

Understanding Reader Takeaways in Thematic Maps Under Varying Text, Detail, and Spatial Autocorrelation

Arlen Fan*

Fan Lei*

Michelle V. Mancenido

Ross Maciejewski

afan5@asu.edu

flei5@asu.edu

mvmancenido@asu.edu

rmacieje@asu.edu

Arizona State University

Tempe, AZ, USA

Alan M. MacEachren

maceachren@psu.edu

Pennsylvania State University

State College, USA

ABSTRACT

Maps are crucial in conveying geospatial data in diverse contexts such as news and scientific reports. This research, utilizing thematic maps, probes deeper into the underexplored intersection of text framing and map types in influencing map interpretation. In this work, we conducted experiments to evaluate how textual detail and semantic content variations affect the quality of insights derived from map examination. We also explored the influence of explanatory annotations across different map types (e.g., choropleth, hexbin, isarithmic), base map details, and changing levels of spatial autocorrelation in the data. From two online experiments with $N = 103$ participants, we found that annotations, their specific attributes, and map type used to present the data significantly shape the quality of takeaways. Notably, we found that the effectiveness of annotations hinges on their contextual integration. These findings offer valuable guidance to the visualization community for crafting impactful thematic geospatial representations.

CCS CONCEPTS

- Human-centered computing → Empirical studies in visualization; Information visualization.

KEYWORDS

Visualization, text, annotation, design, maps

ACM Reference Format:

Arlen Fan, Fan Lei, Michelle V. Mancenido, Ross Maciejewski, and Alan M. MacEachren. 2024. Understanding Reader Takeaways in Thematic Maps Under Varying Text, Detail, and Spatial Autocorrelation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*,

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0330-0/24/05...\$15.00

<https://doi.org/10.1145/3613904.3642132>

May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 17 pages.
<https://doi.org/10.1145/3613904.3642132>

1 INTRODUCTION

Text elements, such as titles, captions, and supplementary information, have been shown to improve accessibility [3], comprehension [8, 29, 37] and the speed of information conveyance[36] in various domains of application (e.g., education, news reporting, online retail). However, the majority of the research studies on the addition of text elements have been focused on simple data visualizations, such as line charts [24, 48]. Design recommendations for augmenting text elements in thematic maps, as representations of complex, geospatial data, are lacking in the literature.

Thematic maps, which use symbols, colors, and patterns to portray statistical data, are widely used in domains that rely on geospatial information (such as urban planning [52], public health [41], environmental science [18], etc.) often revealing interesting trends and patterns across spatial dimensions. Maps, in general, have also become popular in digital platforms. In 2018, it was estimated that maps comprise 30% of all D3.js visualizations found on the internet [2]. Because geospatial visualizations are inherently complex, designing effective and engaging maps for digital consumption poses unique challenges in comparison to other popular visualization techniques (e.g. line, bar, pie charts).

The objective of this work is to provide *design guidelines* for integrating text information into thematic maps to optimize the quality of reader takeaways. To accomplish this objective, we examine the effects of text and map-related elements on reader takeaways and comprehension by running two experiments in the crowdsourcing platform Prolific.co. In contrast with previous work [24, 48], we used a more comprehensive set of text (e.g., semantic level) and map-related attributes (e.g., map type) to reflect the unique complexities of summarizing and presenting geospatial data. The majority of these attributes and their interactions were purposely manipulated in the preparation of the maps being presented to the $N = 103$ participants. Nine major hypotheses on the effect of text elements on the rated quality of participant takeaways were formulated and tested through regression-type analysis of categorical data [49]. The results of the modeling procedure were analyzed and

interpreted to determine which subset of variables had significant impacts on the quality of takeaways.

This study provides the following major contributions:

- It is demonstrated, via empirical data, that the effect of thematic map type on the granularity and semantic level of reader takeaways is dependent on the level of spatial detail and the principled design of text elements;
- It is shown how changes in the geographic detail of thematic maps, in combination with changes in the design of related text elements, affect the granularity and semantic level of reader takeaways;
- It is demonstrated through empirical evidence that a reader's behavioral inclination to consult either the visual or text elements of a thematic map can be influenced by manipulating these elements;
- Finally, we propose practical design guidelines when integrating text elements into thematic maps depending on the desired objectives on reader takeaways.

2 RELATED WORK

This work is grounded on theoretical foundations from visualization and cartography. This section provides a brief overview of relevant prior work on the topic.

2.1 Human Interaction with Text in Visualizations

The role of text elements in visualization remains a polarizing topic, with some work concluding that visual elements are more influential than text (e.g. [24, 34]) while other work suggests the opposite (e.g. [4, 25]). For example, Kim et al. [24] showed that when both chart and caption emphasized a high-prominence feature, it was predominantly considered the main takeaway by readers *i.e.*, readers were found less likely to use information from the text to form their takeaways. Similarly, O'Brien and Lauer [34] found that deceptive techniques (such as truncated or inverted axes) caused readers to misinterpret information even when paired with accurate explanatory text. These works suggest that a reader's attention is more naturally influenced by the visualization rather than textual attributes.

In contrast, several studies support the view that text annotations can have a strong influence on a reader's focus, inferred takeaways, and preferences. Borkin et al.'s eye-tracking study [3] found that titles and text are key elements in a visualization that help with recall, with readers fixating on text even if it did not appear in the title. Kong et al. [25] found that the slant of the title in a visualization influenced the perceived main idea, with readers arriving at divergent takeaways from the same visualization. Ottley et al. [36], also through an eye-tracking study, found that visualizations *primarily* facilitated the identification of main topics. Once the main topics were identified, however, it was found that readers extracted information from the text annotations. Additionally, participants with a preference for visual information over text information still preferred charts with a higher number of annotations. Finally, Hearst and Tory [21] examined individual preferences for visualizations when interacting with chatbots, finding that more than half of

surveyed participants preferred seeing text annotations without visualizations.

A recent work closest in spirit to our study is Stokes et al. [48], which examined the interplay between text and visual elements in line charts. Their work captured reader preferences and takeaways when using these two modalities and distilled major findings into design guidelines for integrating text annotations into line charts. However, this study only examined annotations in univariate line charts and used synthetic datasets with abrupt peaks and trends. In contrast, our study expands on Stokes et al.'s initial findings by addressing several limitations: (1) we explore a more complex visualization technique (maps) and its interplay with text annotations; (2) we use real-world datasets and apply accepted methods in thematic cartography to generate base maps. Through preliminary discussions with a cartography expert, it is hypothesized that the guidelines for line charts as recommended by Stokes et al. may not naturally extend to thematic maps, due to the nuanced complexities of geospatial datasets and related visualization techniques. Although recent research by Stokes et al. [47] supports using text-only options when presenting data information, we excluded this option due to the intricacy of maps and possible biases arising from converting geospatial data to text (this limitation is further discussed in Section 6).

2.2 Thematic Map Design

A thematic map is a type of map that emphasizes the spatial distribution of a particular theme, topic, or attribute within a specific geographic area. Instead of showing traditional geographical features like rivers, mountains, and political boundaries, thematic maps use various visual elements such as colors, symbols, and patterns to represent data related to a specific subject (a theme) [14]. For example, a population density thematic map might use different shades of color to depict the concentration of people in different regions, with darker colors indicating higher population densities. Thematic maps are valuable tools for visualizing and analyzing spatial data for a wide range of applications, including urban planning, environmental studies, and social sciences [43].

Spatial Autocorrelation: A critical visual task when using thematic maps is the identification of spatial clusters. Typically, *color* changes have been used to provide visible clustering among regions that have similarities or differences. The measurement of *spatial autocorrelation* with respect to an attribute of interest has been the focus of previous work on clustering map components. Some popular metrics proposed include: Join count statistic [10] developed for binary variables based on the probability that a unit area belongs to the same class as its adjacent areas; Moran's I [32] which considers pairwise products of deviations; and Geary's C [17] which uses the sum of squared distances. In this study, Moran's I was selected due to its emphasis on the global detection of similarities among regions in contrast with Geary's C, which is more appropriate for detecting spatial heterogeneity.

Classification Scheme: After identifying and calculating appropriate spatial autocorrelation metrics, the analysis naturally segues into choosing a classification scheme, a key aspect of thematic map design. While spatial autocorrelation measures the extent of spatial association among individual observations, classification

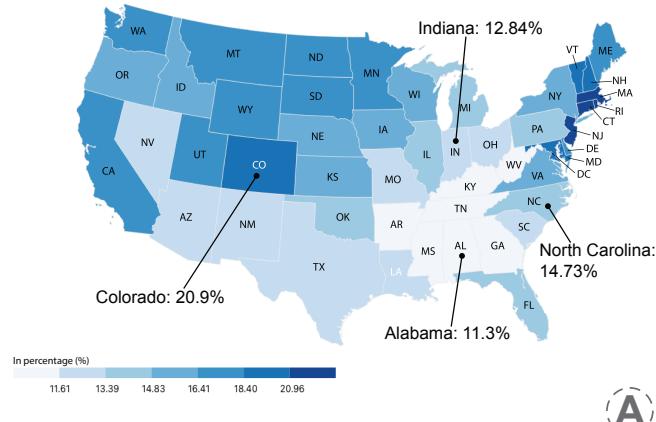
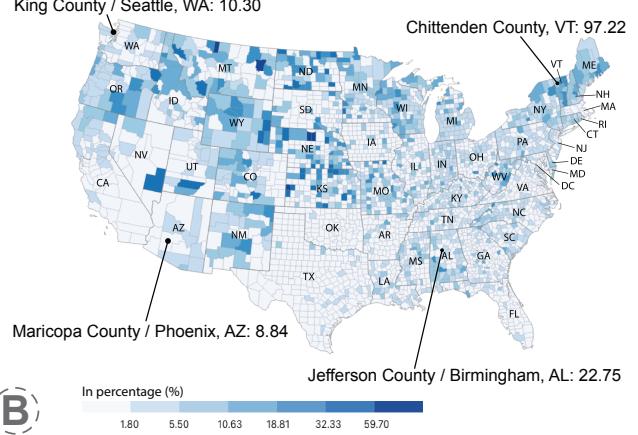
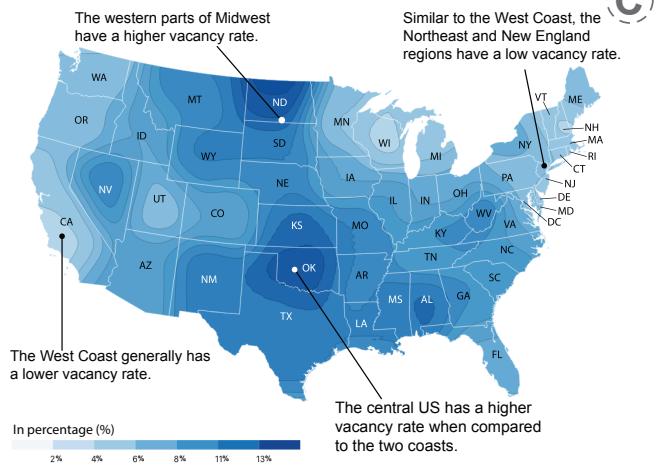
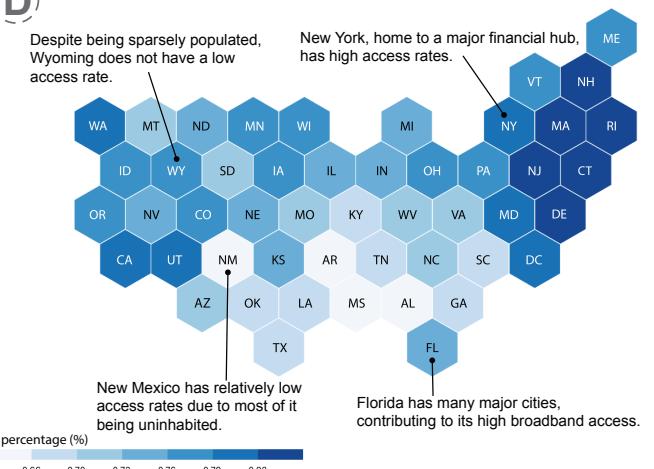
US population rate with at least a bachelor's degree, 2019**Farmer's markets per 1,000,000 population in the U.S.****US apartment vacancy rates 2019****US home broadband access rate in 2017**

Figure 1: Different types of maps (which are bolded in this caption) with varying semantic levels of text as described by Lundgard and Satyanaryan [31] for classifying text in visualizations. A U.S. State-level choropleth map (A) and a county-level choropleth map (B) using descriptive statistics (Semantic Level 2 or perceiver-independent) text annotations. (C) An isarithmic map with text annotations describing complex and overall trends (Semantic Level 3 or perceiver-dependent). (D) A hexbin map, which is a cartogram variant, containing text annotations with external or background information (Semantic Level 4 or perceiver-dependent).

schemes provide a set of rules for assigning each map region to a class. Selection of the classification scheme plays a major role in the appearance of spatial clusters, which could reveal key trends or patterns that would otherwise not be readily apparent [28]. For univariate data, the simplest class selection methods rely on quantiles and standard deviations [30]. More complex methods include Jenks' natural breaks [23], which maximizes interclass variance and minimizes intraclass variance; Cromley's [11] minimum boundary error, which seeks to maximize spatial similarity among adjacent areas in the same class; and Jenks Caspall [23], which minimizes the deviation around class means using a heuristic process. Each method results in a different map appearance, which underlines the importance of the judicious selection of a classification scheme

based on the statistical properties of the dataset. After careful consideration of the attributes of the dataset used in this study, we used the 7-class Fisher-Jenks method as our primary classification scheme for all generated maps.

Text in Maps: In addition to graphical components, it is essential to highlight the distinct roles and importance of textual elements in maps, a notion supported by previous research [15]. The effectiveness of textual elements largely hinges on their presentation. It is important to differentiate between text annotations or narratives, which supply pivotal additional content and explanations, and labels, which identify and delineate the elements they are attached. Traditionally, cartographic research has primarily centered on the

study of text labels, focusing on enhancing their legibility and facilitating the correct association with the respective features they denote; a guidance substantiated in the detailed categorization by Dent et al. [13].

Data Storytelling and Data Journalism: In recent years, accessible and interactive representations of geographic information called *story maps* have also risen in popularity [42]. Story maps integrate maps and text, organized in the form of focused narratives. They have been used as an engaging method for showing compelling evidence of the rise in global sea levels [46] and the spread of COVID-19 [38]. In the context of data journalism, Song et al. [46] studied whether themes (US presidential campaign donations, US coastal sea-level rise), genres (longform infographic, dynamic slideshow), or tropes (color highlighting, leader lines), would influence reader retention or comprehension. They found that the story theme had no influence, while participants performed better using long-form infographics and leader lines. Individual audience differences by expertise and prior beliefs also impacted participant response. Our work continues to build on these findings by closely examining the interplay between visual and textual elements in the comprehension process from a data visualization and cartography standpoint. Our examination will highlight the importance of proper text annotations in conveying clear and concise information and in balancing visual and textual elements for an optimal reader experience.

3 STUDY DESIGN

In this section, we first define the manipulated variables in the study (Section 3.1). Then we state our research questions and hypotheses (Section 3.2). Finally, we explain our survey in detail (Sections 3.3, 3.4, and 3.5)

3.1 Map Design Factors

We considered best practices in map design and analyzed existing gaps to identify map design factors for evaluation. To inform the map design factors tested in this study, we consulted Slocum et al.'s [43] textbook on thematic cartography and selected elements that could yield measurable variations in reader takeaways. Additionally, we drew insights from various other sources, including Stokes et al.'s [48] work to further inform the design of the study. Ultimately, we identified these five elements: Map Type, Map Detail, Semantic Level, Spatial Autocorrelation, and Text-Map Detail Alignment.

Map Type: Map types, as a fundamental design factor, differ with respect to their purpose and distinct representations [33]. In this study, we limit our scope to common map types used for quantitative thematic mapping [14].

Slocum et al.'s textbook [46] on thematic cartography categorizes three primary map types suitable for visualizing a univariate measure in a geographic context, each accompanied by relevant examples:

- **Choropleth Maps:** The most popular type of map for displaying statistical data for enumeration areas using an ordered color palette. A prototypical use of a choropleth map is to visualize population density or other demographic data

derived for census enumeration units such as states or counties. An example representing enumerated data is shown in Fig. 1A, B.

- **Isarithmic Maps:** The second most popular map type, isarithmic maps display continuous data by connecting points or places of equal value with contoured lines. These lines divide the spatial surface into different areas and highlight where data levels change. Isarithmic maps can depict physically continuous surfaces (e.g., elevation or air temperature) or statistically continuous surfaces generated from enumerated data for areas (e.g., population density or average income). The latter use the same input data as choropleth maps and are the focus here. An example is shown in Fig. 1C.
- **Cartograms:** Similar to choropleth maps but the spatial geometry is distorted for visual effect. Cartograms are used in media and news, especially for political campaign analysis and sales data visualization. In this study, we opted for a specific variant of cartogram known as the hexbin map, where equal-area hexagons represent states in the U.S. While there are other types of cartograms, such as area and shape-warping, our focus is solely on the hexbin type since, like both the choropleth and isarithmic maps included in the study, they use color shading to depict univariate data values. An example is shown in Fig. 1D.

Map Detail: A map's geographic detail has been known to influence its legibility and a reader's cognitive load for processing the presented information [1]. While this factor [6] has been theoretically considered in past studies, its effect on reader takeaways has not been studied, specifically in the presence of changing text elements. In this study, we chose to use U.S. maps at the *state* or *county* level, with Alaska and Hawaii omitted.

Semantic Level: We apply Lundgard and Satyanaryan's framework [31], which classifies accompanying text for visualizations into four levels. Originally intended to improve text accessibility for the visually impaired, Stokes et al. [48] utilized this framework to examine different aspects of a visualization's content.

- **Semantic Level 1 (L1):** Refers to elemental and encoded properties, such as chart type, encoding channels, title, and labels. L1 text is considered to be **perceiver-independent** i.e., perceptions or interpretations are not expected to change from one reader to the next.
- **Semantic Level 2 (L2):** Covers descriptive statistics such as outliers, extrema, or point-wise comparisons. Similar to L1, text at this level is **perceiver-independent**.
- **Semantic Level 3 (L3):** Describes perceptual and cognitive aspects, such as trends and patterns. Text at this level is considered to be **perceiver-dependent**, i.e., information inferred is contingent on a reader's interpretation and perception.
- **Semantic Level 4 (L4):** Refers to external information, such as past and current events that supplement the topic. Similar to L3, these types of annotations are **perceiver-dependent**.

Spatial Autocorrelation: Changes in spatial autocorrelation result in changes in map appearance, as well as information regarding associations among adjacent locations [9, 27]. Examining

Dataset	Experiment 1			Experiment 2			Group
	Text-Map Detail Align.	Semantic Level	Map Type	Text-Map Detail Align.	Semantic Level	Map Detail	
1	unaligned	perceiver dependent	choropleth	aligned	level 2	county	A
1	aligned	perceiver independent	isarithmic	unaligned	level 3	state	B
2	aligned	perceiver dependent	isarithmic	unaligned	level 4	county	A
2	aligned	perceiver dependent	hexbin	unaligned	level 2	county	B
3	aligned	perceiver independent	hexbin	aligned	level 3	state	A
3	aligned	perceiver dependent	choropleth	aligned	level 4	county	B
4	unaligned	perceiver dependent	isarithmic	unaligned	level 2	state	A
4	unaligned	perceiver independent	choropleth	unaligned	level 3	county	B
5	unaligned	perceiver independent	isarithmic	aligned	level 4	state	A
5	unaligned	perceiver independent	hexbin	aligned	level 2	state	B
6	aligned	perceiver independent	choropleth	aligned	level 3	county	A
6	unaligned	perceiver dependent	hexbin	unaligned	level 4	state	B

Figure 2: Experimental design constructed using JMP®software. The input variables—comprising text-map detail alignment, semantic level, map type, and map detail—were systematically varied to rigorously assess both individual and interaction effects. JMP®was employed to optimize the design of experiments, ensuring comprehensive evaluation of all interaction effects. In Experiment 1, all of the maps are at the state level. In Experiment 2, all maps are choropleth maps, which means that the variable **Map Type** is constant.

spatial autocorrelation as a factor in map perception allows for the examination of how readers perceive spatial patterns and clusters, which is crucial for understanding spatial cognition. A pioneering study by Olson [35] examined the relationship between people's perception of complexity and autocorrelation and clustering. The experiment found that highly clustered visualizations were generally rated as less complex, while dispersed visualizations were seen as more complex. Participants' open-ended feedback supported these findings. Later, Bunch and Lloyd [6] explored various types of cognitive load in the context of geographic information, distinguishing between intrinsic load (associated with inherent complexity), extraneous load (related to presentation format), and germane load (linked to processing and schema automation). Given these insights, we hypothesize that spatial autocorrelation influences reader understanding.

In this paper, we use the metric Moran's I [32], which calculates the degree of similarity between neighboring observations to measure a map's global spatial autocorrelation. Moran's I is sensitive to broad patterns and is ideal for identifying overarching trends in spatial datasets, making it a suitable measure of spatial autocorrelation for the purposes of our study.

T Text-Map Detail Alignment: This variable focuses on the relationship between the granularity of text annotations and the corresponding map detail. When the level of detail in text annotations matches that of the map, such as county-level text on a county-level map or state-level text on a state-level map, it is considered aligned. Conversely, misalignment occurs when the granularity of text annotations does not correspond with the map detail. For instance, county-level text annotations on a state-level map represent a misalignment. Our interest in this variable stems from a hypothesis that the alignment (or misalignment) of these elements is not merely a design choice, but a factor that can lead to interesting and measurable differences in how readers process and understand thematic maps. We included this variable to primarily

explore cognitive processing in map reading, as misaligned text and map detail may pose cognitive challenges, requiring readers to integrate separate pieces of information.

To effectively author annotations, we consulted Fairbairn's [15] paper, which lists a comprehensive taxonomy of text symbols that can appear on maps. They can be (1) descriptive, which reflects features that are symbolized on the map; (2) analytical, which links the reader with feature attributes; and (3) positional, which describes or confirms location. These text types have further subcategories. However, as our study only utilizes univariate thematic maps, many of these text types are not applicable. For instance, it would be unsuitable to include warning text or longitude and latitude markings on choropleth maps. Consequently, we will use descriptive (L1), determinative (L2), interpretative (L3-L4), and temporal/positional (L4) text in the study.

3.2 Research Questions

While past work [19] has focused on exploring the effects of text and map design factors on a reader's comprehension, our work differs in that we considered the interdependence of these factors on novel measures of design effectiveness via the takeaways provided. First, the participants were asked to formulate a minimum of 1 to a maximum of 5 takeaways, where each takeaway is a set of sentences describing the most important information gleaned from the annotated map. Then we asked participants to rate the perceived influence of map and text in their process of obtaining information from an annotated map. The takeaways were then rated by the project team with respect to the level of geographic detail of information provided (county, state, or region) and semantic level. Thus, the following research questions and related hypotheses focused on the effects of the design variables on these three performance metrics: source, detail, and semantic level of takeaways.

Research Question 1 (RQ1): How does the design of a map influence a reader's primary source for information (map or text annotations)?

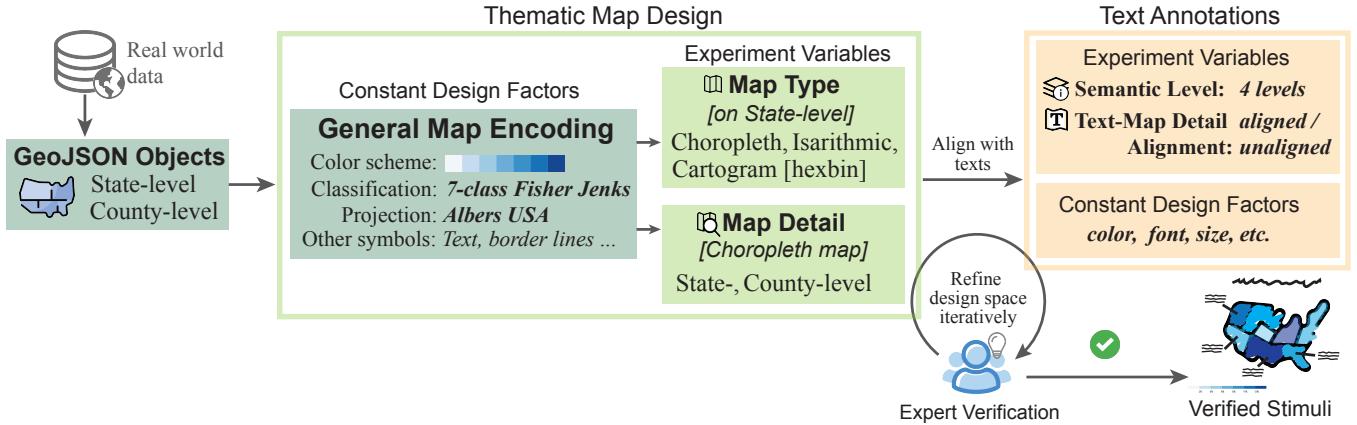


Figure 3: The stimuli construction pipeline. Data were mapped onto GeoJSON objects and thematic map design principles were applied to generate all stimuli. Map design factors were either varied or kept constant as required by our experimental design (Fig. 2). Maps were validated for adherence to best practices in thematic map design. Finally, annotations were added with differing semantic levels and reviewed a final time to produce verified stimuli.

- **Hypothesis 1a (H1a): Map type influences a reader's reliance on text annotations.**
 - *Explanation:* Distorted maps such as cartograms might lead to increased reliance on textual information due to their unfamiliar graphical representations. The semantic level of text annotations could further modulate this effect.
- **Hypothesis 1b (H1b): Increased geographic detail in maps leads to greater reliance on text annotations.**
 - *Explanation:* Finer geographic detail can increase cognitive load, potentially enhancing reliance on high-level semantic text annotations for clearer understanding.
- **Hypothesis 1c (H1c): Higher spatial autocorrelation in maps might reduce reliance on text annotations.**
 - *Explanation:* Maps with high spatial autocorrelation tend to exhibit clearer patterns, possibly reducing the need for text annotations. The map type might alter this effect.

Research Question 2 (RQ2): How does map design affect the granularity of takeaways considering the map type, geographic detail, and text content?

- **Hypothesis 2a (H2a): Map type affects the granularity of a reader's takeaways.**
 - *Explanation:* Different map types like isarithmic maps may draw attention to broader regions, influencing the level of detail in the takeaways.
- **Hypothesis 2b (H2b): The granularity of takeaways varies with map detail.**
 - *Explanation:* The level of detail in both the map and text can influence the specificity of a reader's interpretation, with consistency between the two reinforcing certain levels of detail.
- **Hypothesis 2c (H2c): Higher spatial autocorrelation leads to takeaways with coarser detail.**
 - *Explanation:* Spatially autocorrelated data often forms clusters, steering readers towards broader, cluster-based interpretations.

Research Question 3 (RQ3): How does map design impact the semantic level of takeaways in relation to geographic detail and text content?

- **Hypothesis 3a (H3a): Map type influences the semantic level of a reader's takeaways.**
 - *Explanation:* Certain maps, like isarithmic ones, may encourage higher-level semantic interpretations due to their focus on broader regions.
- **Hypothesis 3b (H3b): Coarser map details lead to higher-level semantic takeaways.**
 - *Explanation:* Describing phenomena across larger areas typically falls into higher semantic levels, with the strength of this effect influenced by the semantic level of text annotations.
- **Hypothesis 3c (H3c): Spatial autocorrelation within a map dataset influences the semantic level of takeaways.**
 - *Explanation:* Higher spatial autocorrelation, particularly in maps like isarithmic ones, can lead to more complex and higher-level semantic interpretations.

All hypotheses and analyses scripts are preregistered on OSF¹.

3.3 Participants

Participants for our study were recruited via Prolific.co, with eligibility criteria including U.S. residency, desktop computer usage, and a minimum 95% task acceptance rate on Prolific.co. The survey, designed for a 20-minute completion time, offered a \$4 compensation upon successful completion, equating to an hourly rate of \$12. We considered 103 participants in our results (65 male, 36 female, 2 non-binary, $M_{age} = 37.2$, $SD_{age} = 23.3$). The majority (61) had at least a 4-year degree.

¹https://osf.io/3pzvj/?view_only=86167a2f0ba041899f6edee0c73c6f68. This is an anonymous, read-only link

Dataset	Variable plotted	Moran's I	
		state	county
1	2021 US Farmer's markets density [USDA]	0.41	0.24
2	2017 US home broadband access rates*	0.43	0.49
3	2019 US income per capita [US Census]	0.13	0.46
4	2019 US apartment vacancy rates [US Census]	0.41	0.55
5	2019 US unemployment rates [US Census]	0.30	0.62
6	2019 US population rates with at least a bachelor's degree [US Census]	0.51	0.44

Figure 4: The 6 datasets used in the study. The Moran's I value was calculated by using the choropleth map. *Dataset 2 is from Tolbert et al. [50]

3.4 Stimuli

The experimental design, as depicted in Fig. 2, was developed using the JMP® statistical analysis tool, where **Map Type**, **Map Detail**, **Spatial Autocorrelation**, **Semantic Level**, and **Text-Map Detail Alignment** were considered to construct an efficient experimental design. We began by assembling six real-world datasets at both the state and county levels.

Generating the stimuli followed a structured pipeline, visually represented in Fig. 3. The pipeline consisted of several key stages: encoding the assembled data into GeoJSON objects, applying thematic map design principles, and introducing variations or constants in map design factors as dictated by our experimental design (Fig. 2). An important consideration was the spatial autocorrelation metric, Moran's I, which is determined by the dataset properties. We documented the values of Moran's I for both the state and county levels as indicators of spatial autocorrelation for each dataset (Fig. 4).

3.4.1 Thematic Map Design. Thematic map design factors not measured in our experiment (e.g. color, classification) must be controlled, as they have an effect on reader perception and introduce biases. Thus, we use the same color scheme and classification on all of our maps. We considered multiple classification division methods for our maps and evaluated their performance using Smith's [44] Goodness of Variance Fit (GVF). After weighing our options, a 7-class Fisher-Jenks [23] was selected, as it consistently was in the top quartile of GVF scores. The legend is divided into 7 data ranges, and the class breaks were defined by the Fisher Jenks classification criteria [39], which minimizes deviation around class means using a heuristic process. We chose ColorBrewer's [20] 7-class sequential blue color scheme to depict the classes that the method defined.

After selecting the color scheme and classification method, we encoded symbols (e.g. border width, stroke, and typeface) onto the map. Following design guidelines from Dent et al. [14], we added states standard two-letter abbreviations using 18pt Arial font and drew state borders with a 0.5pt width. For states that are too small to fit the text, solid black leader lines with 1pt width were used. All maps used the Albers USA projection [45] except for the hexbins. After unifying the general map design space, we generated thematic maps by varying the **Map Type**, which took on one of three possible values: choropleth map, isarithmic map, or hexbin map. All maps were implemented in D3.js [5] and exported into the SVG format.

Choropleth map: For each dataset, we generated state-level (Fig. 1A) or county-level choropleth maps (Fig. 1B) by varying the **Map Detail**. We grouped them into subsets (see Fig. 2).

Isarithmic map: Isarithmic maps, Fig. 1C, are used to illustrate the density of the observations on a map [14]. To convert discrete area data into continuous contour data, we use Kernel Density Estimation (KDE) on county-level choropleth maps. For each isarithmic map, we employed the fixed bandwidth KDE as described by Wang et al. [51].

Cartogram: Cartograms are widely used by media for storytelling. However, due to the absence of strict sorting criteria for grid map layouts used in data journalism, there are numerous variations [7]. In this case, we applied a popular grid layout practice from NPR's Danny Debelius [12] which uses tessellated, equal-area hexagons (Fig. 1D) to represent states. As stated earlier, we use the hexbin variant exclusively in the study.

We recognized that certain combinations of map types and annotations might not be as effective, such as hexbin maps paired with annotations discussing aggregate patterns in regions like the western parts of the Midwest. Our methodology accounted for this, hypothesizing that different map types might have varying suitability for annotations at particular semantic levels.

3.4.2 Annotation Content and Placement. The **Semantic Level** refers to the framework outlined by Lundgard and Satyanaryan [31]. L1 text, such as title and legend, was present in all map stimuli. Annotations were manually authored in order to ensure that the text content was consistent with the semantic level described. For L2 annotations, descriptive statistics were used (See Fig. 1A).

Aware of the potential influence of titles and framing effects on interpretation and takeaways, as discussed in Hullman and Diakopoulos's work on visualization rhetoric [22], we aimed to keep titles impartial, stating only the statistic being plotted without additional commentary or framing. This strategy was intended to minimize any bias. Additionally, our map legends were deliberately left unlabeled, as the titles themselves were designed to directly convey that information.

We used the text templates for L3 annotations as Fig. 5 shows. Templates #1 and #2 focus on perceiver-dependent phenomena or subjective, imprecise descriptions of the data, whereas #3 compares individual points to multiple points or entire regions. Templates #4 and #5 address complex trends and observations. To develop L4 annotations, we searched Wikipedia for relevant articles and chose information that offered suitable explanations for map areas. For

#	Template	Example
1	[point/region] has a [comparison word] than average [value].	Idaho has a lower than average income.
2	[point/region] has a(n) [adjective/adverb] [comparison word] [value].	The West Coast generally has a lower vacancy rate.
3	[point/region] has a [comparison word] [value] than [region].	The central US has a higher vacancy rate than two coasts.
4	[value] are [clustered/scattered] in [region(s)].	People with a college degree are scattered in New England.
5	There are more higher [value] than lower ones in [region(s)].	There are more counties with lower college education rates in the US.

Figure 5: Templates used for L3 annotations.

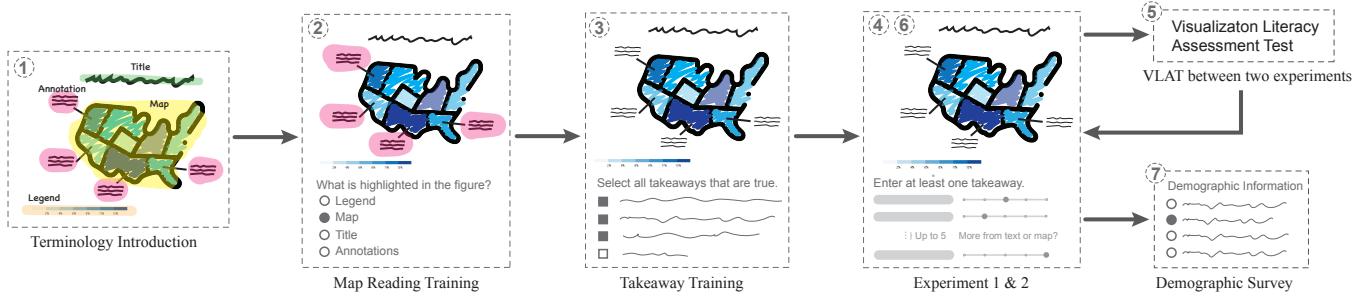


Figure 6: Participants workflow with seven sections. 1) terminology introduction 2) map reading training, 3) takeaway introduction and training, 4) the first experiment investigates how different map types interacted with varying textual elements, 5) VLAT, 6) the second experiment examines the interplay between choropleth maps' varying geographic detail and textual elements, and 7) demographics.

perceiver-dependent annotations, we used either L3 or a mix of L3 and L4, with each map containing four annotations.

We drew leader lines manually, as no practical implementation was identified in the existing literature. We considered creating annotations that coincided with visually salient features as demonstrated in Kim et al.'s study [24]; however, our research lacks a corresponding crowdsourcing component. Still, we followed Dent et al.'s map construction guidelines [14] to place critical elements such as the title, legend, and map itself. Starting with a blank canvas, we placed each element and calculated the remaining workable map space, repeating this process until all elements were situated, resulting in a visually balanced map.

3.5 Survey Measures

The full survey pipeline is detailed in Fig. 6. Before the study starts, the participants are either assigned to Group A or B. The exact stimuli seen for each group are in Fig. 2. The first section was an attention check and terminology introduction. Participants were shown a map with four basic components of thematic maps: legend, title, map, and annotations. After seeing this introduction page, in the training section, they were asked to identify two of the four components at random. Participants must answer both correctly to continue.

As part of our study, we focused on gathering data on specific output variables to understand how participants interact with and interpret the map stimuli. Our primary output variables are:

Source: An ordinal measure of where the reader obtained the takeaway. This variable can assume five different values, representing the source of the reader's takeaway; a value of 1 denotes deriving insight solely from the text, while a value of 5 signifies solely relying on the map. Values between 1 and 5 indicate varying degrees of combined influence from both the text and the map.

Takeaway Granularity: This variable captures the level of detail in the participants' takeaways, categorized into county, state, or region.

Takeaway Semantic Level: Assessing the semantic depth of the takeaways, classified according to the framework outlined by Lundgard and Satyanaryan [31].

The third section was a continuation of the training, primarily designed to inform the participant what the takeaways are and how to form them after reading a visualization. We defined a takeaway as “a key fact, point, or idea to be remembered after viewing the map”, adapted from Stokes et al.’s [48] study. We provide a list of examples of takeaways and a brief explanation of how they were formed. We clarified that a reader can form a takeaway from reading the annotations about the map, reading the map only, or a combination of both. After this training, participants were asked two multiple-choice questions. There were four answer choices and three of them were correct. The incorrect answer was a direct contradiction to one of the annotations.

The fourth section of the study consisted of the stimuli in Experiment 1. Each participant saw six maps and was asked to write their takeaways for each map. They were allowed a maximum of five takeaways and a minimum of one takeaway per map. For each takeaway, participants reported how they formed their takeaways on a 5-point scale (1- all from text, 2- more text, 3- same, 4- more map, 5- all from the map), which constitutes the **Source** variable.

The fifth section of the study was an abridged version of the Visual Literacy Assessment Test [26], a validated way of assessing critical reading of various data visualizations such as line, pie, bar, stacked bar, etc. Since our study is primarily concerned with spatial data, we used questions pertaining to scatterplots, bubble charts,

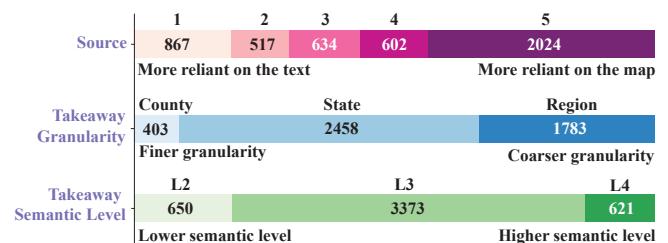


Figure 7: A statistical count of all of the coded takeaways. Horizontal bars stack to 4,644, the total number of takeaways analyzed in the study. **Source** took on 5 values on a 5-point scale (1- all from text, 2- more text, 3- same, 4- more map, 5- all from the map)

Fixed Effect	Experiment 1				Experiment 2		
	Source	Takeaway Granularity	Takeaway Semantic Level	Source	Takeaway Granularity	Takeaway Semantic Level	
Semantic Level	p < 0.0001	x	p < 0.0001	p = 0.0008	p < 0.0001	p < 0.0001	
Map Type	p = 0.431	p = 0.0003	p < 0.0001	x	x	x	
Spatial Autocorr.	x	p = 0.6221	p = 0.9777	x	p < 0.0001	x	
Text-Map Detail Align.	x	p < 0.0001	p = 0.0349	p = 0.0273	p < 0.0001	x	
Map Detail	x	x	x	p = 0.0164	p < 0.0001	p < 0.0001	
SL × MT	p = 0.0004	x	x	x	x	x	
SA × MT	x	p = 0.0001	p < 0.0001	x	x	x	
SA × MD	x	x	x	x	p < 0.0001	x	
T-MDA × SL	x	x	x	x	p < 0.0001	x	
T-MDA × MT	x	x	p < 0.0001	x	p < 0.0001	x	

Figure 8: Table of Fixed Effects. All independent variables are **bolded and colored green** and dependent variables are **bolded and colored purple**. **Semantic Level** was abbreviated to SL, **Map Type** to MT, **Spatial Autocorrelation** to SA, **Text-Map Detail Align** to T-MDA, and **Map Detail** to MD. Cells containing p-values indicate that the particular result was tested in our hypotheses. Grayed-out boxes indicate that the result is not significant at the $\alpha = 0.05$ level. Crossed-out boxes indicate that the result is discarded due to non-significance or is not of interest.

and choropleth maps, for a total of 17 questions. The purpose of this section was to reduce stimuli fatigue by adding variance to the tasks. It is also to properly test participants on their ability to critically read spatial data visualizations in lieu of relying on a self-reported metric. The sixth section of the study consisted of the stimuli in Experiment 2. Mechanics from Experiment 1 were repeated. Finally, the seventh and final section was a demographic survey.

4 RESULTS

Two authors independently rated participant takeaways for the level of detail (county, state, region) and the semantic level. One hundred and three ($N = 103$) participants provided a total of 4,644 takeaways, an average of 45.09 takeaways for each participant. There were 55 Group A participants, who provided a total of 2,534 takeaways total and 46 Group B participants, who provided 2,110 takeaways in total. The coders disagreed on the level of detail in 3.15% of all responses and on the semantic level in 6.72% of all responses. A third coder resolved all of the remaining conflicts except for 9. The remaining conflicts were discussed among the three coders. Fig. 7 shows an overview of the coded takeaways.

Out of the total takeaways, 44 from 15 distinct participants were categorized as off-topic. These were further subdivided into three categories: 19 were classified as blanks, which were zero-character answers submitted for the sole purpose of advancing the survey questions; 4 were related to technical difficulties, such as issues with loading the map ("The image doesn't load and I'm not allowed to refresh the page..."); and 21 were statements that did not qualify as takeaways. These non-takeaways often involved participants merely repeating the title or did not meaningfully relate to the dataset, and they could not be easily classified under semantic levels L2, L3, or L4.

4.1 Model Building

Statistical analysis was conducted using the SAS® statistical software package. Due to the non-continuous nature of some of the variables, such as **Map Detail**, we used a flexible family of models

called generalized linear mixed models (GLMM) for both estimation and tests of significance [49]. The SAS® software's GLIMMIX package was used to estimate and test different GLMM's. We initially fitted models that had active main effects and two-factor interactions (2FI); then, we used F-tests to determine and eliminate effects that had no significant impact on the response (**Source**, **Takeaway Granularity**, **Takeaway Semantic Level**) via the *p*-value criterion. Finally, the final GLMM used for analysis only consisted of main effects and/or 2FI that had *p*-values less than the significance level of 0.05. We believe that these model forms, specifically with the addition of the 2FI, were parsimonious enough to be interpretable but complex enough to capture interesting joint effects of the variables of interest.

Post-hoc analysis of the difference in contrasts between groups (i.e., categories of the independent variables) yielded estimates of odds ratios. For binary responses, odds are the ratios of the probabilities of success vs. failure; in the case of ordered data, the odds are proportional odds which are equal odds (with respect to any two categories) of the outcome being a higher-level vs. a lower level category. Odds ratios, on the other hand, are calculated as the ratios of the odds of two groups. For example, when examining the effect of a specific map design feature such as **Map Type** (independent variable) on the likelihood of achieving a higher **Source** value (outcome), an odds ratio for **Chloropleth** (Group 1) and **Isarithmic** (Group 2) that is greater than 1 would suggest that the Group 1 odds of yielding a high **Source** value is greater than the Group 2 odds. If the odds ratio is not significantly different from 1, this indicates that there is no difference between the two groups with respect to their impact on the **Source** value.

4.2 Effects on the Source of Takeaways

Hypothesis 1a (H1a): Map type influences a reader's reliance on text annotations.

Due to the ordinal nature of the **Source** variable, where higher values indicate more reliance on maps, while lower values show more reliance on text, we used a proportional odds model to analyze the impact of the map and text variables on the comparative

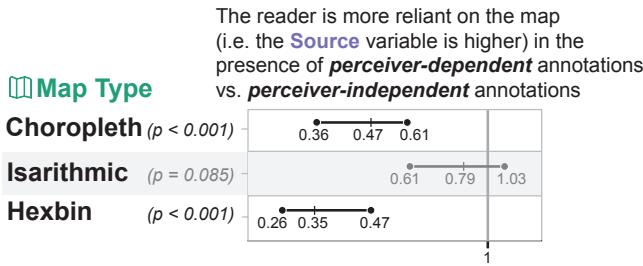


Figure 9: 95% confidence interval chart for the odds ratios model for the H1a hypotheses. The charts summarize the impact of different types of annotations on map reliance in choropleth maps. A notable trend appears: when maps include perceiver-dependent annotations (L3, L4), readers tend to rely less on the map itself, with a likelihood of about 0.47 times compared to maps with perceiver-independent (L2) annotations. In contrast, for maps with perceiver-independent annotations, readers are more than twice as likely ($1/0.47 \approx 2.13$) to rely on the map for their interpretations. This suggests a clear distinction in how different annotations influence reader dependence on the map for understanding the presented data.

odds of being more reliant on map over text (odds ratios). Results from the fitted proportional odds model show that the effect of **Map Type** on **Source** was found to be significantly dependent on the semantic level of text² i.e., the **Map Type** × **Semantic Level** interaction effect was significant ($p = 0.0004$), but not on **Text-Map Detail Alignment** ($p > 0.05$). Estimates of the multiplicative effect on the odds ratios (Fig. 9) show that for the choropleth map, perceiver-independent annotations increase the chance of using the map as the primary source of takeaways by a factor of $2.12 \times (1/0.47 = 2.12)$ when compared to perceiver dependent annotations. This implies that for every 100 people, 32 will report relying more on the map when reading a choropleth map with perceiver-dependent annotations and 68 will report relying more on the map when reading a choropleth map with perceiver-independent annotations. To further clarify, consider a sample of 100 map readers. The proportion of readers that are in the perceiver-dependent group versus perceiver-independent group can be determined by the ratio of 1 to 2.12. When these ratios are normalized (i.e., $\frac{1}{1+2.12}$ for perceiver-dependent and $\frac{2.12}{1+2.12}$ for perceiver-independent annotations), they correspond to approximately 32 and 68 readers. For participants reading an isarithmic map, utilizing perceiver-independent annotations as opposed to perceiver-dependent annotations results in a $1.27 \times (p = 0.085)$ increase in the odds of predominantly relying on the map. In the scenario of the hexbin map interaction, perceiver-independent annotations result in a $2.86 \times (p < 0.001)$ increase in the odds of a reader focusing more on the map compared to when perceiver-dependent annotations are present. Annotations at the dependent level generally increase reliance on text annotations, aligning with the hypothesis's predicted directionality.

²Based on tests on Fixed Effects for main effects and two-factor interaction effects in a proportional odds model with $\alpha = 0.05$

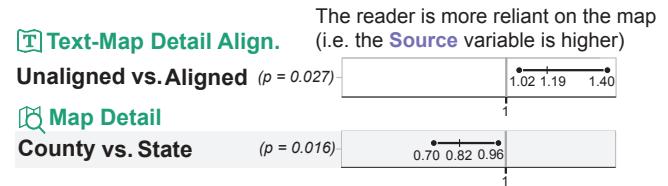


Figure 10: 95% confidence interval chart for the odds ratios model for the H1b hypotheses, which provides an overview of how text alignment and map granularity influence map reliance. This analysis reveals that readers show a $1.19 \times$ higher tendency to rely on maps when text details are unaligned with the map elements, compared to when they are aligned. Furthermore, when comparing state-level choropleth maps with county-level ones, readers show a $1.22 \times$ increased reliance on the state-level maps. These findings highlight the significant role of text alignment and map granularity in shaping how users interact with and interpret map-based information.

These results indicate that the semantic level's effect (dependent vs. independent) on takeaways is the strongest for isarithmic maps, followed by choropleth maps, and is least apparent for hexbins. This indicates that a higher semantic level tends to foster more reliance on text annotations, with this tendency being most prominent in isarithmic maps and least so in hexbins.

Hypothesis 1b (H1b): Increased geographic detail in maps leads to greater reliance on text annotations.

We used a proportional odds model to analyze the impact of the map and text variables on a higher **Source** (i.e. reader being more reliant on map). The Fixed Effect tests (Fig. 8) indicate that **Map Detail** is significant on **Source** ($p = 0.0164$). Readers who are shown county-level maps tend to display a higher reliance on text than on the map itself. The odds of a reader relying more on the map for insights are $1.22 \times$ higher ($p = 0.016$) when the map detail is at the state level compared to the county level (Fig. 10) supporting this hypothesis.

Hypothesis 1c (H1c): Higher spatial autocorrelation in maps might reduce reliance on text annotations.

This hypothesis is rejected. The results reveal that Moran's I is not significant on **Source** for both experiments (Fig. 8). Similarly, text attributes exhibit no interaction with Moran's I. This leads to the conclusion that differences in geographic correlations within the phenomenon mapped do not influence a reader's reliance on annotations.

4.3 Effects on the Takeaway Granularity

Hypothesis 2a (H2a): Map type affects the granularity of a reader's takeaways.

The **Takeaway Granularity** output variable, which takes on the values county, state, region is an ordinal data type. We used a proportional odds model to analyze the impact of the input variables on the comparative odds of having a coarser (regional) takeaway.

We examine the effect of **Map Type** on **Takeaway Granularity**, to which there is a significant effect ($p = 0.0003$). According to

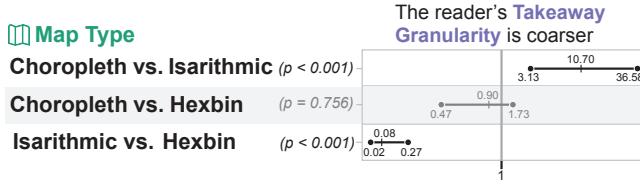


Figure 11: 95% confidence interval chart for the odds ratios model for the H2a hypotheses, which examines the impact of map type on the level of detail in reader takeaways. This chart illustrates a significant trend in how different map types influence the granularity of the information readers extract. When viewing choropleth maps, readers are over 10 times more likely to have a broader, less detailed takeaway compared to isarithmic maps. Similarly, when comparing hexbin maps with choropleth maps, readers lean slightly towards broader takeaways with hexbin maps. The contrast is even more pronounced when hexbin maps are compared with isarithmic maps, where readers are 12.5 times more likely to focus on general, less detailed information. These insights underscore the substantial effect of map type on the level of detail perceived and utilized by readers.

the odds ratio model which is summarized in Fig. 11, readers are $10.70 \times (p < 0.001)$ more likely to produce takeaways at a coarser granularity when reading a choropleth map vs. an isarithmic map. Readers are also $12.5 \times$ more likely to have coarser takeaways when reading a hexbin vs. an isarithmic map. When comparing hexbins to choropleth maps, hexbins were $1.11 \times (p = 0.756)$ more likely to create coarser takeaways, although this result is not significant. Thus, we conclude that people are least likely to have a coarser-detailed takeaway when reading isarithmic maps.

The immediate implications of these results suggest that choropleth and hexbin maps lead to coarser granularity takeaways, while isarithmic maps help readers focus on finer details. These findings underscore the importance of the contextual use of each map type. For example, in teaching geographic patterns, choropleth maps may be more effective in highlighting trends at the regional level, while isarithmic maps may be better suited for showing phenomena at a finer level.

Hypothesis 2b (H2b): The granularity of takeaways varies with map detail.

The test of Fixed Effects (Fig. 8) suggests that both **Text-Map Detail Alignment** ($p < 0.0001$) and **Text-Map Detail Alignment × Semantic Level** ($p < 0.0001$) have a significant influence on the granularity of a reader's takeaway. When viewing a county-level map with text-map detail misalignment, the odds of a reader offering a coarser takeaway granularity are $23.03 \times (p < 0.001)$ greater (Fig. 12) than when reading a map with an aligned text-map detail. This result suggests that detailed maps paired with coarser text detail lead to more generalized takeaways. For less detailed (state-level) maps, text detail, regardless of whether it is coarser or finer, tends to elicit more precise (state or county-level) takeaways (Fig. 12). It is noteworthy that regardless of the level of text detail, county-level maps yield same-level (county) coarser takeaways (at the state or regional level) compared to state-level maps, since there

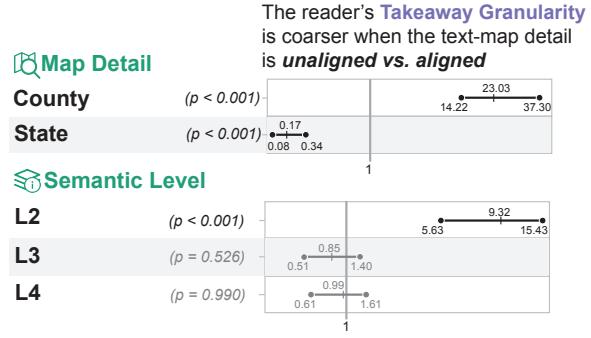


Figure 12: 95% confidence interval chart for the odds ratios model for the H2b hypotheses. All plots model that the reader's takeaway granularity is coarser under different treatments. We show odds ratios for **Map Detail × Text-Map Detail Alignment** and **Semantic Level × Text-Map Detail Alignment**.

is no finer level of detail beyond the county level. This suggests that the production of finer-grained takeaways is not caused by higher-detailed maps.

Hypothesis 2c (H2c): Higher spatial autocorrelation leads to takeaways with coarser detail.

Due to the categorical nature of the dependent variable such as **Takeaway Granularity** and the continuous nature of the independent variable (Moran's I for Spatial Autocorrelation), we use nominal logistic regression to model the odds of having a coarser takeaway. Recall that the ordinal ordering from fine to coarse is county, state, region.

The effect of Spatial Autocorrelation depends only on **Map Type**, but not on text elements (Fig. 8). The results indicate that higher levels of spatial autocorrelation produce finer details for both choropleth and hexbin but higher levels of Moran's I produce coarser level of detail for the isarithmic maps. This is seen by the negative coefficient values for choropleth (-1.8869) and hexbin (-2.2194) maps, but a positive coefficient (3.3453) for the isarithmic map (see Fig. 13). Thus, this hypothesis is only supported for isarithmic maps.

The reader's Takeaway Granularity is at a coarser level		
Map Type	Moran's I	
	p	Coefficient
Choropleth	0.0019	-1.8869
Isarithmic	0.0176	3.3453
Hexbin	0.0001	-2.2194

Figure 13: Table of coefficient estimates for the H2c hypotheses. The coefficient shows the correlation between Moran's I and the probability that the **Takeaway Granularity** is coarser. A positive coefficient, as observed with the isarithmic map, indicates a direct correlation between the variables. In the context of an isarithmic map, this means that as Moran's I increases, the reader's takeaway is increasingly likely to be at a coarser level.

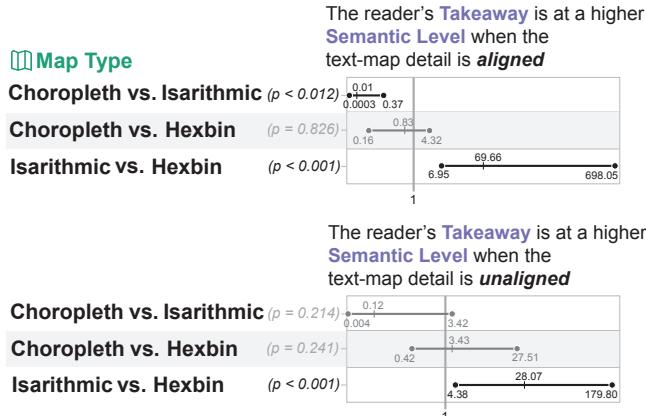


Figure 14: 95% confidence interval chart for the odds ratios model for the H3a hypotheses. The event being modeled is the **Takeaway Semantic Level** having a higher value. Recall that higher semantic levels represent more complex and novel information, while lower levels correspond to basic facts and statistics. The plots show the pairwise comparisons between the three different map types when the text-map detail is aligned (top) and when it is unaligned (bottom).

4.4 Effects on the Semantic Level of Takeaways

Hypothesis 3a (H3a): Map type influences the semantic level of a reader's takeaways.

To better analyze the semantic level of reader takeaways, which is categorized into ordinal levels (L2, L3, or L4), we again employed a proportional odds model. This model is used to assess the influence of various map and text factors on the likelihood of a reader deriving a takeaway at a more advanced semantic level.

Our model (Fig. 14) that estimates the probability that the takeaway will be at a higher semantic level. Fixed Effect tests (Fig. 8) show that **Map Type** ($p < 0.0001$) has an effect on **Takeaway Semantic Level**, but the direction and magnitude of the effect is dependent on the **Text-Map Detail Alignment** (i.e. **Text-Map Detail Alignment** \times **Map Type** is also significant at $p = 0.0001$). The interpretations of the odds ratios are as follows:

When text and map detail were **unaligned**, the odds that the reader has a higher semantic level takeaway is $8.33 \times (p = 0.214)$ when reading an isarithmic map vs. a choropleth map. This difference is more pronounced when text and map detail were **aligned**, here the odds increased to $100 \times (p = 0.012)$ for the same comparison - isarithmic map vs. a choropleth map.

When text and map detail were **unaligned**, the odds that a reader provides a higher semantic level takeaway is $3.43 \times (p = 0.241)$ greater for a choropleth map vs. a hexbin. When text and map detail were **aligned**, the odds of readers providing a higher semantic level takeaway is $1.20 \times (p = 0.525)$ greater for a hexbin vs. a choropleth.

When text and map detail were **unaligned**, the odds that a reader provides a higher semantic level takeaway is $28.07 \times (p < 0.001)$ for an isarithmic map vs. a hexbin. When text and map detail were **aligned**, the odds increased to $69.66 \times (p < 0.001)$ for the same comparison - an isarithmic map vs. a hexbin.

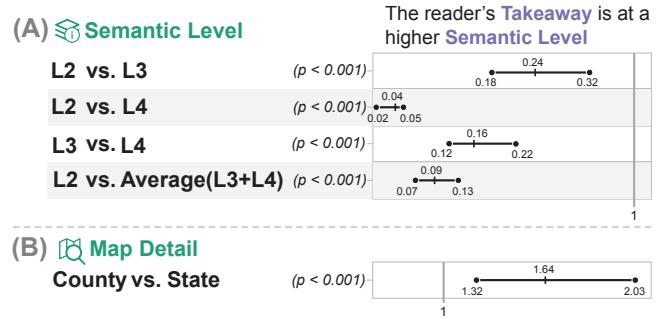


Figure 15: 95% confidence interval chart for the odds ratios model for the H3b hypotheses. The event modeled is that the reader's takeaway will be at a higher semantic level, similar to all other H3 hypotheses. In (A), we compare the odds ratios across different semantic levels. In (B), we examine the odds ratios when comparing map details at the county level vs. the state level, providing insight into how the granularity of map detail influences the semantic level of reader takeaways.

These quantitative results show that regardless of the level of text-map detail alignment, isarithmic maps produce significantly higher semantics than choropleths, with the effect stronger when the text detail is the **aligned**. Thus, this hypothesis is supported.

Hypothesis 3b (H3b): Coarser map details lead to higher-level semantic takeaways. We build an odds model to analyze the impact of map and text factors on the comparative odds of having a higher semantic level takeaway. We see that test of Fixed Effects indicates that **Map Detail** ($p < 0.0001$) has a significant effect on the semantic level of the takeaways (Fig. 8). More specifically, county-level maps produce higher semantic level takeaways over state-level maps by a factor of $1.64 \times (p < 0.001)$ (Fig. 15B). This implies that finer map details are more likely to lead to takeaways at a higher semantic level. The opposite of our hypothesis is true, so it is rejected.

When examining this hypothesis, we also discovered that **Semantic Level** has a highly significant effect on the reader's **Takeaway Semantic Level** ($p < 0.0001$). In experiment 2 (Fig. 15A), the pairwise comparisons in our odds ratios model yielded: Readers have $4.177 \times (p < 0.001)$ higher odds of having a higher level semantic takeaway when reading L3 annotations compared to L2 annotations. When reading L4 annotations compared to L2, the odds increase significantly to $25 \times (p < 0.001)$ higher. Moreover, readers have $6.25 \times (p < 0.001)$ higher odds of experiencing a higher level semantic takeaway when comparing L4 to L3 annotations. From these findings, we conclude that higher semantics in the text annotations result in higher semantics in the takeaways.

Hypothesis 3c (H3c): Spatial autocorrelation within a map dataset influences the semantic level of takeaways.

For this hypothesis, we use nominal logistic regression (Fig. 16) to model how Moran's I affects the semantic level of a reader's takeaway. **Spatial Autocorrelation** \times **Map Type** ($p < 0.0001$) has a significant effect on **Takeaway Semantic Level**. For isarithmic maps, higher Moran's I results in lower semantic levels, as denoted by a significant p-value and negative coefficient(*coefficient* =

Map Type	Moran's I	
	p	Coefficient
Choropleth	0.1305	1.9961
Isarithmic	0.0410	-6.8788
Hexbin	0.0019	4.7801

Figure 16: Table of coefficient estimates for the H3c hypotheses. The coefficient shows the correlation between Moran's I and the odds that the Takeaway Semantic Level is higher. A positive value indicates that as Moran's I increases, the reader's takeaway will be at a higher semantic level.

$-6.8788, p = 0.0410$). Hexbins produce takeaways at a higher semantic level with higher spatial autocorrelation ($\text{coefficient} = 4.7801, p = 0.0019$). Choropleth maps also are more likely to produce takeaways at a higher semantic level with an increasing Moran's I ($\text{coefficient} = 1.9961, p = 0.1305$), but this result was not significant.

4.5 Additional Statistical Analyses

The preceding subsections focused on examining the influence of map variables on participant takeaways. This subsection aims to concisely present insights gained from analyzing participants' VLAT scores. It's important to note that VLAT scores are observed outcomes; hence, any causal relationships cannot be established based on these scores.

The Pearson correlation coefficient between VLAT scores and the variable **Source** is approximately 0.085 ($p < 0.001$), signifying a significant correlation, albeit a slight correlation. This suggests that participants with higher visual literacy are more inclined to extract information directly from the map. However, we reiterate this relationship is correlative and does not imply causation. Furthermore, the nature of the **Source** as ordinal data presents limitations when correlating with VLAT scores, which are cardinal in nature.

Similarly, the Pearson correlation coefficient of -0.033 ($p < 0.023$) shows a weak, yet statistically significant negative relationship between VLAT scores and the **Takeaway Semantic Level**. This indicates that higher VLAT scores correlate with lower levels of takeaway semantics. It further suggests that individuals with better visual literacy are inclined to rely less on personal interpretation, favoring direct data extraction (e.g., reading specific values from a map) rather than making abstract or trend-based inferences.

Regarding the volume of takeaways, the mean number recorded per dataset is 774, with a standard deviation of approximately 11.85. This statistic reflects that each dataset in the study, on average, elicited about 774 takeaways. On an individual level, participants contributed an average of 46 takeaways each, with a standard deviation of 12.48.

5 DISCUSSION

The implications of this study come at an opportune time when digital information sources are more widely used and trusted. Thematic maps, which are useful as statistical reporting tools, are sometimes

used to spatially represent critical information in high-stakes domains. Examples include the spread of infectious diseases during a pandemic [40] and resource distribution in natural disaster mitigation [16]. Thus, optimally integrating textual annotations with different thematic maps at varying levels of detail can produce maps that are tailored to the specific readership in focus. The consequences of the findings from this study are explored in this section to provide actionable recommendations for the visualization community. These findings and their associated design implications are summarized in Fig. 17. It has to be noted that the participants recruited for this study primarily represent a segment of the general population with a background in higher education, with $n = 47$ holding 4-year degrees, while some have post-graduate qualifications. They are not experts in cartographic interpretation. Thus, the insights discussed here are intended to aid the comprehension and usability of thematic maps for the educated population.

5.1 Design Implications

Effects of Geographic Detail: As per **H1b** and **H3b**, readers (without specialized training in map reading and use) who saw maps with higher geographic detail (*i.e.* county-level maps) tend to have a higher reliance on text and produce higher semantic level takeaways. This could be due to the variation in the expertise of the map reader. *Counterintuitive, our results suggest that is the aim is to impart fine-grained information to the reader, less detailed maps are more effective, particularly when combined with text annotations at the same detail as the map.* A focused approach that emphasizes the regions of interest while potentially reducing the prominence of irrelevant areas can enhance the reader's comprehension. This strategy is often adopted in census maps, where a targeted representation not only prevents information overload but also facilitates a more insightful reading experience.

Effects of Semantic Level of Text Annotations: Based on the outcomes from **H2b** and **H3b**, the semantic level of text annotations significantly impacts the granularity and semantic level of takeaways. Annotations with higher semantic levels (dependent) are more likely to produce spatially coarser takeaways and higher semantic level takeaways. Therefore, *when the intention is to elicit broad or high-level understanding, using dependent level annotations can be effective.* To further contextualize, it is worth noting that annotations that succinctly encapsulate trends and delineate regional patterns stand as powerful tools in encouraging a high-level comprehension. This approach is already employed in various real-world scenarios including, but not limited to, weather forecasting and temperature mapping, where conveying overarching trends and patterns to news viewers takes precedence.

Spatial Autocorrelation and Map Type: Results from **H1c** and **H2c** show that spatial autocorrelation does not significantly influence the source of takeaways or reliance on text, but it does influence the detail of takeaways depending on the map type. Higher levels of spatial autocorrelation produce finer details for choropleth and hexbin maps but coarser detail for isarithmic maps. Therefore, *when designing maps, the author should be aware that the type of map chosen will impact the granularity of the takeaways.* Being aware of the spatial autocorrelation existing in the dataset can guide the optimal choice of map type to communicate details more effectively.

Map Design Factors		Takeaways			Design Guidelines
		Source	Takeaway Granularity*	Takeaway Semantic Level	
Map Detail	higher (county-level)	more from text	not significant	higher (dependent) lower (independent)	Aim: to impart fine-grained information to the reader. Guide: less detailed maps are more effective, particularly when combined with text annotations at the same detail as the map.
	lower (state-level)	more from map	not significant	lower (independent)	
Semantic Level	higher (dependent)	not significant	coarser	higher (dependent)	Aim: to elicit broad or high-level understanding from the map. Guide: using dependent level annotations (e.g. succinctly encapsulate trends and delineate regional patterns) can be effective.
	lower (independent)	not significant	finer	lower (independent)	
Spatial Autocorr.	higher	not significant	finer (choropleth & hexbin) / coarser (isarithmic)	higher (dependent, hexbin) / lower (independent, isarithmic)	Aim: to present information with (a) fine detail. / (b) generalized overview. Guide: (a) choropleth or hexbin maps are more productive. / (b) isarithmic maps are better.
	lower	not significant	coarser (choropleth & hexbin) / finer (isarithmic)	lower (independent, hexbin) / higher (dependent, isarithmic)	
Map Type	choropleth & hexbin	more from map	coarser	not significant	Aim: (a) map should be the main information source and the administrative units in data are important. / (b) need a higher depth of understanding of the pattern. Guide: (a) may use choropleth or hexbin maps. / (b) isarithmic maps are more suited.
	isarithmic	more from text	finer	higher (dependent)	

*The takeaway granularity is naturally ordered from coarse to fine as follows (1) region, (2) state, (3) county.

Figure 17: Summary of design implications and design guidelines.

For instance, in scenarios where fine detail is important, leveraging the strengths of choropleth or hexbin maps can be more productive. Conversely, if the intention is to present a more generalized overview, isarithmic maps are a better choice.

Balancing Visual Complexity with Semantic Detail: The results from **H1a** and **H3a** reveal that map type impacts both the reliance on text and the depth of understanding derived from the map. Isarithmic maps, with their de-emphasis of the administrative units for which data were collected, often lead to seek clarity from text annotations, resulting in a higher level of understanding. Conversely, map types that visually emphasize the administrative units for which data were collected, such as choropleth maps, encourage readers to derive information directly from the map, particularly when paired with more accessible annotations. Regardless of the text-map detail alignment, hexbins and choropleths do not differ significantly with respect to their effect on the semantic level of takeaways. Additionally and as expected, dependent level annotations produce higher level semantics in the takeaways, regardless of the Map Type. Therefore, *map creators should strategically choose the map type based on the desired depth of understanding and the reliance on the map itself versus the accompanying text*. Hexbin or choropleth maps are preferable when the map should be the primary source of information, especially in cases where the complexity of information can't be altered. Conversely, isarithmic maps are more suited when a higher depth of understanding is the goal, but they may risk overwhelming the reader. Given this, it is evident why most online news sources leverage choropleth maps or cartograms – these maps are extensively used in representing census, socioeconomic, or demographic data where delineating administrative units clearly is pivotal. This style accommodates a direct, unambiguous visualization of data corresponding to specific regions. In contrast, isarithmic maps naturally fit scenarios requiring the representation of continuous spatial data distributions, like weather forecasts or temperature maps. This is because they effectively convey gradients

and variations, offering a holistic view of data trends across different geographical expanses. Therefore, understanding the inherent strengths of each map type can guide designers in choosing the most effective way to communicate data publicly.

To synthesize, the study illustrates the complex interplay between map design elements and their impact on reader comprehension and information takeaway. Notably, this research highlights the importance of considering the interaction between different map elements, such as type, detail, and spatial autocorrelation, along with the semantic level and detail of accompanying text when designing maps. By strategically utilizing these elements, cartographers can better guide reader comprehension and the granularity and depth of their takeaways.

5.2 Additional Insights

In this subsection, we refer to participants by their ID from P1-P103.

Subjectivity of Takeaways: The interpretation of maps can be influenced by personal context and knowledge. For example, P30 writes “Florida does achieve more employment than Ohio, where I am located. The population might be more homogenous [sic],” while P47 drew on their knowledge about the U.S. agricultural industry: “The agricultural sector is highly dependent on government policies and subsidies, which have a significant impact on the prices of agricultural commodities.” These insights can provide unique perspectives on the data, but also highlight the potential for subjectivity in interpretation.

Some responses provided by participants were not easily verifiable. For example, P2 wrote about the map in Fig. 1C, “less vacancy in large metro areas.” Although this may be true, we surmise that different readers may disagree on the validity of this statement. This example demonstrates how readers may bring their own assumptions to the interpretation of maps, which could lead to unobjective conclusions.

These results indicate that participants bring their own unique perspectives and experiences to the interpretation of visual information. This subjectivity can lead to diverse and creative insights, but can also introduce biases and inaccuracies in interpretation, which can hinder effective decision-making. Therefore, it is important to acknowledge and account for the subjective nature of map reading in research and in practical applications, while also valuing the potential for diverse perspectives and insights.

VLAT Scores: Participant scores on the VLAT were not considered in the results because the inclusion of these results did not yield any meaningful conclusions. We believed that VLAT scores would correlate positively with the number of takeaways per person. However, it was found that participants would provide the same number of takeaways for each map stimulus throughout the entire duration of the study. Thus, the VLAT section in the study was primarily a mechanism for preventing stimuli fatigue.

Repeating content in annotations: Our study observed instances where participants echoed the content of annotations in their responses, reflecting the impact of annotations on the interpretation process. This aligns with findings from Hullman et al., highlighting how the presence or absence of annotations can steer viewers' attention and shape their understanding of the visualization [22]. However, this effect was small, as only 44 takeaways copied the text annotations from the maps, with 2 takeaways copying Semantic Level 2 annotations, 28 copying Level 3, and 14 copying Level 4. This distribution suggests that the majority of these repetitive takeaways are categorized under semantic levels 3 and 4. The higher occurrence in these levels indicates that while some participants directly mirrored the annotations, they predominantly did so in contexts requiring a more complex understanding (levels 3 and 4), rather than merely restating basic facts (level 2).

6 LIMITATIONS AND FUTURE WORK

Design Factors: We examined the effects of map type, map detail, spatial autocorrelation, semantic level, and text-map detail alignment on reader takeaways, which is only a subset of thematic map design parameters. Limiting the design options was necessary due to the overwhelming design space of thematic maps. In Section 3.4.1, we listed other common design factors such as color and classification schemes, both of which were kept constant in our study. Additional design factors such as the underlying data from which isopleth maps are generated (county-level versus state level) and county-level versus state-level hexbins, and the number of annotations can also be considered. Future work can and should explore the effects of these additional factors.

In this study, we focus solely on hexbin cartograms, a specific variant that represents geographical areas as hexagons. Consequently, our findings may not be generalizable to other cartogram types.

Study Randomization and Order Effects: In our study, we implemented a random order of dataset presentation based on a Java randomization function. This approach aimed to minimize potential sequential bias that may be introduced when using a fixed dataset order. However, we recognize that this method may introduce variability due to the lack of control over potential sequence effects.

We considered stratified randomization and Latin square designs as alternatives for their ability to distribute dataset types evenly and control for order effects. However, these methods may introduce artificial structuring or reduce the randomness of the conditions, respectively. We chose to show our datasets using a Java randomized order, but acknowledging that this choice carries increased variability. We suggest that subsequent studies include analyses to examine any order effects and their implications for the study results.

Learning Effects Mitigation: In both experiments, identical datasets were used. To reduce memory-related biases and learning effects among participants, we systematically altered the input variables, such as map design elements and textual semantic levels, for each experiment. This approach ensured that participants were exposed to distinct maps, albeit derived from the same dataset. For instance, a participant who read a hexbin map featuring perceiver-dependent texts in alignment, would encounter a different map in the second experiment—specifically, a county-level choropleth map with perceiver-independent (L2) annotations. The study's design specifics are illustrated in Fig. 2. Additionally, to further minimize the impact of participants' recollections, a Visual Literacy Assessment Test (VLAT) was inserted between the two experiments.

Subjectivity in Source Variable: A potential limitation of our study is the reliance on participants' self-reported assessment to determine the source of their takeaways. This method of subjective evaluation can introduce variability in responses, as individual participants may have differing abilities to accurately recall their cognitive processes in map reading. Future work may address by incorporating more objective measures of source determination, such as eye-tracking or other behavioral indicators that more accurately assess fixation on map elements.

7 CONCLUSION

In an era where public sentiment and response to global and national events are heavily influenced by digital information, it is imperative to establish clear guidelines for visualization tools such as thematic maps. Beyond improving clarity and readability, advancing visualization guidelines is vital for upholding truth and trust in data journalism. This study confirmed how various map configurations with textual annotations affected the quality of reader takeaways. In contrast to previous studies that predominantly examined one variable at a time, our research used a factorial experimental design, which granted insights into more complex effects of both map and textual attributes. Our results therefore provide richer insights into the effect of textual annotations for different map designs, highlighting design synergies or potential antagonisms.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 17STQAC00001-07-00. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

REFERENCES

- [1] Anthony J Aretz. 1991. The Design of Electronic Map Displays. *Human Factors* 33, 1 (1991), 85–101. <https://doi.org/10.1177/001872089103300107>
- [2] Leilani Battle, Peitong Duan, Zachery Miranda, Dana Mukusheva, Remco Chang, and Michael Stonebraker. 2018. Beagle: Automated Extraction and Interpretation of Visualizations from the Web. In *In Proceedings of the ACM Conference on Human Factors in Computing Systems*. 1–8. <https://doi.org/10.1145/3173574.3174168>
- [3] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2015. Beyond Memorability: Visualization Recognition and Recall. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 519–528. <https://doi.org/10.1109/TVCG.2015.2467732>
- [4] Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315. <https://doi.org/10.1109/TVCG.2013.234>
- [5] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D³ Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2301–2309. <https://doi.org/10.1109/TVCG.2011.185>
- [6] Rick L Bunch and Robert Earl Lloyd. 2006. The Cognitive Load of Geographic Information. *The Professional Geographer* 58, 2 (2006), 209–220. <https://doi.org/10.1111/j.1467-9272.2006.00527.x>
- [7] Luz Calvo, Fernando Cucchietti, and Mario Pérez-Montoro. 2022. Measuring the Effectiveness of Static Maps to Communicate Changes over Time. *IEEE Transactions on Visualization and Computer Graphics* (2022), 1–13. <https://doi.org/10.1109/TVCG.2022.3188940>
- [8] Qing Chen, Shixiong Cao, Jiazhe Wang, and Nan Cao. 2023. How Does Automation Shape the Process of Narrative Visualization: A Survey of Tools. *IEEE Transactions on Visualization and Computer Graphics* (2023), 1–20. <https://doi.org/10.1109/TVCG.2023.3261320>
- [9] Andrew David Cliff and J Keith Ord. 1981. *Spatial Processes: Models & Applications*. Taylor & Francis.
- [10] Andrew D Cliff and Keith Ord. 1970. Spatial Autocorrelation: A Review of Existing and New Measures with Applications. *Economic Geography* 46, sup1 (1970), 269–292. <https://doi.org/10.2307/143144>
- [11] Robert G Cromley. 1996. A Comparison of Optimal Classification Strategies for Choroplethic Displays of Spatially Aggregated Data. *International Journal of Geographical Information Systems* 10, 4 (1996), 405–424. <https://doi.org/10.1080/02693799608902087>
- [12] Danny Debelius. 2015. Let's Tesselate: Hexagons for Tile Grid Maps. <http://blog.apps.npr.org/2015/05/11/hex-tile-maps.html>
- [13] Borden D Dent. 1972. Visual Organization and Thematic Map Communication. *Annals of the Association of American Geographers* 62, 1 (1972), 79–93. <https://doi.org/10.1111/j.1467-8306.1972.tb00845.x>
- [14] Borden D Dent. 1999. *Cartography: Thematic Map Design*. Ingram.
- [15] David J Fairbairn. 1993. On the Nature of Cartographic Text. *The Cartographic Journal* 30, 3 (1993), 104–111. <https://doi.org/10.1179/000870493787860157>
- [16] FEMA. 2023. National Risk Index Map. <https://hazards.fema.gov/nri/map> Accessed on Sept. 13, 2023.
- [17] Robert C Geary. 1954. The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician* 5, 3 (1954), 115–146. <https://doi.org/10.2307/2986645>
- [18] Michael F Goodchild. 2003. Geographic Information Science and Systems for Environmental Management. *Annual Review of Environment and Resources* 28, 1 (2003), 493–519. <https://doi.org/10.1146/annurev.energy.28.050302.105521>
- [19] Marilyn M Griffin and Daniel H Robinson. 2005. Does Spatial or Visual Information in Maps Facilitate Text Recall? Reconsidering the Conjoint Retention Hypothesis. *Educational Technology Research and Development* 53, 1 (2005), 23–36. <https://doi.org/10.1007/bf02504855>
- [20] Mark Harrower and Cynthia A Brewer. 2003. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal* 40, 1 (2003), 27–37. <https://doi.org/10.1179/000870403235002042>
- [21] Marti Hearst and Melanie Tory. 2019. Would You Like A Chart With That? Incorporating Visualizations into Conversational Interfaces. In *In Proceedings of IEEE Visualization Conference*. IEEE, 1–5. <https://doi.org/10.1109/VISUAL.2019.8933766>
- [22] Jessica Hullman and Nick Diakopoulos. 2011. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2231–2240. <https://doi.org/10.1109/TVCG.2011.255>
- [23] George F Jenks and Fred C Caspall. 1971. Error on Choroplethic Maps: Definition, Measurement, Reduction. *Annals of the Association of American Geographers* 61, 2 (1971), 217–244. <https://doi.org/10.1111/j.1467-8306.1971.tb00779.x>
- [24] Dae Hyun Kim, Vidya Setlur, and Maneesh Agrawala. 2021. Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 1–11. <https://doi.org/10.1145/3411764.3445443>
- [25] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2018. Frames and Slants in Titles of Visualizations on Controversial Topics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 1–12. <https://doi.org/10.1145/3173574.3174012>
- [26] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2016. Vlat: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2016), 551–560. <https://doi.org/10.1109/TVCG.2016.2598920>
- [27] Pierre Legendre. 1993. Spatial Autocorrelation: Trouble or New Paradigm? *Ecology* 74, 6 (1993), 1659–1673. <https://doi.org/10.2307/1939924>
- [28] Fan Lei, Arlen Fan, Alan M. MacEachren, and Ross Maciejewski. 2023. GeoLinter: A Linting Framework for Choropleth Maps. *IEEE Transactions on Visualization and Computer Graphics* (2023), 1–16. <https://doi.org/10.1109/TVCG.2023.3322372>
- [29] Fan Lei, Yuxin Ma, A. Stewart Fotheringham, Elizabeth A. Mack, Ziqi Li, Mehak Sachdeva, Sarah Bardin, and Ross Maciejewski. 2023. GeoExplainer: A Visual Analytics Framework for Spatial Modeling Contextualization and Report Generation. *IEEE Transactions on Visualization and Computer Graphics* (2023), 1–11. <https://doi.org/10.1109/TVCG.2023.3327359>
- [30] Paul A Longley, Michael F Goodchild, David J Maguire, and David W Rhind. 2005. *Geographic Information Systems and Science*. John Wiley & Sons.
- [31] Alan Lundgard and Arvind Satyanarayan. 2021. Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 1073–1083. <https://doi.org/10.1109/TVCG.2021.3114770>
- [32] Patrick AP Moran. 1950. Notes on Continuous Stochastic Phenomena. *Biometrika* 37, 1/2 (1950), 17–23. <https://doi.org/10.2307/2332142>
- [33] Ian Muehlenhaus. 2013. *Web Cartography: Map Design for Interactive and Mobile Devices*. CRC Press.
- [34] Shaun O'Brien and Claire Lauer. 2018. Testing the Susceptibility of Users to Deceptive Data Visualizations When Paired with Explanatory Text. In *Proceedings of the ACM Conference on the Design of Communication*. 1–8. <https://doi.org/10.1145/3233756.3233961>
- [35] Judy M Olson. 1975. Autocorrelation and Visual Map Complexity. *Annals of the Association of American Geographers* 65, 2 (1975), 189–204. <https://doi.org/10.1111/j.1467-8306.1975.tb01030.x>
- [36] Alvitta Ottley, Aleksandra Kaszowska, R Jordan Crouser, and Evan M Peck. 2019. The Curious Case of Combining Text and Visualization. In *Proceedings of the IEEE VGTC/Eurographics Symposium on Visualization*. 121–125. <https://doi.org/10.2312/evs.20191181>
- [37] Joao Pinheiro and Jorge Poco. 2022. ChartText: Linking Text with Charts in Documents. *arXiv* (2022). <https://doi.org/10.48550/arXiv.2201.05043>
- [38] Timothy Prestby. 2022. Design Techniques for COVID-19 Story Maps: A Quantitative Content Analysis. *Cartography and Geographic Information Science* (2022), 1–20. <https://doi.org/10.1080/15230406.2022.2102077>
- [39] Sergio J Rey, Philip Stephens, and Jason Laura. 2017. An evaluation of sampling and full enumeration strategies for Fisher Jenks classification in big data settings. *Transactions in GIS* 21, 4 (2017), 796–810.
- [40] Mehdi Rezaei, Abbas Ali Nouri, Gyun Seok Park, and Doo Hwan Kim. 2020. Application of Geographic Information System in Monitoring and Detecting the COVID-19 Outbreak. *Iranian Journal of Public Health* 49, Suppl 1 (2020), 114. <https://doi.org/10.18502/ijph.v49iS1.3679>
- [41] Thomas C Ricketts. 2003. Geographic Information Systems and Public Health. *Annual Review of Public Health* 24, 1 (2003), 1–6. <https://doi.org/10.1146/annurev.publhealth.24.100901.140924>
- [42] Robert E Roth. 2021. Cartographic Design as Visual Storytelling: Synthesis and Review of Map-based Narratives, Genres, and Tropes. *The Cartographic Journal* 58, 1 (2021), 83–114. <https://doi.org/10.1080/00087041.2019.1633103>
- [43] Terry A Slocum, Robert M McMaster, Fritz C Kessler, Hugh H Howard, and Robert B Mc Master. 2008. *Thematic Cartography and Geographic Visualization*. Prentice Hall.
- [44] Richard M Smith. 1986. Comparing Traditional Methods for Selecting Class Intervals on Choropleth Maps. *The Professional Geographer* 38, 1 (1986), 62–67. <https://doi.org/10.1111/j.0124-1986.00062.x>
- [45] John Parr Snyder and Philip M Voxland. 1989. *An Album of Map Projections*. Number 1453. US Government Printing Office.
- [46] Zihai Song, Robert E Roth, Lily Houtman, Timothy Prestby, Alicia Iverson, and Song Gao. 2022. Visual Storytelling with Maps: An Empirical Study on Story Map Themes and Narrative Elements, Visual Storytelling Genres and Tropes, and Individual Audience Differences. *Cartographic Perspectives* (2022). <https://doi.org/10.14714/CP100.1759>
- [47] Chase Stokes and Marti A Hearst. 2021. Give Text a Chance: Advocating for Equal Consideration for Language and Visualization. In *IEEE Workshop on Exploring Opportunities and Challenges for Natural Language Techniques to Support Visual Analysis*.
- [48] Chase Stokes, Vidya Setlur, Bridget Cogley, Arvind Satyanarayan, and Marti A Hearst. 2022. Striking a Balance: Reader Takeaways and Preferences when Integrating Text and Charts. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 1233–1243. <https://doi.org/10.1109/TVCG.2022.3209383>
- [49] Walter W Stroup. 2012. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC press.

- [50] Caroline J Tolbert and Karen Mossberger. 2015. US Current Population Survey & American Community Survey Geographic Estimates of Internet Use, 1997–2014. *Harvard Dataverse* 4 (2015). <https://doi.org/10.7910/DVN/UKXPZX>
- [51] Feng Wang, Jose Ibarra, Muhammad Adnan, Paul Longley, and Ross Maciejewski. 2014. What's in a name? Data Linkage, Demography and Visual Analytics. In *Proceedings of the IEEE VGTC/Eurographics Symposium on Visualization in Visual Analytics Workshop*. 7–11. <https://doi.org/10.2312/eurova.20141143>
- [52] Anthony GO Yeh. 1999. Urban Planning and GIS. *Geographical Information Systems* 2, 877–888 (1999), 1.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009