GeoVisA11y: An AI-based Geovisualization Question-Answering System for Screen-Reader Users

Chu Li
Allen School of Computer Science
University of Washington, USA
chuchuli@cs.washington.edu

Rock Yuren Pang
Allen School of Computer Science
University of Washington, USA
ypang2@cs.washington.edu

Arnavi Chheda-Kothary
Allen School of Computer Science
University of Washington, USA
chheda@cs.washington.edu

Ather Sharif
Allen School of Computer Science
University of Washington, USA
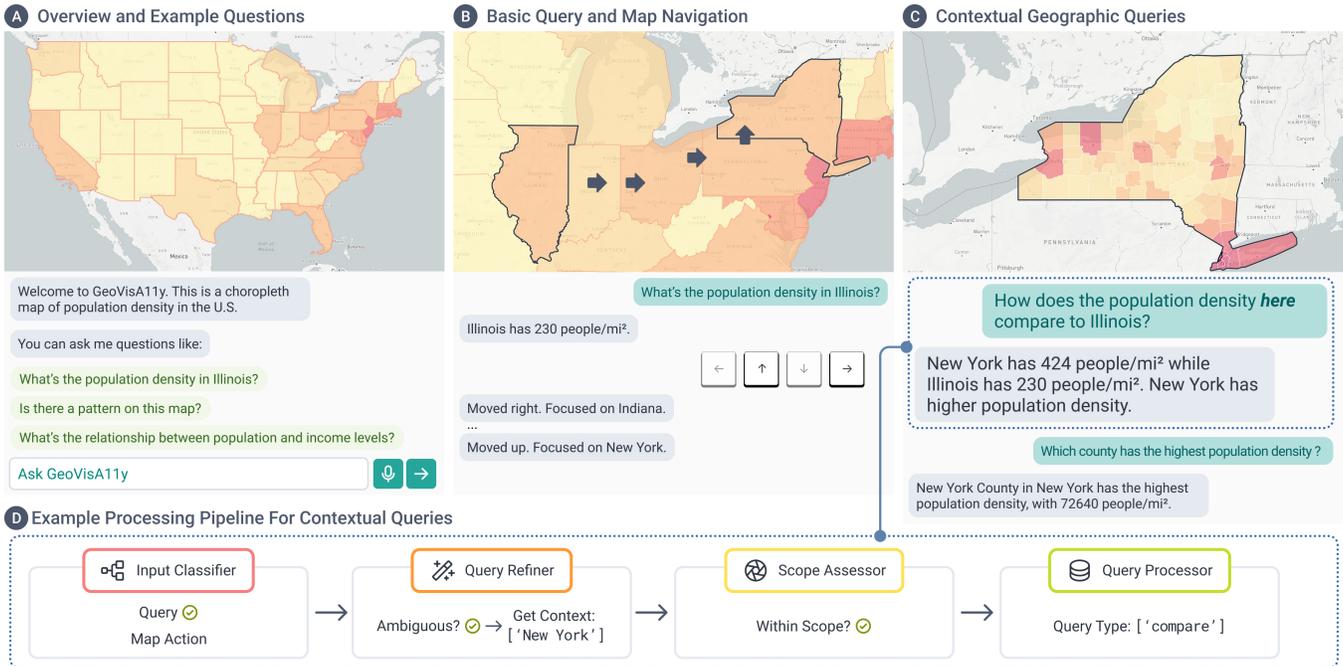asharif@cs.washington.edu

Henok Assalif
Allen School of Computer Science
University of Washington, USA
henok206@uw.edu

Jeffrey Heer
Allen School of Computer Science
University of Washington, USA
jheer@cs.washington.edu

Jon E. Froehlich
Allen School of Computer Science
University of Washington, USA
jonf@cs.washington.edu

Figure 1: GeoVisA11y is an AI-based question-answering system for geovisualizations designed for screen-reader users. (A) Upon loading, users are presented visualization overview and example questions. (B) Querying specific locations automatically focuses the map. Users can also navigate the map via keyboard controls. (C) Contextual queries where the system answers based on current map focus, with automatic geographic unit changes from state to county level as needed. (D) Processing pipeline showing how contextual queries are classified, refined (resolving ambiguous terms), assessed for scope, and processed to generate responses.

Abstract

Geovisualizations are powerful tools for communicating spatial information, but are inaccessible to screen-reader users. To address this limitation, we present GeoVisA11y, an LLM-based question-answering system that makes geovisualizations accessible through

natural language interaction. The system supports map reading, analysis, interpretation and navigation by handling analytical, geospatial, visual and contextual queries. Through user studies with six screen-reader users and six sighted participants, we demonstrate that GeoVisA11y effectively bridges accessibility gaps while revealing distinct interaction patterns between user groups. We contribute: (1) an open-source, accessible geovisualization system, (2) empirical findings on query and navigation differences, and (3) a dataset of geospatial queries to inform future research on accessible data visualization.

CCS Concepts

• **Human-centered computing** → **Accessibility systems and tools**; **Interactive systems and tools**.

Keywords

Geovisualization, visualization question and answering, accessible visualization

1 Introduction

> *"[GeoVisA11y] made it all come to life. Instead of having to muddle through a bunch of numbers and try to find correlations myself, the way it could analyze information and describe patterns is impressive."*
> –B5, a blind participant

Interactive map visualizations, or *geovisualizations*, are powerful tools for discerning patterns, trends, and relationships in spatial data [4, 71]. Geovisualizations support informed decision-making across urban planning [69] and public health [9]. They also communicate essential information in news media, such as COVID case numbers [66] and election results [17, 55]. Yet, most geovisualizations are inaccessible to screen-reader users [21, 60, 70]. Even when accessibility features such as *alt text* and *data tables* are available, they fail to capture the full analytical and interpretive potential of geovisualizations [26, 34]. A study of fifteen geovisualizations found that *none* communicated any high-level spatial patterns to screen-reader users [21].

While recent work like *AltGeoViz* [38] and *Data Navigator* [19], offer alt text for geovisualizations based on user interactions, these approaches remain largely descriptive, akin to "map reading" rather than "map analysis" [48, 52, 53]. We aim to support deeper spatial interpretation—allowing screen-reader users to identify patterns, relationships, and geometric characteristics that are crucial for insight building [48, 52, 53]. Emerging question-answering (QA) systems for geovisualizations such as *VoxLens* [58] represent progress but rely on keyword matching, limiting interactive queries.

In this paper, we introduce GeoVisA11y, an AI-based geovisualization QA system that advances the field by enabling complex spatial interpretation and analysis for screen-reader users. By integrating geostatistical analysis with large language models, GeoVisA11y moves beyond keyword-matching and rule-based approaches to support previously inaccessible analytical tasks. Moreover, GeoVisA11y uniquely handles geospatial queries about *spatial patterns* [56], *spatial relationships*, and *geometric characteristics*. To ease interaction, users can fluidly switch between keyboard-based navigation and conversational commands for map exploration. Geo-VisA11y was designed iteratively in collaboration with two screen-reader users and drew upon design principles from accessible QA literature, including disambiguating deictic references [28, 33] and supporting general/contextual queries [26, 33].

Compared with simple data visualizations such as bar charts, line graphs and scatter plots, geovisualizations are inherently more complex: scale, distance, location and direction all carry complex real-world meanings [3, 6, 20, 43]. This visual and semantic complexity affects all users: even sighted individuals misinterpret geovisualizations due to misrepresented scales, distorted projections, misleading color schemes, and inappropriate classification methods [31, 50, 56]. Following the universal design paradigm [63], where accessibility features initially developed for specific populations subsequently benefit broader populations (*e.g.*, subtitles for deaf and hard-of-hearing users, curb ramps for wheelchair users), we hypothesize that accessible geospatial interfaces will serve both screen-reader users and sighted users experiencing difficulty with complex geovisualizations. Building on this motivation, we use GeoVisA11y to investigate the following research questions:

RQ1 What *types of queries* do screen-reader users make when interacting with a geovisualization QA system?

RQ2 How can a geovisualization QA system effectively support *screen-reader users* in engaging with geovisualizations?

RQ3 To what extent can such a system enhance *sighted users'* engagement with geovisualizations?

RQ4 What are the key differences in *querying strategies and interaction patterns* between sighted and screen-reader users when using such a system?

To answer our research questions, we conducted a user study with six screen-reader users and six sighted participants. Importantly, our goal is not to position the sighted users as a *baseline* group [44] to derive confirmatory conclusions, but rather to explore how both groups interact with GeoVisA11y and investigate how visualization QA tools designed for screen-reader users could benefit broader populations. Participants performed exploratory data analysis for two tasks: (1) distributing digital equity funding based on underserved populations and limited digital access, and (2) identifying predominant energy sources across U.S. regions and their potential causes. We found that GeoVisA11y effectively supports both user groups across the three levels of geovisualization engagement (map *reading, analysis, interpretation* [52]). Screen-reader participants successfully employed diverse querying strategies for spatial analysis and valued GeoVisA11y's informative and clear responses. They particularly highlighted how the system enabled

self-directed spatial exploration. Sighted users demonstrated varying levels of system usage based on their domain expertise and geographic knowledge, with increased usage when working with more complex visualizations. For querying strategies, screen-reader users relied primarily on verbal queries combined with keyboard navigation, whereas sighted users typically performed visual assessment before querying specific details. Despite the different approaches, participants from both groups successfully identified similar patterns in the data, demonstrating that GeoVisA11y represents a step toward creating a shared understanding of geovisualizations. We also observed differences in how information was processed: screen-reader users tended to place greater trust in system-provided descriptions, while sighted users' judgments appeared influenced by visual biases such as region size. These findings reveal opportunities to further bridge perceptual gaps and create more robust shared understanding of geovisualizations.

In summary, we contribute: (1) *GeoVisA11y*, an open-source system for accessible geovisualization interaction[1]; (2) *an empirical qualitative evaluation* with 12 participants that highlights the differences between the blind and low-vision (BLV)[2] and sighted user groups in terms of query formation and map navigation; and (3) a *dataset* containing queries asked by both user cohorts to inform future research in accessible, geovisualization QA tools.

2 Related Work

Our work builds on prior work in geospatial visualization, accessible data visualization, and visualization QA systems.

2.1 Geospatial Visualization and Analysis

MacEachren *et al.* [42] defined geovisualization as using concrete visual representations to make spatial contexts and problems visible. Scholars have established three progressive levels of engagement: (1) *map reading*: identifying what elements the cartographer has encoded, (2) *map analysis*: recognizing significant patterns among these elements, and (3) *map interpretation*: explaining observed patterns by integrating additional knowledge[48, 52, 53]. The engagement framework highlights pattern recognition as a critical step from basic reading to comprehensive analysis. However, the concept of a *pattern* remains challenging to define: it is often subjectively understood and intuitively recognized, with scholars frequently discussing patterns without explicit definitions (*e.g.* [48, 61]).

While verbalizing visual patterns is essential for screen-reader users, explicit pattern descriptions benefit all users. Geovisualizations are more complex than traditional visualizations, making them more susceptible to misinterpretation [50, 61] and requiring geospatial data literacy for meaningful interaction [30, 56]. Shiewe [56] identified three perceptual biases in choropleth map interpretation: associating darker colors with higher values, neglecting of smaller geographic areas, and misinterpreting patterns based on data classification methods. Despite the need for pattern articulation, few studies have generated comprehensive textual descriptions that fully capture visual patterns [38, 64, 65]. AltGeoViz [38] generates dynamic alt text of basic statistics (min, max, average), viewport

boundaries, zoom levels, and spatial patterns using grid-based techniques. However, it does not account for polygon geometry or spatial autocorrelation. Atlas.txt [64, 65] offers comprehensive descriptions including maximum/minimum clusters, trends, averages, boundary information, human/physical geography, and cardinal directions. Despite these capabilities, their clustering and trend identification methods lack documentation, and the system does not support interactive exploration or query-based interactions. GeoVisA11y addresses this gap by adopting Schiewe's [56] definition of *pattern* as the *detection of hot/cold spots*, and implementing Local Indicators of Spatial Association (LISA) [7], a widely used geostatistical approach to identify geographical clustering of high and low values.

2.2 Accessible Data Visualization

Research on visualization accessibility has introduced various modalities, including tactile, haptic and multimodal techniques [35], but these custom hardware solutions remain difficult to access. Screen-readers continue to be the most widely adopted and affordable assistive technology [35]. For online visualizations, data tables and alt text are the most commonly used and recommended accommodations for screen-reader users [16, 22, 29, 70]. However, both approaches have significant limitations. Data tables lose the abstraction benefits of visualizations, making it difficult to identify broader patterns or trends [70]. Static alt text requires users to accept the creator's interpretation without allowing personal exploration [41]. More recent approaches include automatic alt text generation through computational pipelines that extract visualization features [16, 49]. Despite these advances, alt-text-based solutions remain largely unidirectional, *i.e.,* systems present information to users who must passively accept it.

Question-answering (QA) capabilities overcome this unidirectional limitation by letting users actively query the system. Sharif *et al.* developed *VoxLens* [58], a JavaScript plugin whose QA module responds to predefined keywords like "maximum," "minimum," and "median." In subsequent work, they extended VoxLens to support geospatial visualization queries and more detailed information extraction [59, 60]. These approaches were limited by the technology available at the time, supporting only simple, keyword-based queries rather than flexible, robust interactions. Recent AI advances and improved model accessibility enable more responsive systems. Our work builds on this foundation by making geovisualizations accessible to BLV users through natural language interactions.

2.3 Visualization Question-Answering

QA systems for data visualization have gained popularity in helping users understand and complete complex analytical tasks [23, 28, 32, 57, 58]. These systems respond to natural language queries by offering direct answers [32, 58], highlighting relevant information in existing visualizations [28, 57], or creating new visualizations tailored to the query [23]. For geovisualizations specifically, *MapQA* offers a large-scale dataset of approximately 800K QA pairs covering about 60K choropleth maps [14]. Despite the scale, MapQA addresses only three basic question categories—legend interpretation, data retrieval, and relational queries—none going beyond the elementary *map reading* stage of geovisualization engagement. Furthermore,

---

Figure 2: GeoVisA11y UI overview. (A) Interactive map component. (B) Chat component where users can ask questions and receive answers through text or voice input. (C) Welcome section with an overview of the visualization, example questions, and shortcut key instructions. (D) Status indicator that announces changes to screen-reader users, *e.g.,* *"Map interaction enabled. Focused on Kansas."* **or** *"Chat interaction enabled. Type a question to ask GeoVisA11y."* **(E) Legend of the current interactive map.**

these systems depend on vision-language models, which are more susceptible to errors as visualization QA is particularly vulnerable to slight pixel-level variations compared to natural image QA [13]. Recent studies demonstrate that incorporating chart specifications directly into language models yields superior performance compared to vision-language approaches [13, 26]. Hence, GeoVisA11y employs a language-based pipeline to achieve greater accuracy and reliability. LLM-based systems are particularly promising for geovisualization QA because they inherently possess geographic knowledge that enables understanding of how geospatial insights emerge from the interplay of multiple contextual layers—such as population density, infrastructure, and natural features [40, 45]. Moreover, while vision-language models struggle with spatial reasoning, LLMs paired with structured geographic data can achieve precise spatial analysis by directly processing coordinate information, topological relationships, and spatial autocorrelation measures rather than relying on less reliable visual inference.

Although visualization QA systems often cite accessibility as a potential application, they rarely focus on BLV users' specific needs, incorrectly assuming their questioning behavior mirrors that of sighted users [33]. Research in both image and visualization QA has rejected this assumption, demonstrating that BLV users pose significantly different questions with distinct phrasing patterns [18, 27, 33]. To address this gap, Kim *et al.* [33] conducted a Wizard of Oz study with BLV participants and developed design considerations for accessible visualization QA systems. *VizAbility* [26] represents perhaps the closest approach to our work—a QA system that enhances chart content navigation through conversational interaction, allowing users to query visual data trends using natural language. While VizAbility can process choropleth maps, it works only with static visualizations; our work differs by focusing on interactive, comprehensive geovisualizations.

3 Designing GeoVisA11y

To design and build GeoVisA11y, we drew on the above literature and followed an iterative, human centered design process involving six hours of co-design sessions with two blind participants over three weeks. Our early-stage interactive prototype consisted of

a chat interface with basic AI query capabilities and an interactive map that responded to chat-based navigation commands. We selected participants who exemplify our target user population: individuals who use screen-readers to actively engage with data visualization and analysis in professional or personal contexts. Participants included one individual blind from birth (C1) and another who experienced gradual vision loss and is now completely blind (C2), representing different experiences with vision loss (Table 3). Participants were compensated $40 per hour.

3.1 System Iteration

Through our iterative co-design process, we identified key areas for system improvement across query handling, navigation paradigms, and interface design.

Initial Reactions. Both co-design participants had prior experience with online map visualizations but found them challenging to access. C1 noted: *"I really liked how intuitive GeoVisA11y was. It surprised me... I threw some more nuanced questions at it and it understood."* When asked about the usefulness of the prototype, C1 explained: *"I think it could definitely improve my geography knowledge, and just being able to read and understand [map visualizations]. Because before, reading maps was always a big undertaking."* C2 stated *"I love this"* and *"I could see so much potential in this, for people who are visually impaired like myself. Even people who are sighted can use this because it's a great way to access data points easily without having to sift through the traditional statistical maps."*

Query Types. While our initial prototype could handle query types identified by prior work (retrieve, compare, find extremum, aggregate, filter, sort, cluster) [33] and summarize spatial patterns, our co-design sessions revealed additional geovisualization-specific query types. These included queries about *shapes* of geographic entities (*e.g.,* *"What is the shape of Alabama"*) and *spatial relationships* between geographic entities (*e.g.,* *"What are the neighboring states of Tennessee?"*). Participants also used *context-aware querying*, asking questions like *"What's the population density here?"* without explicitly specifying the geographic context when they had already focused on a particular region.
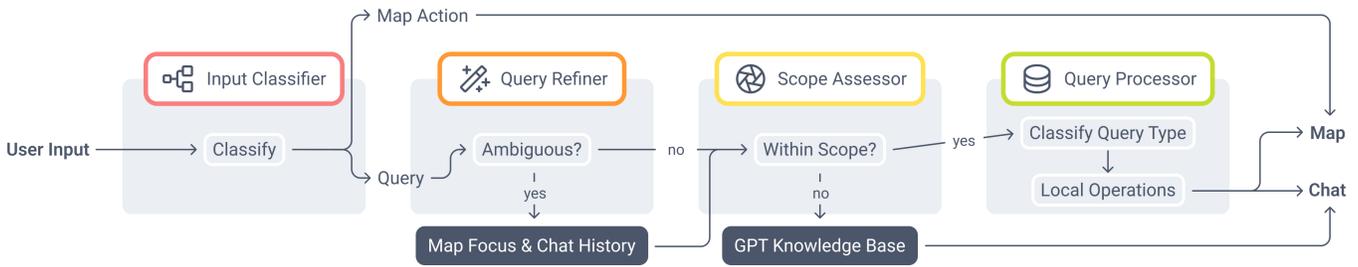
Figure 3: System pipeline diagram. All four components uses GPT-4o-mini via few-shot prompting. (1) The Input Classifier separates map action commands from information queries. (2) The Query Refiner resolves ambiguities (*e.g.,* replacing *"here"* with *"Ohio"*, *"it"* with *"population density"*) by ingesting current map focus and chat history. (3) The Scope Assessor determines if the query can be answered from the local database. (4) The Query Processor classifies the query and triggers relevant operations, then updates the Chat and/or Map interface.

Map Navigation. We began our co-design sessions with a map display that updated based on queries but did not support direct user interaction (*e.g.,* pan, zoom, select). C1 explicitly requested *simple keyboard-based navigation mechanisms* to explore different states and counties. Prior research [38] shows screen-reader users expect arrow keys to facilitate discrete movement between regions (*e.g.,* jumping from one state to its neighbor) rather than continuously panning as in interfaces designed for sighted visual exploration, so we implemented this discrete navigation approach using arrow keys. Participants also demonstrated preferences for natural language navigation, typing commands like *"Focus on Washington"* or *"Take me to Massachusetts"* and expecting the map to update accordingly.

3.2 Design Considerations

From our co-design sessions and key literature, we synthesized the following design considerations:

Support geovisualization-specific questions. Beyond query types identified by prior research [26, 33], our co-design revealed the need for queries about geographic entity shapes and spatial relationships between regions. In addition, to create a robust visualization QA experience for screen-reader users, an accessible geoanalytical system should handle disambiguation, context tracking, guided questions, semantic breadth, and contextual querying.

Support direct and accessible map exploration. Navigation should align with screen-reader users' mental models, enabling discrete movement between geographic regions rather than continuous panning designed for visual exploration.

Support synchronization between map and chat. The map and chat components should work in concert through synchronous focus (map updates based on chat interactions), natural language-driven navigation, and context-aware querying that supports implicit references to focused geographic entities.

Follow standard accessibility guidelines. In line with WCAG [67] principles, designs should prioritize clear focus state awareness, consistent interaction feedback, flexible input modalities, accessible conversation history, plain language communication, and intuitive keyboard navigation patterns.

4 The GeoVisA11y System

Informed by our iterative co-design process and prior work, we designed GeoVisA11y to make geovisual analytics accessible to screen-reader users through natural language interaction. GeoVisA11y has two primary components: (1) a screen-reader compatible UI with an interactive map and an AI-based chat that supports analytical, geospatial, visual, and contextual queries; (2) a custom QA pipeline that combines chat questions with map interactions to form queries and uniquely combines geo-statistical analysis with LLM-based summaries. We begin by describing the QA pipeline as it is a central technical contribution of our work. GeoVisA11y is available as open source software at: https://github.com/makeabilitylab/geovisa11y.

4.1 QA Pipeline

To transform natural language questions into geoanalytical responses and map interactions, we implemented a structured pipeline that parses user queries, determines appropriate response types, and coordinates both analytical processing and map state updates. We first detail the pipeline's four-component architecture, then examine how each query type is processed and analyzed.

4.1.1 Pipeline Architecture. The GeoVisA11y QA pipeline consists of four components (Figure 3): the *Input Classifier, Query Refiner, Scope Assessor,* and *Query Processor*. All components use GPT-4o-mini with few-shot learning [10], which offers strong classification performance [10, 24], low computational cost for rapid prototyping, and inference latency suitable for real-time applications [54]. We acknowledge that fine-tuning these models (*e.g.,* on our dataset artifact) may improve task performance, but we leave this exploration to future work. Additionally, we designed separate prompts for each component as LLM performance degrades with longer input length [37, 39]. See Supplementary Materials for specific prompts.

Input Classifier. Upon receiving text or voice input from users, the Input Classifier employs few-shot prompting to categorize the input as either an *action command* or an *information query*. Action commands trigger direct map manipulation, information queries proceed to Query Refiner for further processing.

Query Refiner. This component addresses potential ambiguities in user queries. The system identifies and resolves two types of ambiguity: (1) *location ambiguity*, where deictic references such

Action/Query	Detailed Type	Example
Map Action		Go to Boston
Analytical Query	Retrieve	What's the population density of Vermont?
	Compare	Which state has higher population density, Louisiana or South Dakota?
	Find Extremum	Which state has the highest population density?
	Aggregate	What's the average population density?
	Filter	Which states have density over 300 people/sqm?
	Sort	Top 5 states with the highest population density?
	Cluster	Which state has a similar population density to Oregon?
Geospatial Query	Pattern	Is there a pattern on the map? Can you describe it?
	Outlier	What are the outliers?
Visual Query	Legend	Can you tell me more about the legend?
	Shape	What is the shape of Wisconsin?
	Spatial Relationships	What are the neighboring states of Illinois?
Contextual Query	Visualization Knowledge	What is a choropleth map?
	General Knowledge	Is there a relationship between income and population density?

Table 1: Supported query types handled by the Query Processor with example questions.

as *"here"* or *"this state"* are resolved using the current map focus or previous conversation, *e.g., "What's the population density here?"* when a specific state is highlighted; and (2) *topic ambiguity*, where pronouns like *"that"* or *"it"* are resolved using conversation history, *e.g., "How does that compare to Ohio?"* where *"that"* refers to previously discussed population density. Some queries exhibit both types, *e.g., "How does it compare to its neighbors?"* where *"it"* refers to a previously discussed metric and *"its"* refers to the focused state.

Scope Assessor. Following disambiguation, the system evaluates whether the query falls within the scope of local data capabilities. The system provides GPT with a compact textual description of the dataset schema (*e.g.,* variable names, metric definitions, and geographic units). Using this information, GPT acts as a binary classifier to determine whether the query can be answered using local data and geostatistical operations. *Within-scope queries* proceed to the Query Processor for classification and local operations, while *beyond-scope queries* are routed to GPT's knowledge base.

Query Processor. The Query Processor uses GPT with few-shot prompting to classify queries into one of 14 predefined categories (Table 1). As with the Scope Assessor, no map imagery or raw datasets are sent to GPT—it serves strictly as a text-based classifier. Once classified, *analytical* and *geospatial* queries are processed using local database operations and spatial analysis algorithms. This design ensures consistent, reproducible results while minimizing latency. Only *visual* and *contextual* queries that cannot be answered by the local database are handled by GPT's knowledge base. The specific processing approaches for each query type are detailed in the following section.

Output Generation. In the final stage, the pipeline delivers results through coordinated updates to both interface components. Textual answers appear in the chat interface, and for locally processed queries, the system also highlights relevant geographic entities on the map.

4.1.2 Query Taxonomy & Analysis Approach. Building upon prior work in accessible chart QA system [33], we identified five primary query categories for geovisualization interactions. See Table 1 for

supported query types and example questions. Here, we detail the data analysis approach for each query type.

Map Actions directly manipulate the map visualization, letting users navigate to specific geographic locations (*e.g., "Go to Ohio"*). The system processes these commands by looking up the target location's coordinates and repositioning the map view accordingly. Screen-reader users receive audio announcements of their current location (*e.g., "Now focused on Ohio."*)

Analytical Queries include: (1) retrieval queries that access specific data points, (2) comparison queries that evaluate relationships between entities, (3) extremum queries that identify maximum or minimum values, (4) aggregation queries that compute summary statistics, (5) filter queries that identify geographic entities meeting specific criteria, (6) sort queries that rank geographic entities, and (7) cluster queries that identify similar value entities.

Most analytical queries are processed through direct database operations, with results formatted using templates that include metric units and contextual information. Filter and sort queries first use GPT to extract numerical conditions or result counts (*e.g.,*extracting *">100"* from *"Which states have density greater than 100?"*) before executing database operations. For cluster queries, the system identifies entities with values within 20% margin of a reference entity.

Geospatial Queries include: (1) pattern identification queries that describe overall geographic trends and (2) outlier identification queries that highlight anomalous regions.

To summarize visual spatial patterns, we integrated global pattern assessment with local anomaly detection. Global pattern identification uses *Moran's I* [51] analysis to assess spatial autocorrelation. Moran's I is a statistical measure that quantifies the degree of spatial clustering, indicating whether similar values tend to be located near each other more than expected by chance [51]. Our pipeline constructs spatial weight matrices using Queen contiguity [5], then computes Moran's I with 999 permutations to determine statistical significance. Results are interpreted as clustered ($I>0$, $p<.05$), dispersed ($I<0$, $p<.05$), or random patterns ($p>.05$).

Outlier identification employs *Local Indicators of Spatial Association* (LISA) [7] analysis, an exploratory tool used to statistically
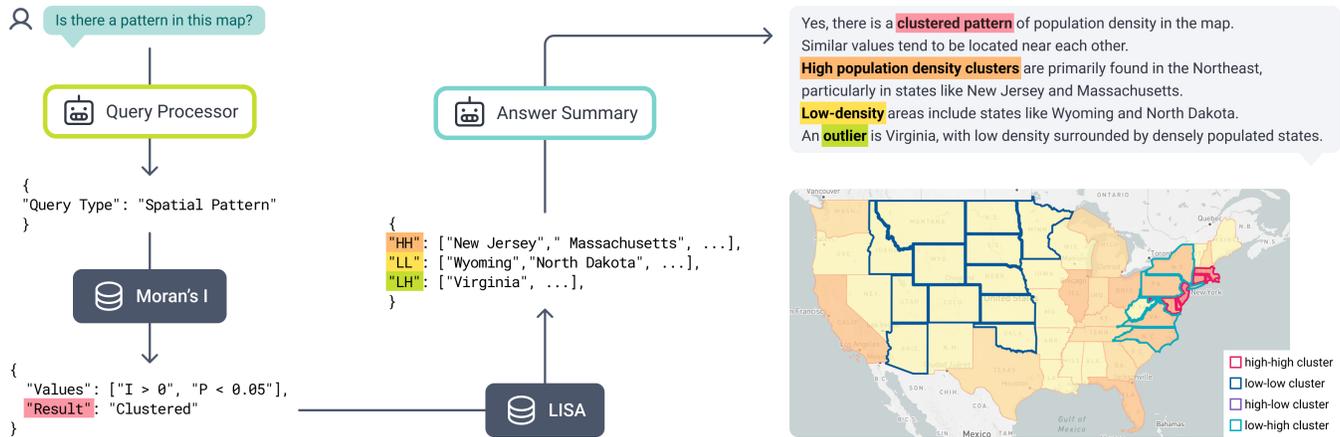
Figure 4: Pipeline workflow for spatial pattern queries. When users ask *"Is there a pattern on the map?"*, the system runs global Moran's I analysis to detect significant patterns, then performs LISA cluster analysis if patterns exist. LISA outputs are summarized by GPT with representative examples, while the map displays colored outlines for cluster types. Color highlights in the answer text indicate which part of the pipeline produced that information: the red highlight on "clustered pattern" corresponds to the Moran's I result, while orange, yellow, and green highlights correspond to specific cluster types (HH, LL, LH) identified by the LISA analysis.

assess geographic clustering of high and low values in a dataset. LISA calculates local spatial autocorrelation at each individual location using a single variable, enabling the quantitative estimation of local spatial clustering by indicating how similar an observation is to all other observations within a defined region [7]. LISA identifies four cluster types: *High-High* (high values surrounded by high values), *Low-Low* (low values surrounded by low values), *High-Low* (high values surrounded by low values), and *Low-High* (low values surrounded by high values) [7]. Only statistically significant clusters (p<.05) are reported. GeoVisA11y then synthesizes these statistical results into natural language summaries, providing two representative examples from each cluster type when available. The map synchronously displays outlines highlighting the four cluster types. This pipeline is illustrated in Figure 4.

Visual Queries address the visual elements of the geovisualization rather than the underlying data, including: (1) *legend or color queries*, (2) *shapes of geographic entities queries*, and (3) *spatial relationship queries* that explore the layout and adjacency of geographic entities. Legend queries are processed by providing GPT with information about legend color mappings and value specifications from the map current visualization. Shape and spatial relationship queries are handled entirely by GPT using its built-in geographic knowledge, as these queries typically concern general geographic facts rather than dataset-specific information.

Contextual queries are questions that extend beyond the immediate dataset, including: (1) *visualization concept queries*, and (2) *general knowledge queries* that connect the visualization to broader concepts. All contextual queries are processed by GPT, prompted to act as a geovisualization expert providing concise answers that connect the visualization to broader geographic knowledge and visualization principles.

4.2 User Interface

GeoVisA11y's UI comprises a map visualization and an AI chat subsystem—both feed into the QA pipeline. Screen readers parse the map and messages and convey them through audio to participants. Though customizable for other regions, we use a sample dataset about U.S. population density to demonstrate the system's functionality in this section. See Figure 2 and video for demonstration, and Table 2 for all available shortcuts and functions.

4.2.1 Interactive Map. To start map interactions, users can press `Ctrl + M` and then `Tab` to focus on a state, and use arrow keys to move between states. Since the four cardinal directions are insufficient for all spatial relationships(*e.g.,* both Ohio and Indiana are north of Kentucky), we developed an algorithm that identifies adjacent states, determines cardinal directions by comparing centroids, and selects the closest neighbor in each direction. For states with complex geometries and atypical centroid positions, we manually adjusted them to ensure natural navigation (*e.g.,* New York, DC, Rhode Island). When users attempt to navigate in a direction where

Shortcut / Key	Function
`Ctrl` + `M`	Toggle between map and chat interaction
`←` `↑` `→` `↓`	Navigate between states/counties
`+`	Zoom in to county level within a state
`−`	Zoom out to state level
`Ctrl` + `L`	Hear the last response again
`Ctrl` + `H`	Show/hide the help window
`Ctrl` + `I`	Refresh question suggestions

**Table 2: List of all shortcut keys and associated behaviors.**
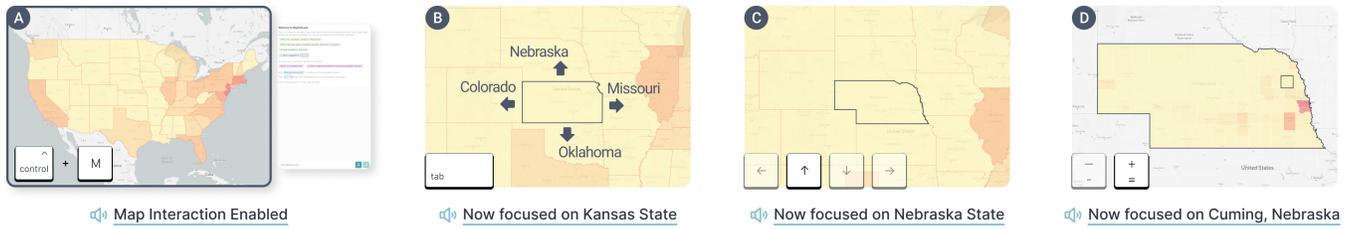
Figure 5: Interactive map keyboard navigation. (A) Ctrl + M activates the map interface. (B) TAB focuses on an initial state, then arrow keys navigate to adjacent states in cardinal directions. (C) Example navigation from Kansas to Nebraska using the up arrow. (D) + and - keys zoom between county and state levels, shown here focused on Cuming County, Nebraska.

no neighboring state exists, the system indicates the boundary condition, *e.g., "There is no state south of Texas"*. When focused on a state, users can press `+` to zoom to county-level. The system automatically focuses on the county closest to the state's centroid, with similar arrow key navigation at the county level. Users can return to the state-level view using the `-` key (Figure 5).

4.2.2 AI Chat. The chat interface follows conversational UI design standards [36] while accommodating screen-reader interaction. Users can toggle focus between map and chat using `Ctrl + M`. Upon focusing on the chat, GeoVisA11y announces, *"Type your question here, press enter to submit."*. After submission, the system repeats the user query followed by a status indicator (*"Looking for answers..."*). Users can navigate to previous conversation history using `Tab`. `Ctrl + L` repeats the most recent system response while maintaining the input field focus.

On load, GeoVisA11y provides an introduction and three selectable example geovisualization questions. Below these examples, a *More Suggestions* button (accessible via `Ctrl + I`) refreshes the question set, cycling through a list of 12 predefined example questions. We also encourage users to ask contextual questions that extend beyond the immediate dataset, *e.g., "What is a choropleth map?"*. For additional assistance, users can ask *"What else can you do?"* to retrieve supported query types. `Ctrl + H` displays a comprehensive list of shortcuts (Figure 6).

4.2.3 Map and Chat Synchronization. To support a tightly integrated, holistic interactive experience, the map and chat components are bidirectionally synchronized. When users query specific geographic entities through AI chat, the system automatically updates the map to provide relevant spatial context. For example,
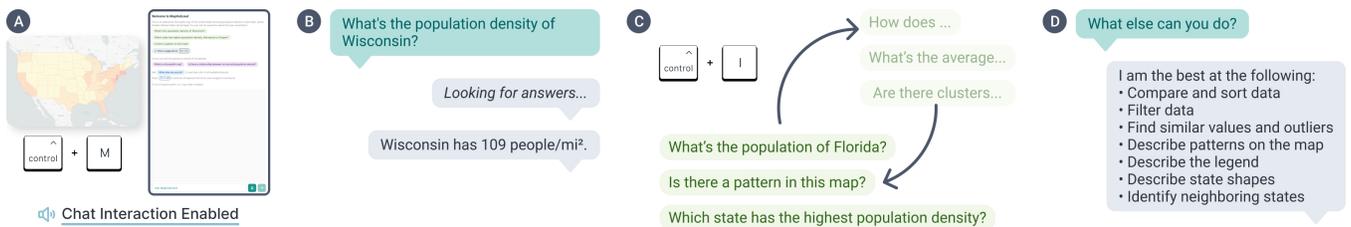
when a user asks about a single state's value (*e.g., "What is the population density of Illinois?"*), the map centers on Illinois and highlights the boundary (Figure 7). Beyond queries, GeoVisA11y also responds to explicit navigation commands in natural language, such as *"Take me to Wyoming"*, *"Focus on Cook County, Illinois"*, or *"Go to Sacramento"*, by immediately updating the map focus accordingly. Moreover, GeoVisA11y maintains the context of the user's current map focus during free exploration, enabling implicit geographic referencing in AI Chat. For example, when focused on Colorado, users can ask location-specific questions such as *"What's the population density here?"*, *"What are the neighboring states?"*, or *"What's the shape of this state?"* (Figure 7).

4.3 Implementation

Data Preparation. We obtained geographic boundaries for 48 contiguous U.S. states and 3,222 counties from TIGER shapefiles [12]. Our demographic and socioeconomic datasets are drawn from state- and county-level data from American Community Survey five-year estimates [11]. These datasets require only minimal processing—merging the shapefiles with the demographic datasets—before being stored in DuckDB to enable efficient queries and analysis.

System Architecture. GeoVisA11y's frontend is implemented in MapboxJS and React.js, with a backend built using Python's Flask framework and DuckDB for geospatial data management. The system uses Google Cloud Platform for hosting and MongoDB for user log data storage.

Speech Integration. Speech input is implemented by integrating the browser's WebAudio API with a server-side pipeline, using OpenAI's Whisper model for accurate speech-to-text conversion.

Figure 6: AI chat key interactions. (A) Ctrl + M activates the chat if previously focused on the map. (B) Upon receiving a query, system repeats the user query followed by a status indicator. (C) The three example questions can be refreshed through Ctrl + I. (D) Asking *"What else can you do?"* can see/hear all the available query types.
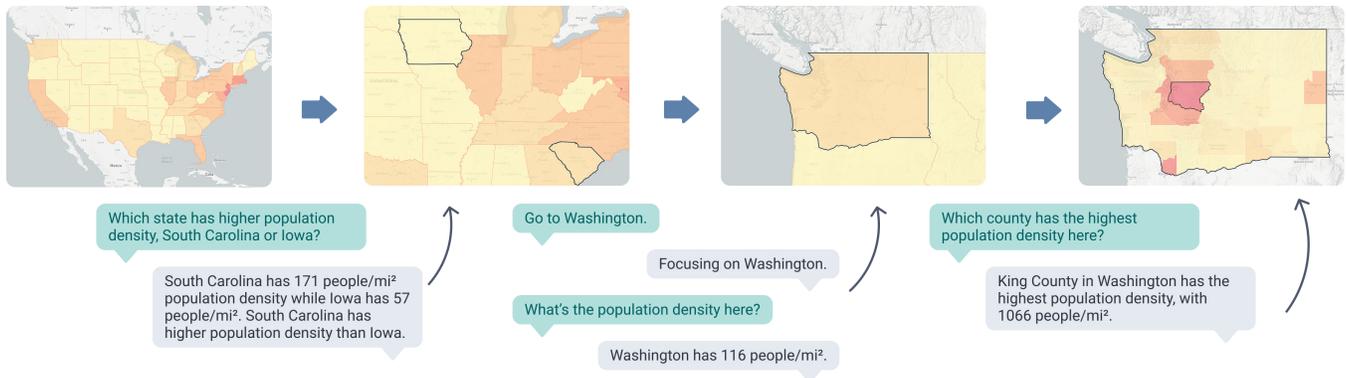
Figure 7: Bidirectional synchronization between map and chat interface. (A) Comparative queries automatically adjust map zoom to display all referenced states, with textual responses providing enumerated values and answers. (B) Users navigate via natural language commands and ask location-specific questions using deictic references like "here." (C) Users can zoom to county level and query detailed local information.

5 User Study

To examine user engagement with an accessible geovisualization QA system and compare interaction patterns and query strategies between blind and sighted users (RQ1–4), we conducted a user study with both groups. As the first AI-based accessible geovisualization QA system, our study goals were twofold: first, to explore what *types of queries* do screen-reader users make when interacting with an accessible geovisualization; and second, to examine *how well* the current GeoVisA11y prototype supports these queries, builds appropriate mental models of the underlying data, and leads to accurate insights and takeaways. All study sessions were audio and video recorded. We used a combination of qualitative and quantitative measures to address our research goals.

5.1 Participants

We recruited six screen-reader users (referred to as B1-B6) through mailing lists and six sighted researchers and practitioners in human-computer interaction, computer science, and urban planning, with varying levels of U.S. geography knowledge (S1-S6). Before the study, participants filled out a questionnaire to record their demographic information, education level, daily computer usage, frequency of interaction with map visualizations, AI chatbots, and
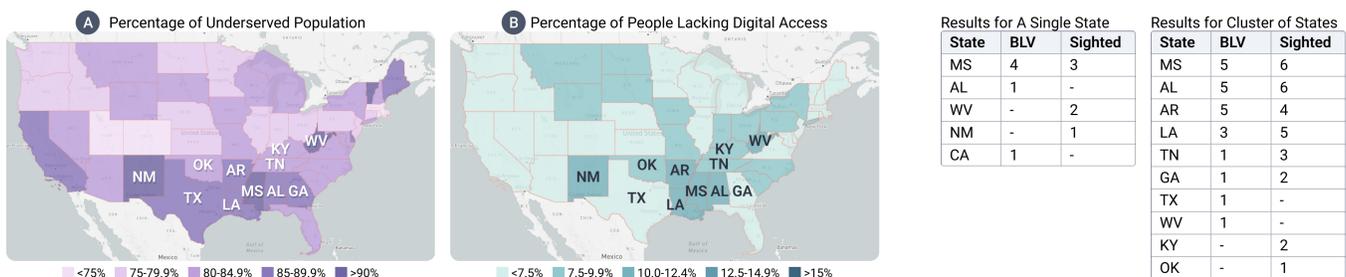
voice assistants. Screen-reader users answered additional questions about their screen-reader software and vision-loss level (Table 3).

5.2 Study Design

The core of our study comprised two data analysis tasks. We intentionally designed tasks requiring complex, multi-step analytical processes rather than single-question interactions. To create a natural exploratory experience and ensure participant engagement, we based our tasks on existing data stories and real-life scenarios. Prior to the study sessions, we conducted pilot studies with co-design participants and one sighted member of the research team.

In Task 1, participants were asked to imagine themselves as decision-makers responsible for distributing *State Digital Equity Planning Grant* funding across states—a task adapted from the U.S. Census on mapping digital equity, which originally contained inaccessible geovisualizations[3]. Participants were asked to select both a single state and a cluster of four to six geographically contiguous states most deserving of funding allocation.

---

[3]https://www.census.gov/library/stories/2022/05/mapping-digital-equity-in-every-state.html

State	BLV	Sighted
MS	4	3
AL	1	-
WV	-	2
NM	-	1
CA	1	-

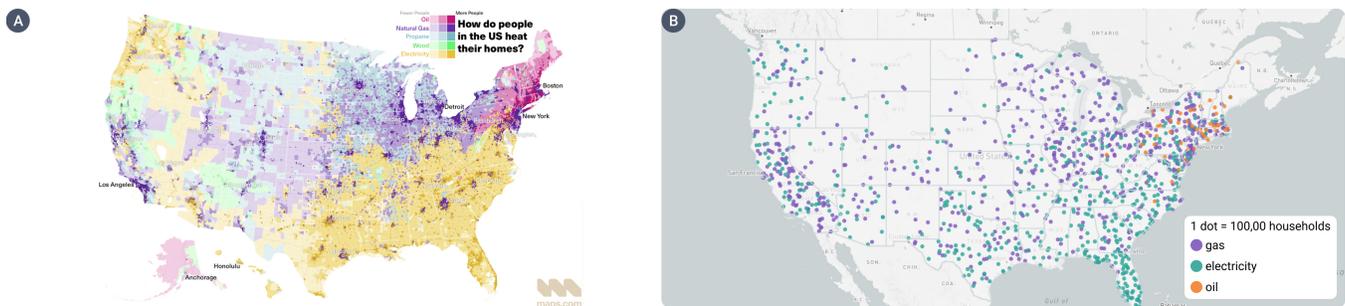| State	BLV	Sighted
MS	5	6
AL	5	6
AR	5	4
LA	3	5
TN	1	3
GA	1	2
TX	1	-
WV	1	-
KY	-	2
OK	-	1

Figure 8: Task 1 visualizations and results. (A) Choropleth map showing the percentage of underserved population across states. (B) Choropleth map showing the percentage of people lacking digital access. Tables present participant funding choices for both a single state and a cluster of states, revealing convergence in pattern identification between BLV and sighted participants.

PID	Age	Gender	Education Level	Computer Usage	Online Map Usage	AI Chatbot Usage	Voice Assistant Usage	Screen-reader	Vision-loss Level
C1	25-34	Woman	Masters	5 hours+	Occasionally	Occasionally	Daily	NVDA	Blind since birth
C2	45-54	Man	Masters	3-5 hours	Frequently	Occasionally	Daily	JAWS	Lost vision gradually, Complete blindness
B1	25-34	Man	Bachelor	5 hours+	Rarely	Regularly	Occasionally	NVDA	Blind since birth
B2	55-64	Woman	Bachelor	5 hours+	Rarely	Regularly	Daily	JAWS	Lost vision gradually, Complete blindness
B3	55-64	Man	Masters	5 hours+	Regularly	Daily	Daily	JAWS	Lost vision gradually
B4	35-44	Woman	High school	0-1 hours	Occasionally	Never used	Daily	JAWS/Voice Over	Complete blindness
B5	55-64	Man	Bachelor	5 hours+	Frequently	Frequently	Frequently	JAWS/Voice Over/NVDA	Lost vision gradually, Complete blindness
B6	25-34	Man	High school	5 hours+	Occasionally	Daily	Daily	NVDA	Blind since birth
S1	25-34	Woman	Masters	5 hours+	Occasionally	Daily	Rarely		
S2	25-34	Woman	Masters	5 hours+	Occasionally	Regularly	Frequently		
S3	25-34	Man	Doctoral	5 hours+	Daily	Daily	Never used		
S4	25-34	Man	Doctoral	5 hours+	Frequently	Frequently	Occasionally		
S5	18-24	Woman	Bachelor	5 hours+	Occasionally	Daily	Never used		
S6	25-34	Man	Masters	5 hours+	Occasionally	Frequently	Frequently		

Table 3: Participant demographics and technology experience. C1- C2 are co-design participants, B1- B6 are screen-reader users, S1-S6 are sighted participants. Computer usage is based on daily usage hours. For all other technology usage, frequently = multiple times per week, regularly = weekly, occasionally = few times a month, rarely = less than once a month.

For Task 2, we adapted a *Washington Post* article on U.S. home heating[4]. To explore how well our system supports sighted users when immediate visual patterns were unavailable, we transformed the original choropleth maps into dot density maps. Each dot represented 100K households and was randomly distributed across its corresponding geographic region (Figure 9). Participants were asked to identify predominant heating fuel sources in different U.S. regions and to explain these patterns.

[4]https://www.washingtonpost.com/climate-environment/interactive/2023/home-electrification-heat-pumps-gas-furnace/

5.3 Procedure

Study sessions were conducted individually via Zoom, with each session lasting between 90 and 120 minutes. Every session followed a four-part protocol.

Introduction. Sessions began with an introduction to the project and an overview of the study agenda. After obtaining informed consent, we asked participants about their experience with AI chatbots and geovisualizations.

Tutorial & Free Exploration. The researcher then guided participants through a tutorial using the U.S. population density dataset, covering GeoVisA11y's capabilities: answerable question types, keyboard navigation, zoom controls, shortcut keys, and voice input. Participants were then encouraged to freely explore the system.

Figure 9: Task 2 comparison of (A) original Washington Post choropleth map showing predominant home heating fuel sources and (B) our adapted dot density version. The dot density map removes the immediate visual patterns available in the choropleth map, allowing us to explore how well our system could support sighted users when immediate visual patterns are unavailable.

Study Tasks. Participants were given up to 30 minutes to complete the two analytical tasks. As an exploratory study focused on analytical approaches rather than completion rates, we allowed participants sufficient time to explore and demonstrate reasoning. If the combined tutorial, exploration, and Task 1 phases exceeded one hour, participants were not asked to proceed to Task 2, ensuring adequate time remained for the interview and debrief. For both tasks, we emphasized our interest in studying participants' analytical processes rather than their derived answers. When researchers observed immediate inaccuracies in system responses during the session, they notified participants to ensure the study could continue effectively and to prevent participants from building an analysis on incorrect information.

Semi-structured Interview & Debrief. Following task completion, the researcher conducted semi-structured interviews to gather qualitative insights about participants' experiences, challenges, and suggestions for improvement. At the conclusion of each session, participants received a comprehensive debrief explaining the technical architecture of GeoVisA11y and were shared the original sources of the data stories upon request.

5.4 Evaluation Measures

We used both objective and subjective evaluation measures to assess system performance. The objective measure consisted of the percentage of questions that the system answered correctly during user evaluation. We determined correctness by comparing system responses against ground truth data from the original datasets (American Community Survey) and verified geographic information. For interpretive questions, two researchers independently assessed answer accuracy based on whether responses were factually supported by the underlying data. In cases of disagreement, the two researchers discussed their assessments until reaching agreement.

For subjective evaluation, we adapted measures from prior work on map visualization usage: *reading, analysis*, and *interpretation* [52]. Given our focus on interactive maps, we also included *navigation* as a fourth dimension. Additionally, we evaluated chat functionality and the relative importance of both map and chat components. All subjective measures were captured using 7-point Likert scale questions followed by open-ended questions (Figure 10).

5.5 Data and Analysis

All study sessions were audio and video recorded. Our analysis focused on summarizing high-level themes through qualitative open coding [15]. One researcher developed a set of themes based on video transcripts and observational notes, then thematically coded the responses. A second researcher verified these themes [62]. For system performance, two researchers independently validated all answers; disagreements were discussed until consensus.

6 Findings

Overall, participants reacted positively to the GeoVisA11y prototype. BLV participants particularly valued the system's clear and detailed spatial descriptions, the ability to navigate maps independently using natural language commands, and access to analytical capabilities previously unavailable in accessible geovisualization

tools. Half of the BLV participants requested access to GeoVisA11y for continued use after the study session.

All participants completed at least one study task. On average BLV participants spent 30.9 minutes (*SD* = 12.2) interacting with the tutorial site, 18.2 minutes (*SD* = 5.3) on Task 1, and 17.3 minutes (*SD* = 0.75) on Task 2. Sighted participants spent 18.6 minutes (*SD* = 5.4) interacting with the tutorial site, then 14.2 minutes (*SD* = 4.5) on Task 1 and 19.0 minutes (*SD* = 5.3) on Task 2. Below we organize our findings around our research questions.

6.1 Types of Query Asked & System Performance (RQ1)

A key function of GeoVisA11y is the ability to pose natural language geoanalytic questions or enact UI commands via the chat interface. We begin by analyzing the types of queries posed and how well GeoVisA11y responded.

Types of Query. Our participants (*N* = 12) asked GeoVisA11y a total of 346 questions. Table 4 details the query type distribution by user group. Both BLV and sighted users shared the same top three most frequent query categories. These were *general knowledge* queries (*e.g., "Why do Midwestern states use predominantly gas?"*), followed by *pattern* queries (*e.g., "Is there a pattern in this region regarding broadband access?"*), and *action* queries (*e.g., "Take me back to Arizona"*). Usage patterns then diverged. BLV users most frequently asked *retrieval* (*e.g., "How many households use gas heating in Texas? "*) and *find extremum* (*e.g., "Which state in the United States has the highest percentage of people lacking broadband access? "*) queries, while sighted users favored *aggregate* (*e.g., "Give me the population density for each of these and the standard deviation compared to the mean."*) and *sort* (*e.g., "Tell me the top 5 states lacking broadband access"*) queries. As expected, *visual queries* (shape and spatial relationship) were asked almost exclusively by BLV users (*e.g., "What's the shape of Alabama?"* and *"What are Florida's neighboring states?"*).

Impressively, we found that GeoVisA11y supported 92% of the total questions asked. The remaining 8% fell into two categories. 2% of the questions were outside of our defined query types (Table 4); these included requests to create additional visualizations (*e.g., "Can you highlight the six states on the map?"*) or to recolor visualizations based on different data (*e.g., "Can you color the states based on the most common fuel source?"*). The other 6% were correctly classified into supported query types (*e.g., Aggregate*), but GeoVisA11y could not provide an answer because the specific calculation required was not yet implemented in our backend operations (*e.g.,"Give me the population density for each of these and the standard deviation compared to the mean"*).

Accuracy. Because GeoVisA11y incorporates statistical analyses, database queries, and LLM-based analytics, evaluating response accuracy is critical. In total, 289 (83.8%) questions were answered correctly. Of the 56 incorrect responses, we identified specific failure points within our pipeline (Table 5). Query Refiner made 15 (26.8%) mistakes, primarily failing to retrieve the current user focus or relevant chat history for disambiguation. For example, when a user focused on Washington state and asked *"What's the population density here?"*, the system returned a generic definition of population density: *"To determine population density, divide the total*

Query Type	BLV	%	Sighted	%	Total	%	Example
Action	27	18.0%	29	14.9%	56	16.2%	*"Take me back to Arizona."* (B4)
Retrieve	22	14.7%	10	5.1%	32	9.3%	*"How many households use gas heating in Texas?"* (B2)
Compare	4	2.7%	4	2.1%	8	2.3%	*"Compare OK and NE populations."* (B5)
Find Extremum	12	8.0%	9	4.6%	21	6.1%	*"Which state in the United States has the highest percentage of people lacking broadband access?"* (B3)
Aggregate	3	2.0%	16	8.2%	19	5.5%	*"Give me the population density for each of these and the standard deviation compared to the mean."* (S6)
Filter	0	0.0%	4	2.1%	4	1.2%	*"Which counties in this state have higher population density than the average population density of the state?"* (S3)
Sort	3	2.0%	10	5.1%	13	3.8%	*"Tell me the top 5 states lacking broadband access."* (B1)
Cluster	0	0.0%	0	0.0%	0	0.0%	
Pattern	28	18.7%	50	25.6%	78	22.6%	*"Is there a pattern in this region regarding broadband access?"* (B1)
Outlier	0	0.0%	2	1.0%	2	0.6%	*"Can you find significant outliers to population density?"* (S5)
Legend	0	0.0%	0	0.0%	0	0.0%	
Shape	5	3.3%	1	0.5%	6	1.7%	*"What's the shape of Alabama?"* (B6)
Spatial Relationships	8	5.3%	0	0.0%	8	2.3%	*"What are Florida's neighboring states?"* (B5)
Visualization Knowledge	0	0.0%	2	1.0%	2	0.6%	*"What's a choropleth map?"* (S4)
General Knowledge	38	25.3%	51	26.2%	89	25.8%	*"Why do Midwestern states use predominantly gas?"* (B2)
Others	0	0.0%	7	3.6%	7	2.0%	*"Can you highlight the six states on the map? (S1)"*

Table 4: The distribution of query types by user group, participants (N =12) asked GeoVisA11y 346 questions in total. Both user groups showed similar usage patterns for the three most common query types (General Knowledge, Pattern, and Action), while BLV users more frequently asked visual queries (Shape, Spatial Relationships) and Retrieve queries compared to sighted users. This table categorizes all questions based on query type, regardless of whether the system successfully answered them.

population of an area by its land area (in square miles or kilometers)..." Scope Assessor made 12 (21.4%) mistakes, incorrectly routing questions either to GPT when they should have been processed locally or to local processing when they required external knowledge. For instance, when asked *"What is the percentage of underserved populations here?"*, the system incorrectly determined this to be out of scope and provided a generic definition rather than analyzing the available demographic data. Query Processor made 3 (5.4%) classification errors, such as misclassifying *"Why do rural counties use more electricity than gas?"* as a comparison query rather than general knowledge, resulting in data comparisons instead of explanatory content. The remaining 26 (46.4%) incorrect responses stemmed from limitations beyond our current system capabilities, including requests for visualization creation (e.g., *"Can you create a heat map for the gas heating fuel type layer?"*), GPT hallucinations (such as incorrectly claiming geographic connections), ambiguously phrased user questions (e.g., incomplete queries like *"states with lower population density"*), and server errors.

6.2 Geovisualization Engagement (RQ2 & RQ3)

Here we analyze how both BLV and sighted participants engaged with GeoVisA11y across four categories of geovisualization usage: reading, analysis, interpretation, and navigation.

Map Reading. When asked to rate the ease of identifying specific values or regions on the map (on a scale of 1-7), BLV participants reported a median of 5 (*IQR* = 0.75), while sighted participants reported a median of 6 (*IQR* = 0.75) (Figure 10). For BLV participants,

GeoVisA11y effectively supported basic map reading tasks while helping recall their geographic knowledge. As B3 noted: *"This is really neat ... because sometimes, I'll go through all 50 states in my head, and I can't remember things like, 'Oh, Rhode Island is south of Massachusetts.'"* Regarding the quality of answers provided for map reading, BLV participants described the system's responses as *"very informative"* (B5), *"clear, and concise"* (B5). However, some participants (B3, B4) raised concerns about verifying the provided information. As B4 noted: *"I'd probably ask Alexa and see if their answers match; if they do for a couple of questions, I would trust [GeoVisA11y] more."* Beyond data-related queries, participants appreciated GeoVisA11y's ability to describe state shapes and geographic relationships, with B6 commenting *"It really describes things well, like the shape of Alabama."*

Sighted participants approached map reading differently, valuing how GeoVisA11y prevented misinterpretation through natural language queries. S1 observed: *"I think being able to ask natural questions really helps because otherwise I would have thought that 5% represents the population that has broadband access, which is incorrect."* Having both the map and the chat let participants verify information across channels; as S6 mentioned, *"I was able to corroborate information given by the chat with the map."*

Both user groups benefited from GeoVisA11y's highlighting function, though for different reasons. For sighted users with limited U.S. geographic knowledge, the feature provided immediate orientation: *"I like that it highlighted the different states related to the question, because I don't really know where New Jersey is"* (S2). For BLV users,

Error Type	Count	%	Example Question	GeoVisA11y's Incorrect Answer	Error Analysis
Input Classifier	0	0.0%			
Query Refiner	15	26.8%	*"Whats the population density here?"*	To determine population density, divide the total population of an area by its land area (in square miles or kilometers). For example, if a city has 100,000 people and covers 50 square miles, the population density is 2,000 people per square mile.	Failed to catch the focus state ['Washington'].
Scope Assessor	12	21.4%	*"What is the percentage of underserved populations here?"*	Underserved populations refer to groups lacking adequate access to services, often including low-income, minority, or rural communities. The percentage varies by region and context, but specific statistics can be found in reports from organizations like the U.S. Census Bureau or local health departments.	Incorrectly determined this question to be out of scope.
Query Processor	3	5.4%	*"Why do rural counties used more electricity than gas?"*	Here's a breakdown of predominantly used heating fuels in Georgia. For urban counties: 75.0% predominantly use electricity, 25.0% predominantly use gas. For rural counties: 97.2% predominantly use electricity, 2.8% predominantly use gas.	Incorrectly classified this question as *compare* when it should have been *general knowledge*.
Others	26	46.4%	*"Can you create a heat map for the gas heating fuel type layer?"*	I can't create visual content like heat maps.	User asking for additional visualizations that is beyond the system current capabilities.

Table 5: Error analysis of incorrectly answered questions by pipeline component. The table shows 56 errors (16.2% of 346 total questions) across system components, with representative examples of failed queries, incorrect responses, and explanations.

highlighting served as a trigger for screen-reader focus on the map, creating a multimodal signal of spatial information.

Map Analysis. When asked to rate the ease of identifying spatial trends and patterns on the map (on a scale of 1-7), BLV participants reported a median of 5.5 (*IQR* = 1.75), while sighted participants reported a median of 7 (*IQR* = 0.75) (Figure 10). BLV participants appreciated the system's ability to translate complex spatial data into accessible language. B2 noted *"It uses language that's not complicated. It's very general, easy to understand."* B5 highlighted GeoVisA11y's analytical advantage over other tools: *"This is much better than most maps that I've seen, because you can't ask those kind of questions and get that feedback. Mostly, ... [other maps are] just a superficial layer of the map."*

Sighted participants similarly valued the system's analytical capabilities, but with greater emphasis on its dynamic visualization features. S3 appreciated that *"it automatically prepares these additional visualization layers to help me [identify patterns],"* while S5 noted that *"the coolest thing is that you can add layers to help you visualize certain patterns."* S1 noted GeoVisA11y's advantage compared to accessing only the visualization: *"If I only had the map, the information will all be scattered, and I'll need to manually collect them, note them down, then compare the numbers since the color doesn't really show much when the numbers are very close."*

Map Interpretation. When asked to rate how well the system helped them understand why certain patterns or relationships exist (on a scale of 1-7), BLV participants reported a median of 6 (*IQR* = 0.75), while sighted participants reported a median of 5 (*IQR* = 1.5) (Figure 10). B5 expressed how GeoVisA11y facilitated engaging pattern-finding: *"What I enjoyed the most was the ability to query across datasets, to find relevancy from the information within each and come up with an answer that gave me more specific targets."* However, they sometimes requested more specific explanations: *"I remember asking the system to explain the [heating fuel] pattern, and it mentioned climate factors. If we're able to get specific data like average temperatures for different regions, it would provide me an immediate snapshot of why those patterns exist"* (C2).

Sighted participants demonstrated diverse approaches to map interpretation based on their background knowledge. S4 relied on the system for explanations after identifying the patterns independently, asking questions like, *"Why does Washington use more electricity than gas for heating homes?"* While S5 used the system for pattern identification, then drew on personal expertise for interpretation, stating, *"Ah, I can definitely think of the potential reasons for this. In Washington, electricity is generated from various renewable sources, like hydropower and wind."* The differences in interpretation strategies highlight how GeoVisA11y accommodated diverse knowledge levels, providing necessary context for users with limited domain expertise (both BLV and sighted) while facilitating
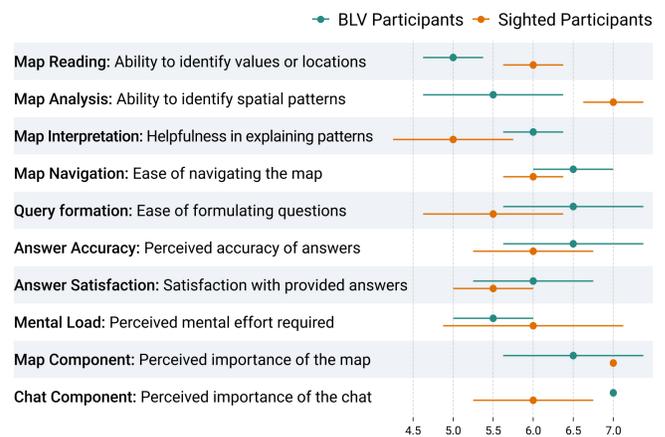
Figure 10: Likert scale ratings (1-7) across multiple dimensions, where 7 represents better outcomes (*i.e.,* easier, more accurate, more important). Plot shows median values with IQR for each user group. BLV participants unanimously rated the chat component as 7/7, while sighted participants unanimously gave the map component 7/7.

more efficient pattern identification for users who could then interpret findings using their specialized knowledge.

Map Navigation. When asked to rate their navigation experience, BLV participants reported a median of 6.5 (*IQR* = 1), and sighted participants reported a median of 6 (*IQR* = 0.75). BLV participants highlighted the unprecedented autonomy offered by GeoVisA11y's navigation: *"This is pretty cool because I have never understood mapping or tactile graphs this way before. When you're studying history, you're always feeling tactile graphs just to fill in the blanks or feel what the teachers ask you to. But now I can explore with free will, using the arrow keys to see for myself"* (B1). Some participants reported that navigation triggered their latent spatial knowledge: *"Oh, wow. I just realized what I did because of my knowledge of the Northeast. I arrowed around because I know Vermont is pretty close to New York"* (C2).

Navigation preferences differed between groups, with BLV participants primarily using keyboard navigation and sighted participants preferring direct mouse interaction. As S6 explained: *"I was expecting that I would be able to click on the states as well because I'm more of a mouse user than a keyboard user."* Despite these different interaction preferences, both groups particularly valued the semantic navigation commands. B6 mentioned: *"Being able to tell it where I wanted to go, like 'go to Georgia,' that's not something most programs use. It's really nice."* (B6), while sighted users found it helpful when dealing with unfamiliar geography: *"I think this go-to function is very helpful because I don't actually know where the states are given my limited knowledge of U.S. geography"* (S2).

6.3 Task Completion & Querying Strategies (RQ4)

Here we analyze how BLV and sighted participants completed Task 1 (distributing federal broadband funding) and Task 2 (identifying regional heating fuel patterns), focusing on the querying approaches they employed. Due to time constraints, only two of the six BLV participants completed Task 2, so our quantitative analysis focuses on Task 1, with qualitative insights drawn from Task 2 where relevant.

BLV Participants. Overall, five of the six BLV participants came up with reasonable answers for Task 1, completing the task in an average of 18.2 minutes (*SD* = 5.8 minutes). For individual state funding recommendations, participants selected Mississippi and Alabama. For the cluster funding recommendation, participants identified Mississippi, Arkansas, Alabama, Louisiana, Texas, Georgia, West Virginia, and Tennessee (Figure 8). One participant insisted on funding California despite the data, explaining it was their home state. BLV participants employed diverse query strategies, including finding extremum, sorting, comparing, and inquiring about geographic patterns and spatial relationships. They used a combination of verbal queries and key navigation to determine spatial relationships between target states.

Participants demonstrated adaptability when encountering system misinterpretations. B5 initially used ambiguous pronouns beyond the system's capabilities by asking *"What are **my** neighboring states?"*. When they recognized the system struggled to answer such a question, they self-corrected by specifying state name. When B2

asked about *"fuel patterns"* and received information about transportation fuels, they refined their query to specify *"heating fuels for homes"*. In another instance, B3 requested *"Select a group of six contiguous states that have low access to broadband,"* which the system incorrectly answered *"Mississippi, Arkansas, Louisiana, West Virginia, Kentucky, Alabama"*. The participant recognized the geographical error, noting *"West Virginia isn't connected right?"* and then verified using keyboard navigation to confirm the mistake.

Notably, BLV participants demonstrated greater reliance on the researcher's task descriptions when formulating queries and higher expectations of the system's analytical capabilities. For example, B3 directly asked: *"Which state do you think funding should be appropriated to based on the percentage of lack of broadband access?"*

Sighted Participants. All six sighted participants developed reasonable answers for Task 1, recommending Mississippi, West Virginia, and New Mexico as individual states, and identifying a southeastern cluster including Mississippi, Alabama, Louisiana, and Arkansas as priority regions (Figure 8). Comparing results from BLV participants with sighted participants reveals a general convergence in pattern identification between the two groups. For single state selection, both BLV and sighted participants identified Mississippi most frequently. When analyzing clusters of states, Mississippi, Alabama, Louisiana and Arkansas were chosen most frequently in both groups. The results begin to diverge for secondary states, with BLV participants selecting Tennessee, Georgia, Texas, and West Virginia (each chosen once), while sighted participants identified Tennessee (3), Georgia (2), Kentucky (2), and Oklahoma (1).

Although sighted participants used similar query types (finding extremum, sorting, comparing, and geographic patterns), they visually narrowed selections and determined geographical connectivity before making specific queries. Their system usage varied based on domain expertise and map complexity. S4 (proficient with gaming visualizations) identified patterns independently but used the system for interpretations, while S5 (urban planning researcher) primarily applied their domain knowledge after using the system to locate patterns. Geographic familiarity influenced interaction approaches—S1 (unfamiliar with U.S. geography) relied heavily on text commands like *"Go to Vermont"*. All sighted participants demonstrated increased chat usage when working with dot density maps compared to choropleth maps.

While the system could not answer certain questions from BLV participants due to ambiguity or misinterpretation, most system limitations encountered by sighted participants were related to requests for expanded capabilities. Sighted participants frequently requested enhanced visualization features, such as *"Can you plot the size of each state?"* (S3) and *"Can you make the dots bigger?"* (S5). Additionally, sighted users requested more sophisticated statistical analysis capabilities, including standard deviation calculations and percentile-based rankings.

6.4 Suggested Improvements & Applications.

We next summarize participant feedback on potential system enhancements for the map component, the chat component, and the synchronization between these two.

Enhanced Navigation. BLV participants frequently (*N* = 4/6) requested improved navigation capabilities. B1 and B2 specifically

cited navigation as both their most enjoyable and most challenging aspect of the system. The current implementation supports only cardinal direction navigation, which falls short for the complex spatial relationships. For instance, moving south from Vermont directs users to Massachusetts, but navigating back north also directs them to New Hampshire (since both Vermont and New Hampshire border Massachusetts to the north, with New Hampshire's centroid closer to Massachusetts's centroid). This asymmetric directionality created confusion for BLV participants unfamiliar with the intricate spatial relationships between states.

More Specific, Structured Answers with Source Attribution. Participants from both groups observed that answers related to map interpretation lacked specificity (S1, S4, B2). Unstructured responses or responses structured differently than previous responses made participants question answer accuracy, as S1 noted, *"I paused a little and wondered is this part of information that's hallucinated?"* B3 suggested incorporating hyperlinks along with answers to facilitate verification of information sources.

More Transparent Reasoning Processes. When posing higher level synthesis questions that required integration across multiple datasets, participants wanted greater transparency into the system's analytical processes. S3 noted that the system *"doesn't explicitly outline why it made the decision"*, while S6 commented, *"I was looking to understand why the system would consider certain counties to be outliers."*

Improved Map-Chat Synchronization. S2 identified a mismatch between the map and the chat: *"Sometimes, there's a mismatch between what I think the chat is selecting versus what the map is actually focusing on."* This mismatch occurred when GeoVisA11y couldn't answer using local data, resulting in GPT responses that didn't trigger a geographic unit on the map. S1 wanted more direct interaction between the visual and conversational interfaces: *"I wish I can just click on the map to highlight four states and then ask what are the average income of these four."*

Potential Applications. Although we did not explicitly solicit input on potential applications, BLV participants spontaneously identified various use cases, including interpreting election maps, accessing supplementary information after listening to the news, and understanding resource distribution following natural disasters. B5 drew on professional experience to suggest applications: *"I worked for [a state government department]. There is an audience for this application in public service, especially among those who are in charge of distributing funds or assessing taxes."* B6 highlighted the educational value: *"This could be really good for college students who are going into any kind of social sciences."*

7 Discussion & Future Work

Below, we discuss key findings, identify challenges, and suggest directions for future work.

Capturing and Verbalizing Spatial Patterns. Our evaluation revealed substantial overlap in pattern recognition between BLV and sighted participants, demonstrating that our approach represents a step toward creating shared understanding of geospatial patterns. Nevertheless, we observed important differences highlighting improvement opportunities. For instance, S5 incorrectly identified New Mexico as having the *"highest percentage (of underserved population)"* when West Virginia actually held this distinction. This misinterpretation is likely influenced by New Mexico's large visual area on the map (Figure 8). BLV users had to perform additional queries about spatial relationships and navigate multiple times around the map to confirm spatial connections, as the pattern summaries only included two states per cluster (a design choice to avoid information overload). While our implementation relied on LISA clusters and Moran's I [7], all traditional cluster identification methods, (Moran's I [7], Geary's C [8], Getis-Ord Gi* [25]) have limitations in accurately representing patterns that would be visually apparent to sighted users [68]. Future work should explore pattern detection approaches that incorporate estimate error and allow user-defined thresholds, making descriptions more intuitively meaningful and practically valuable [68].

Communicating Accuracy. Our evaluation revealed that although BLV participants generally trusted the system's answers, they lacked easy methods to verify responses. Despite prior work [33] emphasizing the importance of communicating uncertainty levels in AI-generated responses, we deliberately chose not to explicitly distinguish which answers were obtained from the local dataset versus those generated by GPT for two reasons: (1) our pipeline classifiers rely on GPT, which itself can make errors in assessment; highlighting only those answers that originated entirely from GPT might inadvertently cause users to place excessive trust in other responses; and (2) many answers combined local raw data with GPT summaries, making it challenging to communicate the varying uncertainty levels between different components in the same response. A more effective approach would incorporate source links in responses as suggested by B3 and increasingly implemented by commercial AI tools such as Perplexity [2].

Incorporating Dynamically Guided Prompting. Our study revealed that meaningful interaction with visualization QA systems requires some level of data literacy. In several cases, participants struggled not because of system limitations but due to how they formulated their queries. For example, multiple participants asked about "fuel" instead of "heating fuel" and consequently received explanations about automotive fuel rather than household energy sources. These challenges highlight the need for dynamic guided prompting to avoid query ambiguity, where the system proactively clarifies user intent during the querying process. As S1 observed, the system has the potential to correct misinterpretations. Such prompting could be implemented with simple clarifying questions such as: *"Are you asking about fuel consumption in general or specifically about household heating fuel?"* The system could also consider users' current map focus to offer contextually relevant guided questions.

Supporting Direct Manipulation Both screen-reader and sighted participants expressed interest in more direct manipulation capabilities, such as selecting multiple states simultaneously for comparison and analysis. This desire for additional direct manipulation reflects broader limitations in purely language-based interfaces. These limitations echo challenges identified decades ago with command-line interfaces: indirect engagement, semantic distance, and articulatory distance [47]. Recent work such as Direct-GPT [47] and Textoshop [46] has begun exploring ways to integrate direct manipulation with LLM-powered interfaces. Implementing

higher-level direct manipulation capabilities for geovisualization QA systems is an exciting direction for future work and would benefit all user groups.

Supporting Users with Diverse Abilities and Expertise. Our study revealed that while screen-reader and sighted users have distinct needs, there was considerable overlap in how both groups benefited from GeoVisA11y. We observed that sighted participants demonstrated varying levels of system usage based on their domain expertise and geographic knowledge. This finding challenges a binary view of accessibility and suggests that the ability to accurately read, analyze, and interpret geovisualizations applies to a broad spectrum of users. When asked to rate the additional benefits of GeoVisA11y compared to viewing the visualizations alone, sighted participants gave it a median score of 6 on a scale of 1 (no additional benefit) to 7 (significant improvement). This aligns with prior work on designing voice interfaces, which argues that *"when the blind lead the sighted through voice interface design, both blind and sighted users can benefit"* [1]. We acknowledge that our screen-reader user group included only blind users and no low-vision users. Low-vision individuals might rely on hybrid strategies such as combining screen magnification or high-contrast modes with conversational queries. Future work should investigate how users across this fine-grained spectrum of visual abilities interact with geovisualization QA systems differently, as these insights could inform more nuanced design recommendations for assistive technologies that support diverse perceptual needs and preferences.

Broader Deployment and Scalability. While our study focused on U.S. datasets, GeoVisA11y's architecture is designed for broader applicability across diverse geospatial domains. Currently, the system requires minimal data preprocessing—for all the interactive maps used in the study, researchers simply matched geospatial boundary data with census tables. Deployment barriers could be further reduced by automating the system's configuration phase. Specifically, this could involve standardizing a metadata schema to automatically ingest data tables, variable definitions and metric types, as well as automating the manual adjustments currently needed to ensure natural navigation interactions.

GeoVisA11y's modular pipeline architecture can be deployed as a browser plugin for existing geovisualization platforms. By extracting underlying data and geographic boundaries from web-based maps, the system could augment any choropleth or statistical map with natural language querying, allowing users to retrofit accessibility features without complete interface redesigns.

The system's reliance on general-purpose LLMs for query processing and pattern detection, rather than domain-specific trained models, supports scalability across different contexts through few-shot prompting with domain-appropriate examples. Whether applied to urban planning, environmental monitoring, or economic datasets, our findings suggest that this fundamental approach—combining spatial data analysis with natural language interfaces—provides a scalable foundation for making geovisualizations accessible across diverse user populations.

8 Conclusion

We introduced GeoVisA11y, an LLM-based question-answering system that makes geovisualizations accessible to screen-reader users while also benefiting sighted users. Our evaluation with twelve participants demonstrated that both user groups successfully engaged with geospatial data through different but complementary interaction strategies—screen-reader users combining verbal queries with keyboard navigation, and sighted users leveraging visual assessment with targeted queries. Despite these different approaches, both groups identified similar patterns, showing that our system creates a shared understanding of geovisualizations. GeoVisA11y demonstrates that inclusive design enhances spatial data analysis for all users, regardless of visual ability.

Acknowledgments

References

[1] Ali Abdolrahmani, Kevin M. Storer, Antony Rishin Mukkath Roy, Ravi Kuber, and Stacy M. Branham. 2019. Blind Leading the Sighted: Drawing Design Insights from Blind Users towards More Productivity-oriented Voice Interfaces. *ACM Transactions on Accessible Computing* 12, 4 (Dec. 2019), 1–35. doi:10.1145/3368426

[2] Perplexity AI. 2022. Perplexity. https://www.perplexity.ai

[3] G. Andrienko, N. Andrienko, P. Jankowski, D. Keim, M.-J. Kraak, A. MacEachren, and S. Wrobel. 2007. Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science* 21, 8 (Sept. 2007), 839–857. doi:10.1080/13658810701349011

[4] Natalia Andrienko and Gennady Andrienko. 2006. *Exploratory analysis of spatial and temporal data: a systematic approach.* Springer Science & Business Media.

[5] Luc Anselin. 1988. *Spatial econometrics: methods and models.* Vol. 4. Springer Science & Business Media.

[6] Luc Anselin. 1989. What Is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis. (1989).

[7] Luc Anselin. 1995. Local Indicators of Spatial Association—LISA. *Geographical Analysis* 27, 2 (April 1995), 93–115. doi:10.1111/j.1538-4632.1995.tb00338.x

[8] Luc Anselin. 2019. A Local Indicator of Multivariate Spatial Association: Extending Geary's *c*. *Geographical Analysis* 51, 2 (April 2019), 133–150. doi:10.1111/gean.12164

[9] Francis P. Boscoe and Linda W. Pickle. 2003. Choosing Geographic Units for Choropleth Rate Maps, with an Emphasis on Public Health Applications. *Cartography and Geographic Information Science* 30, 3 (Jan. 2003), 237–248. doi:10.1559/152304003100011171

[10] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, and Amanda Askell. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

[11] US Census Bureau. 2023. 2018-2022 ACS 5-year Estimates. https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2022/5-year.html Section: Government.

[12] US Census Bureau. 2024. TIGER/Line Shapefiles. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html Section: Government.

[13] Victor Soares Bursztyn, Jane Hoffswell, Eunyee Koh, and Shunan Guo. 2024. Representing Charts as Text for Language Models: An In-Depth Study of Question Answering for Bar Charts. In *2024 IEEE Visualization and Visual Analytics (VIS)*. 266–270. doi:10.1109/VIS55277.2024.00061

[14] Shuaichen Chang, David Palzer, Jialin Li, Eric Fosler-Lussier, and Ningchuan Xiao. 2022. MapQA: A Dataset for Question Answering on Choropleth Maps. doi:10.48550/arXiv.2211.08545 arXiv:2211.08545.

[15] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis.* sage.

[16] Jinho Choi, Sanghun Jung, Deok Gun Park, Jaegul Choo, and Niklas Elmqvist. 2019. Visualizing for the Non-Visual: Enabling the Visually Impaired to Use Visualization. *Computer Graphics Forum* 38, 3 (2019), 249–260. doi:10.1111/cgf.13686 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13686.

[17] Saurabh Datar, Ilana Marcus, Eli Murray, Ethan Singer, Alex Lemonides, and Christine Zhang. 2025. An Extremely Detailed Map of the 2024 Election. *The New*

York Times (Jan. 2025). https://www.nytimes.com/interactive/2025/us/elections/2024-election-map-precinct-results.html

[18] Pierre Dognin, Igor Melnyk, Youssef Mroueh, Inkit Padhi, Mattia Rigotti, Jarret Ross, Yair Schiff, Richard A. Young, and Brian Belgodere. 2022. Image captioning as an assistive technology: Lessons learned from vizwiz 2020 challenge. *Journal of Artificial Intelligence Research* 73 (2022), 437–459. http://www.jair.org/index.php/jair/article/view/13113

[19] Frank Elavsky, Lucas Nadolskis, and Dominik Moritz. 2023. Data Navigator: An Accessibility-Centered Data Navigation Toolkit. *IEEE Transactions on Visualization and Computer Graphics* (2023), 1–11. doi:10.1109/TVCG.2023.3327393

[20] Arlen Fan, Fan Lei, Michelle Mancenido, Alan M. Maceachren, and Ross Maciejewski. 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*. ACM, Honolulu HI USA, 1–17. doi:10.1145/3613904.3642132

[21] Danyang Fan, Alexa Fay Siu, Hrishikesh Rao, Gene Sung-Ho Kim, Xavier Vazquez, Lucy Greco, Sile O'Modhrain, and Sean Follmer. 2023. The Accessibility of Data Visualizations on the Web for Screen Reader Users: Practices and Experiences During COVID-19. *ACM Transactions on Accessible Computing* 16, 1 (March 2023), 1–29. doi:10.1145/3557899

[22] WGBH National Center for Accessible Media. 2025. Effective Practices for Description of Science Content within Digital Talking Books. https://www.wgbh.org/foundation/services/ncam/tools-resources/effective-practices-for-description-of-science-content-within-digital-talking-books

[23] Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G. Karahalios. 2015. DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, Charlotte NC USA, 489–500. doi:10.1145/2807442.2807478

[24] Tianyu Gao, Adam Fisch, and Danqi Chen. 2021. Making Pre-trained Language Models Better Few-shot Learners. doi:10.48550/arXiv.2012.15723 arXiv:2012.15723 [cs].

[25] Arthur Getis and J. K. Ord. 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* 24, 3 (July 1992), 189–206. doi:10.1111/j.1538-4632.1992.tb00261.x

[26] Joshua Gorniak, Yoon Kim, Donglai Wei, and Nam Wook Kim. 2024. VizAbility: Enhancing Chart Accessibility with LLM-based Conversational Interaction. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. ACM, Pittsburgh PA USA, 1–19. doi:10.1145/3654777.3676414

[27] Danna Gurari, Qing Li, Abigale J. Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P. Bigham. 2018. Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3608–3617. http://openaccess.thecvf.com/content_cvpr_2018/html/Gurari_VizWiz_Grand_Challenge_CVPR_2018_paper.html

[28] Enamul Hoque, Vidya Setlur, Melanie Tory, and Isaac Dykeman. 2018. Applying Pragmatics Principles for Interaction with Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (Jan. 2018), 309–318. doi:10.1109/TVCG.2017.2744684 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[29] W3C Web Accessibility Initiative (WAI). 2022. Complex Images. https://www.w3.org/WAI/tutorials/images/complex/

[30] Carsten Juergens. 2020. Digital Data Literacy in an Economic World: Geo-Spatial Data Literacy Aspects. *ISPRS International Journal of Geo-Information* 9, 6 (June 2020), 373. doi:10.3390/ijgi9060373 Number: 6.

[31] Carsten Juergens. 2020. Trustworthy COVID-19 Mapping: Geo-spatial Data Literacy Aspects of Choropleth Maps. *Kn - Journal of Cartography and Geographic Information* 70, 4 (2020), 155–161. doi:10.1007/s42489-020-00057-w

[32] Dae Hyun Kim, Enamul Hoque, and Maneesh Agrawala. 2020. Answering Questions about Charts and Generating Visual Explanations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. doi:10.1145/3313831.3376467

[33] Jiho Kim, Arjun Srinivasan, Nam Wook Kim, and Yea-Seul Kim. 2023. Exploring Chart Question Answering for Blind and Low Vision Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3544548.3581532

[34] N. W. Kim, G. Ataguba, S. C. Joyner, Chuangdian Zhao, and Hyejin Im. 2023. Beyond Alternative Text and tables: Comparative Analysis of Visualization Tools and Accessibility Methods. *Computer Graphics Forum* 42, 3 (2023), 323–335. doi:10.1111/cgf.14833 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14833.

[35] N. W. Kim, S. C. Joyner, A. Riegelhuth, and Y. Kim. 2021. Accessible Visualization: Design Space, Opportunities, and Challenges. *Computer Graphics Forum* 40, 3 (June 2021), 173–188. https://onlinelibrary.wiley.com/doi/10.1111/cgf.14298

[36] Raina Langevin, Ross J Lordon, Thi Avrahami, Benjamin R. Cowan, Tad Hirsch, and Gary Hsieh. 2021. Heuristic Evaluation of Conversational Agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. doi:10.1145/3411764.3445312

[37] Mosh Levy, Alon Jacoby, and Yoav Goldberg. 2024. Same Task, More Tokens: the Impact of Input Length on the Reasoning Performance of Large Language Models. doi:10.48550/arXiv.2402.14848 arXiv:2402.14848 [cs].

[38] Chu Li, Rock Yuren Pang, Ather Sharif, Arnavi Chheda-Kothary, Jeffrey Heer, and Jon E. Froehlich. 2024. AltGeoViz: Facilitating Accessible Geovisualization. In *2024 IEEE Visualization and Visual Analytics (VIS)*. 61–65. doi:10.1109/VIS55277.2024.00020

[39] Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2024. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics* 12 (2024), 157–173. https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00638/119630

[40] Bastien Liétard, Mostafa Abdou, and Anders Søgaard. 2021. Do Language Models Know the Way to Rome? *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP* (2021), 510–517. doi:10.18653/v1/2021.blackboxnlp-1.40

[41] Alan Lundgard and Arvind Satyanarayan. 2022. Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (Jan. 2022), 1073–1083. doi:10.1109/TVCG.2021.3114770 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[42] A.M. MacEachren and B.P. Campbell. 1992. Visualization. In *Geophraphy's Inner Worlds: Pervasive Themes in Contemporary American Geography*. Rutgers University Press, 99–137.

[43] Alan M MacEachren and Menno-Jan Kraak. 2000. Research Challenges in Geovisualization. *Cartography and Geographic Information Science* 28, 1 (2000).

[44] Kelly Mack, Emma McDonnell, Dhruv Jain, Lucy Lu Wang, Jon E. Froehlich, and Leah Findlater. 2021. What Do We Mean by "Accessibility Research"? A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3411764.3445412

[45] Gengchen Mai, Weiming Huang, Jin Sun, Suhang Song, Deepak Mishra, Ninghao Liu, Song Gao, Tianming Liu, Gao Cong, Yingjie Hu, Chris Cundy, Ziyuan Li, Rui Zhu, and Ni Lao. 2024. On the Opportunities and Challenges of Foundation Models for GeoAI (Vision Paper). *ACM Transactions on Spatial Algorithms and Systems* 10, 2 (June 2024), 1–46. doi:10.1145/3653070

[46] Damien Masson, Young-Ho Kim, and Fanny Chevalier. 2025. Textoshop: Interactions Inspired by Drawing Software to Facilitate Text Editing. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (2025).

[47] Damien Masson, Sylvain Malacria, Géry Casiez, and Daniel Vogel. 2024. DirectGPT: A Direct Manipulation Interface to Interact with Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. doi:10.1145/3613904.3642462

[48] Janet E. Mersy. 1990. Choropleth map design - a map user study. *Cartographica* 27, 3 (Sept. 1990), 33–50. doi:10.3138/1928-QQ57-3625-L024

[49] Prerna Mishra, Santosh Kumar, Mithilesh Kumar Chaube, and Urmila Shrawankar. 2022. *ChartVi*: Charts summarizer for visually impaired. *Journal of Computer Languages* 69 (April 2022), 101107. doi:10.1016/j.cola.2022.101107

[50] Mark Monmonier. 2018. *How to lie with maps*. University of Chicago Press.

[51] Patrick AP Moran. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)* 10, 2 (1948), 243–251.

[52] Phillip Muehrcke. 1978. Functional map use. *Journal of Geography* 77, 7 (Dec. 1978), 254–262. https://www.proquest.com/docview/1290933193/citation/D967AF4B20D247BCPQ/1 Num Pages: 9 Place: Macomb, Ill., etc., United States.

[53] Judy M. Olson. 1976. A Coordinated Approach to Map Communication Improvement. *The American Cartographer* 3, 2 (Jan. 1976), 151–160. doi:10.1559/152304076784080177

[54] OpenAI. 2024. GPT-4o mini: advancing cost-efficient intelligence. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/

[55] Alice Park, Charlie Smart, Rumsey Taylor, and Miles Watkins. 2021. An Extremely Detailed Map of the 2020 Election. *The New York Times* (Feb. 2021). https://www.nytimes.com/interactive/2021/upshot/2020-election-map.html

[56] Jochen Schiewe. 2019. Empirical Studies on the Visual Perception of Spatial Patterns in Choropleth Maps. *KN - Journal of Cartography and Geographic Information* 69, 3 (Sept. 2019), 217–228. doi:10.1007/s42489-019-00026-y

[57] Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. 2016. Eviza: A Natural Language Interface for Visual Analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, Tokyo Japan, 365–377. doi:10.1145/2984511.2984588

[58] Ather Sharif, Olivia H. Wang, Alida T. Muongchan, Katharina Reinecke, and Jacob O. Wobbrock. 2022. VoxLens: Making Online Data Visualizations Accessible with an Interactive JavaScript Plug-In. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–19. doi:10.1145/3491102.3517431

[59] Ather Sharif, Andrew M. Zhang, Katharina Reinecke, and Jacob O. Wobbrock. 2023. Understanding and Improving Drilled-Down Information Extraction from Online Data Visualizations for Screen-Reader Users. In *20th International Web*

for All Conference. ACM, Austin TX USA, 18–31. doi:10.1145/3587281.3587284

[60] Ather Sharif, Andrew Mingwei Zhang, Anna Shih, Jacob O. Wobbrock, and Katharina Reinecke. 2022. Understanding and Improving Information Extraction From Online Geospatial Data Visualizations for Screen-Reader Users. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility.* ACM, Athens Greece, 1–5. doi:10.1145/3517428.3550363

[61] Terry A. Slocum, Robert B. McMaster, Fritz C. Kessler, and Hugh H. Howard. 2022. *Thematic cartography and geovisualization.* CRC Press.

[62] Sharon Spall. 1998. Peer Debriefing in Qualitative Research: Emerging Operational Models. *Qualitative Inquiry* 4, 2 (June 1998), 280–292.

[63] Edward Steinfeld and Jordana Maisel. 2012. *Universal Design: Creating Inclusive Environments.* John Wiley & Sons.

[64] Kavita Thomas and Yaji Sripada. 2008. What's In a Message? Interpreting Georeferenced Data for the Visually-impaired. In *Proceedings of the Fifth International Natural Language Generation Conference,* Michael White, Crystal Nakatsu, and David McDonald (Eds.). Association for Computational Linguistics, Salt Fork, Ohio, USA, 113–120. https://aclanthology.org/W08-1115

[65] Kavita E. Thomas and Somayajulu Sripada. 2007. Atlas.txt: linking geo-referenced data to text for NLG. In *Proceedings of the Eleventh European Workshop on Natural Language Generation - ENLG '07.* Association for Computational Linguistics,

Germany, 163. doi:10.3115/1610163.1610193

[66] The New York Times. 2020. Coronavirus in the U.S.: Latest Map and Case Count. *The New York Times* (March 2020). https://www.nytimes.com/interactive/2023/us/covid-cases.html

[67] W3C. 2024. Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/

[68] David W. S. Wong. 2021. Issues in the Current Practices of Spatial Cluster Detection and Exploring Alternative Methods. *International Journal of Environmental Research and Public Health* 18, 18 (Jan. 2021), 9848. doi:10.3390/ijerph18189848 Number: 18.

[69] Anthony GO Yeh. 1999. Urban planning and GIS. *Geographical information systems* 2, 877-888 (1999), 1.

[70] Jonathan Zong, Crystal Lee, Alan Lundgard, JiWoong Jang, Daniel Hajas, and Arvind Satyanarayan. 2022. Rich Screen Reader Experiences for Accessible Data Visualization. *Computer Graphics Forum* 41, 3 (June 2022), 15–27. doi:10.1111/cgf.14519

[71] Arzu Çöltekin, Halldór Janetzko, Sara Fabrikant, University of Zurich, and University of Zurich. 2018. Geovisualization. *Geographic Information Science & Technology Body of Knowledge* 2018, Q2 (April 2018). https://gistbok-topics.ucgis.org/CV-05-035