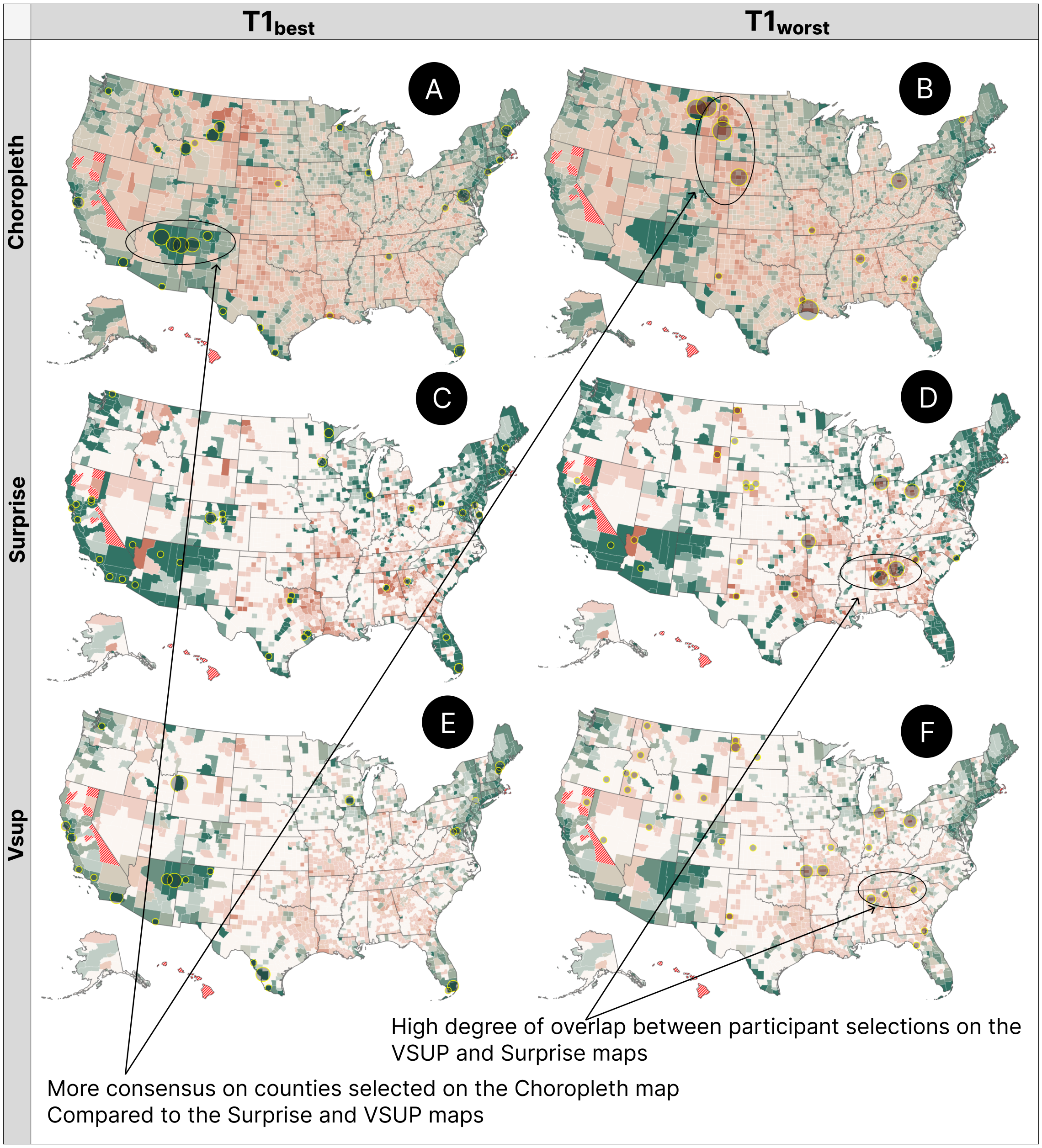

Figure 3.8: Pilot Analysis of participants' county selection for conditions $T1_{Best}$ and $T1_{Worst}$. **A)** Choropleth - $T1_{Best}$ **B)** Choropleth - $T1_{Worst}$ **C)** Surprise - $T1_{Best}$ **D)** Surprise - $T1_{Worst}$ **E)** VSUP - $T1_{Best}$ **F)** VSUP - $T1_{Worst}$ maps. We find selection consensus on the Choropleth map compared to the Surprise and VSUP maps. We also note a high degree of selection overlap between Surprise and VSUP map counties compared to the Choropleth map.

political affiliation. Here are two examples:

Response 1: *"I think it's going okay. In the beginning everyone was reluctant* *since it's so new, there's hardly and research. But as time has gone on more and more people are getting vaccinated [...] It seems like there'll be a lot more people vaccinated by the end of the year".*

Response 2: *"It's going pretty poorly because our country is full of anti-science [...] who still believe in Creationism [...]".*

Given these results, we rephrased our tasks to a product sales and marketing decision-making problem using both the Covid-19 vacinations and poverty datasets (see Table 3.2). We further developed two task categories across the metrics and conditions to be considered by summarizing map analyses tasks and objectives used in the studies by Roth [48] and Besançon [11] (see Figure 3.9). We developed additional "scrollytelling" training for all conditions to help reduce sources of noise in the full experiment (see Figure 3.6). We also added legend interaction to make it easier for participants to identify smaller counties on the map (see Figure 3.9).

3.4 Task and Procedure

We used a between subjects design across two datasets (Covid-19 vaccinations and poverty). We designed 3 stimuli (conditions) and 3 tasks $T1_{Best}$, $T1_{Worst}$ and $T2_{Explore}$ (see Figure 3.1). We randomly assigned 25 participants to each condition. The total number of participants for the study was therefore, 2 experiments $\times 3$ conditions (Choropleth, VSUP and Surprise stimuli) $\times 2$ tasks ($T1_{Best}$, $T1_{Worst}$) $\times 25$ participants $= 300$. Of our participants, 160 identified as female, 136 identified as male, and 4 participants chose not to disclose their gender. Participants' age ranged from 18 to 76 with an average of 35. The study was IRB-reviewed and we required a consent form before participation. Participants were not constrained to a completion time, however, we estimated an average completion time of 7 minutes, used to calculate a payment of \$1.40 to exceed US Minimum Wage. We collect

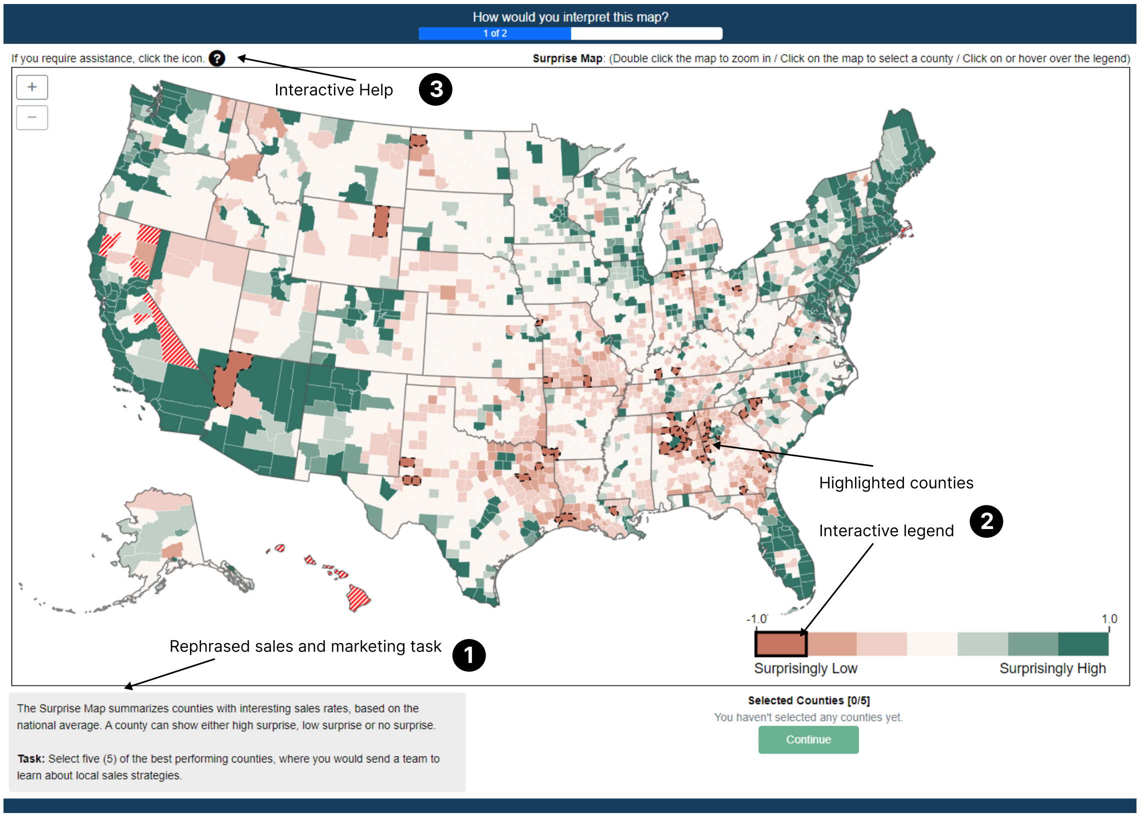

Figure 3.9: We overcome the challenges observed in the pilot study findings by 1) Rephrasing the task to a sales and marketing decision making problem 2) Adding legend interactivity to highlight counties, making it easier for participants to identify and explore smaller counties 3) Adding an interactive training section.

meta-data on counties of interest for each participant (e.g. population), as well as feedback regarding their perception of the study.

3.5 Summary

In this chapter, we detail the experiment design considerations, task design, datasets and, the participant demographic. We explain how we refine the user experience for the study by analyzing results from the pilot study. In the next chapter, we discuss the implications of our quantitative analysis based on participants' responses.

Chapter 4

Results

We used a Kruskal-Wallis test to detect overall effects in data across the three different mapping techniques (see Figure 4.3 and **supplemental material for full results**). For post hoc tests, we use Dunn's test with Bonferroni correction. We also compute and report 95% confidence intervals using bootstrapping. For geo-spatial analysis, we create point maps of participant county selections across the three conditions (see Figure 4.1 and Figure 4.2).

In the identify tasks, performance across conditions differed across several metrics, including the rates, the population and Surprise values of the selected counties (see Figure 4.3).

4.1 Identify Tasks $\mathbf{T1_{Best}}$ and $\mathbf{T1_{Worst}}$ - Vaccinations Dataset

Rate: While we anticipated that Choropleth maps would result in selection of counties with higher rates, since the Choropleth map directly visualized rate, we find that the VSUP and Surprise maps have an effect on the selection of counties with high rates (see Figure 4.3A and C). We find overall differences between the map conditions for vaccine best task $KW = 12.75$ $p = 0.0017$ $H = 0.02749$ and the

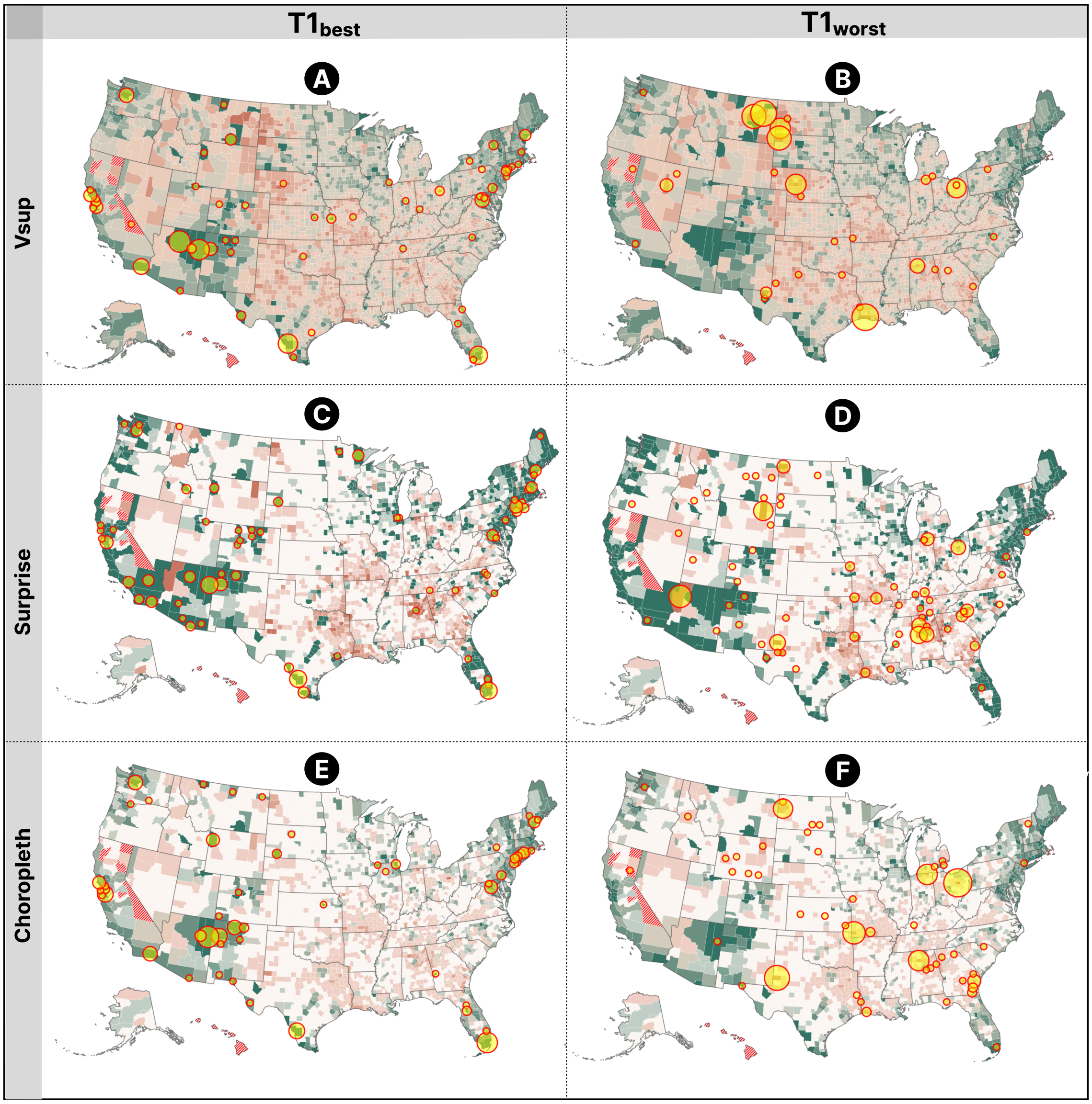

Figure 4.1: Participants' county selections for vaccination data tasks $T1_{Best}$ and $T1_{Worst}$. **A)** Choropleth map ($T1_{Best}$) **B)** Choropleth map ($T1_{Worst}$) **C)** Surprise ($T1_{Best}$) **D)** Surprise map ($T1_{Worst}$) **E)** ($T1_{Best}$) VSUP map **F)** VSUP ($T1_{Worst}$). Visual analysis shows a high degree of consensus on the VSUP maps, particularly in **F** (VSUP). We see some consensus on **B** (Choropleth-Worst) compared to **A** (Choropleth-Best). We also see a high degree of dispersion on the Surprise maps **C** and **D** compared to both the Choropleth and VSUPs.

vaccine worst tasks $KW = 122.4$ $p = 2.88e - 27$ $H = 0.3479$. Post-hoc comparisons suggest that the VSUP performs best in the vaccine best task, and the Choropleth map performs best in the vaccine worst task. In the latter case, VSUPs appear

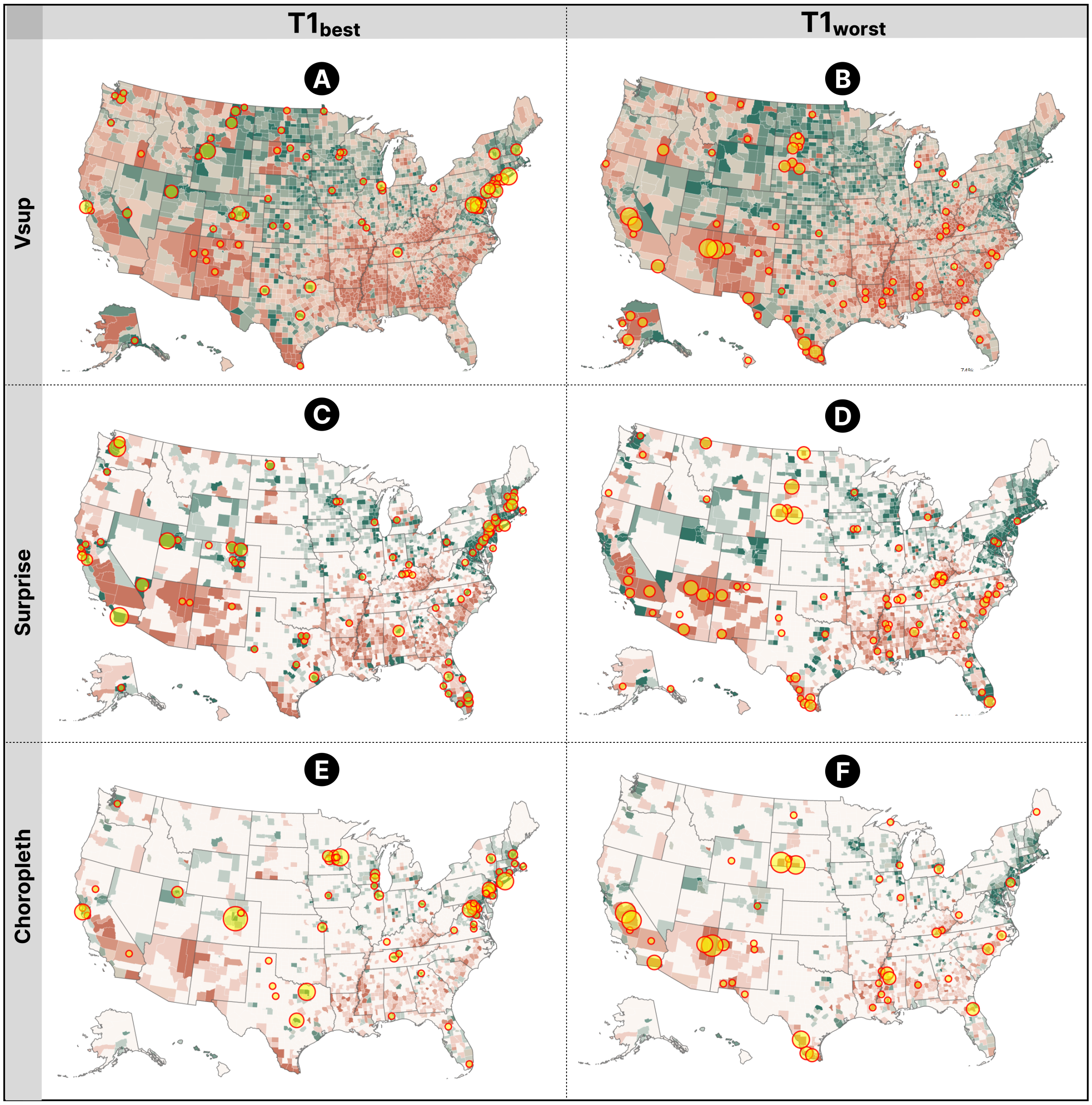

Figure 4.2: Participants' county selections for poverty data tasks $T1_{Best}$ and $T1_{Worst}$. **A)** Choropleth map ($T1_{Best}$) **B)** Surprise map ($T1_{Best}$) **C)** VSUP ($T1_{Best}$) **D)** Choropleth map ($T1_{Worst}$) **E)** ($T1_{Worst}$) Surprise map **F)** VSUP ($T1_{Worst}$). Visual analysis shows a high degree of consensus on the VSUP maps compared to both the Choropleth and Surprise. The lack of consensus in Choropleth in this dataset compared to vaccine dataset may be due to skewed rate.

to balance the differences between the Surprise and Choropleth maps. However, the differences observed between the vaccine-best and vaccine-worst tasks may be attributed to the variability observed among counties with high levels of surprise.

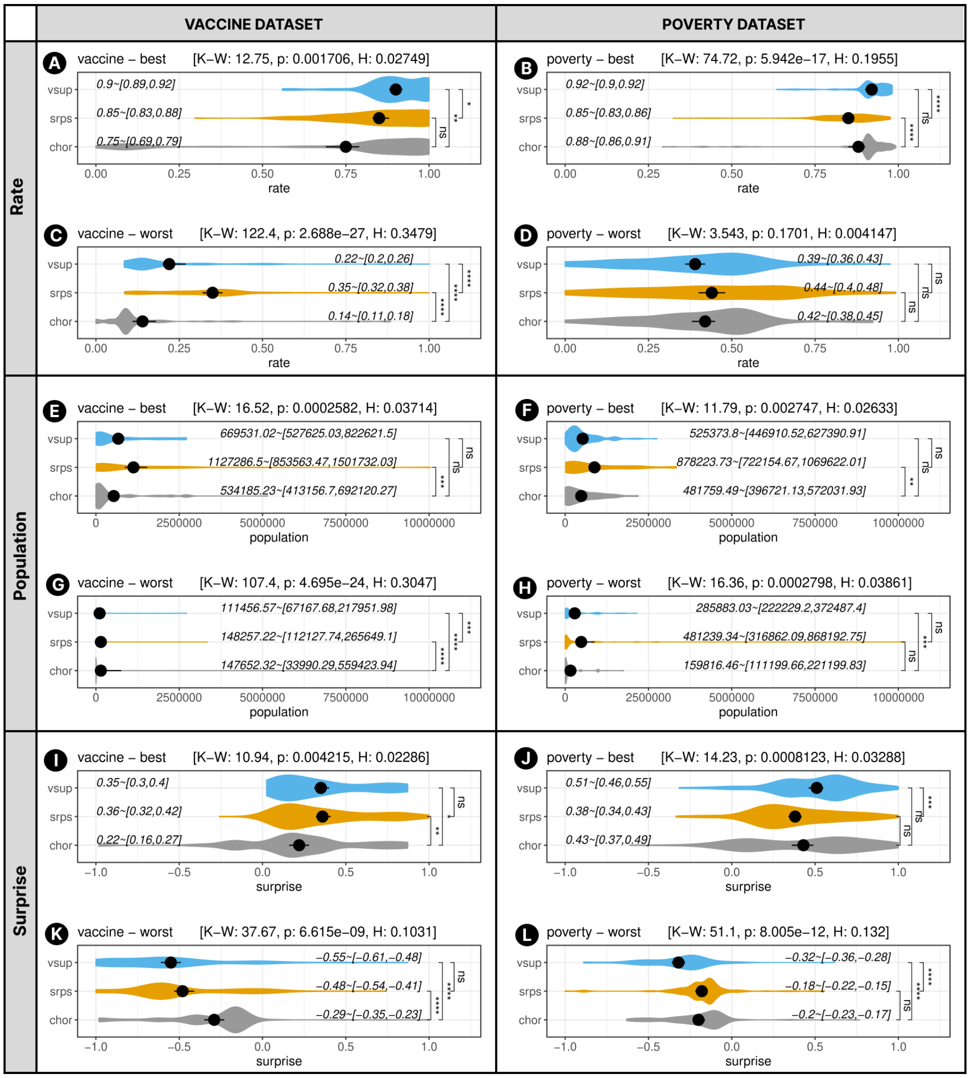

Figure 4.3: Sup-3: Quantitative results for rate, population and surprise metrics for both dataset and all conditions. Left column is based on vaccine dataset and right column is based on poverty dataset. **A)** T1$_{\text{Best}}$-Vaccine (Rate) **B)** T1$_{\text{Best}}$-Poverty (Rate) **C)** T1$_{\text{Worst}}$-Vaccine (Rate) **D)** T1$_{\text{Worst}}$-Poverty (Rate) **E)** T1$_{\text{Best}}$-Vaccine (Population) **F)** T1$_{\text{Best}}$-Poverty (Population) **G)** T1$_{\text{Worst}}$-Vaccine (Population) **H)** T1$_{\text{Worst}}$-Poverty (Population) **I)** T1$_{\text{Best}}$-Vaccine (Surprise) **J)** T1$_{\text{Best}}$-Poverty (Surprise) **K)** T1$_{\text{Worst}}$-Vaccine (Surprise) **L)** T1$_{\text{Worst}}$-Poverty (Surprise) We use the Kruskal-Wallis test to find differences in the significance of the data collected using the stimuli. We calculated a 95% confidence interval using a bootstrap method. Quantitative analyses suggests that VSUPs lead to the selection of counties with high rates and high surprise whilst Surprise maps lead to the selection of highly populated counties.

Population: In terms of selected counties, the Surprise maps tended to lead participants towards counties of higher population (see Figure 4.3B and G). In particular, we find in the vaccine best task an overall effect $KW = 16.52$ $p = 0.00025$ $H = 0.037$. However, we note that these effects tend to place Surprise maps above Choropleth maps, but not above VSUPs, which appear to balance the effects of the other two. Similar effects and trends are found in the vaccine worst $KW = 107.4$ $p = 4.695e - 24$ $H = 0.3047$, and in the poverty worst $KW = 16.36$ $p = 0.0002798$ $H = 0.03861$ tasks.

Surprise: Results suggest that VSUPs and Surprise maps led participants to select counties with high surprise values for the vaccine best and low surprise values for the vaccine worst task as shown in Figure 4.3I and K with $KW = 10.94$ $p = 0.004$ $H = 0.02286$ and Figure 4.3F with $KW = 37.67$ $p = 6.615e - 09$ $H = 0.1031$. However, although we note significant differences in the selection of counties that are surprisingly high or low between Choropleth and Surprise maps as well as VSUPs, we observe insignificant differences between VSUPs and Surprise maps. We hypothesize that a reduction in the visual search space for both maps, may lead to insignificant differences in participants selections.

4.2 Identify Tasks $\mathbf{T1_{Best}}$ and $\mathbf{T1_{Worst}}$ - Poverty Dataset

Rate: We anticipated some disparity in findings between both experiments due to the characteristics of each dataset, we note somewhat consistent results for both the Covid and Poverty datasets which show that VSUP and Surprise maps have an effect on the selection of counties with high event rates (see Figure 4.3B and D). We find overall differences between the map conditions for poverty best task $KW = 74.72$ $p = 5.942e - 17$ $H = 0.1955$ and the poverty worst tasks $KW = 3.543$ $p = 0.1701$ $H = 0.0042$. Post-hoc comparisons suggest that the VSUP performs best in both the poverty best and worst tasks. However, we attribute our failure to realize significant post-hoc differences for the poverty worst task to a negative skew in the dataset (see Figure 3.2).

Population: Surprise maps tended to lead participants towards counties of higher population (see Figure 4.3F and H). We find in the poverty best task an overall effect $KW = 11.79$ $p = 0.00275$ $H = 0.026$. However, we note that these effects tend to place Surprise maps above Choropleth maps, but not above VSUPs, which similarly to the vaccination dataset results, appear to balance the effects of the other two. Similar effects and trends are found in the poverty worst task with $KW = 16.36$ $p = 0.0002798$ $H = 0.03861$.

Surprise: Results suggest that VSUPs led participants to select counties with high surprise values for both the poverty best and worst tasks as shown in the Figure 4.3J with $KW = 14.23$ $p = 0.0008$ $H = 0.03288$ and Figure 4.3L with $KW = 51.1$ $p = 8.005e - 12$ $H = 0.132$. We realize somewhat disimilar post-hoc comparison results between the Covid and Poverty datasets with VSUPs outperforming Surprise maps maps in the selection of surprisingly high and surprisingly low counties. However, we observe insignificant effects between VSUPs and Choropleth maps as well as Surprise and Choropleth maps for task $\mathrm{T1}_{\mathrm{Best}}$-Poverty. In addition, we observe significant effects between VSUPs and Surprise maps as well as VSUPs and Choropleth maps, whilst observing insignificant differences between Surprise and Choropleth maps, for task $\mathrm{T1}_{\mathrm{Worst}}$-Poverty. We also attribute these disparities to a negative skew in the poverty dataset (see Figure 3.2.

4.3 Explore Task T2

We considered participants' feedback based on relevance, similarity and identified keywords such as population, color, surprisingly high and surprisingly low. In the Discussion we expand on our takeaways from participant responses which suggest that:

1. Participants only consider a subset of the metrics presented (§ 5.2.1).

2. Visual encodings like color influence how people interpret surprise (§ 5.2.2).

3. County size may skew peoples' takeaways (§ 5.2.3).

4.4 Summary

In this chapter, we explore a quantitative analysis of our findings for tasks $\mathrm{T1}_{\mathrm{Best}}$ and $\mathrm{T1}_{\mathrm{Worst}}$ for both the vaccine and poverty datasets. We use the Kruskal-Wallis and Dunn's post-hoc tests with Bonferroni correction to detect effects across three different mapping techniques (Choropleth, Surprise and VSUP maps). Our findings show overall differences in participants' responses for both experiments. We find VSUP and Surprise maps leading to the selection of counties with high rates. Further analysis suggests that Surprise maps lead to the selection of counties with high population as well as high surprise values. We hypothesize that the characteristics of our datasets may contribute to deviations from our expected results. In the next chapter, we discuss our takeaways from participant responses.

Chapter 5

Discussion

In this chapter, we analyse the implications of participants' map exploration by creating point maps from their geospatial metadata. Further analysis of participants' feedback suggests that visual encodings and metrics influence participants' takeaways when exploring map visualizations.

5.1 Visual Saliency of High/Low performing counties

Results from participant ranking selections are aggregated by county (see Figure 4.1 and Figure 4.2). We infer the following takeaways from participants' interactions and selections on the maps.

5.1.1 Spatial Analysis of Identify Task $\mathbf{T1_{Best}}$ and $\mathbf{T1_{Worst}}$ - Covid Dataset

County selection on the Choropleth map show a high level of dissimilarity compared to the Surprise and VSUP maps. We attribute this to the narrowed visual search space when visualizing Surprise compared to event-rates on a Choropleth map. We support our claim by a further analysis of the Surprise and VSUP maps, where we see much stronger clusters and consensus of participants ranking selections. However, our findings show a higher degree of ranking consensus by participants' on the VSUP map compared to the Surprise map. We attribute this to the fact that VSUPs further suppress highly uncertain values by combining color cells in a palette using a tree structure [32]. To assess whether participants considered population in their decision-making process, we conducted further analysis of the population of counties they selected. Our findings suggest that participants consider highly populated areas when making task based decisions using both Surprise and VSUP maps, minimizing the effect of coincidence in our findings. This claim is further validated by our qualitative analysis, for example:

Response 1: *"I would focus our marketing efforts on counties with a large population that have seen surprisingly low sales"*

Response 2: *"I would focus my efforts on areas that have both a high surprise rate and low sales. [...] I would also look for areas that have a higher population as that means there are more people that can be targeted."*

5.1.2 Spatial Analysis of Identify Task $T1_{Best}$ and $T1_{Worst}$ - Poverty Dataset

Spatial analysis of participants' county selections for the poverty dataset show slighlty polarized findings when compared to county selections for the Covid-19 vaccinations dataset. Our findings show a high level of dissimilarity in county selections on the Choropleth map compared to the Surprise and VSUP maps. However, our findings suggest much stronger consensus by participants on the Surprise map compared to the VSUP map. We attribute this not only to a narrowed visual search space when visualizing Surprise compared to event-rates on a Choropleth map, as hypothesized for the vaccinations dataset, but to a skewed distribution for the poverty dataset (see Figure 3.2), as well as a suppression effect for the VSUP map. Similar to the

Covid-19 vaccinations spatial analysis, we find that participants consider highly populated areas when making task based decisions using both Surprise and VSUP maps, minimizing the effect of "chance" in our findings. For example:

Response 1: *“I would focus my efforts on the areas that are have surprisingly low sales rates and high populations. Given that these areas have higher populations I think efforts to increase sales rates could lead to overall higher sales than focusing on low population areas."*

Response 2: *“I would focus marketing efforts in the most densely populated counties that had surprising low sales in the most densely populated states in order to increase product sales [...]"*

5.2 Explore Task T2

We summarize feedback from an open-ended response task, where we ask participants to explore the map and give insights on where they would focus their marketing efforts to increase the sales of the product.

5.2.1 Participants only consider a subset of the metrics presented on the maps.

Participant comments suggest that some appeared to have difficulty in making use of all the available metrics (Surprise, Rate and Population), instead they used only 1 or 2 of the available metrics to select high or low performing counties. These findings are supported by summarizing participants feedback on the strategies they used in selecting counties on the maps and contribute to insights on the challenges associated with comprehending Surprise without the consideration of other metrics (Population and Rate).

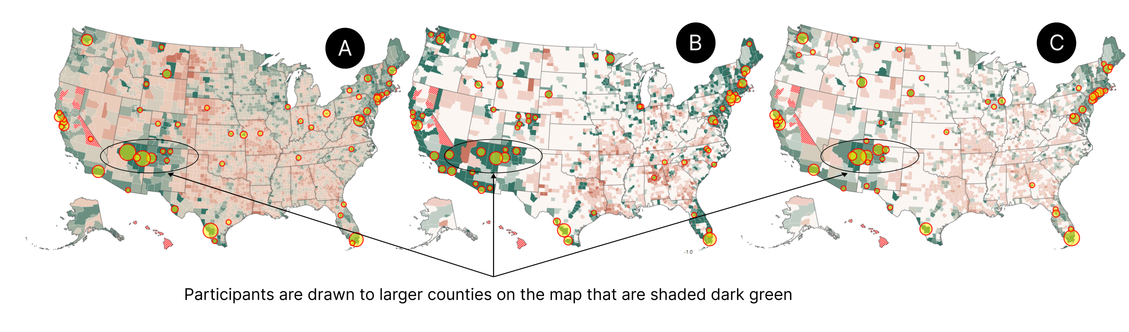

Figure 5.1: Geospatial analysis of participants' county selections on A) Choropleth, B) Surprise and C) VSUP maps. Findings suggest that participants' are drawn to large counties on the map shaded using darker colors.

Response 1: *"I looked for areas with high Surprise metrics (or low) and considered that most areas could be converted because of their proximity to areas with good sales"*

Response 2: *"I paid more attention to counties with low sales rates and populations over 100,000. I didn't pay much attention to the surprise metric."*

However, other participants effectively used interactions to explore smaller counties by hovering over the legend and counties. This allowed them to carry out more complex queries on the maps, suggesting that they could gain more insights by carrying out other tasks, for example:

Response 1: *"I selected areas using the proportion scale/bar on the lower right. Then, I'd hover over these areas to understand the surprise metric relative to the sales success rate and population. Being able to compare the data helped me to understand what the surprise metric meant, and then helped me develop a hypothesis on why these are high success/high surprise areas."*

Response 2: *"Hovered over the key to get the map highlights and then focus on areas from that. Lowest areas actually seem to also have a low population."*

5.2.2 Color influences how people interpret surprise

We observed the influence of color in how participants interpret either event-rates or surprise. These findings suggest that some participants consider dark green and dark brown as high or low sales rate counties respectively [49]. For example:

Response 1:*"In the areas that were darker green, sales are already excellent [...]"*

Response 2:*"Darker green colors show positive and more response to the marketing and the darker pink color is the opposite [...]"*

While interpretation is true for standard choropleth maps, it is not true for Surprise and VSUP maps which show complex values.

5.2.3 Size influences how people interpret of uncertainty

Similar to findings by Schiewe [49], both our qualitative and point pattern analysis suggest that some participants neglect smaller counties and are drawn to larger counties or states on the maps. For example:

Response 1: *"I will look for big states and big counties that are surprisingly low"*

Response 2: *"looked at sizes of areas, counties, populations, etc.[...]"*

However, our findings also suggest that VSUPs and Surprise maps suppress low-population counties with high rates (*e.g.* unsurprising), which may help alleviate one aspect of this bias. How to ensure small yet high population counties also receive sufficient attention remains a challenge for maps geared towards the general public.

5.3 Limitations and Future Work

Our analysis of the Surprise, Rate and Population metrics shows clear differences between the mapping techniques used in this study (Choropleth, Surprise and VSUP maps). However, we hypothesize that the use of datasets with different distributions (normal and left skewed) may impact findings of our study. Future experiments may investigate directly the impact of skews in rates (*e.g.* through simulation) on participant exploration and takeaways. Furthermore, metrics and interaction techniques that build on existing work like Surprise and VSUPs may further enrich map analysis for the general public.

Another limitation is noise itself in the experiment. While the sales scenario worked well overall by allowing us to ask the same task across Choropleth, Surprise, and VSUP maps, one key issue arose in the "worst" tasks. We observed outlier participants across all conditions who, when prompted to select the worst performing counties, instead selected the best performing counties. This may be a bias with the framing of sales, which could be addressed by experimenting with other scenarios or by additional design or feedback mechanisms.

5.3.1 Future Work

Future work should consider:

- Collecting prior probability distributions from participants.

  Prior research by [5, 34], suggests that visualization users should be able to capture and update their prior beliefs based on new observations. However, our current study does not provide map users with a technique that allows them to capture or update their prior beliefs. In addition, when dealing with large geo-spatial datasets, belief elicitation may be a daunting task. Therefore, future research may focus on the design of suitable belief elicitation techniques for Surprise or VSUP maps.

- Experimenting with other representational techniques.

  For this study, we focus on the use of uni-variate and bi-variate map representation techniques. However, future work may consider the use of other techniques such as map pairs, which may improve the accessibility of highly technical maps for the general public.

- The use of decision based models as suggested in work by Kay [32] and Fumeng *et al.* [55].

  Future work may also explore the implementation of alternative approaches for the suppression of uncertainty such as shrinkage and perceptual VSUPs [32]. Such techniques may reduce biases associated with the inconsistent suppression of uncertainty values inherent in tree VSUPs.

5.4 Summary

In this chapter, we discuss the geospatial analysis findings of participants' map exploration metadata. Our takeaways show a high degree of ranking consensus by participants on the VSUP map compared to the Surprise and Choropleth maps. In addition, qualitative analysis of participants' feedback, suggests that visual encodings influence participants' takeaways when exploring map visualizations. We conclude the chapter by identifying areas that may provide further insights on participants' takeaways from maps that encode multiple metrics such as rates, population and Surprise.

Chapter 6

Conclusion

Despite the pervasive use of choropleth map visualizations, especially when communicating critical data to the public (*e.g.* vaccine trends or election results), they suffer from well-documented biases and limitations. In this study, we design an experiment to test two recently proposed techniques, Surprise maps and VSUPs, in a crowdsourced setting similar to how participants might encounter such maps online. Results generally indicated that Surprise maps and VSUPs do indeed offset some of the issues of traditional Choropleth maps. However, close inspection also reveals opportunities for addressing confusion and misconceptions of these new techniques. Going forward, designers may benefit from knowing that Choropleths perform similarly to these new techniques (*i.e.* reducing the risk of harm), while results that indicate these new techniques can lead people to more surprising or populous counties may give designers the confidence to experiment with new and innovative ways of communicating with the general public.

Appendix A

Appendix

A.1 Sup-1: Kruskal-Wallis test results

Table A.1: Kruskal-Wallis test results for $T1_{Best}$, $T1_{Worst}$ - vaccination dataset and $T1_{Best}$, $T1_{Worst}$ - poverty dataset

Dataset	**Task**	**Metric**	**p_val**	**K-W**	**Effsize**
Vaccine	Best	Rate	0.0017	12.75	0.0275
Vaccine	Best	Population	0.00026	16.52	0.0372
Vaccine	Best	Surprise	0.0042	10.94	0.0229
Vaccine	Worst	Rate	2.687e-27	122.4	0.348
Vaccine	Worst	Population	4.965e-24	107.4	0.305
Vaccine	Worst	Surprise	6.615e-09	37.67	0.103
Poverty	Best	Rate	5.942e-17	74.72	0.1955
Poverty	Best	Population	0.00274	11.79	0.0263
Poverty	Best	Surprise	0.0008	14.23	0.03287
Poverty	Worst	Rate	0.1701	3.543	0.004
Poverty	Worst	Population	0.00028	16.36	0.0386
Poverty	Worst	Surprise	8.005e-12	51.1	0.132

A.2 Sup-6: Dunn's test post-hoc results

Table A.2: Post-Hoc results for $T1_{\text{Best}}$ - Rate (Vaccine)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	0.5779	0.5633	1.0	ns
chor	vsup	3.4051	0.0006	0.0019	**
srps	vsup	2.6769	0.0074	0.02228	*

Table A.3: Post-Hoc results for $T1_{\text{Best}}$ - Population (Vaccine)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	4.0639	0.00005	0.0001	**
chor	vsup	1.7556	0.079	0.2374	ns
srps	vsup	-2.2341	0.0254	0.0764	ns

Table A.4: Post-Hoc results for $T1_{\text{Best}}$ - Surprise (Vaccine)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	3.0006	0.00027	0.008	**
chor	vsup	2.5949	0.0094	0.0284	*
srps	vsup	-0.417	0.6766	1	ns

Table A.5: Post-Hoc results for $T1_{\text{Worst}}$ - Rate (Vaccine)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	11.06	1.9606e-28	5.88e-28	****
chor	vsup	5.9705	2.365e-09	7.095e-09	****
srps	vsup	-5.376	7.613e-08	2.283e-07	****

Table A.6: Post-Hoc results for $T1_{\text{Worst}}$ - Population (Vaccine)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	10.307	6.536e-25	1.96e-24	****
chor	vsup	6.62	3.485e-11	1.0455e-10	****
srps	vsup	-3.885	1.02e-04	3.06e-04	***

Table A.7: Post-Hoc results for $T1_{\text{Worst}}$ - Surprise (Vaccine)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	-4.487	0.0000007	2.164e-05	****
chor	vsup	-5973	0.000000002	6.998e-09	****
srps	vsup	-1.584	0.11325	3.3976e-01	ns

Table A.8: Post-Hoc results for $T1_{\text{Best}}$ - Rate (Poverty)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	-6.5829	4.6136e-11	1.384e-10	****
chor	vsup	1.605	1.863e-01	3.5589e-01	ns
srps	vsup	8.1434	3.8418e-16	1.1525e-15	****

Table A.9: Post-Hoc results for $T1_{\text{Best}}$ - Population (Poverty)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	3.4234	0.0006	0.0019	**
chor	vsup	1.9489	0.0513	0.1539	ns
srps	vsup	-1.474	0.1404	0.421	ns

Table A.10: Post-Hoc results for $T1_{\text{Best}}$ - Surprise (Poverty)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	-1.7753	0.07585	0.22755	ns
chor	vsup	1.995	0.046	0.1381	ns
srps	vsup	3.77	0.00016	0.0004	***

Table A.11: Post-Hoc results for $T1_{\text{Worst}}$ - Rate (Poverty)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	1.0217	0.3068	0.92065	ns
chor	vsup	-0.858	0.39086	1	ns
srps	vsup	-1.8798	0.1804		ns

Table A.12: Post-Hoc results for $T1_{\text{Worst}}$ - Population (Poverty)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	1.84445	0.0651	0.1953	ns
chor	vsup	4.04	0.00005	0.00016	***
srps	vsup	2.1956	0.0281	0.084	ns

Table A.13: Post-Hoc results for $T1_{\text{Worst}}$ - Surprise (Poverty)

Group 1	Group 2	Statistic	p	p.adj	p.adj.signif
chor	srps	0.645	5.189e-01	1e+00	ns
chor	vsup	-5.843	5.124e-09	1.5372e-09	****
srps	vsup	-6.488	8.6939e-11	2.608e-10	****

Bibliography

[1] COVID-19 vaccinations in the united states,county. https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh/data. Accessed: 2023-4-16.

[2] Data sets. https://www.openintro.org/data/index.php?data=county. Accessed: 2023-4-16.

[3] Zahraa Said Abdallah, Lan Du, and Geoffrey I Webb. Data preparation., 2017.

[4] Gregor Aisch. Data visualization guidelines – by gregor aisch – international journalism festival: School of data - evidence is power, Apr 2013.

[5] Gregor Aisch, Amanda Cox, and Kevin Quealy. "you draw it: How family income predicts children's college chances". *NY Times*, May 2015.

[6] Koko Alberti. Web-based visualization of uncertain spatio-temporal data. Master's thesis, 2018.

[7] Gennady L Andrienko and Natalia V Andrienko. Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, 13(4):355–374, 1999.

[8] William Bajjali and William Bajjali. Map classification and layout. *ArcGIS for Environmental and Water Issues*, pages 25–40, 2018.

[9] Sarah Battersby. How to make effective bivariate choropleth maps with tableau, Mar 2018.

[10] Enrico Bertini and Moritz Stefaner. Data stories: Surprise maps with michael correll and jeff heer on apple podcasts, Jun 2017.

[11] Lonni Besançon, Matthew Cooper, Anders Ynnerman, and Frédéric Vernier. An evaluation of visualization methods for population statistics based on choropleth maps. *arXiv preprint arXiv:2005.00324*, 2020.

[12] Steve Blenkinsop, Pete Fisher, Lucy Bastin, and Jo Wood. Evaluating the perception of uncertainty in alternative visualization strategies. *Cartographica:*

The International Journal for Geographic Information and Geovisualization, 37(1):1–14, 2000.

[13] Elizabeth Borneman. Types of map projections, Jul 2023.

[14] Mike Bostock. D3/d3-geo: Geographic projections, spherical shapes and spherical trigonometry., Dec 2022.

[15] Nadia Boukhelifa, Anastasia Bezerianos, Tobias Isenberg, and Jean-Daniel Fekete. Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2769–2778, 2012.

[16] Cynthia A Brewer. Color use guidelines for mapping. *Visualization in modern cartography*, 1994(123-148):7, 1994.

[17] Cynthia A Brewer, Alan M MacEachren, Linda W Pickle, and Douglas Herrmann. Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers*, 87(3):411–438, 1997.

[18] Michael Correll and Michael Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, 2014.

[19] Michael Correll and Jeffrey Heer. Surprise! bayesian weighting for de-biasing thematic maps. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):651–660, 2016.

[20] Michael Correll, Dominik Moritz, and Jeffrey Heer. Value-suppressing uncertainty palettes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2018.

[21] Jack Dougherty and Ilya Ilyankou. *Hands-on data visualization.* " O'Reilly Media, Inc.", 2021.

[22] Lucas Dworschak. *Semantically Zoomable Choropleth Map*. PhD thesis, Technische Universität Wien, 2016.

[23] Martin E Elmer. Symbol considerations for bivariate thematic maps. In *Proceedings of 26th International Cartographic Conference*, 2013.

[24] Angela Fagerlin, Catharine Wang, and Peter A Ubel. Reducing the influence of anecdotal reasoning on people's health care decisions: is a picture worth a thousand statistics? *Medical decision making*, 25(4):398–405, 2005.

[25] Rocio Garcia-Retamero and Mirta Galesic. Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. *American journal of public health*, 99(12):2196–2202, 2009.

[26] LAURE Genebes, AURIANE Renaud, and FRANÇOIS Sémécurbe. Spatial smoothing. *Handbook of Spatial Analysis: Theory and Application with R.; Loonis, V., Bellefon, MP, Eds*, pages 205–229, 2018.

[27] Lydia E Gerharz and Edzer J Pebesma. Usability of interactive and non-interactive visualisation of uncertain geospatial information. *Geoinformatik*, 4:223–230, 2009.

[28] Christine Graeff. Ethical implications of biases and errors in geographic information systems. In *2006 IEEE International Symposium on Technology and Society*, pages 1–8. IEEE, 2006.

[29] Amy L Griffin. Trustworthy maps. *Journal of Spatial Information Science*, 2020(20):5–19, 2020.

[30] Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):903–913, 2019.

[31] Carsten Juergens. Trustworthy covid-19 mapping: Geo-spatial data literacy aspects of choropleth maps. *KN-journal of cartography and geographic information*, 70(4):155–161, 2020.

[32] Matthew Kay. How much value should an uncertainty palette suppress if an uncertainty palette should suppress value? statistical and perceptual perspectives. 2019.

[33] Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 2016 chi conference on human factors in computing systems*, pages 5092–5103, 2016.

[34] Yea-Seul Kim, Paula Kayongo, Madeleine Grunde-McLaughlin, and Jessica Hullman. Bayesian-assisted inference from visualized data. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):989–999, 2021.

[35] Christoph Kinkeldey, Alan M MacEachren, and Jochen Schiewe. How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4):372–386, 2014.

[36] Petr Kubíček and Čeněk Šašinka. Thematic uncertainty visualization usability–comparison of basic methods. *Annals of GIS*, 17(4):253–263, 2011.

[37] Shahid Latif, Siming Chen, and Fabian Beck. A deeper understanding of visualization-text interplay in geographic data-driven stories. In *Computer Graphics Forum*, volume 40, pages 311–322. Wiley Online Library, 2021.

[38] Le Liu, Alexander P Boone, Ian T Ruginski, Lace Padilla, Mary Hegarty, Sarah H Creem-Regehr, William B Thompson, Cem Yuksel, and Donald H House. Uncertainty visualization by representative sampling from prediction ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2165–2178, 2016.

[39] Grace Lochhead. Choropleth maps, and the different methods associated., Apr 2021.

[40] Alan M MacEachren. Visualizing uncertain information. *Cartographic perspectives*, (13):10–19, 1992.

[41] Alan M MacEachren, Cynthia A Brewer, and Linda W Pickle. Visualizing georeferenced data: representing reliability of health statistics. *Environment and planning A*, 30(9):1547–1561, 1998.

[42] Alan M MacEachren, Anthony Robinson, Susan Hopper, Steven Gardner, Robert Murray, Mark Gahegan, and Elisabeth Hetzler. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3):139–160, 2005.

[43] Peter Mooney and Levente Juhász. Mapping covid-19: How web-based maps contribute to the infodemic. *Dialogues in Human Geography*, 10(2):265–270, 2020.

[44] Zachary Munn, Sara Twaddle, Duncan Service, Elaine Harrow, Patrick Mbah Okwen, Holger Schünemann, and Per Olav Vandvik. Developing guidelines before, during, and after the covid-19 pandemic, 2020.

[45] AM Nivala, LT Sarjakoski, A Jakobsson, and E Kaasinen. Usability evaluation of topographic maps in mobile devices. In *Proceedings of the 21st International Cartographic Conference*, volume 10, page 2003, 2003.

[46] Noëlle Rakotondravony, Yiren Ding, and Lane Harrison. Probablement, wahrscheinlich, likely? a cross-language study of how people verbalize probabilities in icon array visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[47] Anthony C Robinson, Urška Demšar, Antoni B Moore, Aileen Buckley, Bin Jiang, Kenneth Field, Menno-Jan Kraak, Silvana P Camboim, and Claudia R Sluter. Geospatial big data and cartography: research challenges and opportunities for making maps that matter. *International Journal of Cartography*, 3(sup1):32–60, 2017.

[48] Robert E Roth. An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2356–2365, 2013.

[49] Jochen Schiewe. Empirical studies on the visual perception of spatial patterns in choropleth maps. *KN-Journal of Cartography and Geographic Information*, 69(3):217–228, 2019.

[50] Michail Schwab, Sicheng Hao, Olga Vitek, James Tompkin, Jeff Huang, and Michelle A Borkin. Evaluating pan and zoom timelines and sliders. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–12, 2019.

[51] Bettina Speckmann and Kevin Verbeek. Necklace maps. *IEEE Trans. Vis. Comput. Graph.*, 16(6):881–889, 2010.

[52] Georgianna Strode, John Derek Morgan, Benjamin Thornton, Victor Mesev, Evan Rau, Sean Shortes, and Nathan Johnson. Operationalizing trumbo's principles of bivariate choropleth map design. *Cartographic perspectives*, (94):5–24, 2019.

[53] The New York Times. See how vaccinations are going in your county and state, Dec 2020.

[54] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131, 1974.

[55] Fumeng Yang, Maryam Hedayati, and Matthew Kay. Subjective probability correction for uncertainty representations. pages 836:1–836:17, 2023.

[56] Yifan Zhang and Ross Maciejewski. Quantifying the visual impact of classification boundaries in choropleth maps. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):371–380, 2017.

[57] Brian J Zikmund-Fisher, Holly O Witteman, Mark Dickson, Andrea Fuhrel-Forbis, Valerie C Kahn, Nicole L Exe, Melissa Valerio, Lisa G Holtzman, Laura D Scherer, and Angela Fagerlin. Blocks, ovals, or people? icon type affects risk perceptions and recall of pictographs. *Medical decision making*, 34(4):443–453, 2014.