Cartography and Geographic Information Science

ISSN: 1523-0406 (Print) 1545-0465 (Online) Journal homepage: www.tandfonline.com/journals/tcag20

MapColorAI: designing contextually relevant choropleth map color schemes using a large language model

Nai Yang, Yijie Wang, Zhiwei Wei & Fan Wu

To cite this article: Nai Yang, Yijie Wang, Zhiwei Wei & Fan Wu (23 Jul 2025): MapColorAI: designing contextually relevant choropleth map color schemes using a large language model, Cartography and Geographic Information Science, DOI: 10.1080/15230406.2025.2531055

To link to this article: https://doi.org/10.1080/15230406.2025.2531055

CARTOGRAPHY AND GEOGRAPHIC INFORMATION SCIENCE
https://doi.org/10.1080/15230406.2025.2531055

MapColorAI: designing contextually relevant choropleth map color schemes using a large language model

Nai Yang[a], Yijie Wang[a], Zhiwei Wei[b,c] and Fan Wu[a]

[a]School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei, China; [b]School of Geographic Sciences, Hunan Normal University, Changsha, Hunan, China; [c]Hunan Key Laboratory of Geospatial Big Data Mining and Application, Hunan Normal University, Changsha, Hunan, China

ABSTRACT

Choropleth maps are fundamental tools for geographic data analysis, primarily relying on color to convey information. Consequently, the design of their color schemes is important in choropleth map production. However, the traditional coloring methods offered by GIS tools such as ArcGIS and QGIS are not user-friendly enough for nonprofessionals. These tools provide numerous color schemes, making selection difficult, and cannot also easily fulfill personalized coloring needs, such as requests for "summer-like" map colors. To address these shortcomings, we develop a novel system that leverages a large language model and map color design principles to generate contextually relevant and user-aligned choropleth map color schemes. The system follows a three-stage process: Data processing, which provides an overview and classification of the data; Color Concept Design, where color theme and mode are conceptualized based on data characteristics and user intentions; and Color Scheme Design, where specific colors are assigned to classes. Our system incorporates an interactive interface for choropleth map color design and allows users to customize color choices flexibly. Through user studies and evaluations, the system demonstrates acceptable usability, accuracy, and flexibility, with users highlighting its efficiency and ease of use.

ARTICLE HISTORY
Received 25 January 2025
Accepted 2 July 2025

KEYWORDS
Choropleth map; color design; large language model; data classification; cartography

1. Introduction

Color is a fundamental perceptual element in maps that enhances their effectiveness for communication, information organization, and spatial understanding (Brewer, 1994; Shive & Francis, 2013). Among various map types, choropleth maps are particularly popular for thematic data visualization, using color gradients to represent data values across geographic regions and effectively highlight spatial patterns (Lei et al., 2024). The success of choropleth maps heavily relies on selecting color schemes that align with both the data and the map's purpose (Brewer et al., 1997; Sun et al., 2014). Therefore, developing tools that help users design tailored choropleth map colors while following cartographic conventions has long been a priority in map design.

To support effective map color design, numerous guidelines have been developed, primarily based on the visual variables first proposed by Bertin (1974) and later refined by scholars like MacEachren (2004), Shive and Francis (2013), and Zhou and Hansen (2016). These studies emphasize aligning color choices with data measurement levels: using hue for qualitative data, and value or saturation for ordinal or numerical data (Armstrong et al., 2003; He et al., 2016). Furthermore, the color selection should follow conventional associations (e.g. blue for water, green for vegetation) and adhere to official standards (Wei et al., 2018; Xue et al., 2024), while maintaining sufficient visual contrast for graphical symbols (Brychtova & Coltekin, 2015; Chesneau, 2011). These principles have become essential guidelines for map color design.

Despite these guidelines, users often struggle to create their preferred color schemes. To address this challenge, researchers like Brewer et al. (2003) introduced color templates that offer practical solutions by summarizing widely accepted color combinations. These templates include qualitative schemes with distinct hues for categories and sequential schemes varying in lightness to differentiate data classes. Harrower and Brewer (2003) further advanced this approach by developing ColorBrewer.org, an online tool allowing designers to explore and compare color schemes for thematic maps. Later, Christophe (2011) introduced a knowledge base for map color specifications, integrating insights from visual perception, cognitive science, graphic semiotics, cartography, and art to provide

CONTACT Zhiwei Wei 2011301130108@whu.edu.cn
Supplemental data for this article can be accessed online at https://doi.org/10.1080/15230406.2025.2531055

tailored guidelines for color design. Many modern GIS tools, including ArcGIS and QGIS, have also incorporated these templates to assist users in the color design process (Kristian, 2020). Moreover, the advent of deep learning has led to data-driven techniques, such as transferring color schemes from images to maps (Wu, Sun, et al., 2022) and evaluating the aesthetic quality of map colors (Wu et al., 2024).

However, these automated approaches often fall short of being user-friendly, particularly for nonprofessionals. On the one hand, users are frequently confronted with a vast array of color schemes, making it challenging to select the one that best matches the map's thematic context. On the other hand, these systems struggle to address more nuanced and personalized color requests. In many cases, users may only have a vague idea, rather than a clearly defined color palette, such as a user seeking a "summer-like" color scheme for a map. Conventional tools like ArcGIS, and ColorBrewer, while robust, are not designed to handle such open-ended or abstract input. Specifically, traditional tools and methods are unable to dynamically adapt color schemes based on evolving user preferences or contextual cues during the design process. This lack of contextual awareness limits their ability to provide truly personalized and flexible solutions. This highlights the need for more flexible and interpretable solutions that bridge the gap between automated color generation and user-driven customization. Given that user intentions are frequently conveyed through natural language, the advent of the Large Language Model (LLM) opens up new possibilities for innovative color design. Leveraging their powerful generative capabilities and advanced language understanding (Vázquez, 2024), LLM has the potential to greatly enhance the map color design process by delivering color schemes that are both contextually relevant and closely aligned with user preferences. Thus, we can use the LLM to actively interpret and respond to contextual information throughout the iterative design process in our approach, ensuring that each color scheme suggestion is tailored to the evolving dialogue with the user. However, two challenges still need to be solved to apply LLM. First, LLM is also inherently an end-to-end approach, producing homogeneous content for identical inputs, which may not be conducive to controllability issues in the design process (Shi et al., 2023). Second, being a generic model, LLM lacks domain-specific knowledge, such as color scheme types, and color psychology, all of which are crucial aspects of choropleth map color design.

To address these challenges, we propose the MapColorAI in this approach, (1) Granting users flexibility in expressing their design intent throughout the process (address challenge I), and (2) Integrating diverse domain-specific knowledge in the design process for choropleth map colors (address challenge II). To achieve this goal, we conducted an in-depth investigation and summary of domain-specific knowledge and design workflows for choropleth map color schemes (Section 2). Next, we decompose the choropleth map color design process into a series of sequential steps, applying domain knowledge and LLM at each stage. Additionally, we incorporate multi-level interactive features that allow user intervention at each step to enhance system controllability (Section 3). Finally, a user study is implemented to demonstrate the usability of the proposed system (Section 4) and discussions are provided in (Section 5).

2. Knowledge of choropleth map color design

In this section, we outline the key components required for designing color schemes for choropleth maps, including classification methods, color theory, and design practices. These elements serve as essential guidelines for developing our map color design system using an LLM.

2.1. Classification method

(1) Number of classes

Before applying a data classification method, cartographers need to determine the number of classes while making a choropleth map. The human cognitive system, which has limitations on processing distinct chunks of information, suggests that the optimal number of classes should be around 7 ± 2 (Miller, 1956). However, this number can be adjusted based on the complexity of the data. For simple datasets, the number of classes is typically limited to between 3 and 7 (Slocum et al., 2022; Wei et al., 2018), while for more complex datasets, the number can be increased. For example, the ColorBrewer tool supports up to 11 colors for diverging schemes and 9 colors for sequential schemes (Brewer et al., 2003). Based on these guidelines, our framework recommends using between 3 and 11 classes, which is consistent with the approach adopted by Lei et al. (2024). Additionally, the system allows users to adjust the number of classes to suit specific needs or preferences.

(2) Classification method

Numerous classification methods for choropleth maps have been developed for various purposes. We instantiate 6 widely used methods according to Lei et al. (2024) in the proposed system based on the cartographic literature, as follows (Table 1).

Table 1. Basic information of six classification methods.

Method Name	Description	Advantages	Disadvantages
Equal-Intervals	Divides data into equal-width intervals Evans (1977).	Simple and easy to understand.	Ignores data distribution; may result in sparse or crowded intervals.
Quantiles	Divides data into equal parts based on distribution Evans (1977).	Reflects data distribution; equal data points per class.	Large numeric differences between classes; may obscure trends.
Jenks-Caspall	Determines natural breaks using optimization Jenks and Caspall (1971).	Minimizes variance within classes; maximizes variance between classes.	Computationally intensive for large datasets.
Fisher-Jenks	Improved Jenks-Caspall with reduced computation time Jenks and Caspall (1971).	Efficient; minimizes internal variance; maximizes external variance.	Still computationally intensive, but less so than Jenks-Caspall.
Max-*p*	Uses cluster analysis for high internal consistency Duque et al. (2012).	High internal consistency; well-defined boundaries.	Computationally intensive; not suitable for all datasets.
Pretty-Breaks	Selects "neat" numbers as classification boundaries Lei et al. (2024).	Easy to understand and remember; user-friendly.	May not reflect data distribution; less precise.

In Slocum et al.'s textbook on Cartography and Geographic Visualization, the authors provide a detailed introduction to the principles, applications, advantages, and disadvantages of the above methods (Slocum et al., 2022).

(3) Evaluate the classification quality

Given the variety of data classification methods available, it is important to assess the results of classification to help users select the most appropriate method. To achieve this, we utilize the Goodness of Variance Fit (*GVF*), a numerical indicator that is related to perceptual accuracy and measures the quality of classification results in choropleth map design (Brewer et al., 1997; Lei et al., 2024). The *GVF* is defined as follows:

$$GVF = 100 - \frac{SSW}{SST} \times 100 \qquad (1)$$

Where SST(Sum of Squares Total) represents the sum of squared deviations of individual data values from the overall mean, and SSW(Sum of Squares Within) is the sum of squared deviations of data values within each class from the class mean, with the sum taken across all classes. A *GVF* value of 9.5 or higher is considered indicative of a "satisfactorily accurate classification" (Declerq, 1995). While other indices – such as Moran's *I* (Lei et al., 2024) – can also be employed to evaluate classification quality, they often involve more complex statistical interpretation. Given that our target users are non-experts, incorporating such metrics may increase cognitive load and hinder usability. Therefore, our system adopts the simpler and more interpretable *GVF* score to support intuitive decision-making in the classification process.

2.2. Color theory

2.2.1. Color system

Different applications may require different color systems for maps. For example, the RGB color system is typically used for screen displays, while the CMYK color system is preferred for print publications (He et al., 2016). Additionally, for human color perception, commercial software such as ArcPro and QGIS integrate color systems like HSV, HSB, HSL, and CIELab (Wei et al., 2018). Since our system is web-based and designed for screen display, we have also adopted the RGB color system. Notably, the well-known ColorBrewer tool also provides color schemes with this color system.

2.2.2. Color concept

(1) Color themes

In many design fields, such as architecture and interior design, color themes play a crucial role in defining the overall style (Hou et al., 2024). For instance, an "elegant" color theme often utilizes muted or desaturated tones, which have been shown in psychological studies to evoke sedate and calming emotions (Levy, 1984). Similarly, color themes are also integral to map design. He et al. (2016) have outlined commonly used color themes for maps, including "light," "moderate," "strong contrast," "elegant," and "classical." Building on this, we apply these themes to choropleth map color design, with the potential for further expansion in future applications.

(2) Color moods

Psychological effects and emotional responses may be evoked by different colors. A thoughtful application of color psychology can enhance the expressive impact of maps, guiding user attention and emotions. From a psychological perspective, the influence of color on choropleth map design is primarily reflected in three aspects also known as color moods.

- ***Temperature perception.*** Colors evoke a sense of warmth or coolness, which can influence how map areas are interpreted (Cuff, 1973). Warm colors, such as red, orange, and yellow, typically evoke feelings of warmth, energy, and forward movement, making them ideal for representing regions with high temperatures, high densities, or greater significance. In contrast, cool colors – such as blue, green, and purple – convey calmness, retreat, and stability, making them more appropriate for depicting areas with lower temperatures, low densities, or background regions.

- ***Spatial perception.*** Colors can influence the observer's subjective judgment of spatial dimensions, including area size and distance (Holtzschue, 2016). Colors with higher brightness and saturation tend to create a sense of forward movement, making regions appear larger and more prominent. In contrast, colors with lower brightness and saturation tend to recede, giving the impression that the areas are smaller and farther away. This spatial perception effect can be strategically employed in choropleth map design to establish a visual hierarchy, highlight key areas, and de-emphasize less significant regions, thereby effectively guiding the viewer's attention to the most relevant parts of the map.

- ***Weight perception.*** In choropleth maps, darker colors are often perceived as heavier and more stable, while lighter colors convey a sense of lightness and freshness. This weight perception is crucial for illustrating hierarchical differences in data, helping to establish visual prominence and subordination (Alexander & Shansky, 1976). By using this effect, cartographers can enhance the clarity of data classification, ensuring that higher-priority regions stand out, while less significant data is visually subdued.

2.2.3. Quantitative representation of color

The choropleth map uses color to represent the data-associated areas. A foundational principle in designing color schemes for choropleth maps is that people tend to associate higher values with darker colors. It follows that if the choropleth needs to show logical ordering of value then rainbow colors must be avoided (Golebiowska & Coltekin, 2022). To better reflect the data, two kinds of data or kind of color schemes have been provided by ColorBrewer and have been designed principles in choropleth design (Brewer et al., 2003). For sequential data, smooth transitions from light to dark shades effectively represent low-to-high values (sequential scheme). In diverging data, balanced midpoints and contrasting extremes ensure clarity (diverging scheme). A well-chosen color scheme highlights subtle variations while maintaining overall visual coherence. In the proposed system, we also adopted the two-color scheme (sequential vs diverging) in our system.

2.3. Design practice

Based on previous research (Lei et al., 2024; Slocum et al., 2022) and interviews with two experts, the workflow for color design in choropleth maps can be divided into three key steps: data processing, color concept design, and color scheme design. Our MapColorAI system incorporates tailored interactions at each of these stages, enabling users to flexibly adjust relevant parameters throughout the design process.

(1) **Data processing**. In designing a choropleth map color scheme, the mapmaker first gains an overview of the data, including its content and range. The next step involves selecting an appropriate classification method to group the data into discrete intervals or classes, as well as determining the color scheme type (sequential or diverging) based on the data characteristics. The choice of classification methods – such as equal intervals, quantiles, or natural breaks – can significantly influence the map's visual representation and interpretation (Slocum et al., 2022). This decision depends on the data distribution and the intended focus, whether it's to highlight trends, outliers, or variations. For instance, when visualizing population density, using natural breaks may be preferable to emphasize meaningful differences between regions, ensuring the map highlights relevant variations effectively (Jenks & Caspall, 1971). To ensure the classification aligns with user needs, our system allows users to freely select the classification method and the number of classes based on their specific requirements. Users can evaluate the classification results through the histogram and *GVF* score, enabling them to make informed decisions that best suit their map objectives.

(2) **Color concept design**. In the second step, designers conceptualize the color concepts including color themes and color moods based on the data and specific user requirements. First, designers associate specific color moods with user requirements, keeping in mind color psychology. Second, designers establish the overall color theme, such as "light" or "elegant," providing a cohesive style for the map that aligns with the intended message and user preferences. For instance, a vague user request, such as "a lively atmosphere for the map," may be interpreted as favoring a warm color mood and a "strong contrast" color theme, which are known to evoke energy and vibrancy. Our system allows users to freely adjust style parameters such as color themes and color moods according to their needs. The LLM responds to these adjustments in real-time, generating new color schemes that reflect the user's preferences and requirements.

(3) **Color scheme design**. The final step involves designing the specific color scheme to be applied to the map. At this stage, the designer selects appropriate colors for each class defined in the classification step. Based on the concept developed in Step 2, the color scheme is refined to reflect the color themes and color moods. For example, if a warm color mood was

selected in the concept design phase, colors such as red or orange may be incorporated to visually represent higher values. This stage ensures that the map's color scheme aligns with both the data content and the overall design concept, enhancing the map's clarity and interpretability. To provide users with greater flexibility, our system supports fine-tuning of individual colors within the scheme. It provides a comparative analysis between the LLM-generated scheme and the most similar ColorBrewer scheme, facilitating optimal selection. Furthermore, users can interact with the system through natural language input, specifying their requirements, and the LLM will generate a new color scheme tailored to these inputs.

2.4. LLM and map color design integration

LLM have emerged as powerful tools capable of generating coherent and contextually relevant text, driven by advancements in artificial intelligence and natural language processing (Hou et al., 2024). An LLM is characterized by its extensive parameter count, enabling it to capture and generate diverse linguistic patterns from vast amounts of textual data. Recent developments, such as DeepSeek (Liu et al., 2024), Qwen (Bai et al., 2023), ChatGPT (Achiam et al., 2023), and other generative models, have demonstrated their potential to assist in various creative and analytical tasks, including those related to cartography and map design (Kang et al., 2024).

On this basis, we explored the application of LLM in the field of map color design. Traditional map color scheme design often relies on manual selection or rule-based systems, which require professional experience and may lack flexibility. In contrast, LLM can provide appropriate content based on user input. Leveraging this capability, we designed prompt templates through prompt engineering to guide the model in generating the expected output (Liu et al., 2023). Inspired by PromptChainer (Wu et al., 2022) and AI Chains (Wu et al., 2022), we have deconstructed the map color design process into steps consistent with common cartographic workflows. Additionally, we added user interaction at each stage of the design process to ensure that the user can generate their preferred colors. By integrating design principles and domain-specific knowledge into the prompt templates at different stages, we effectively constrained the model's behavior, ensuring its output aligns with cartographic standards and principles.

3. Methodology

3.1. System overview

This study aims to investigate the use of LLM-based assistance in designing color schemes for choropleth maps, focusing on transforming cartographers' vague mapping intentions into color schemes while adhering to the fundamental principles of map color design and providing language interaction. Based on the design practices outlined in Section 2.3, we developed the "MapColorAI" system, which supports choropleth map color scheme creation through three stages: data processing (Section 3.2), color concept design (Section 3.3), and color scheme design (Section 3.4), as illustrated in Figure 1. However, unlike traditional workflows, our approach incorporates LLM functionalities into each stage to streamline and enhance the design process. Each stage integrates LLM capabilities with user interactions (Section 3.5), enabling the creation of personalized color schemes while adhering to the specific design principles of choropleth maps.

3.2. Data processing

In this stage, two sub-processes are implemented, as discussed in Section 2.3: data analysis and data classification. The data analysis process identifies potential data errors, provides an overview of the data, and suggests the appropriate color scheme type (sequential vs. diverging) based on the data's characteristics. This is achieved through the use of LLM. The data classification process applies commonly used classification methods to categorize the data and recommends the highest-rated method based on an evaluation of the classification result's quality.

3.2.1. Data analysis

Recent research has demonstrated that LLM is highly effective in understanding statistical data (Tian et al., 2025). Thus, we directly specify the tasks in the prompt for the LLM, without the need for special design to implement the data analysis. The primary objective of this process is to provide the user with an overview of the data and to extract key information that will inform subsequent color design.

The tasks include: (1) Check possible data errors in the upload data, such as missing data or abnormal values. (2) Provide as detailed a description as possible based on the uploaded data, including topics, range, acquisition time, etc., as much as possible. This information such as data topic may be necessary for later color design, for example, "blue" tone may be preferred for a data topic related to "water." The general data

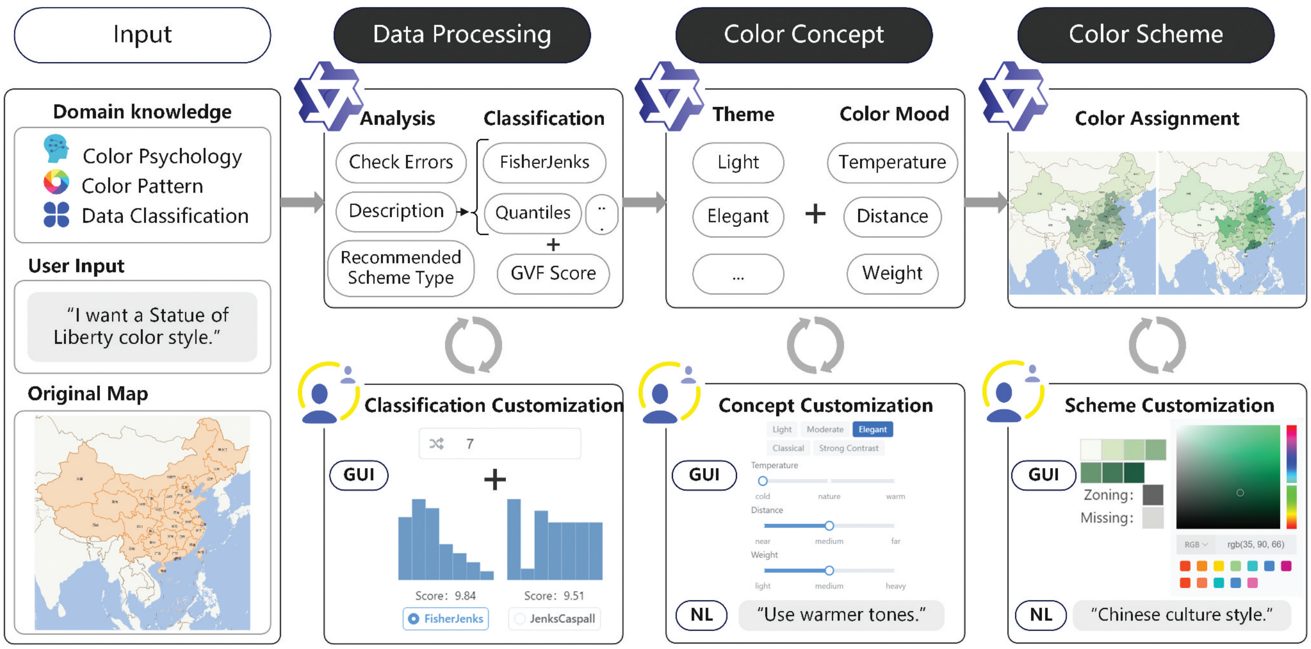

Figure 1. MapColorAI encompasses three stages: (1) data processing: data comprehension, data classification, and determining color scheme type; (2) color concept design: customization of color themes and color moods; (3) color scheme design: generation of specific color schemes and real-time preview.

description is also an input for later color design. (3) Suggest color scheme type (sequential vs diverging) based on data characteristics. As the LLM is not specifically designed for choropleth map creation, we integrate domain knowledge to guide this task. Domain knowledge is set according to He et al. (2016) and Peterson (2020) as follows:

> Sequential scheme is ideal for visualizing data with a clear order or magnitude, such as population density or income levels. Diverging scheme is ideal for visualizing data that deviates in two opposite directions from a meaningful midpoint, such as temperature anomalies or percentage change. You can determine the scheme type based on—Does the ranking have a "center" or "middle"? If it does, a diverging scheme is appropriate; if not, a sequential scheme is preferred.

Here, we take the "2023 GDP of Chinese Provinces" (excerpt) as an example. The input data of our system is organized in GeoJSON format, with the names of each region and their corresponding GDP values within the same array item, as shown in the left part of Figure 2. Once the data is uploaded, our system will automatically determine the applicable region based on the uploaded data and the data analysis module then outputs the following response for the above three tasks, as shown in the right part of Figure 2. For Task 1: LLM found no errors in the data; for Task 2: The LLM summarized information such as data coverage, maximum and minimum values, and significant differences between regions; for Task 3: It recommends using the "Sequential" color scheme type. As demonstrated in this example, the module provides a comprehensive overview of the data, allowing users to obtain key insights without needing to manually inspect the JSON file.

3.2.2. Data classification

After obtaining an overview of the data, the user proceeds to the data classification step. This is a semi-automatic process in which the LLM is not involved. The user begins by selecting the number of classes, which is restricted to a range between 3 and 11, as determined by the analysis in Section 2.1 Once the classification number is chosen, the system applies six commonly used classification methods (Equal-Intervals, Quantiles, Jenks-Caspall, Fisher-Jenks, Max-*p*, and Pretty-Breaks, as described in Section 2.1) to categorize the data. The *GVF*, defined by Equation 1, is then used to evaluate the quality of the classification (Lei et al., 2024). This score helps the user identify their preferred classification method, with the method achieving the highest score being automatically selected. For instance, the Fisher-Jenks method, as shown in Figure 3, may be selected based on its superior score. To further assist users, the classification results are visualized using histograms, providing a more intuitive understanding of the classification and allowing users to select the method that best meets their specific mapping needs.

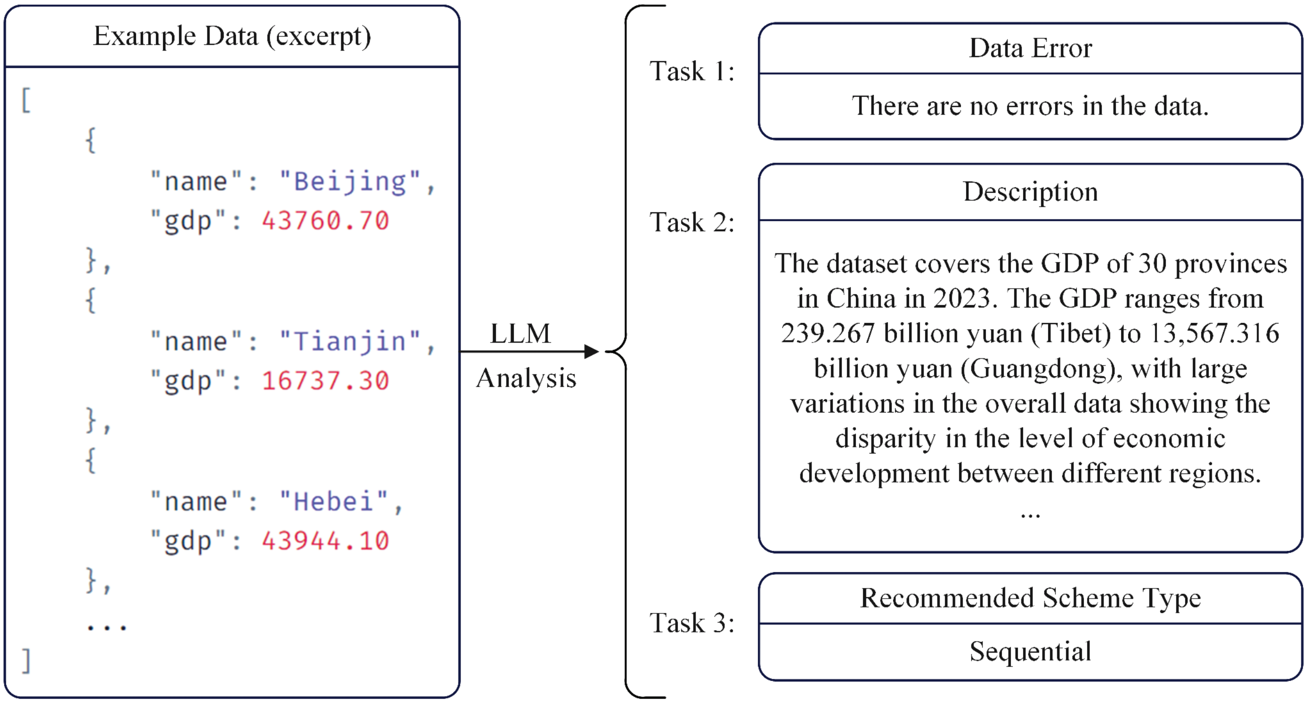

Figure 2. The left side illustrates an example of input data, while the right side presents the results of the LLM completing three data analysis tasks.

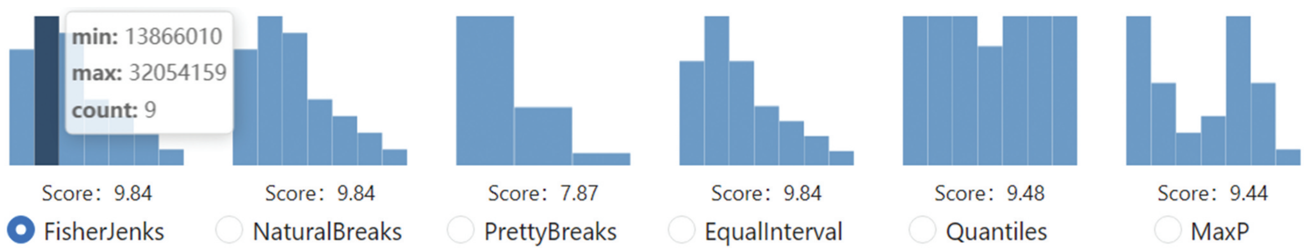

Figure 3. Visualization of the classification results. The histogram illustrates the distribution of data points across various outcomes, with the horizontal axis representing progressively increasing classification intervals and the vertical axis indicating the number of data points in each category. The "score" below the histogram denotes the GVF score. Once a user hovers over an individual bar in the histogram, additional information is displayed, including the left and right boundaries of the class and the number of data points within that class.

3.3. Color concept design

3.3.1. Color concept generation

This step involves translating the user's vague mapping intentions into color concepts, including specific color themes and corresponding color moods. This is achieved by using the LLM with a structured prompt template to guide the generation translation process. The prompt follows the common practice which consists of five parts (Caelen & Blete, 2024), as illustrated in Figure 4.

- Data Input: As shown in Figure 4a), the input for this stage includes the user's vague intent and the data description information generated in the previous stage, enabling the LLM to output content that meets the user's requirements based on the data characteristics.
- Profile Setting: It has been empirically observed that assigning a specific role to an LLM significantly enhances its performance (Caelen & Blete, 2024). As we aim to design a color scheme for the choropleth map, we assign the LLM with a role as map designer, briefly outlining its character's background and skills, to ensure the activation of the model's role-playing capabilities.
- Domain Knowledge: The color concept design, driven by user intentions, is typically carried out by professionals with domain-specific expertise. To ensure that the output aligns with design requirements, we summarize the commonly considered

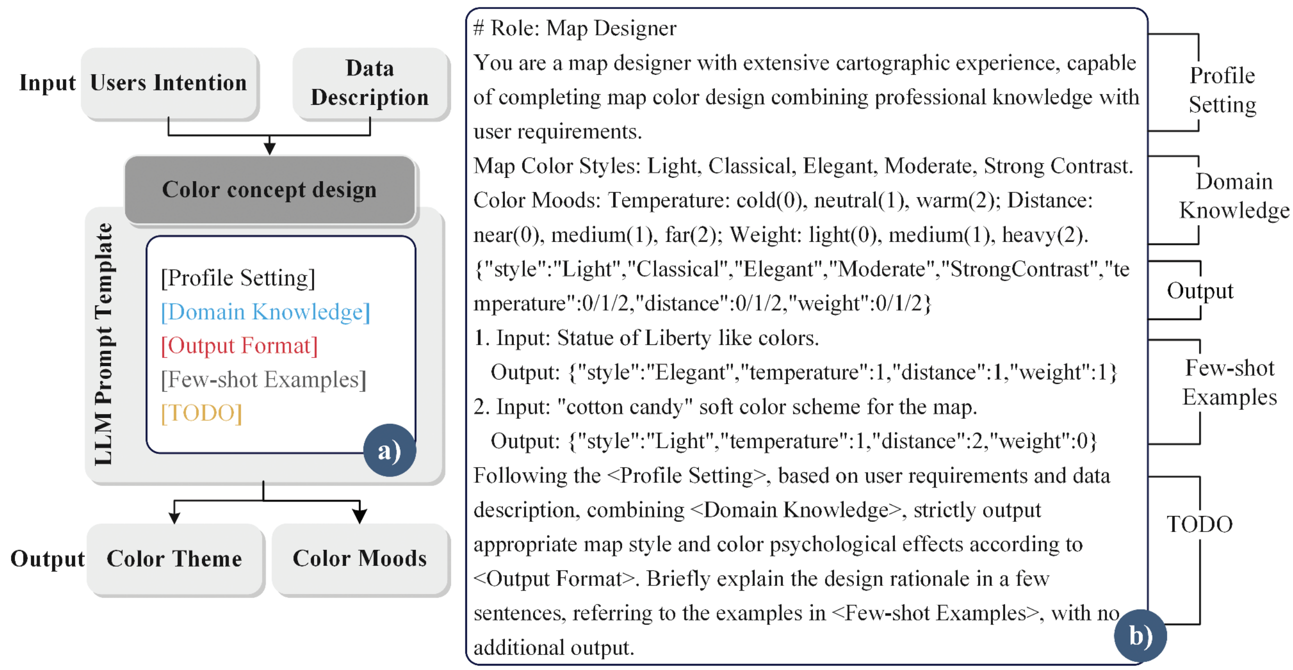

Figure 4. Color concept generation. a) The prompt template; b) a specific prompt that integrates data descriptions to transform user vague input into color theme and color moods, taking into account domain knowledge of map color design.

color themes and color moods in choropleth map design, as analyzed in Section 2.2.2 The color themes include “strong contrast,” “light,” “moderate,” “elegant,” and “classical,” which can also be extended in practice. The color moods include “Temperature Perception,” “Spatial Perception,” and “Weight Perception.” Each attribute is quantified into three levels. To ensure accuracy and consistency, we align these levels’ natural language descriptions with specific numerical values, such as “cold (0), neutral (1), warm (2)” for temperature, “near (0), medium (1), far (2)” for distance, and “light (0), medium (1), heavy (2)” for weight.

- Output Format: It is a constraint on the output format of LLM, which ensures a standardized output to facilitate subsequent parsing and utilization.
- Few-shot Example: Research has found that LLM is capable of generating higher-quality results when provided with a small number of sample examples (Liu et al., 2023). In map color design, the users’ intentions and their related themes or color moods are usually various and vague, which are difficult to summarize into high-level design rules, such as mapping from the “Statue of Liberty color style” to an “elegant” map style. Therefore, these mappings cannot be included in the domain knowledge section. To address this issue, we supplement the domain knowledge by providing users with fuzzy inputs paired with corresponding design concepts, as shown in Figure 4b). An example of this includes user intentions, map themes, and three key attributes that reflect the desired color moods.
- TODO: In this section, a comprehensive framework for the behavior of the LLM has been established, ensuring that it follows the outlined procedures to accomplish tasks based on the aforementioned content. Additionally, the framework requires the LLM to provide a design rationale for its outputs as well as propose three alternative modifications to enhance user customization options. First, the rationale serves to explain the reasoning behind the model’s design choices, ensuring that the generated solutions are not only contextually relevant but also align with established principles of map design, user intentions, and color theory. This transparency helps users understand the underlying logic of the LLM’s decisions, fostering trust and improving the overall user experience. Second, the three suggested modifications act as actionable recommendations for users during the later stages of design customization. These suggestions empower users to articulate more refined and professional design intentions, such as “increase the brightness,” rather than vague or overly subjective requests like “make it feel like summer.”

As shown in Figure 4b), an example-generated prompt is illustrated. Here the user wants a “Statue of Liberty like" map color design, and the LLM uses the prompt to transfer it into design concepts, including themes such as "elegant," and color mood attributes including "neutral tones," "medium distance," and "medium weight."

3.3.2. Color concept design customization

The system supports interactive modifications for color concept design through both graphical interfaces and natural language. On the one hand, map themes are presented in the form of tags, while color mood attributes (temperature, distance, and weight) are displayed via sliders. Users can select other theme tags or modify the intensity of color moods through the interface directly.

Additionally, we also provide a natural language-based interaction method for users with more ambiguous needs, enabling them to refine input information or make adjustments. For example, the "classic" color theme can be further interpreted as using "light weight" or "heavy weight" color schemes and will result in two distinct choropleth maps. Users can clarify their preference by specifying "classic soft tones," indicating a desire for a more gentle color combination that conveys a sense of stability and tradition, thus modifying the color mode into "light weight." Furthermore, users can also select one of the three suggested language intentions (shown in Figure 6a) if they feel that one aligns more closely with their true intentions.

3.4. Color scheme design

3.4.1. Color scheme generation

During this phase, the classified results and color concepts derived from the previous two steps are transformed into specific color schemes. In traditional map color design, this stage is typically domain-specific, where the map maker, drawing on experience or referencing existing color schemes or websites, designs the final color scheme (He et al., 2016, Xi et al., 2023). To address this, two sub-processes are implemented in this stage. First, a color scheme is generated using the LLM, leveraging the data and color concepts from the first two stages (**color scheme generation via LLM**). Second, to ensure the output color scheme aligns with professional standards and optimally utilizes existing color scheme resources, we provide users with matching color schemes from a database of well-established options (**color scheme matching**).

(1) Color scheme generation via LLM

This process is implemented using the LLM through a specialized prompt template, similar to the one used in the color concept design stage. As a result, we provide only a concise description of the prompt in this stage, omitting detailed explanations. The prompt includes domain knowledge, output format specifications, few-shot examples, and behavioral settings, as illustrated in Figure 5a). Since the dialogue at this stage builds upon

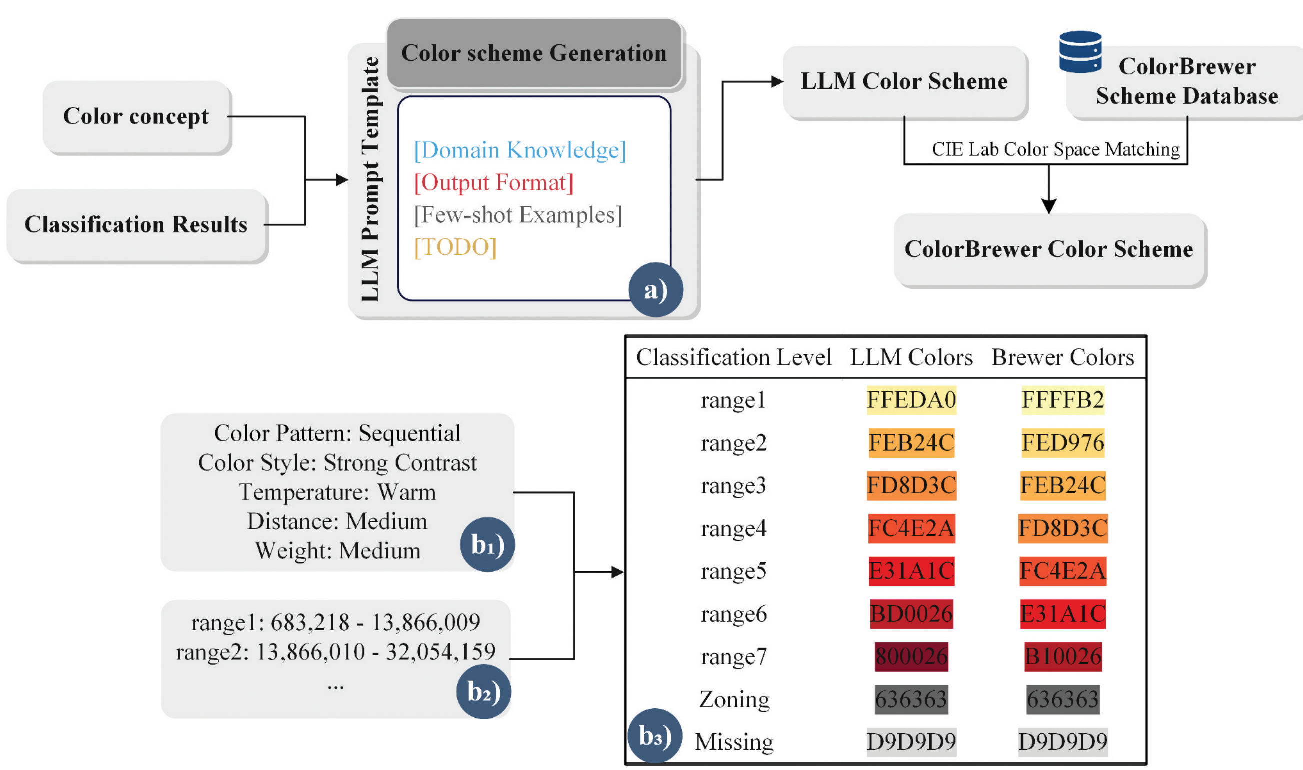

Figure 5. Color scheme generation. a) The prompt template used for color scheme generation; b_1) and b_2) an example of input data; b_3) the LLM-generated color scheme and its most similar ColorBrewer color scheme.

the content from Section 3.2 (Color Concept Design), role settings are also incorporated within the prompt. The generation process capitalizes on the LLM's ability to map semantically similar words, a feature shaped by both the model's pre-training and the specific context of the prompt. This enables the LLM to effectively align color concepts and data topics with established knowledge in the field of map color design, ensuring both semantic and thematic consistency. For instance, as shown in Figure 5b$_1$), based on the "warm" color mood in the color concept, the LLM proposed a color scheme dominated by red and yellow tones, avoiding cooler shades like blue or green. This selection aligns with the "Sequential" color scheme type, transitioning from light yellow to deep red to represent different data levels through varying shades of a similar hue. Additionally, the "Strong contrast" color theme is evident in the results, which accentuates the differences between data levels, making them more visually distinct.

(2) Color scheme matching

The ColorBrewer color schemes are widely recognized for their effectiveness in choropleth map design (Lei et al., 2024). To enhance the professionalism of the output color schemes, we incorporate ColorBrewer's color schemes as a reference database. A total of 207 ColorBrewer color schemes, downloaded from the ColorBrewer website, are used as the base for comparison. The system calculates the difference between the output color scheme generated by the LLM and each scheme from the ColorBrewer database, with the most similar color scheme being recommended to the user, as shown in Figure 5. To compare the color schemes, we first filter out those with the same number of colors as the output scheme, ensuring a more accurate match. As both the LLM-generated and ColorBrewer schemes are represented in RGB format (which is commonly used for screen displays but not aligned with human color perception), we convert both sets of colors from RGB to the CIELab color space – a color system designed to reflect human color perception (Harrower & Brewer, 2003). The comparison is then conducted by calculating the color differences between the two schemes in the CIELab color space. The difference (ΔE) between two colors is quantified as the Euclidean distance between their respective points in the CIELab color space, and defined as follows.

$$\Delta E = \sqrt{(L1 - L2)^2 + (a1 - a2)^2 + (b1 - b2)^2} \quad (2)$$

Where (L_1, a_1, b_1) and(L_2, a_2, b_2) are the two colors in the CIELab color system.

3.4.2. Scheme customization

After generating a color scheme, we provide an interactive customization interface for users. This graphical interface enables users to adjust each color individually, with real-time updates that allow for immediate visual feedback. For example, if a user finds it difficult to distinguish between two colors, they can fine-tune attributes such as brightness or saturation to create clearer differentiation. To further simplify the customization process, especially for nonprofessional designers, we also incorporate natural language interaction, which is similar to the color concept design stage. Users can adjust the color scheme with simple commands. For instance, if a user feels that the overall color scheme is too monotonous or lacks vibrancy, they can simply request, "Make these colors more vivid," and the system will interpret this intent by enhancing the color saturation or introducing contrasting hues to increase visual impact.

3.5. Implement details and interaction design

3.5.1. Implement details

Our system is designed to support interactive and intelligent choropleth map color design, utilizing a frontend-backend architecture integrated with LLM. It follows a modular design, breaking the map design process into distinct, sequential stages where the output of one stage feeds into the next. This modularity enhances the system's flexibility and maintainability and makes each stage of the map design process controllable.

Web Interface:

- The user interface is a web-based application crafted using Vue.js, a progressive JavaScript framework that simplifies the development of interactive web applications.
- The map is rendered using Leaflet, a popular open-source JavaScript library for mobile-friendly interactive maps. The leaflet provides a wide range of functionalities for displaying and interacting with geospatial data.
- Some other components of the interface use the Ant Design Vue component library, which provides a rich set of pre-built UI components to enhance the user experience and streamline development.

Backend Architecture:

- The system backend is built using Flask, a lightweight web framework. Flask is used to handle web requests and route them to the appropriate functions.
- Data classification and preprocessing tasks are accomplished on the backend to ensure the efficiency of processing.

LLM Integration:

- The data processing stage utilizes the Qwen-long model (Bai et al., 2023), which has a maximum context length of 10,000,000 tokens, enabling the processing of large-scale text data files.
- The color concept and scheme generation stages utilize the Qwen-plus model (Bai et al., 2023), which, compared to Qwen-long, offers a more balanced performance in terms of effectiveness, speed, and cost, enabling it to comprehend users' simple or abstract requirements for map colors. Its context length is 131,072 tokens.
- The temperature of all Qwen models is set to the default value of 1.0. For each stage, we use some few-shot examples that cover map color design styles. These examples are carefully selected to provide a comprehensive yet concise context for the LLM. All outputs of the LLM are strictly formatted to facilitate parsing and utilization by the system.

3.5.2. Interaction design

We have developed an interactive mapping system that integrates LLM into every step of the mapping process, as illustrated in Figure 6. For more detailed interactive demonstrations, please see the supplementary materials. The interface primarily consists of three views:

(1) **Conversation View** (Figure 6a). This view serves as the primary interface for users to interact with the system using natural language. Users can guide the color design process by inputting instructions, such as "I want a Statue of Liberty like map" for generating the color scheme directly, or "make the colors brighter" or "increase the contrast" to refine the color concepts or color schemes. The results are then previewed in the right-hand view. To handle complex tasks, this view supports multi-turn dialogs, allowing users to gradually refine their requirements and receive coherent assistance. In cases where there is a conflict between the system's initial data-driven color suggestions and user preferences, the system prioritizes user inputs. Specifically, while the initial color scheme is informed by an analysis of the thematic characteristics of the data, subsequent user modifications guide the LLM to adapt the color scheme accordingly, ensuring that the final output aligns predominantly with user requirements.

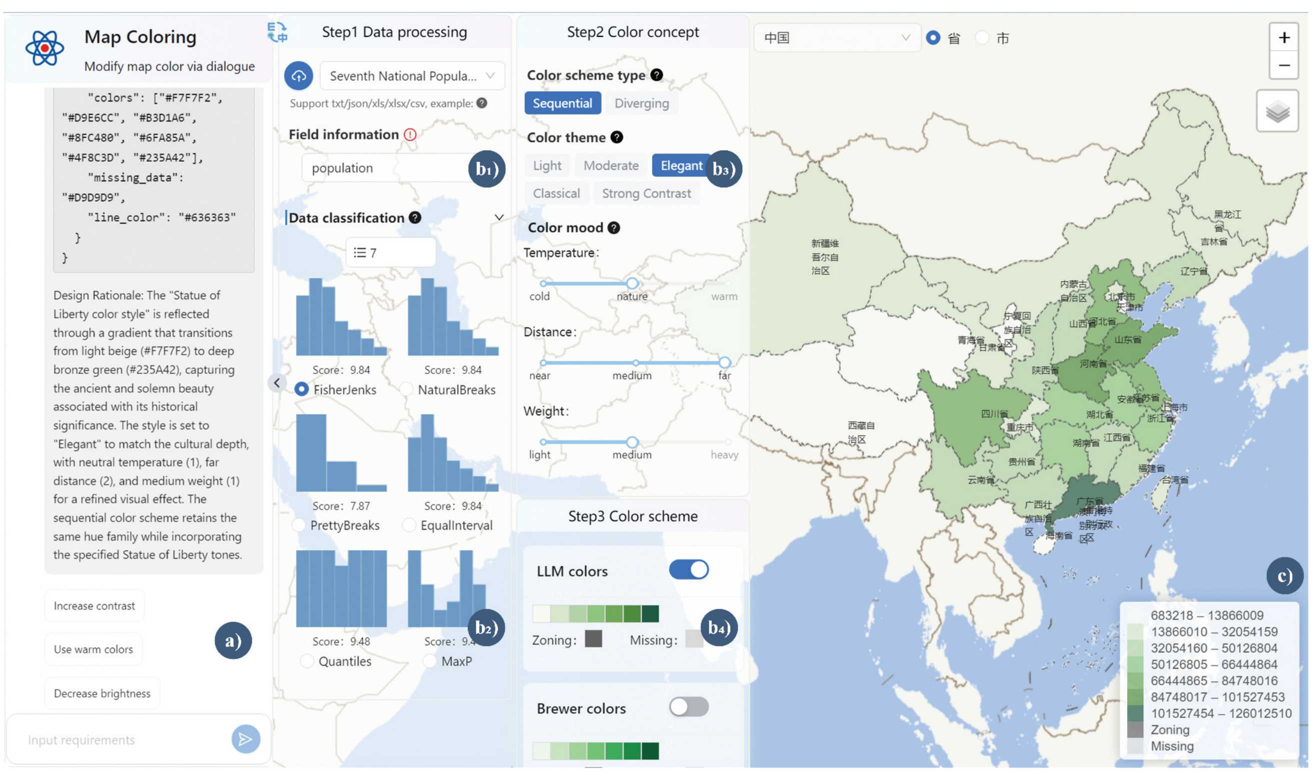

Figure 6. System interface (an example of "Statue of Liberty" map color design). a) Conversation view: users can input their initial design intents and customizations in natural language. b) Color design view: users can fine-tune the intermediate results generated by the LLM. c) Map view: users can evaluate the results with reasoning and refine the color assignments based on their preferences.

(2) **Color design view** (Figure 6b). This view focuses on the interactive modification of various attributes during the color design process. As mentioned earlier, we divide the choropleth map color creation process into three stages. In the first data processing stage, users can upload their mapping data (Figure 6b$_1$), classify the data, and we provide visual results of six classification methods for users to choose from, automatically recommending the result with the highest GVF score (Figure 6b$_2$). In the second stage of color creative design, users can select the desired color scheme types and color themes through tags and adjust color psychological effects such as temperature, weight, and distance using sliders (Figure 6b$_3$). When users adjust these elements, prompts appear below the sliders and to the left in the "Conversation view," asking users whether they wish to apply the current parameters to generate a corresponding color scheme. Upon users' confirmation, the system triggers the LLM to regenerate the color scheme based on the updated parameters. Throughout this process, the system ensures that user preferences dynamically override earlier data-driven suggestions when inconsistencies arise, maintaining flexibility in adapting the color outputs. The newly generated color scheme is then displayed in both the "Color scheme" and "Map view." In the third stage of color schemes, specific colors generated by LLM in hierarchical order are displayed, along with the closest ColorBrewer color scheme, allowing users to switch between these two schemes for viewing or adjust the specific color using a modified color plate (Figure 6b$_4$).

(3) **Map view** (Figure 6c). This view displays the resulting map using the generated color scheme, allowing users to zoom in, zoom out, and pan the map to inspect the coloring details of different areas. The hierarchical division of the mapping area and the corresponding colors are placed in the legend at the bottom right corner for user reference. The map view is tightly integrated with the other two views, and any color changes are instantly reflected in the map view, facilitating user adjustments and refinements.

4. User study

To assess the usability of the proposed system, we conducted a user study to assess the system's usability and satisfaction, including outcome relevance (Q1 and Q2), tool completeness (Q3), ease of use (Q4-Q6), system flexibility (Q7), and overall user satisfaction (Q8-Q10), with the tasks illustrated in Figure 7. Participants were instructed to use the system to generate color schemes for a given dataset. Feedback was then collected through a combination of structured questionnaires and semi-structured interviews.

4.1. Participants

We conducted a user study with 60 participants (P1-P60) aged between 18 and 44 years, a sample size determined based on relevant research conclusions to ensure result reliability (Julious, 2005; Lancaster et al., 2004; Teresi et al., 2022). Participants self-identified their gender, with 32 identifying as male and 28 as female. Given that professional knowledge or background in designing color schemes for maps could significantly influence feedback, we assessed participants' familiarity with this subject. They responded to a Likert-scale question: "*Please indicate*

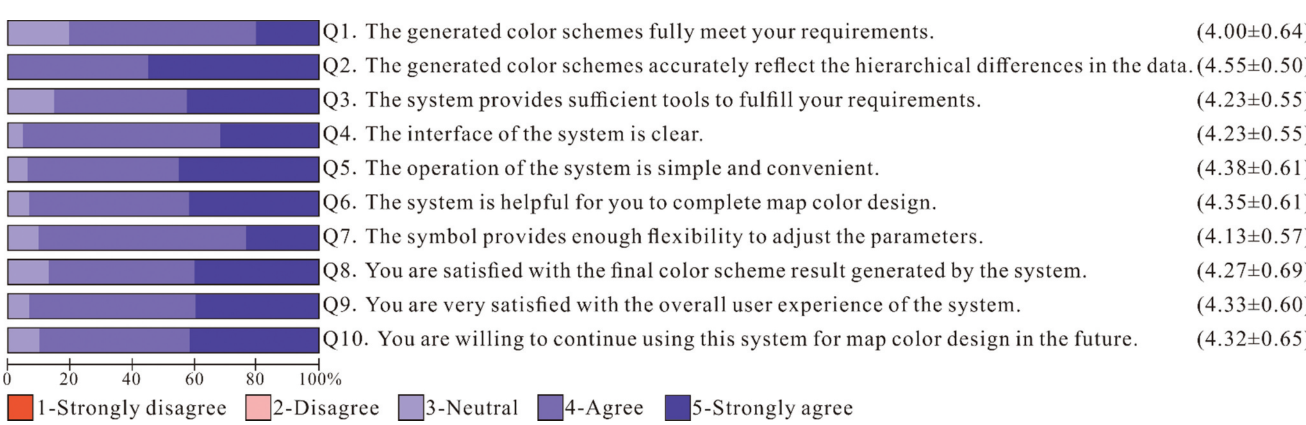

Figure 7. Ratings for system usability and satisfaction on a 5-point Likert scale. The middle column shows the detailed questions. The right column displays the average and standard deviations.

your level of knowledge [1, 2, 3, 4] regarding color scheme design for a map." Responses were categorized into two groups: unfamiliar (levels 1 and 2) and familiar (levels 3 and 4). The results showed an equal distribution, with 30 participants classified as familiar and 30 as unfamiliar.

4.2. Material

Participants will uniformly use the administrative division map of China as the base map. Four sets of data are prepared for users to create maps, and users can freely choose to create at least one map, with a brief introduction to the data as follows:

- "Per Capita Disposable Income by Region in 2022 of China," unit (yuan), data field "income;"
- "GDP of Provinces in 2023 of China," unit (billion yuan), data field "gdp;"
- "Seventh National Population Census in China," unit (person), data field "population."
- "Total Electricity Generation by Region in 2022 of China," unit (100 million kWh), data field "generation."

Our system is suitable for creating choropleth maps across various themes and is not limited to the themes listed above. Participants with professional backgrounds and specific needs can still refer to the system's data format guidelines to craft and upload their data. These datasets were selected because they represent common and widely used themes in choropleth mapping, particularly in the context of socio-economic and demographic analysis. By focusing on these themes, we aimed to ensure the study's relevance to practical applications and to facilitate a clear evaluation of our system's performance across familiar and interpretable datasets.

4.3. Procedure

The procedure contains three steps, as follows: (1) a tutorial session to make users familiarized with the proposed system, (2) a creation session to experience the authoring process, and (3) a post-study evaluation to collect their feedback on the utility of the system.

(1) **Tutorial**. Participants were introduced to the proposed system through a 15-minute presentation explaining the motivation behind the work. This was followed by a 10-minute demonstration using slides, showcasing the system's key functionalities, such as inputting user requirements, setting the number of classifications, selecting a color mood, designing a color scheme, and interacting with related features. Participants were then encouraged to freely explore the system's functions and interactions, asking questions as needed. Once participants confirmed their familiarity with the tool, we introduced the formal dataset for the user study.

(2) **Creation**. After completing the tutorial, participants were instructed to use the system according to the system user manual and demonstration videos, and request additional guidance when necessary. They could refer to the tutorial slides and request additional guidance if necessary. Upon completing their designs, each participant shared and explained their map. This creation phase lasted approximately 5 to 10 minutes.

(3) **Post-study Survey and Interview**. Following the creation phase, participants completed a post-study questionnaire featuring 10 questions (Figure 7) using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The 10 questions were used to assess the system's usability and satisfaction, including outcome relevance (Q1 and Q2), tool completeness (Q3), ease of use (Q4-Q6), system flexibility (Q7), and overall user satisfaction (Q8-Q10). Additionally, participants provided demographic information, including age, gender, and familiarity levels, before responding to the system evaluation questionnaire. Finally, we conducted semi-structured interviews to collect qualitative feedback from each participant, offering deeper insights into their experiences and perspectives.

4.4. Result analysis

As we use the Likert scale to assess the system's usability and satisfaction, the reliability and validity of the survey were first assessed using Cronbach's α coefficient, which was 0.888 (standardized $\alpha = 0.887$), indicating good internal consistency, and a KMO value of 0.894 along with Bartlett's test of sphericity ($\chi^2 = 251.341$, $p < 0.001$), confirming significant correlations among variables and the suitability of the data for factor analysis. The results indicate the reliability and validity of the survey and the detailed results of different tasks are analyzed as follows.

4.4.1. The overall results

The average scores, standard deviations (SD), and response distributions for each question are presented in Figure 7. As shown, there was a strong level of agreement among participants, with the overall scores indicating high usability of the proposed system. The mean score for outcome relevance (Q1 and Q2) was 4.00 (SD: 0.64) and 4.55 (SD: 0.50), respectively, reflecting that participants found the tool highly relevant to their needs. For tool completeness (Q3), the average score was 4.23 (SD: 0.55), suggesting that users perceived the tool as comprehensive and capable of meeting their requirements. As P2 remarked, "The system is relatively convenient to use and the tools can meet most color matching needs."

When evaluating ease of use (Q4-Q6), the average scores ranged from 4.23 to 4.38, with standard deviations between 0.55 and 0.61, indicating that participants found the tool intuitive and user-friendly. Regarding the interface, P22, P25, and P53 specifically commented, "The interface is concise and aesthetically pleasing." P3 noted, "The interface is straightforward, making it easy for me to start using it without a steep learning curve." For system flexibility (Q7), the average score was 4.13 (SD: 0.57), suggesting that users appreciated the tool's adaptability. As P24 observed, "The tool can be customized to suit my specific needs, which is an important feature for non-experts."

Finally, overall user satisfaction (Q8-Q10) yielded average scores of 4.27 (SD: 0.69), 4.33 (SD: 0.60), and 4.32 (SD: 0.65), reflecting a high level of user contentment. P41 remarked, "Using large language models for color matching is both interesting and convenient, saving a lot of time and effort."

In summary, the average score for all questions exceeded 4.0, with participants highlighting the tool's accuracy, ease of use, and flexibility as key factors contributing to their positive experience.

4.4.2. The influence analysis of familiarity (familiar vs unfamiliar)

The statistical analysis was conducted to compare the performance between two groups – Familiar and Unfamiliar – across 10 questions (Q1-Q10). The Shapiro-Wilk test was first employed to assess the normality of the data. Although the *p*-values for all 10 questions were below 0.05, indicating a departure from normality, further analysis based on skewness and kurtosis revealed that each question has a bell-shaped distribution with a peak in the middle and lower ends. These suggest the data could be considered approximately normal. Consequently, a test for homogeneity of variance was performed to assess the equality of variances between the two groups. In cases (Q1-Q3, Q5, Q6, and Q8-Q10) where the homogeneity assumption was met, an independent samples t-test was used to determine if there were significant differences between the groups. In cases (Q4 and Q7) where the homogeneity assumption was violated, Welch's t-test, which does not assume equal variances, was used. The results are shown in Table 2.

The results show that the Unfamiliar and Familiar groups exhibited significant differences in performance in Q1 and Q9, with *p*-values less than 0.05 and Cohen's d values greater than 0.5. These results suggest that the differences, while statistically significant, have moderate practical significance. Such differences may be attributed to the contrasting experiences of the two groups. For example, professionals familiar with preset templates may find it challenging to generate color schemes that cater to personalized needs, whereas non-experts may not perceive this issue. In contrast, the new system addresses these challenges effectively. As noted by Participant P5:

Table 2. The influence analysis of familiarity.

Questions	Factor	Average score	Homogeneity test (*p*-value)	t-Test (*p*-value)	Welch's t-Test (*p*-value)	Cohen's d value
Q1	Familiar	4.2	0.259	**0.014****	____	**0.655**
	Unfamiliar	3.8				
Q2	Familiar	4.633	0.182	0.201	____	0.334
	Unfamiliar	4.467				
Q3	Familiar	4.2	0.159	0.472	____	0.187
	Unfamiliar	4.333				
Q4	Familiar	4.333	**0.001*****	____	0.351	0.243
	Unfamiliar	4.2				
Q5	Familiar	4.4	0.866	0.835	____	0.054
	Unfamiliar	4.367				
Q6	Familiar	4.267	0.898	0.290	____	0.275
	Unfamiliar	4.433				
Q7	Familiar	4.2	**0.099***	0.366	____	0.235
	Unfamiliar	4.067				
Q8	Familiar	4.333	0.336	0.456	____	0.194
	Unfamiliar	4.2				
Q9	Familiar	4.5	0.029**	____	**0.031****	**0.572**
	Unfamiliar	4.167				
Q10	Familiar	4.433	0.149	0.1677		0.361
	Unfamiliar	4.2				

***, **, and * denote significance levels of 1%, 5%, and 10%, respectively.

This tool can effectively address the lack of color knowledge among beginners in map design. Many beginners tend to use random color schemes from tools like ArcGIS or QGIS, and even when there are customization options, adjusting them can be difficult. However, this tool integrates large language models for custom color generation, solving the problem of inaccurate map color choices for non-professionals and demonstrating excellent application effectiveness. This feedback highlights that experts recognize the potential of the proposed system, particularly its practical value for future use.

4.5. Semi-structured interview

To gain a deeper understanding of the user experience with MapColorAI, particularly in comparison to professional cartographic tools (ArcGIS, QGIS, etc.), we conducted follow-up semi-structured interviews with a subset of participants from the initial user study. While the initial study provided quantitative data on utility and usability, this interview aimed to capture more nuanced perspectives and direct comparisons from users with experience in professional cartographic workflows. The specific interview design can be referenced in the Appendix.

Participants. We invited 12 participants (P1-P12) from our original study who had indicated experience with professional cartographic software (ArcGIS, QGIS, etc.) to participate in follow-up interviews. The interviews lasted approximately 15 minutes and were structured around three main themes: color scheme workflow, user interaction for adjustments, and comparative analysis between MapColorAI and professional tools.

The interviews were audio-recorded with participants' consent and later transcribed for analysis. We employed thematic analysis to identify recurring patterns and insights across the interviews. The semi-structured format allowed us to explore predefined topics while giving participants the freedom to elaborate on aspects they found most relevant.

Workflow Comparison. Participants generally perceive that MapColorAI's process of generating initial color schemes through LLM is simpler and more efficient. Unlike traditional tools that require users to select from a large number of preset color palettes, MapColorAI can automatically generate initial schemes based on data characteristics and classification results, reducing the user's selection burden and trial-and-error time. For instance, P1 mentioned, "It can directly generate color schemes through the large model in one step, without the need to choose from multiple schemes as in ArcGIS." Additionally, MapColorAI's visualization of classification results and GVF scoring function make the data classification process more intuitive and convenient, further simplifying the workflow. However, P9 pointed out that MapColorAI's generation process is relatively opaque, and in some cases, multiple attempts may be required to achieve satisfactory results. Overall, MapColorAI excels in initial color scheme generation and classification method selection, with participants assigning it an average score of 4.42, significantly higher than the 3.00 for professional cartographic tools. However, there remains room for improvement in the transparency and stability of the generated schemes.

Interaction Design Evaluation. Regarding the flexibility and controllability of interaction methods, participants generally agreed that MapColorAI offers more flexible modification approaches. They rated the color adjustment interaction provided by MapColorAI at an average of 4.50, higher than the 3.42 given to professional tools. Natural language interaction was frequently highlighted as a core feature, enabling users to express color scheme intentions in a manner closer to human thought processes, facilitating more macro-level style adjustments (P2, P6, P7, and P12), or addressing needs that are difficult to quantify (P8). Style parameter adjustments were also viewed as an innovative control method (P8, and P9). Although P11 noted that natural language, while highly flexible, is not always the most controllable due to potential deviations from expected interpretations by LLM, the overall consensus was that the diversified interaction methods are superior to those of specialized tools.

Overall Preference. In a comprehensive comparison of overall preferences for completing map colorization tasks between MapColorAI and professional cartography tools, participants showed a stronger inclination toward using MapColorAI. MapColorAI received an average rating of 4.58, while professional cartography tools averaged 3.92. Participants identified the most appealing features of MapColorAI as natural language interaction, initial scheme recommendations based on data characteristics, and style parameter adjustments. Participants also highlighted areas where MapColorAI falls short and requires improvement: limited support for only choropleth maps and restricted functional coverage (P3, P4, P9 andP12); occasional slower or unstable generation speeds by the LLM (P2 and P9). Overall, MapColorAI excels in rapid map generation and creative exploration but still needs further enhancement in professional depth and functional breadth.

5. Discussion

5.1. Broader map color knowledge

The current system demonstrates substantial capabilities in utilizing LLM for generating map color schemes based on predefined parameters. However, there remains considerable potential to expand the system's functionality by incorporating a broader and more nuanced understanding of map color knowledge. As remarked by P3 "The number of available themes now is a bit limited." This would allow for the development of more sophisticated and contextually relevant color schemes, addressing a wider range of map types and user needs.

(1) Integration of Data-Driven Insights. The system could further expand its capabilities by integrating data-driven insights into the color generation process. By analyzing the themes of maps and the data they represent, the system could adapt its color generation to be more suitable for specific themes or data. For instance, by analyzing the underlying data, the system could apply more suitable color schemes, improving the clarity and utility of the final map.
(2) Context-Aware Color Generation. A key aspect of enhancing the system's color knowledge is the integration of context-aware color generation. While the existing system generates color schemes primarily based on standard color theory, it could benefit from considering additional factors such as cultural preferences, color blindness, and regional color conventions. For example, the use of certain color palettes may vary across different cultures and geographic regions, with specific colors having symbolic meanings (e.g. red representing danger or urgency in some regions). By incorporating region-specific and culturally relevant color data, the system could generate more culturally appropriate and effective color schemes for a global user base. Incorporating such nuanced knowledge could increase the system's versatility, ensuring that the generated maps are not only visually appealing but also relevant and culturally sensitive.
(3) Expansion of the Professional Color Scheme Library. The system currently relies on a limited professional color scheme library – namely, ColorBrewer – for benchmarking and user comparison. While ColorBrewer is widely accepted in the cartographic community, expanding the scheme database to include color systems from other domains (e.g. print design, branding, or climate visualization) could improve the richness and creativity of LLM-generated outputs. This expansion may also improve alignment with domain-specific needs or aesthetics.

5.2. Extend multimodal and user interaction

While the system currently generates map colors based on text input and preset parameters, there is significant potential to expand its interactivity and multimodal capabilities. Future versions of the system could allow users to interact with the tool through more diverse modes, such as visual feedback and voice commands. This would provide users with a more intuitive experience, particularly for those who may find textual interfaces less comfortable or accessible. For example, many users now prefer generating color schemes by providing an image, from which colors are then transferred to their maps (Wu, Sun, et al., 2022). However, current image-to-map color transfer methods are typically end-to-end processes, which lack flexibility. Integrating this feature into our system, alongside the existing text input option, could offer greater flexibility, allowing users to apply colors directly from an image to a map while receiving immediate visual feedback that can aid decision-making. Two modes were planned to be added to this system:

(1) **Visual interaction**. Unlike previous approaches where the input image is solely used as a color source, our system will allow users to interactively select specific objects or areas within the image for color extraction. This selective approach gives users more control over which elements of the image are used to generate the color scheme for the map. Additionally, to enhance personalization, users will also have the option to combine text input with image input, allowing them to refine the color selection process by specifying textual preferences alongside visual cues. This capability leverages the power of advanced visual large language models, which can understand and process images. For example, models like CLIP (Radford et al., 2021) can describe the content of an image, and DINO v2 (Oquab et al., 2023) can segment the image and recognize objects. This flexibility allows for the more nuanced and context-aware color generation, catering to both expert and non-expert users.
(2) **Voice interaction**. Similar to text-based interactions, voice input would be processed in two stages: the spoken input would first be converted into text through speech recognition technology, and the text would then be used to drive the system's response. Advances in large language models, such as GPT-4, now support voice-based interactions, which can be seamlessly integrated into the system. Given that voice interactions may take slightly longer to process than text inputs, careful design of the user interface is necessary to avoid frustration. To enhance responsiveness, the system could stream output progressively as it receives voice commands, providing users with immediate feedback while they interact with the tool. Furthermore, the system could offer prompt acknowledgments, such as

confirming receipt of the command and displaying intermediate results as they are generated.

5.3. Improve the system via open user evaluation

Though we evaluated the usability of the proposed system with a dataset across 60 participants, the sample size and test cases are still limited. It is crucial to further test the system in a wider range of use scenarios and with a more diverse group of users to better understand its limitations and areas for improvement. The current evaluation provides valuable insights, but additional testing across different contexts – such as varying map data, user expertise levels, and geographical regions – is necessary to assess the system's generalizability and effectiveness in real-world applications. Furthermore, while the system is functional, it is still in the early stages of development, and significant opportunities exist to enhance its performance, versatility, and overall user experience. As P39 claimed "The AI has shown a decent understanding capability, and the generated color schemes basically meet the requirements, but the functionality is still relatively rudimentary." Continuous development is critical to ensure the tool's long-term success and relevance.

5.4. Other limitations and future works

While MapColorAI demonstrates promising capabilities in enhancing user-centered map color design, beyond the above limitations in *Sections* 5.1 to 5.3, several other limitations also remain that highlight opportunities for future development.

First, although the system offers contextually adaptive color schemes via LLM-based interaction, it currently focuses solely on choropleth maps. This narrow scope restricts the applicability of the system to other thematic map types, such as proportional symbol maps, dot density maps, isopleth maps, or other map elements, such as the color of point symbols or layouts, which involve different design considerations and data representation challenges. Future work could explore extending the system's underlying framework and interaction paradigms to accommodate these formats.

Second, although the LLM enables flexible, natural language-based interactions, it may still produce inconsistent or unstable outputs in response to similar prompts – an inherent limitation of current generative models. In some cases, generating satisfactory color suggestions requires multiple rounds of iteration, which can affect the efficiency and reliability of the user experience. Enhancing the consistency, transparency, and controllability of the model remains a critical direction for future improvement. Furthermore, the integration of more advanced models with stronger reasoning capabilities may help address these limitations and support more stable and interpretable design outcomes.

Third, while the system supports multi-modal input (e.g. natural language and image inspiration), this function is still in early-stage implementation. Future enhancements could deepen the integration of image-based prompts, allowing the LLM to extract visual semantics (e.g. atmosphere, palette tones) and translate them into data-driven color schemes for maps.

Fourth, reproducibility remains an important challenge in LLM-based systems. Although we have provided detailed prompt templates, user instructions, and supplementary files (see Appendix and Figshare repository), we acknowledge that LLM behaviors may evolve with model updates or even hallucinations, potentially affecting long-term consistency. We plan to integrate version control and logging mechanisms into the system to support better reproducibility and user traceability.

6. Conclusion

We have introduced a novel system for choropleth map color design, which leverages LLM to align color schemes with user intentions and data semantics. This system addresses the limitations of traditional mapping tools by providing a more adaptable and user-centric approach to map design. Through a three-stage process – Data Classification, Color Concept Design, and Color Scheme Design – the system integrates design principles and color psychology to produce relevant and visually effective color schemes. The interactive interface further enhances the system's usability, allowing users to customize and refine their color choices based on specific needs and preferences. Results from user studies highlight the tool's high usability, flexibility, and alignment with user expectations, with participants emphasizing its efficiency and ease of use. However, it is important to note that the system is still in its early stages, and several areas require improvement. Future work will focus on enhancing the accuracy of color mapping in diverse contexts and expanding the interaction modes to provide users with a more intuitive experience. Despite its current limitations, we believe this work provides valuable insights into how LLM can be used to enhance map design processes, and we look forward to continued improvements that will support creative and data-driven decision-making in spatial data visualization.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was financially supported by the National Natural Science Foundation of China [No. 42171438].

ORCID

Nai Yang http://orcid.org/0000-0001-5306-1163
Zhiwei Wei http://orcid.org/0000-0002-3494-3686

Data and code availability statement and data deposition

Some experimental data and detailed system interaction demonstrations are available at the following link: https://doi.org/10.6084/m9.figshare.28279850. Additional supplementary materials can be found in the Appendix.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L. . . . McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv: 2303.08774*. https://doi.org/10.48550/arXiv.2303.08774

Alexander, K. R., & Shansky, M. S. (1976). Influence of hue, value, and chroma on the perceived heaviness of colors. *Perception and Psychophysics*, *19*(1), 72–74. https://doi.org/10.3758/bf03199388

Armstrong, M. P., Xiao, N., & Bennett, D. A. (2003). Using genetic algorithms to create multicriteria class intervals for choropleth maps. *Annals of the Association of American Geographers*, *93*(3), 595–623. https://doi.org/10.1111/1467-8306.9303005

Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X. . . . Zhu, T. (2023). Qwen technical report. *arXiv preprint arXiv: 2309.16609*. https://doi.org/10.48550/arXiv.2309.16609

Bertin, J. (1974). *Graphische semiologie: Diagramme, netze, karten*. De Gruyter. https://doi.org/10.1515/9783110834901

Brewer, C. A. (1994). Color use guidelines for mapping and visualization. In A. M. MacEachren & D. F. Taylor (Eds.), *Modern cartography series* (pp. 123–147). Academic Press. https://doi.org/10.1016/b978-0-08-042415-6.50014-4

Brewer, C. A., Hatchard, G. W., & Harrower, M. A. (2003). ColorBrewer in print: A catalog of color schemes for maps. *Cartography and Geographic Information Science*, *30*(1), 5–32. https://doi.org/10.1559/152304003100010929

Brewer, C. A., MacEachren, A. M., Pickle, L. W., & Herrmann, D. (1997). Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers*, *87*(3), 411–438. https://doi.org/10.1111/1467-8306.00061

Brychtova, A., & Coltekin, A. (2015). Discriminating classes of sequential and qualitative colour schemes. *International Journal of Cartography*, *1*(1), 62–78. https://doi.org/10.1080/23729333.2015.1055643

Caelen, O., & Blete, M. A. (2024). *Developing apps with GPT-4 and ChatGPT : Build intelligent chatbots, content generators, and more*. O' Reilly Media.

Chesneau, E. (2011). A model for the automatic improvement of colour contrasts in maps: Application to risk maps. *International Journal of Geographical Information Science*, *25*(1), 89–111. https://doi.org/10.1080/13658811003772926

Christophe, S. (2011). Creative colours specification based on knowledge (ColorLegend system). *The Cartographic Journal*, *48*(2), 138–145. https://doi.org/10.1179/1743277411y.0000000012

Cuff, D. J. (1973). Colour on temperature maps. *The Cartographic Journal*, *10*(1), 17–21. https://doi.org/10.1179/caj.1973.10.1.17

Declerq, F. (1995). Choropleth map accuracy and the number of class intervals. *Proceedings of the 17th Conference and the 10th General Assembly of the International Cartographic Association* (Vol. 1, pp. 918–922). Institute Cartogràfic de Catalunya.

Duque, J. C., Anselin, L., & Rey, S. J. (2012). The Max- P -Regions Problem. *Journal of Regional Science*, *52*(3), 397–419. Portico. https://doi.org/10.1111/j.1467-9787.2011.00743.x

Evans, I. S. (1977). The selection of class intervals. *Transactions of the Institute of British Geographers*, *2*(1), 98. https://doi.org/10.2307/622195

Golebiowska, I. M., & Coltekin, A. (2022). Rainbow dash: Intuitiveness, interpretability and memorability of the rainbow color scheme in visualization. *IEEE Transactions on Visualization and Computer Graphics*, *28*(7), 2722–2733. https://doi.org/10.1109/tvcg.2020.3035823

Harrower, M., & Brewer, C. A. (2003). ColorBrewer. org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, *40*(1), 27–37. https://doi.org/10.1179/000870403235002042

He, Z., Song, Y., & Li, L. (2016). *Cartography*. Wuhan University Press.

Holtzschue, L. (2016). *Understanding Color: An introduction for designers*. John Wiley & Sons. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=4169664

Hou, Y., Yang, M., Cui, H., Wang, L., Xu, J., & Zeng, W. (2024). C2Ideas: Supporting creative interior color design ideation with a large language model. In F. F. Mueller, P. Kyburz, J. R. Williamson, C. Sas, M. L. Wilson, P. Toups Dugas, & I. Shklovski (Eds.), *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–18). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642224

Jenks, G. F., & Caspall, F. C. (1971). Error on choroplethic maps: Definition, measurement, reduction. *Annals of the Association of American Geographers*, *61*(2), 217–244. https://doi.org/10.1111/j.1467-8306.1971.tb00779.x

Julious, S. A. (2005). Sample size of 12 per group rule of thumb for a pilot study. *Pharmaceutical Statistics*, *4*(4), 287–291. https://doi.org/10.1002/pst.185

Kang, Y., Gao, S., & Roth, R. E. (2024). Artificial intelligence studies in cartography: A review and synthesis of methods, applications, and ethics. *Cartography and Geographic Information Science*, *51*(4), 599–630. https://doi.org/10.1080/15230406.2023.2295943

Kristian, E. (2020). *Better colors for better mapping*. Esri. https://www.esri.com/arcgis-blog/products/js-api-arcgis/mapping/better-colors-for-better-mapping/

Lancaster, G. A., Dodd, S., & Williamson, P. R. (2004). Design and analysis of pilot studies: Recommendations for good practice. *Journal of Evaluation in Clinical Practice*, *10*(2), 307–312. https://doi.org/10.1111/j.2002.384.doc.x

Lei, F., Fan, A., MacEachren, A. M., & Maciejewski, R. (2024). GeoLinter: A linting framework for choropleth maps. *IEEE Transactions on Visualization and Computer Graphics*, *30*(2), 1592–1607. https://doi.org/10.1109/tvcg.2023.3322372

Levy, B. I. (1984). Research into the psychological meaning of color. *American Journal of Art Therapy*, *23*(2), 58–62. https://psycnet.apa.org/record/1986-23642-001

Liu, A., Feng, B., Xue, B., et al. (2024). Deepseek-v3 technical report. *arXiv preprint arXiv: 2412.19437*. https://doi.org/10.48550/arXiv.2412.19437.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, *55*(9), 1–35. https://doi.org/10.1145/3560815

MacEachren, A. M. (2004). *How maps work: Representation, visualization, and design*. Guilford Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V. . . . Bojanowski, P. (2023). Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv: 2304.07193*. https://doi.org/10.48550/arXiv.2304.07193

Peterson, G. N. (2020). *GIS cartography: A guide to effective map design*. CRC Press.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *arXiv preprint arXiv: 2103.00020*. https://doi.org/10.48550/arXiv.2103.00020

Shi, C., Cui, W., Liu, C., Zheng, C., Zhang, H., Luo, Q., & Ma, X. (2023). NL2Color: Refining color palettes for charts with natural language. *IEEE Transactions on Visualization and Computer Graphics*, 1–11. https://doi.org/10.1109/tvcg.2023.3326522

Shive, J., & Francis, G. (2013). Choosing colors for map display icons using models of visual search. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(2), 373–396. https://doi.org/10.1177/0018720812459341

Silva, S., Santos, B. S., & Madeira, J. (2011). Using color in visualization: A survey. *Computers & Graphics*, *35*(2), 320–333. https://doi.org/10.1016/j.cag.2010.11.015

Slocum, T. A., McMaster, R. B., Kessler, F., & Howard, H. H. (2022). *Thematic cartography and geovisualization*. CRC Press, Taylor & Francis Group. https://doi.org/10.1201/9781003150527

Sun, M., Wong, D. W., & Kronenfeld, B. J. (2014). A classification method for choropleth maps incorporating data reliability information. *The Professional Geographer*, *67*(1), 72–83. https://doi.org/10.1080/00330124.2014.888627

Teresi, J. A., Yu, X., Stewart, A. L., & Hays, R. D. (2022). Guidelines for designing and evaluating feasibility pilot studies. *Medical Care*, *60*(1), 95–103. https://doi.org/10.1097/mlr.0000000000001664

Tian, Y., Cui, W., Deng, D., Yi, X., Yang, Y., Zhang, H., & Wu, Y. (2025). ChartGPT: Leveraging LLMs to generate charts from abstract natural language. *IEEE Transactions on Visualization and Computer Graphics*, *31*(3), 1731–1745. https://doi.org/10.1109/tvcg.2024.3368621

Vázquez, P. P. (2024). Are LLMs ready for visualization? *2024 IEEE 17th Pacific Visualization Conference (PacificVis)* (pp. 343–352). IEEE. https://doi.org/10.1109/pacificvis60374.2024.00049

Wei, Z., Guo, Q., Yan, F., & Wang, Y. (2018). Backtracking method of coloring administrative maps considering visual perception rules. *Acta Geodaetica et Cartographica Sinica*, *47*(3), 396. https://doi.org/10.11947/j.AGCS.2018.20170148

Wu, M., Sun, Y., & Li, Y. (2022). Adaptive transfer of color from images to maps and visualizations. *Cartography and Geographic Information Science*, *49*(4), 289–312. https://doi.org/10.1080/15230406.2021.1982009

Wu, M., Sun, Y., & Xia, X. (2024). Computational assessment of the aesthetic quality of map colors. *Transactions in GIS*, *28*(6), 1556–1576. https://doi.org/10.1111/tgis.13206

Wu, T., Jiang, E., Donsbach, A., Gray, J., Molina, A., Terry, M., & Cai, C. J. (2022). PromptChainer: Chaining large language model prompts through visual programming. In S. Barbosa, C. Lampe, C. Appert, & D. A. Shamma (Eds.), *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery* (Article 359, pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3491101.3519729

Wu, T., Terry, M., & Cai, C. J. (2022). AI chains: Transparent and controllable human-AI interaction by chaining large language model prompts. In S. Barbosa, C. Lampe, C. Appert, & D. A. Shamma (Eds.), *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3491102.3517582

Xi, D., Hu, X., Yang, L., Yang, N., Liu, Y., & Jiang, H. (2023). Research on map emotional semantics using deep learning approach. *Cartography and Geographic Information Science*, *50*(5), 465–480. https://doi.org/10.1080/15230406.2023.2172081

Xue, D., Vazquez-Corral, J., Herranz, L., Zhang, Y., & Brown, M. S. (2024). Palette-based color harmonization via color naming. *IEEE Signal Processing Letters*, *31*, 1474–1478. https://doi.org/10.1109/lsp.2024.3401612

Zhou, L., & Hansen, C. D. (2016). A survey of colormaps in visualization. *IEEE Transactions on Visualization and Computer Graphics*, *22*(8), 2051–2069. https://doi.org/10.1109/tvcg.2015.2489649