SAGE: Segmenting and Grouping Data Effectively using Large Language Models

by

Isabella Pedraza Pineros

S.B., Computer Science and Engineering, Massachusetts Institute of Technology, 2023

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Authored by:	Isabella Pedraza Pineros Department of Electrical Engineering and Computer Science May 17, 2024
Certified by:	Arvind Satyanarayan Assistant Professor, Thesis Supervisor
Accepted by:	Katrina LaCurts Chair, Master of Engineering Thesis Committee

SAGE: Segmenting and Grouping Data Effectively using Large Language Models

by

Isabella Pedraza Pineros

Submitted to the Department of Electrical Engineering and Computer Science on May 17, 2024 in partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

ABSTRACT

Grouping is a technique used to organize data into manageable pieces, reducing cognitive load and enabling users to focus on discovering higher-level insights and generating new questions. However, creating groups remains a challenge, often requiring users to have prior domain knowledge or an understanding of the underlying structure of the data. We introduce **SAGE**, a novel technique that leverages the knowledge base and pattern recognition abilities of large language models (LLMs) to segment and group data with domain-awareness. We instantiate our technique through two structures: *bins* and *highlights*; bins are contiguous, non-overlapping ranges that segment a single field into groups; highlights are multi-field intersections of ranges that surface broader groups in the data. We integrate these structures into Olli, an open-source tool that converts data visualizations into accessible, keyboard-navigable textual formats to facilitate a study with 15 blind and low-vision (BLV) participants, recognizing them as experts in assessing agency. Through this study, we evaluate how SAGE impacts a user's interpretation of data and visualizations, and find our technique provides a rich contextual framework for users to independently scaffold their initial sensemaking process.

Thesis supervisor: Arvind Satyanarayan
Title: Assistant Professor

Acknowledgments

First and foremost, I want to say thank you to my parents and to family. Your kindness, wisdom, and love are the reason I've become the person I am today.

Thank you to all the professors I've learnt from over the years – Angela Blake, Sister Pauline, Carolyn Dorn, Michael Hutsko, Daniel Jackson – your passion for education and your encouragement changed my life.

Thank you to my thesis supervisor, Arvind Satyanarayan, for your mentorship. Your ongoing support spurred me to seek out new, difficult, yet interesting problems to explore. Thank you for always encouraging a culture of curiosity and empathy.

Thank you to my thesis mentor, Jonathan Zong, for your guidance over the years. Your perspective of the world and the way you approach a problem have inspired me to think in new ways.

Thank you to the people at the MIT Visualization Group. You are all the reason why I learnt so much and had so much fun this year.

Thank you Anthony, for all the care you've given me. Thank you for the matcha lattes, home cooked meals, and advice. Thank you for always listening.

Thank you to all of my wonderful friends.

I would like to acknowledge Arvind, Jonathan, and Katie Chen for contributing to this thesis. I would also like to acknowledge ChatGPT for helping me fix grammar mistakes and wording issues on this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Segmenting data into distinguishable groups is crucial for making complex datasets manageable and facilitating the sensemaking process [1]. Grouping helps reduce cognitive load, enabling users to focus on uncovering higher-level insights and generating new hypotheses to test on the data [1], [2]. However, effectively identifying, creating, and leveraging groups presents several challenges. Grouping is a non-trivial task; as the scale and complexity of the data increases, grouping becomes increasingly time-consuming and laborious, with no guarantee of surfacing groups relevant to the domain of the data [3], [4]. Furthermore, forming groups requires prior understanding of the structure of the data in order to choose which fields to group as well as substantial knowledge on the domain of the data to understand what each field represents [4], [5].

At its core, grouping is a technique that simplifies complex data by segmenting it into manageable intervals or selectively filtering it based on specific fields and values. For a specific field in the data, tools like Tableau or Microsoft Excel can automatically segment a field into intervals, groups based on the field's intrinsic values, or groups based on different statistical properties [6]. For multiple fields in the data, visualizations are commonly used to identify multi-field groups [1]. This entails using various graphical mark types and visual channels to analyze the spatial positioning and features of data points in the visualizations to identify groups [4], [5]. To reduce the complexity of generating groups manually, different algorithms have emerged to automate the grouping process for data. These algorithms, such as the k-means algorithm, use mathematical optimization to efficiently partition data into distinct clusters based on various arbitrary similarity measures [7]. However, these techniques treat data as an abstract entity – failing to leverage the fact data serves as a representation of real world phenomena. Consequently, they cannot group data in meaningful ways, meaning users must possess a deep prior understanding of the data and its domain to make connections between the groups surfaced by these techniques and real-world insights.

In this paper, we introduce **SAGE**, a novel grouping technique that utilizes the context embedded within data to intelligently form domain-specific groups. This technique is operationalized through two key structures: *bins* and *highlights*. Bins segment a single field into contiguous, non-overlapping ranges to capture domain-specific aspects of that field; highlights create logical compositions of ranges across multiple fields to surface broader, interconnected groups within the data, thereby capturing the larger domain context. To create these structures, we employ the knowledge base and pattern recognition abilities of large language models (LLMs). Acknowledging people with disabilities as authorities on agency and trust in artificial intelligence contexts [8], we instantiate our technique within Olli, an open-source platform that translates data visualizations into accessible, keyboard-navigable hierarchical textual representations. This setup enables us to gather feedback from blind and low-vision (BLV) screen-reader users by providing them the opportunity to interact directly with our contribution. Through our iterative co-design process with our blind co-author Hajas, we identify the following design dimensions to ensure SAGE affords equitable interactions: *relevance*, which ensures the groups are directly applicable to their specific data domains for enhanced contextual understanding; *clarity*, which ensures the groups are presented clearly and simply to make complex data accessible and understandable to all users; and *agency*, which upholds the autonomy of users by enabling independent data interpretation to foster deeper personal engagement with the data.

To evaluate our contribution, we conducted 100-minute Zoom interviews with 15 blind and low-vision (BLV) participants using three prototypes. Our findings confirm that the three foundational design dimensions of SAGE—relevance, clarity, and agency—are important in enabling users to independently scaffold their initial sensemaking processes. However, the utility of our approach is contingent on the context of the data and is most effective with semantically meaningful visualizations. We conclude with a discussion on the broader applications of SAGE beyond data and visualizations to emphasize the importance of developing A.I. systems that augment human intelligence, rather than replace human decision-making.

Chapter 2

Related Work

2.1 Accessible Textual Data Representations of Visualizations

To enhance the accessibility of data visualizations through textual means, several representations have been developed. Textual descriptions, commonly integrated as alternative text in web visualizations [9], serve as one fundamental approach. Substantial progress has been made in automating the generation of textual descriptions [10]–[14] for visualizations; however, these automated descriptions remain static and limited to linear navigation, which restrict users from directly engaging with the underlying data, thereby hampering independent analysis [15], [16]. Systems such as ChartSense [17] improve on being able to access individual data points by convert chart images into structured data tables, but present their own set of challenges, such as cognitive overload [15], when navigating these textual data representations.

To address these limitations, a newer category of textual data representations has emerged, known as hierarchical textual data representations. Zong, Lee, Lundgard, et al. identify three design dimensions for hierarchical textual data representations: structure (the arrangement of the representation's individual components), navigation (the method by which the user transitions between components), and description (the content that is vocalized at each component) [15]. These representations introduce unique mechanisms for interacting with data visualizations through text; by leveraging structure, these representations afford users the flexibility to explore the visualization's data at various levels of granularity, thereby enhancing the overall accessibility and usability of data visualizations [15]. Notable systems that incorporate hierarchical textual data representations are Olli [18], Data Navigator [19], ASVG [20], VizAbility [21], and Chart Reader [22]. Jones et al. furthermore improve on these hierarchical textual data representations by introducing content tokens [23] to support customization of data descriptions across the following dimensions: presence, or the content that is communicated; verbosity, or the brevity and density of the content's delivery; ordering, or the arrangement of tokens used to present the content; and duration, or the length of time a specific customization is maintained.

So far, there is no textual data representation that captures the semantic meaning of the data it represents to effectively afford a user the ability to independently interact with all four levels of semantic content [16] as defined by Lungard et al.: elemental and encoded details like chart types and labels; statistical and relational data such as outliers and correlations; perceptual and cognitive insights into trends and exceptions; and contextual and domain-specific knowledge, enhancing understanding through tailored, in-depth explanations.

2.2 A.I. Tools for Accessibility

People with disabilities are some of the earliest adopters of artificial intelligence [8]. As such, they've been able to positions themselves as experts of agency and trust of these systems through their lived experiences. This section examines the landscape of A.I. tools designed for accessibility, particularly those built for blind and low-vision (BLV) users. We categorize accessibility-focused applications of A.I. into two distinct approaches: one aimed at making the world more accessible to users, and the other at enhancing the capabilities of the users themselves.

Various systems have been developed to enhance accessibility by enabling users to better interact with their environment. These technologies include automated tools for generating descriptions for blind and low-vision (BLV) users, covering videos [24], art [25], images on social media [26], hint-text [27], and charts [11], [13], [21], [28], [29]. To help users integrate A.I. into their everyday lives, these technologies have been deployed as phone applications such as Be My Eyes or SeeingAI [30], [31], screen-reader extensions such as PictureSmart for JAWS [32], and wearables [33]–[35].

Other systems designed to augment users' capabilities foster independent interaction and authoring. SPICA, for example, focuses on augmenting audio descriptions to enable BLV users to interactively explore video content [36]. GenAssist makes text-to-image generation accessible to enable BLV creators to use images to communicate with sighted audiences [37]. EasySnap and PortraitFramer assist BLV photographers in capturing better-composed photos [38]. BLVRUN provides BLV developers with streamlined, insightful error overviews to enhance their programming workflows [39]. PeopleLens uses more open-ended A.I. experiences to enrich social sensemaking [40].

These tools exemplify the potential of A.I. in fostering greater independence and accessibility for BLV users across various fields and use cases.

Chapter 3

Design Dimensions

Through an iterative co-design process led by co-authors Pedraza Pineros and Hajas, we identified three key design dimensions that underscore our commitment to enhancing human interaction through the use of large language models (LLMs). The central aim of these dimensions is to augment rather than replace the role of human judgment, ensuring that our technology supports users' independence and autonomy in drawing their own conclusions.

3.1 Relevance

Relevance refers to the ability to segment and group data in ways that are meaningful within its larger domain. This dimension also prioritizes adaptability, meaning our technique must be able to create bins and highlights across various contexts. Additionally, bins and highlights must be created in such a way where they align with the domain-specific frameworks used by experts analysts of the domain of the data. By connecting data to its broader domain context, relevance enhances a user's ability to understand patterns and relationships, leading to more informed decision-making and meaningful insights.

3.2 Clarity

Clarity refers to the ability to present information in an easily understandable manner. This design dimension focuses on simplifying and structuring data to reduce a user's cognitive load when exploring and interpreting information. By prioritizing clarity, we ensure that users can contextualize different bins and highlights without needing prior knowledge of the data. To achieve clarity, we recognize that we must structure the bins, highlights, and additional context in ways that best help a user orient themselves within the domain of the data.

3.3 Agency

Agency refers to empowering users to make their own interpretations of the data. This design dimension prioritizes the user's ability to critically engage with the information presented by our technique. We aim to augment, not replace, a user's capacity for meaningful data interpretation. This is particularly crucial within the accessibility domain, where users are often more sensitive to encroachments on their autonomy when interacting with the world [41]. To ensure agency, bins and highlights must be integrated into Olli transparently to allow users to easily evaluate the usefulness of a bin or highlight. This way, a user can retain control over their exploration process by being able to determine if the bin or highlight should be used in their interpretation of the data.

Chapter 4

Implementing SAGE in Olli

We've instantiated SAGE within Olli, an open source accessible visualization toolkit [18]. Our technique uses the data visualized by an Olli adapter to augment the hierarchical textual description created by Olli – also known as the Olli tree view – with *bins* and *highlights*. Through an iterative and collaborative co-design process, we've integrated the feedback of our blind co-author, Hajas, into our implementation throughout the course of the project. With his feedback, we've converged to two prototypes, the *independent exploration* prototype and the *guided exploration* prototype, each of which prioritize different design dimensions. In this section, we've elaborated the process and design considerations of our implementation.

4.1 Bins

Bins are contiguous, non-overlapping ranges that segment a single field into groups. An example of how we define a bin can be seen in Figure 4.1. Each bin is composed of three fields: `bin_name`, `reasoning`, and `pred`. The `bin_name` field serves as a short domain-specific label, the `reasoning` field connects the data in the bin to domain-specific knowledge, and the `pred` field is a Vega-Lite predicate that defines what data belongs in the bin. With respect to our first design dimension, relevance, bins use domain-specific labels and additional domain-specific knowledge to ensure users can use a bin to relate the data within the bin to broader

```
{
    "bin_name": "Short",
    "reasoning": "Flipper lengths in this range might be
       associated with slower swimming but greater
       maneuverability, typically found in species
       navigating through densely packed ice or prey-rich
       waters where agility trumps speed",
    "pred": {
        "field": "Flipper Length (mm)",
        "lte": 190
    }
}
```

Figure 4.1: An example bin in SAGE. A bin is instantiated as a JSON object with fields "bin_name", "reasoning", and "pred". This particular example is the "Short" bin for the field "Flipper Length (mm)." It segments flipper lengths that are less than or equal to 190 mm as short. The reasoning bridges the bin to a larger set of a field's domain-specific implications and characteristics.

patterns and relationships in the domain. For our second design dimension, clarity, bins use short labels and reasoning to reduce cognitive load. Lastly, for our third design dimension, agency, to make it clear what data belongs to a certain bin, we include a predicate to show how we've defined the boundaries for a bin. This enables users with the autonomy to evaluate the bin produced by our technique.

To implement bins, we focused segmenting a field within it's own domain to surface groups relevant to the specific field. To do this, we pre-processed the overall dataset to filter for the field's data. Then, we use a large language model (LLM), specifically Open AI's `gpt-4-turbo-preview` [42], to query for a field's bins. We experimented with various prompting techniques such as zero-shot prompting [43], chain-of-thought prompting [44], and few-shot prompting [45] to fine-tune the LLM to create useful bins. After multiple weeks of experimentation, we were able to converge on a set of prompts, including a system prompt and three user prompts.

Our system prompt encourages the large language model to first understand the name of the field and it's values to create a bin. We've included a snippet of the prompts used to create the bins in Figure 4.2. We feed the pre-processed data to the large language model as a JSON string in our first user prompt. Then, we ask the large language model to create the bins by providing the field's name and a brief example on how it might segment a temporal field. Lastly, we provide a few examples of how to format the bins given either quantitative or nominal data since the Vega-Lite predicate format uses a different grammar for both types of data.

```
{role: "system", content: ‘Please analyze the uploaded
   dataset.‘}
```

```
{role: "user", content: ‘Here is the data we’ll be
   analyzing: ${JSON.stringify(dataset)}’},
```

```
{role: "user", content: ‘For the ${field} field, come up
   with a meaningful, non-obvious way to partition the
   field. Feel free to use outside knowledge. It’s ok to
   have overlap. For example, if you have the field "Year",
    you could make bins such as:
    {
        "bins": ["PreIndustrial Era", "Early Industrial Era
           ", "Wheat Boom",
        "Agricultural Revolution" ]
    },
’}
```

```
{role: "user", content: ‘For each bin, use the Vega-Lite
   predicate schema to create one Vega-lite predicate for
   each field in the bin. Each predicate must include the
   field and only ONE property to specify what data from
   that field belongs in the bin: equal, range, lt (less
   than), lte (less than or equal), gt (greater than), gte
   (greater than or equal), or oneOf. Make sure to give a
   full response in a JSON format. Do not change the names
   of the fields in your answer.’}
```

Figure 4.2: The prompts used in our technique in order to segment a field into bins. We use one system prompt and three user prompts to create bins for a field. The last prompt's examples are omitted for brevity, but usually include an example bin output for both quantitative and nominal data.

4.2 Highlights

```
{
    "bin_name": "High-Performance Sports Vehicles",
    "reasoning": "This category encapsulates vehicles
       engineered for maximum performance and power. They
       exhibit high horsepower, usually have a lower fuel
       mileage (miles per gallon), and often utilize 8
       cylinders to deliver the acceleration and towing
       capability expected from sports cars and luxury
       models",
    "pred":
        {
            "and": [
                {
                    "field": "Horsepower",
                    "gt": 200
                },
                {
                    "field": "Miles_per_Gallon",
                    "lte": 20
                },
                {
                    "field": "Cylinders",
                    "equal": 8
                }
            ]
        }
}
```

Figure 4.3: An example highlight in SAGE. A highlight is instantiated as a JSON with fields "bin_name", "reasoning", and "pred". This particular example is the " High-Performance Sports Vehicles" highlight for the fields "Horsepower," "Miles_per_Gallon," and "Cylinders." It groups vehicles with horsepower greater than 200, miles per gallon less than or equal to 20, and 8 cylinders as high-performance sports vehicles. The reasoning bridges the highlight to the larger domain of the entire dataset.

Highlights are multi-field intersections of ranges that surface broader groups in the data. An example of how we define a highlight can be seen in Figure 4.3. Each highlight is composed of three fields: `bin_name`, `reasoning`, and `pred`. The `bin_name` field serves as a short domain-specific label, the `reasoning` field connects the highlight to domain-specific knowledge, and the `pred` field (predicate) is a logical composition of Vega-Lite predicates that define what data from each field belongs in the highlight. With respect to our first design dimension, relevance, highlights use domain-specific labels and additional domain-specific knowledge to ensure users can use a highlight to relate multi-field intersections of data to broader patterns and relationships in the domain. For our second design dimension, clarity, highlights use short labels and reasoning to reduce cognitive load. Lastly, for our third design dimension, to make it clear what fields and their respective data belongs to a certain highlight, we include a predicate composition to show how we've defined the boundaries for a highlight. This enables users with the agency to evaluate our technique.

To implement highlights, we focused on identifying multi-field groups in the data to surface broader groups. To do this, we leveraged the bins created for each field in the dataset to form a highlight. We use a large language model (LLM), specifically Open AI's `gpt-4-turbo-preview` [42], to prompt for these larger multi-field highlights. Our system prompt prepares the large language model to utilize the bins for each field to create the highlights, as seen in Figure 4.4. We first feed the large language model with the JSON object representations of the bins for each field. Then, we and asked the large language model to create "meta-bins" that can overlap and include multiple fields to build the highlights. Lastly, we provide a few examples of how to format the highlights using different logical compositions of Vega-Lite predicate formats for both quantitative and nominal data.

4.3 Prototypes

4.3.1 Olli Prototype

The Olli prototype, seen in Figure 4.5, acts as a control to teach users the structure and how to navigate the hierarchical textual data representation used in the Olli tree view. Additionally, this prototype allows us to study our users sensemaking process with state-of-

```
{role: "system", content: ‘You will analyze bins for
   different fields to create meaningful meta-bins.’}
```

```
{role: "user", content: ‘Here are the bins we’ll be
   analyzing, separated by field: ${JSON.stringify(bins)
   }’},
```

```
{role: "user", content: ‘Come up with a meaningful, non-
   obvious meta-bins of the bins. Feel free to use outside
   knowledge. It’s ok to have multiple fields and overlaps
   across meta-bins’}
```

```
{role: "user", content: ‘Each bin uses the Vega-Lite
   predicate schema to create one Vega-lite predicate for
   each field in the bin. To create meta-bins, you must
   format your answer using predicate compositions. Each
   predicate composition must include the predicates for
   different fields and only ONE property to specify what
   data belongs in that meta-bin: "and", "not", or "or".
   Make sure to give a full response in a JSON format. Do
   not change the names of the fields in your answer.’}
```

Figure 4.4: The prompts used in our technique in order to identify highlights in a dataset. We use one system prompt and three user prompts to create highlights for the fields in a dataset. The last prompt’s examples are omitted for brevity, but include an example highlight output for both quantitative and nominal data.

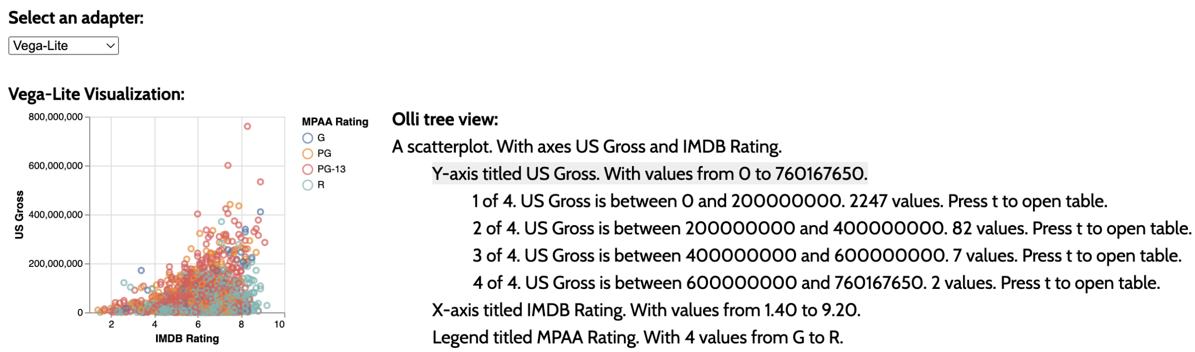

Figure 4.5: On the left, we show the visualization used in our Olli prototype. On the right, we show the un-augmented tree view for the Olli prototype.

the-art techniques that don't integrate SAGE. We've simplified the information displayed for each interval and node in this version to reduce cognitive load and expedite the learning process. This prototype does not include additional statistical information or data type details for fields and their intervals.

4.3.2 Independent Exploration Prototype

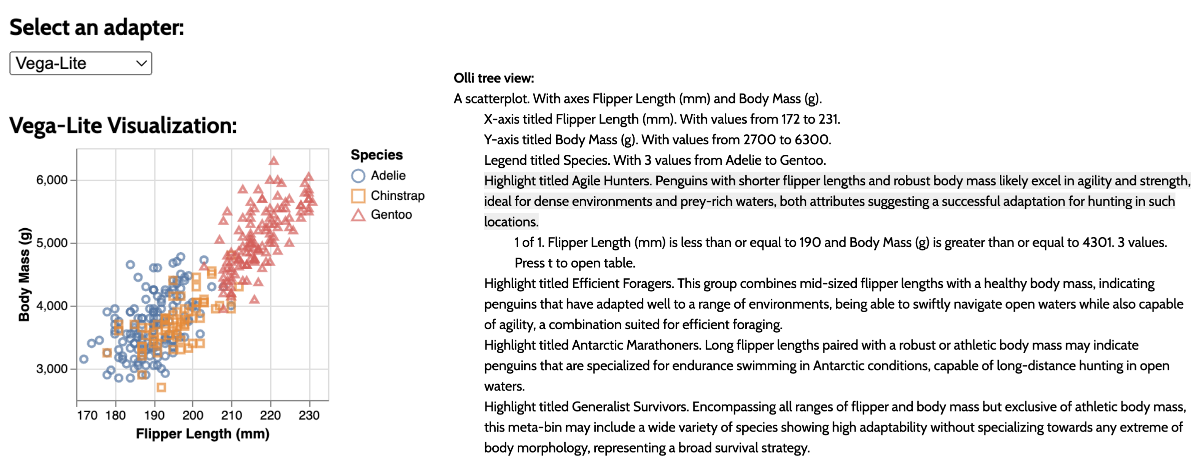

Figure 4.6: On the left, we show the visualization used in our independent exploration prototype. On the right, we show the augmented tree view for the independent exploration prototype. The highlights for the fields "Flipper Length (mm)," "Body Mass (g)," and "Species" are appended to the bottom of the Olli tree view's encodings level.

The independent exploration prototype was designed to prioritize user agency in the data exploration and interpretation process. As shown in Figure 4.6, it integrates information from

SAGE as augmentations to the original Olli structure. This prototype deprioritizes highlights by appending them to the end of the encodings level of the Olli tree view to allow users to first understand the data on their own before receiving additional context. The original Olli structure is preserved, with a highlight's predicate information included as a child interval node. To help reduce cognitive load and make the domain-specific information easier to understand, the prototype incorporates reasoning within the initial description of highlights to address the potential complexity and jargon that may arise from multi-field intersections. Additionally, this prototype uses the bins as additional labels to the original Olli intervals for a field to provide "a different lens on certain sections of the data as a preset," as described by our collaborator Hajas, without completely shaping the user's initial exploration and interpretation of the data. Lastly, to prevent information overload, the prototype hides the reasoning for each bin within the table view of an interval as seen in Figure 4.7, as the bin names are usually descriptive enough on their own.

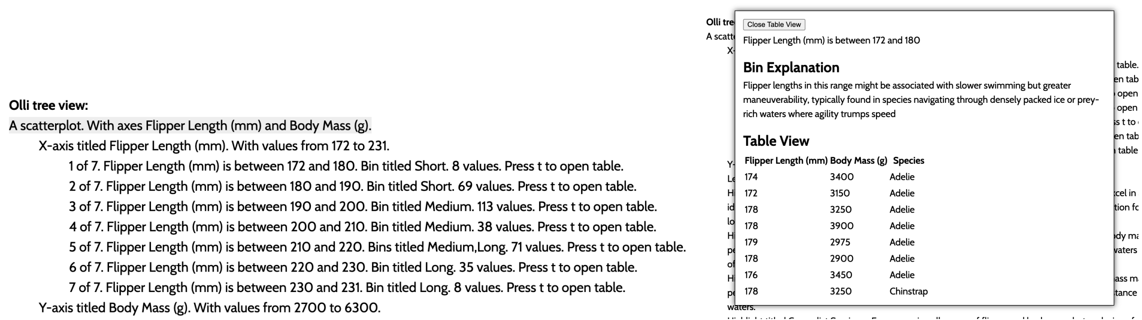

Figure 4.7: On the left, the bins for the field "Flipper Length (mm)" in the independent exploration prototype. On the right, we show how we've included the reasoning for the short bin within the interval's table view for this prototype.

4.3.3 Guided Exploration Prototype

The guided exploration prototype was designed to prioritize learning for users during their initial data interpretation process by modifying the Olli structure to use domain-specific intervals, as seen in Figure 4.8. It enhances the user's experience by providing an overview of the data domain at the top of the Olli encodings level. All highlights are grouped under

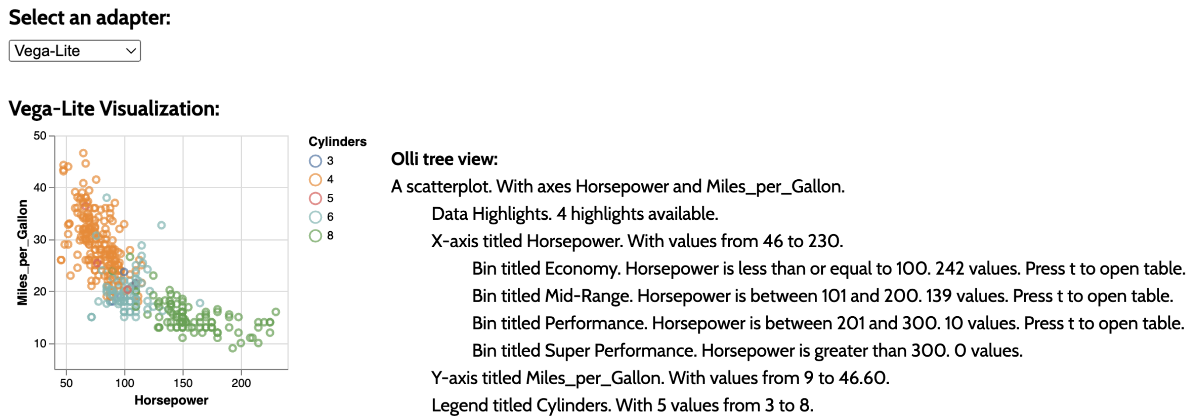

Figure 4.8: On the left, we show the visualization used in our guided exploration prototype. On the right, we show the augmented tree view for the guided exploration prototype. The highlights for the fields "Horsepower," "Miles_per_gallon" and "Cylinders" are appended to the top of the Olli tree view's encodings level under a "Data Highlights" node. The bins are used to structure the intervals of a field in the guided exploration prototype.

one node at the top, allowing users to quickly skip them if they already understand the data or prefer not to have an overview. This feature also makes it easier to find the axes and legend without hearing too much extra information. The prototype leverages context to structure a field's intervals with bins, providing a more intuitive scaffolding for users who may not have a prior understanding of the data's structure or domain. The highlights' predicate information is included in the description of each highlight to reduce keystrokes and present boundaries more clearly, as seen in Figure 4.9. To reduce cognitive load, the prototype incorporates reasoning within the initial description of highlights, addressing potential complexity and jargon associated with multi-field intersections. Lastly, the prototype hides the reasoning for each bin within the table view of a bin, since the bin names are usually descriptive enough on their own, preventing information overload.

Olli tree view:

A scatterplot. With axes Horsepower and Miles_per_Gallon.

- Data Highlights. 4 highlights available.
  - Highlight titled Eco-Friendly Compact Vehicles. Vehicles that are designed with a focus on fuel efficiency and minimal environmental impact, usually embodying characteristics from economy bins across horsepower, miles per gallon, and cylinder count. These vehicles are often small to medium in size, making them suitable for city driving and offering lower maintenance costs. Horsepower is less than or equal to 100 and Miles_per_Gallon is greater than or equal to 35 and Cylinders equals 4. Press t to open table.
  - Highlight titled High-Performance Sports Vehicles. This category encapsulates vehicles engineered for maximum performance and power. They exhibit high horsepower, usually have a lower fuel mileage (miles per gallon), and often utilize 8 cylinders to deliver the acceleration and towing capability expected from sports cars and luxury models. Horsepower is greater than 200 and Miles_per_Gallon is less than or equal to 20 and Cylinders equals 8. Press t to open table.
  - Highlight titled Balanced Family Vehicles. Vehicles that offer a compromise between power and efficiency, aimed at satisfying the requirements of daily commutes and occasional heavy-duty use. Generally, these are family sedans, crossovers, and smaller trucks with a moderate level of horsepower, average fuel economy, and typically 6 cylinders to ensure both performance for varied driving needs and acceptable fuel efficiency. Horsepower is between 101 and 200 and Miles_per_Gallon is between 21 and 34 and Cylinders equals 6. Press t to open table.
  - Highlight titled Unique Engine Choices. Vehicles that stand out due to unconventional choices in engine cylinder counts, representing manufacturers' efforts to balance power, efficiency, and compactness. This group might intersect with various performance, economy, and fuel mileage categories but is uniquely characterized by their uncommon cylinder numbers. Cylinders is equal to one of 3,5. Press t to open table.
- X-axis titled Horsepower. With values from 46 to 230.
- Y-axis titled Miles_per_Gallon. With values from 9 to 46.60.
- Legend titled Cylinders. With 5 values from 3 to 8.

Figure 4.9: The highlights for the fields "Horsepower," "Miles_per_gallon" and "Cylinders" in the guided exploration prototype.

Chapter 5

Evaluation

To evaluate the SAGE, we conducted 100-minute Zoom interviews with 15 blind and low-vision participants. The goal of this evaluation was to identify how our technique, SAGE, can impact user exploration and analysis by studying the differences in how users might utilize bins in contrast to highlights and user preferences regarding the integration of these groupings into a hierarchical textual data representation.

5.1 Study Design

Each interview had Pedraza Pineros act as a guide through the different tasks and prototypes while Chen, Zong, and Patterson served as notetakers.

5.1.1 Study Setup

. Each interview lasted 100 minutes and had users explore three prototypes. We began each interview with a brief 10 minute introduction to learn more about the user's interactions with data, data visualizations, and large language models. Then, we presented users our first prototype for 15 minutes to familiarize them with Olli and how to navigate its hierarchical textual data representation. After a 5 minute break, we had users explore the first and second prototype, each for 20 minutes, whilst describing their sensemaking process when completing a series of intra-field and inter-field data exploration and interpretation tasks. Afterwards, we had users complete the prototype's corresponding Likert survey. In the final 15 minutes, we collected users' overall takeaways and them complete a post-study survey.

5.1.2 Prototypes

. Each prototype tasked users with exploring a scatterplot visualization of a dataset to facilitate skill transfer. Scatterplots were selected for their capacity to represent three distinct fields: two quantitative and one nominal. Each dataset, drawn from various real-world domains, encompassed a mix of nominal, ordinal, and quantitative fields to provide users with a diverse range of contexts to explore to avoid having users "learn" the data. The following three prototypes capture different aspects of our design dimensions:

- **Olli Prototype**. This simplified version of Olli is used a control to understand how state-of-the-art hierarchical textual data representations are used by BLV users to contextualize data. This prototype uses the Movies dataset and "US Gross," "IMDB Ratings," and "MPAA Ratings" fields.

- **Independent Exploration Prototype**. This prototype prioritizes user agency to interpret the data before revealing additional context. It conserves the original Olli structure and treats bins as annotations to a quantitative field's intervals or a nominal field's categories. It deprioritizes highlights by appending them to the end of the encodings level of the Olli tree view. This prototype uses the Palmer penguins dataset [46] and the "Flipper Length (mm)," "Body mass (g)," and "Species" fields.

- **Guided Exploration Prototype**. This prototype focuses on leveraging context as a scaffolding to user interpretation by using the bins to structure a quantitative field's intervals or a nominal field's categories. It highlights the highlights by grouping them into a larger group and appending this group to the beginning of the encodings level of the Olli tree view. This prototype uses the Cars dataset and the "Horsepower," "Miles per gallon," and "Cylinders" fields.

5.1.3 Participants

. We recruited 15 blind and low-vision participants by sending a participant call to a blind programmers' community mailing list and reaching out to BLV previous participants in our local community. Each participant was compensated $60 for 100 minutes. To protect user privacy, we've included anonymized and aggregated demographic information to provide background context on our users [47] whilst recognizing our users' identities extend far beyond demographic attributes. The majority (67%) of our participants were totally blind (n=10), while 27% identified as low-vision with some light perception (n=4) and 7% of participants did not respond (n=1). More than half (60%) of our participants have been blind since birth. 53% of participants were JAWS users (n=8) and 47% were NVDA users (n=7), consistent with screen reader statistics [48]. Demographically, 80% of participants use he/him pronouns (n=12) and 20% of participants use she/her pronouns (n=3). Participants were based across multiple continents, including North America, Europe, and Asia. Participants self-reported their ethnicities (Asian, Black/African, Caucasian/white, Hispanic/Latinx, Other), represented a diverse range of ages (20–50+), and and had a variety of educational backgrounds (high school through to undergraduate and graduate school). With the exception of one participants that did not respond, almost all participants (n=14) self-reported as slightly, somewhat, or moderately familiar with statistical concepts. 13 participants self-reported as slightly, somewhat, or moderately familiar with data visualization methods and 2 participants did not respond. Participants reported a high variety of frequency interacting with data or visualizations, from 1-2 times/year to 3 or more times/week. Most participants (n=9) reported using data analysis tools or visualizations either outside of their professional work or sometimes, but 2 participants reported data analysis tools or visualizations being an important part of their workflow. 4 participants reported rarely use

Table 5.1: Rating scores for each prototype (Independent Exploration Prototype, Guided Exploration Prototype) on a five point Likert scale. Median scores are shown in boldface, averages in brackets, standard deviations in parentheses.

Prompt: After understanding how the [prototype] works...	Independent Exploration	Guided Exploration
How understandable were the data highlights in the prototype?	**4** [4.13] (0.64)	**5** [4.53] (0.64)
How understandable were the bins in the prototype?	**4** [4.00] (1.20)	**5** [4.67] (0.49)
How much influence did the data highlights have on your interpretation of the data?	**4** [3.60] (1.30)	**4** [3.33] (1.45)
How much influence did the bins have on your interpretation of the data?	**3** [3.13] (1.25)	**3** [3.00] (1.36)
How much additional context did the data highlights provide about the data?	**4** [3.93] (1.10)	**4** [3.60] (1.24)
How much additional context did the bins provide about the data?	**4** [3.47] (0.92)	**4** [3.47] (0.83)
How effectively did the prototype explain and justify its data highlights and bins?	**3** [2.93] (0.96)	**4** [3.73] (0.88)
How often did you feel the need to double-check the accuracy of the data highlights and bins?	**2** [2.33] (1.18)	**2** [2.47] (1.06)

data analysis tools.

5.2 Quantitative Results

5.2.1 Likert Scales

To evaluate how well the bins and highlights followed each of our design dimensions, we designed a Likert survey to understand participants' preferences across the independent exploration and guided exploration prototypes. Participants responded on a five point scale where 1 = Very Difficult to Understand/No Influence/No Additional Context/Not Effective/Never Felt the Need and 5 = Very Easy to Understand/Significant Influence/Extensive Additional Context/Extremely Effective/Always Felt the Need as seen in Table 5.1.

The Likert scale scores provide insights into the user experiences and preferences for each prototype. The guided exploration prototype consistently scored higher in terms of understandability for both highlights and bins, indicating that users found this prototype more intuitive and easier to use. Despite this, the influence of highlights and bins on data interpretation was similar between the two prototypes, suggesting that while the guided exploration prototype was easier to understand, it did not significantly change how users interpreted the data compared to the independent exploration prototype. Additionally, both prototypes provided a comparable amount of additional context about the data, as indicated by similar scores in those categories. However, the guided exploration prototype was more effective in explaining and justifying its data highlights and bins, as reflected by higher scores in that category. Users also felt a slightly greater need to double-check the accuracy of data highlights and bins in the guided exploration prototype, though this difference was minimal. These results suggest that while the guided exploration prototype offers better usability and clarity, it does not necessarily reduce the need for users to verify the information provided.

5.2.2 Study Limitations

The current hierarchical textual data representation structure in Olli makes data and data visualizations accessible but requires time to learn, especially for those with prior experience in tactile graphics. This learning curve can affect the initial user experience and potentially bias the results towards those more familiar with the interface. Additionally, because we tested the prototypes in the context of a study and not a real-world scenario, users may not have felt as inclined to double-check the accuracy of the highlights and bins since it was a low risk environment.

5.3 Qualitative Results

Once interviews were complete, Pedraza Pineros, Chen, and Zong performed open coding on the notes taken from each interview. We then jointly reviewed and synthesized other's codes into the following themes.

5.3.1 Traditional large language model interactions are tedious but are “better than nothing”

Traditional question-and-answer interactions with LLMs are described by participants as “time consuming” (P9), “tedious” (P4), “not seamless” (P15), “complicated” (P10), and “a lot of effort” (P15). For participants, part of this frustration stems from the effort it takes to pose a question that guarantees them to receive the output they wanted without having to repeat the process over several trials. However, participants expressed they persisted through these frustrations because it’s “better than having nothing” (P15) or “than not having any information” (P7).

5.3.2 Verification is an essential part of large language model interactions.

Large language models are becoming increasingly more integrated into participants’ lives and workflows for tasks such as summarization, programming, ideation, learning, and as visual aids through tools like Be My Eyes AI [30]. To participants, LLMs serve as tools that enable them to connect with the world around them. One participant shared, “I mainly use it to interpret data and images; to interact with other people. It made a really big change in how I interact with screenshots and images” (P9). And although participants expressed they “tend to want to trust the A.I.,” (P8) due to the frequency of erroneous output, participants see verification as an essential part of their interactions with LLMs: “Since we know it’s a language model and everyone says it sometimes give erroneous information, I verify sometimes with Google or my own domain knowledge” (P3).

5.3.3 Contextualizing data is an important, but difficult step in the sensemaking process.

When encountering unfamiliar data, participants relied on two main techniques: a joint exploration and association approach, and an external research approach. P2 described their process as: "First, I get the domain of the data, understand what the columns are, what properties they represent, and the types of data. Then I dig down and ask questions, and depending on the questions, I'll try to find answers in different ways." In the first prototype, few participants recognized what the legend titled "MPAA Rating" meant. To understand it, participants explored the legend's data, and once they heard the sequence "G, PG, PG-13, R" as the categories, as P3 described, "I didn't know exactly what MPAA rating was, but through association, I was able to make it out" and realize it referred to a movie's censorship rating. When asked how often she used this technique, P10 responded, "A lot - it happens a lot when I have to really dig and search to find out what the context of the data is." In the cases where our participants could not make a personal association to the data, P1 and P13 expressed that they would most likely search on Google for the field names.

After understanding what "MPAA Rating" meant, participants inferred that the domain of the data was movies, which grounded their sensemaking process around this context. This emphasizes how, during exploratory data analysis, BLV users contextualize data by understanding its domain. As P10 described, "A lot of times people overthink about describing a picture - they get into describing but forget why we are talking about it in the first place. I really want context." However, contextualization often involves using their own domain knowledge, which may discourage users from exploring unfamiliar domains.

5.3.4 Agency is an important aspect of interpreting data for users.

Users expressed a strong desire for autonomy in their data analysis process, which current techniques for contextualizing data and visualizations often fail to provide. P14 highlighted this issue, noting that when presented with a bar chart online, the brief descriptions provided often dictate what users should see: "The problem with those interpretations is they tell you what they think you should see, but there might be things in that data representation that you haven't even thought of." This prescriptive approach leaves out potential perspectives and insights that users might have had if they were allowed to explore the data on their own terms. P14 further elaborated, stating, "That data is just data - their caption is their interpretation of that data." This underscores the urge among BLV users to have control over their interpretation of data, rather than relying on predefined captions or summaries that might not align with their specific needs or questions. The feedback from participants indicates a significant gap in current data visualization techniques: the need for tools that enable BLV users to independently navigate and interpret data, thus fostering a sense of ownership and confidence in their analytical processes.

5.3.5 Bins are a useful scaffold for structuring the initial sensemaking process.

Bins play an essential role in contextualizing data by providing meaningful groupings that tie the data to real-world concepts. As P3 described, bins are "very helpful because it gives us the outline of how to categorize it ... we can easily sort out the performance." They provided useful information that users might not figure out independently: "Without it telling me I would have no idea that this range is economy/mid-range" (P8). As P5 described "The bins and highlights can simplify the information to make more connections, in more of a layman's terms." A bin's reasoning, also known as it's explanation in the prototypes, also helped simplify the process of navigating complex data. Participants enjoyed having a bin's reasoning hidden in the table view, suggesting that bin explanations provided a starting point for deeper exploration: "this is interesting, if you want clarification, you can dig into it" (P3). Furthermore, bin explanations helped non-experts feel more confident navigating the data. "In this situation, it was useful for me because I'm not really a specialist in penguins. This helps me understand" (P6). P5 expressed increased confidence, stating, "it gives me a sense of confidence now I have access to information that I didn't have before". However, while bin explanations were helpful for non-experts, domain experts might find them less useful. "I think they are because when you are tackling a subject for the first time...if you are tackling a subject you already know about, you may not need them" (P14). Overall, bin explanations significantly enhance users' ability to understand and work with data, especially for those who are not domain experts. They provide meaningful context and simplify complex information, making data more accessible and interpretable.

5.3.6 Highlights provide an overview of the data's domain and structure.

Participants found that highlights provided a valuable contextual overview, helping align their expectations about what to explore within the data. P1 noted, "Highlights were VERY useful - they were well stated and talked about something that I could have some understanding of." This was echoed by P2: "Yeah, they provide information that is not encoded in the data itself." This is because highlights helped participants make connections with the data's context: "Definitely knowing the highlights makes you more interested in the data and opens your mind - it makes you understand the data better" (P9). P13 appreciated the clarity provided, saying, "When I saw body mass, I didn't know what it was talking about, but now I can definitely see there are different ranges." Users also emphasized the importance of verifying the information in these highlights, often using their own understanding, digging into the table data, cross-checking with predicate information, or consulting external resources like Google: "I might go to Google and re-read on these things, but at least I will take the first impressions here" (P3). This verification is crucial, especially in professional contexts where accuracy is paramount. Lastly, highlights also enhanced user agency, helping them generate relevant questions about the data, similar to a data scientist's approach. P9 noted, "When having these highlights, it's sort of a summary of what's happening and based on that you can generate questions." P11 added highlights "Give me some ideas about what questions I can ask about it."

5.3.7 The independent exploration prototype offers minimal guidance, which benefits users preferring less structure but increases cognitive overload and leads to overlooked highlights.

The independent exploration prototype has both strengths and limitations. One key limitation is that it forces users to memorize which ranges belong in which bin, rather than presenting the whole chunk together, increasing cognitive overload. P1 highlighted that this approach required a lot of time to remember which intervals belonged to which bin. Similarly, P10 expressed a preference for seeing the entire bin rather than annotations on each interval, saying, "I had to go back and check what was classified as short." However, once users became familiar with the structure, they found it useful for subtly adding more information without heavily structuring their interpretation. P3 appreciated the extra information, noting, "You're giving me a lot of information I would otherwise have to dig for." Despite this, the ungrouped highlights at the end were often overlooked by users, as P9 mentioned, "I did not notice the highlight thing."

5.3.8 Participants preferred the guided exploration prototype for its intuitive data grouping and upfront highlights, which made it easier to understand and explore despite potential bias.

Participants found the guided exploration prototype was more intuitive since it split the data into meaningful and interpretable groups, rather than using the original Olli intervals: "This is quick and great – who cares what we are counting by?" (P10). Additionally, by putting the highlights first, participants found it easier to ground their understanding of what to explore in the data. As P3 explained, "It gives an overall picture of what to expect before I dig into the data." However, some participants felt that this approach could bias their perception. As P2 noted, "By having them first, you are subconsciously changing your perception." Grouping the highlights together at the beginning was also appreciated for its organization and efficiency, allowing users to skip over extra information if they didn't want it. P6 preferred this layout, stating, "It's easier and reduces keypresses needed to get info." The guided exploration prototype also made it easier for users to surface their own patterns within the data. P4 stated, "The thing I like the most is that I can find a pattern right away unlike the previous model." In general, the guided exploration prototype was the preferred choice for participants – "I like this better than the last one" (P3) – because participants found it easier to understand, noting, "I don't feel dumb looking at this data" (P4).

Chapter 6

Limitations

Our study revealed limitations in A.I. interpretation accuracy, lack of customization, and the need for equal information access to ensure fairness and effective collaboration between BLV and sighted users. Participants noted the challenges posed by A.I. interpretation and non-determinism. While A.I. can expedite conclusions, it may not always provide accurate evaluations. P10 observed, "A.I. is wonderful because it helps you jump to conclusions faster, but it can NOT give the right evaluation." The context dependency of A.I. outputs and their variability upon refresh raised concerns about reliability, with P6 stating, "You have to take what is good from it, but with a grain of salt." Participants emphasized the need to verify A.I.-generated insights independently, as P6 mentioned, "I would always double-check it if it were really important."

The lack of customization in creating groupings was another significant limitation. Participants expressed a desire for more control over the data presentation and the ability to tailor insights to their specific needs. P6 highlighted the need for personal relevance, saying, "The values don't correspond to my knowledge because I use different units, so it's hard for me to understand." P4 suggested that user options across prototypes would enhance usability, noting, "People try to sort and filter things they want, so certain categorizations will not hurt but the options can be given to certain users on what to do with it."

There is an ongoing debate about whether BLV tools should provide the same level of information to sighted users to ensure fairness and collaboration. Participants were divided on this issue, with some advocating for parity in information access. P10 captured this sentiment, stating, "Is it helpful information? Heck yeah! Is it fair? No...you would need to make sure the sighted person gets that too." The potential for BLV-specific tools to create an imbalance in collaborative settings was also highlighted, with P10 expressing concerns about different levels of understanding: "If you're not getting that write-up, I shouldn't be getting that much more either."

Chapter 7

Discussion and Future Work

Our technique, SAGE, has shown significant promise in helping users, especially those who are blind or have low vision (BLV), access more domain-specific information. Participants expressed enthusiasm about its potential. For example, P9 remarked, "It's all completely new to me, and I had in my mind something similar but didn't know there was a way to get a textual summary of data and move through it easily." Importantly, SAGE has the potential to go beyond the accessibility problem space. Sighted users can also benefit from the detailed insights provided. P13 noted, "Anyone can find what's going on in this data—regardless if you know about the horsepower, mileage, etc." This sentiment was echoed by P5, who stated, "this would be beneficial even for individuals without visual impairment." This inclusivity ensures that all users can access the same level of detailed information, enhancing collaboration and understanding. As P14 mentioned, "Whenever you can put sighted and blind individuals on the same playing field and equal footing, you have accomplished something really marvelous."

The applicability of highlights as a concept extends beyond the current prototypes and can be beneficial in various domains. For instance, P5 noted that the bins and highlights could simplify information in business contexts, making products and services more user-friendly. P3 mentioned the potential for these tools in tracking COVID-19 trends, business and employment trends, and even e-commerce, stating, "Labeling could be useful outside of this context, particularly in shopping for big purchases like fridges or washing machines." Other potential applications include weather tracking, network troubleshooting, and educational contexts. P4 emphasized the usefulness of highlights in troubleshooting, stating, "Highlighting is a big issue for blind users because it could be working with hundreds of rows of data." Similarly, P5 suggested that such tools could enhance experiences in social spaces like museums and botanical gardens by providing detailed visualizations of exhibits. Participants also expressed interest in using data highlights with different types of data, such as time series and seasonal data. P9 stated, "I would be interested to see how we can use it with different kinds of data such as time series and seasonality, and how it can give us these unique insights."

Future research should explore the scalability and customization of data highlights to ensure they meet diverse user needs. More work needs to be done on exploring the usefulness of our technique in real world applications: "I guess I never thought of using an LLM this way—it's really interesting. I'm curious now how effective this would be on real-world data that a real person needs to digest" (P11).

In summary, SAGE has transformative potential for making data more accessible and actionable for all users. By continuing to refine and expand this technology, we can create a more inclusive data exploration experience that benefits a broad range of users in various contexts.

References

[1] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko, "Understanding and characterizing insights: How do people gain insights using information visualization?" en, in *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*, Florence Italy: ACM, Apr. 2008, pp. 1–6, ISBN: 978-1-60558-016-6. DOI: 10.1145/1377966.1377971. [Online]. Available: https://dl.acm.org/doi/10.1145/1377966.1377971 (visited on 05/11/2024).

[2] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," en, *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, Sep. 1999, ISSN: 0360-0300, 1557-7341. DOI: 10.1145/331499.331504. [Online]. Available: https://dl.acm.org/doi/10.1145/331499.331504 (visited on 05/11/2024).

[3] D. A. Szafir, S. Haroz, M. Gleicher, and S. Franconeri, "Four types of ensemble coding in data visualizations," *Journal of Vision*, vol. 16, no. 5, p. 11, Mar. 2016, ISSN: 1534-7362. DOI: 10.1167/16.5.11. [Online]. Available: https://doi.org/10.1167/16.5.11 (visited on 05/11/2024).

[4] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. Van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver, "Data, Information, and Knowledge in Visualization," en, *IEEE Computer Graphics and Applications*, vol. 29, no. 1, pp. 12–19, Jan. 2009, ISSN: 0272-1716. DOI: 10.1109/MCG.2009.6. [Online]. Available: http://ieeexplore.ieee.org/document/4736452/ (visited on 05/11/2024).

[5] Y. Luo, X. Qin, N. Tang, and G. Li, "DeepEye: Towards Automatic Data Visualization," en, in *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, Paris: IEEE, Apr. 2018, pp. 101–112, ISBN: 978-1-5386-5520-7. DOI: 10.1109/ICDE.2018.00019. [Online]. Available: https://ieeexplore.ieee.org/document/8509240/ (visited on 05/11/2024).

[6] A. Sarvghad, B. Saket, A. Endert, and N. Weibel, "Embedded Merge & Split: Visual Adjustment of Data Grouping," en, *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 800–809, Jan. 2019, ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2018.2865075. [Online]. Available: https://ieeexplore.ieee.org/document/8440855/ (visited on 05/11/2024).

[7] A. E. Ezugwu, A. K. Shukla, M. B. Agbaje, O. N. Oyelade, A. José-García, and J. O. Agushaka, "Automatic clustering algorithms: A systematic review and bibliometric analysis of relevant literature," en, *Neural Computing and Applications*, vol. 33, no. 11, pp. 6247–6306, Jun. 2021, ISSN: 1433-3058. DOI: 10.1007/s00521-020-05395-4. [Online]. Available: https://doi.org/10.1007/s00521-020-05395-4 (visited on 05/11/2024).

[8] J. P. Bigham and P. Carrington, "Learning from the Front: People with Disabilities as Early Adopters of AI," en,

[9] *Web Content Accessibility Guidelines (WCAG) 2.2*, 2023. [Online]. Available: https://www.w3.org/TR/WCAG22/ (visited on 05/07/2024).

[10] G. Kim, J. Kim, and Y.-S. Kim, ""Explain What a Treemap is": Exploratory Investigation of Strategies for Explaining Unfamiliar Chart to Blind and Low Vision Users," en, in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg Germany: ACM, Apr. 2023, pp. 1–13, ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3581139. [Online]. Available: https://dl.acm.org/doi/10.1145/3544548.3581139 (visited on 05/07/2024).

[11] E. Kim, C. Onweller, and K. F. McCoy, "Information Graphic Summarization using a Collection of Multimodal Deep Neural Networks," en, in *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy: IEEE, Jan. 2021, pp. 10 188–10 195, ISBN: 978-1-72818-808-9. DOI: 10.1109/ICPR48806.2021.9412146. [Online]. Available: https://ieeexplore.ieee.org/document/9412146/ (visited on 05/07/2024).

[12] *ChartVi: Charts summarizer for visually impaired - 10814993.pdf*, 2022. [Online]. Available: https://www.reprintsdesk.com/userv3/fulltextreader.aspx (visited on 05/07/2024).

[13] A. Balaji, T. Ramanathan, and V. Sonathi, "Chart-Text: A Fully Automated Chart Image Descriptor," en, 2018.

[14] L. Ferres, P. Verkhogliad, G. Lindgaard, L. Boucher, A. Chretien, and M. Lachance, "Improving accessibility to statistical graphs: The iGraph-Lite system," en, in *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, Tempe Arizona USA: ACM, Oct. 2007, pp. 67–74, ISBN: 978-1-59593-573-1. DOI: 10.1145/1296843.1296857. [Online]. Available: https://dl.acm.org/doi/10.1145/1296843.1296857 (visited on 05/07/2024).

[15] J. Zong, C. Lee, A. Lundgard, J. Jang, D. Hajas, and A. Satyanarayan, *Rich Screen Reader Experiences for Accessible Data Visualization*, en, arXiv:2205.04917 [cs], May 2022. [Online]. Available: http://arxiv.org/abs/2205.04917 (visited on 05/07/2024).

[16] A. Lundgard and A. Satyanarayan, "Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content," en, *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 1073–1083, Jan. 2022, ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2021.3114770. [Online]. Available: https://ieeexplore.ieee.org/document/9555469/ (visited on 04/21/2024).

[17] D. Jung, W. Kim, H. Song, J.-i. Hwang, B. Lee, B. Kim, and J. Seo, "ChartSense: Interactive Data Extraction from Chart Images," en, in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver Colorado USA: ACM, May 2017, pp. 6706–6717, ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025957. [Online]. Available: https://dl.acm.org/doi/10.1145/3025453.3025957 (visited on 05/12/2024).

[18] M. Blanco, J. Zong, and A. Satyanarayan, "Olli: An Extensible Visualization Library for Screen Reader Accessibility," en, 2022.

[19] F. Elavsky, L. Nadolskis, and D. Moritz, "Data Navigator: An Accessibility-Centered Data Navigation Toolkit," en, *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2023, ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2023.3327393. [Online]. Available: https://ieeexplore.ieee.org/document/10301522/ (visited on 05/07/2024).

[20] M. Weninger, G. Ortner, T. Hahn, O. Drümmer, and K. Miesenberger, "ASVG Accessible Scalable Vector Graphics: Intention trees to make charts more accessible and usable," en, *Journal of Assistive Technologies*, pp. 239–246, Nov. 2015, ISSN: 1754-9450. DOI: 10.1108/JAT-10-2015-0124. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/JAT-10-2015-0124/full/html (visited on 05/07/2024).

[21] J. Gorniak, J. Ottiger, D. Wei, and N. W. Kim, "VizAbility: Multimodal Accessible Data Visualization with Keyboard Navigation and Conversational Interaction," en, in *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, San Francisco CA USA: ACM, Oct. 2023, pp. 1–3, ISBN: 9798400700965. DOI: 10.1145/3586182.3616669. [Online]. Available: https://dl.acm.org/doi/10.1145/3586182.3616669 (visited on 05/07/2024).

[22] J. R. Thompson, J. J. Martinez, A. Sarikaya, E. Cutrell, and B. Lee, "Chart Reader: Accessible Visualization Experiences Designed with Screen Reader Users," en, in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg Germany: ACM, Apr. 2023, pp. 1–18, ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3581186. [Online]. Available: https://dl.acm.org/doi/10.1145/3544548.3581186 (visited on 05/07/2024).

[23] S. Jones, I. P. Pineros, D. Hajas, J. Zong, and A. Satyanarayan, *"Customization is Key": Four Characteristics of Textual Affordances for Accessible Data Visualization*, arXiv:2307.08773 [cs], 2024. DOI: 10.48550/arXiv.2307.08773. [Online]. Available: http://arxiv.org/abs/2307.08773 (visited on 08/26/2023).

[24] A. Bodi, P. Fazli, S. Ihorn, Y.-T. Siu, A. T. Scott, L. Narins, Y. Kant, A. Das, and I. Yoon, "Automated Video Description for Blind and Low Vision Users," en, in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 2021, pp. 1–7, ISBN: 978-1-4503-8095-9. DOI: 10.1145/3411763.3451810. [Online]. Available: https://dl.acm.org/doi/10.1145/3411763.3451810 (visited on 05/07/2024).

[25] S. A. Doore, D. Istrati, C. Xu, Y. Qiu, A. Sarrazin, and N. A. Giudice, "Images, Words, and Imagination: Accessible Descriptions to Support Blind and Low Vision Art Exploration and Engagement," en, *Journal of Imaging*, vol. 10, no. 1, p. 26, Jan. 2024, Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2313-433X. DOI: 10.3390/jimaging10010026. [Online]. Available: https://www.mdpi.com/2313-433X/10/1/26 (visited on 05/07/2024).

[26] C. Gleason, A. Pavel, E. McCamey, C. Low, P. Carrington, K. M. Kitani, and J. P. Bigham, "Twitter A11y: A Browser Extension to Make Twitter Images Accessible," en, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–12, ISBN:

978-1-4503-6708-0. DOI: 10.1145/3313831.3376728. [Online]. Available: https://dl.acm.org/doi/10.1145/3313831.3376728 (visited on 05/07/2024).

[27] Z. Liu, C. Chen, J. Wang, M. Chen, B. Wu, Y. Huang, J. Hu, and Q. Wang, *Unblind Text Inputs: Predicting Hint-text of Text Input in Mobile Apps via LLM*, en, arXiv:2404.02706 [cs], Apr. 2024. DOI: 10.1145/3613904.3642939. [Online]. Available: http://arxiv.org/abs/2404.02706 (visited on 05/07/2024).

[28] B. Tang, A. Boggust, and A. Satyanarayan, "VisText: A Benchmark for Semantically Rich Chart Captioning," en, in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Toronto, Canada: Association for Computational Linguistics, 2023, pp. 7268–7298. DOI: 10.18653/v1/2023.acl-long.401. [Online]. Available: https://aclanthology.org/2023.acl-long.401 (visited on 05/08/2024).

[29] E. Kim and K. F. McCoy, "Multimodal Deep Learning using Images and Text for Information Graphic Classification," en, in *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, Galway Ireland: ACM, Oct. 2018, pp. 143–148, ISBN: 978-1-4503-5650-3. DOI: 10.1145/3234695.3236357. [Online]. Available: https://dl.acm.org/doi/10.1145/3234695.3236357 (visited on 05/07/2024).

[30] *Be My Eyes - See the world together*, en, 2015. [Online]. Available: https://www.bemyeyes.com/ (visited on 05/05/2024).

[31] *Seeing AI*, en-US, 2017. [Online]. Available: https://www.microsoft.com/en-us/garage/wall-of-fame/seeing-ai/ (visited on 05/07/2024).

[32] *New and Improved Features in JAWS – Freedom Scientific*, en, 2024. [Online]. Available: https://www.freedomscientific.com/training/jaws/new-and-improved-features/ (visited on 05/12/2024).

[33] R. Cohen, *Experience the Power of Assistive Technology with OrCam's AI Devices*, en, Aug. 2023. [Online]. Available: https://www.orcam.com/en-us/home (visited on 05/07/2024).

[34] M. A. Khan, P. Paul, M. Rashid, M. Hossain, and M. A. R. Ahad, "An AI-Based Visual Aid With Integrated Reading Assistant for the Completely Blind," en, *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 6, pp. 507–517, Dec. 2020, ISSN: 2168-2291, 2168-2305. DOI: 10.1109/THMS.2020.3027534. [Online]. Available: https://ieeexplore.ieee.org/document/9234074/ (visited on 05/07/2024).

[35] B. Li, J. P. Munoz, X. Rong, Q. Chen, J. Xiao, Y. Tian, A. Arditi, and M. Yousuf, "Vision-Based Mobile Indoor Assistive Navigation Aid for Blind People," en, *IEEE Transactions on Mobile Computing*, vol. 18, no. 3, pp. 702–714, Mar. 2019, ISSN: 1536-1233, 1558-0660, 2161-9875. DOI: 10.1109/TMC.2018.2842751. [Online]. Available: https://ieeexplore.ieee.org/document/8370712/ (visited on 05/07/2024).

[36] Z. Ning, B. L. Wimer, K. Jiang, K. Chen, J. Ban, Y. Tian, Y. Zhao, and T. J.-J. Li, *SPICA: Interactive Video Content Exploration through Augmented Audio Descriptions for Blind or Low-Vision Viewers*, en, arXiv:2402.07300 [cs], Feb. 2024. [Online]. Available: http://arxiv.org/abs/2402.07300 (visited on 05/07/2024).

[37] M. Huh, Y.-H. Peng, and A. Pavel, *GenAssist: Making Image Generation Accessible*, en, arXiv:2307.07589 [cs], Jul. 2023. [Online]. Available: http://arxiv.org/abs/2307.07589 (visited on 05/07/2024).

[38] C. Jayant, H. Ji, S. White, and J. P. Bigham, "Supporting blind photography," en, in *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, Dundee Scotland, UK: ACM, Oct. 2011, pp. 203–210, ISBN: 978-1-4503-0920-2. DOI: 10.1145/2049536.2049573. [Online]. Available: https://dl.acm.org/doi/10.1145/2049536.2049573 (visited on 05/07/2024).

[39] C. Saben and P. Chandrasekar, *Enabling BLV Developers with LLM-driven Code Debugging*, en, arXiv:2401.16654 [cs], Jan. 2024. [Online]. Available: http://arxiv.org/abs/2401.16654 (visited on 05/07/2024).

[40] C. Morrison, E. Cutrell, M. Grayson, A. Thieme, A. Taylor, G. Roumen, C. Longden, S. Tschiatschek, R. Faia Marques, and A. Sellen, "Social Sensemaking with AI: Designing an Open-ended AI Experience with a Blind Child," en, in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 2021, pp. 1–14, ISBN: 978-1-4503-8096-6. DOI: 10.1145/3411764.3445290. [Online]. Available: https://dl.acm.org/doi/10.1145/3411764.3445290 (visited on 05/12/2024).

[41] A. Lundgard, C. Lee, and A. Satyanarayan, "Sociotechnical Considerations for Accessible Visualization Design," en, in *2019 IEEE Visualization Conference (VIS)*, Vancouver, BC, Canada: IEEE, Oct. 2019, pp. 16–20, ISBN: 978-1-72814-941-7. DOI: 10.1109/VISUAL.2019.8933762. [Online]. Available: https://ieeexplore.ieee.org/document/8933762/ (visited on 05/14/2024).

[42] OpenAI, *GPT-4 Technical Report*, arXiv:2303.08774 [cs], Mar. 2023. [Online]. Available: http://arxiv.org/abs/2303.08774 (visited on 12/11/2023).

[43] X. Wan, R. Sun, H. Dai, S. O. Arik, and T. Pfister, *Better Zero-Shot Reasoning with Self-Adaptive Prompting*, en, arXiv:2305.14106 [cs], May 2023. [Online]. Available: http://arxiv.org/abs/2305.14106 (visited on 05/16/2024).

[44] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*, en, arXiv:2201.11903 [cs], Jan. 2023. [Online]. Available: http://arxiv.org/abs/2201.11903 (visited on 05/16/2024).

[45] T. B. Brown, B. Mann, N. Ryder, *et al.*, *Language Models are Few-Shot Learners*, en, arXiv:2005.14165 [cs], Jul. 2020. [Online]. Available: http://arxiv.org/abs/2005.14165 (visited on 05/16/2024).

[46] A. Horst, *Allisonhorst/palmerpenguins*, original-date: 2020-06-05T14:57:15Z, Apr. 2024. [Online]. Available: https://github.com/allisonhorst/palmerpenguins (visited on 05/05/2024).

[47] B. Saunders, J. Kitzinger, and C. Kitzinger, "Anonymising interview data: Challenges and compromise in practice," en, *Qualitative Research*, vol. 15, no. 5, pp. 616–632, Oct. 2015, ISSN: 1468-7941, 1741-3109. DOI: 10.1177/1468794114550439. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1468794114550439 (visited on 05/05/2024).

[48] *WebAIM: Screen Reader User Survey #9 Results*, 2021. [Online]. Available: https://webaim.org/projects/screenreadersurvey9/ (visited on 05/05/2024).