“Explain What a Treemap is”: Exploratory Investigation of Strategies for Explaining Unfamiliar Chart to Blind and Low Vision Users

Gyeongri Kim*
Jiho Kim*
University of Wisconsin-Madison
Madison, Wisconsin, USA
gyeongrikim3037@gmail.com,kim999@wisc.edu

Yea-Seul Kim
University of Wisconsin-Madison
Madison, Wisconsin, USA
yeaseul.kim@cs.wisc.edu

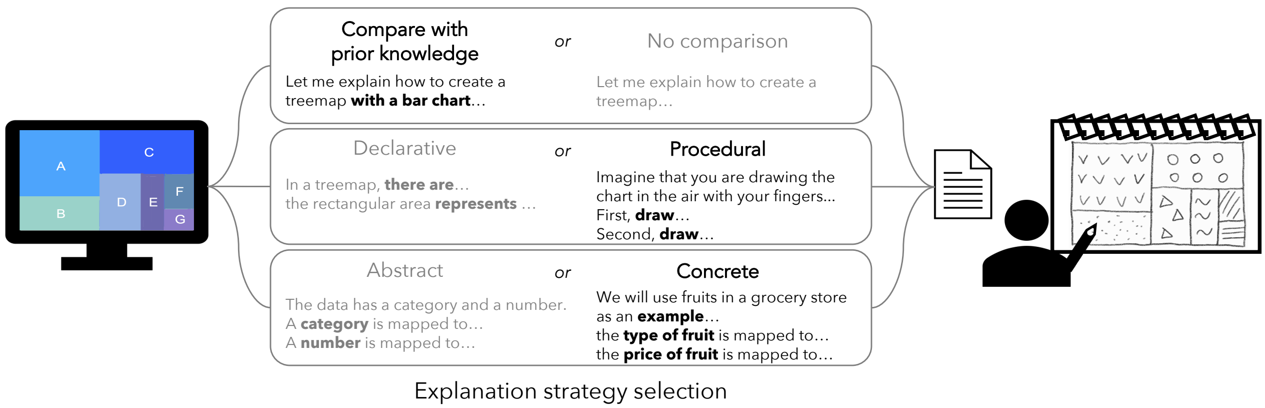

Figure 1: The blind user does not know what a treemap is. The system generates an explanation of treemap using three explanation strategies. The user listens to the explanation to understand how a treemap looks like.

ABSTRACT

Visualization designers increasingly use diverse types of visualizations, but assistive technologies and education for blind and low vision people often focus on elementary chart types. We explore textual explanation as a more generalizable solution. We define three dimensions of explanation strategies based on education theories: comparing to a familiar chart type, describing how to draw one, and using a concrete example. We develop a prototype system that automatically generates text explanations from a given chart specification. We conduct an exploratory study with 24 legally blind people to observe both the effectiveness and the perceived effectiveness of the strategies. The findings include: description of visual appearance is overall more effective than instructions for drawing, effective strategies differ by each chart type and by each participant, and the user's perceived effectiveness does not always lead to better performance. We demonstrate the feasibility of an explanation generation system and compile design considerations.

*Both authors contributed equally to this research.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581139

CCS CONCEPTS

• **Human-centered computing → Empirical studies in accessibility**; **Empirical studies in visualization**.

KEYWORDS

Visualization Accessibility, Blind Users, Explanation Strategy

ACM Reference Format:
Gyeongri Kim, Jiho Kim, and Yea-Seul Kim. 2023. “Explain What a Treemap is”: Exploratory Investigation of Strategies for Explaining Unfamiliar Chart to Blind and Low Vision Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3544548.3581139

1 INTRODUCTION

Data visualizations rely on vision to convey information. Due to this nature, people who cannot leverage vision, such as blind and low vision (BLV) users, are at risk of being disenfranchised from the benefits visualizations offer. To support accessibility, the visualization must be translated into other modalities beyond vision. For example, converting visual charts to audio charts [16, 40, 46] as well as tactile material, including embossed paper [10, 41] and specialized devices [5, 34] has been extensively investigated.

Despite the benefits, audio and tactile charts are not only challenging for BLV users to accurately perceive [40] but also take considerable effort for the authors to create [13]. Describing visualizations in a text can be a more generalizable and feasible solution than using other modalities. Communicating visualization through texts has been investigated to promote accessibility, allowing BLV users to read visualization online with a widespread software, namely screen readers [14, 20, 30].

When communicating visualization to BLV users, *chart type* is crucial information to convey [14]. By knowing the type of chart, BLV users can mentally prepare themselves and be ready to fill in specific details, such as axes and encoding, to understand the chart's content. However, BLV users are not familiar with more complex visualization types such as violin plots and area charts [40]. Many BLV users reported that they had touched some simple charts with embossed materials in school but not their variants. Also, the chart types on the web are becoming increasingly diverse, while many assistive technologies only support the most basic chart types.

To address this problem, we explore the space of chart type explanations and investigate which explanation methods work most effectively for BLV users. We review literature in the field of education to identify three different dimensions of explanation strategies and create explanation templates that can be applied to any chart type. We also explore the feasibility of automatically generated explanations by creating a prototype system that can explain more than 50 chart types. We conduct a user study to observe how BLV participants perceive the different explanation strategies and conduct an exploratory evaluation of the effectiveness of the strategies.

We found that (1) description of visual appearance as a static image is overall more effective than instructions for drawing, (2) effective strategies differ by individuals, (3) participants' traits, including spatial ability and locus of control as well as whether they have functional vision affect their comprehension of chart types from explanations, and (4) user perceived effectiveness does not always lead to better comprehension.

Our contributions are as follows.

- Based on education theories, we identify three dimensions of strategies for verbally explaining chart types and test their effectiveness for BLV users.
- We provide a similarity metric between chart types that can be used to leverage the prior knowledge of the user when explaining an unfamiliar chart type.
- We find out the wants and needs of the BLV users for accessible chart type explanations.
- Based on the findings from the user study, we derive the design considerations for a future explanation system.

2 RELATED WORKS

2.1 Accessible Visualization

Visualizations use human vision to convey a large amount of information effectively. People who cannot fully use their vision cannot access the information presented in visualization without additional accessibility support [21]. A common practice is to translate the information to other sensory modalities such as sound, texture, and text. Each has different advantages. Using sonic dimensions, including pitch and volume, can promptly convey dense time-series data [24, 38, 46], but its perception can vary drastically by individual [40]. Tactile visualizations use embossed paper or specialized devices to enable BLV people to touch the appearance of the visualization [4, 9, 45], but to produce one takes time and requires specialized devices that might not be readily available or affordable for all users. In comparison, text modality is a practical way to communicate visualizations that can be applied to many types of visualizations. An alternative text (alt text) is a textual summary of a visualization that can be read off to a user via a screen reader and is especially practical for visualizations on the Web [14]. Question answering systems that support natural language interactions are becoming more and more sophisticated as well [12].

Lundgard et al. [20] have categorized information in a visualization into 4 semantic levels. Level 1 is the visual components that comprise a graphical representation's design and construction. Level 2 is descriptive statistics or "data facts". Level 3 is complex trends and unforeseen patterns that can be obtained by fully leveraging the high bandwidth of human visual perception. Level 4 is contextual and domain-specific information, the high-level "message" that the visualization conveys. Among them, we are interested in helping BLV people understand the elemental and encoded properties (level 1). These include the chart type (e.g., bar chart, line graph, scatter plot), its title and legend, its encoding channels, axis labels, and the axis scales. Among them, we are most interested in chart type. This information is independent of the perceiver and is most relevant to BLV people because it requires sight, whereas higher levels concern information about the data that is independent of the visualization.

Jung et al. also argue that a description of a visualization should start with its *chart type* before providing any details [14]. They report that BLV users often use the chart type as a starting template and then fill in the specific details such as axes and encoding. However, the name of the chart type is meaningful to a BLV user only if they already know that chart type. Therefore, the fact that assistive technologies until now have mostly focused on supporting basic chart types, including bar charts, line charts, and pie charts [15], and that existing K-12 curricula in mainstream and special schools only teach basic chart types to BLV students [31], puts BLV users at a disadvantage from accessing different chart types.

2.2 Visualization Typology

As new visualization techniques are being developed and used in the wild, the task of classifying them is more difficult than ever before. For example, Sarikaya & Gleicher [29] report that they found 62 different strategies related to scatterplot design.

Several works have attempted to taxonomize visualizations according to different standards. One method is to classify visualizations based on the analytic tasks that they mainly support. Munzner [23] focuses on the task that target users should perform when designing visualizations. Amar et al. [1] provides the breakdown of tasks, while Lee et al. [17] takes a similar approach but focuses on network visualizations. These taxonomies treat visualizations as a means that facilitates a specific data analysis task. A more graphics-based view of visualizations also exists. Wilkinson [43] provided the building blocks of visualization as a set of graphical elements. A visualization can then be defined as the collection and organization of these elementary graphical elements. Visualization libraries that are being used today such as ggplot2 [42] and Tableau [33] adopt

Element	Examples used in our collection
data type	nominal, quantitative, temporal
channel	position, length, size, central angle, color(hue), color(brightness), width, count, shape
mark	bar, line, area, circle, rectangular area, pie slice, donut slice, icon
coordinate system	cartesian, polar, linear, parallel

Table 1: Examples of visualization elements from our specifications of 50 chart types.

this systemic view. With this system, visualizations can be classified by the low-level graphical elements that make up the whole.

Defining the distance between two visualizations has also been studied. Veras & Collins use pixel-level comparisons of charts [37]. The chart type is defined as a design specification whose visuals depend only on the provided data. They then define a similarity metric between visualizations. It compares two different instances of the same chart type to evaluate the design of a given chart type.

Visualizations are also classified by their complexity. To classify visualization based on complexity, Lundgard and Satyanarayan [20] use three degrees (easy, medium, hard), reflecting how close a visualization type is to the primary instance of the type. For instance, a simple bar chart is classified as easy, whereas a stacked bar chart is classified as medium.

Despite the continued efforts to classify visualizations, the nomenclatures used for those categories and how people actually understand them are not well understood. This is particularly true for names people use for chart types, which are highly abstract concepts that contain implications about a visualization's analytic task, data type, and visual appearance. As a result, no standards exist for translating a chart type into text. To address this problem, we investigate the what and how of explaining a chart type in text.

3 EXPLAINING UNFAMILIAR CHART TYPES

We investigated how to formulate an effective text explanation for BLV users to communicate chart types they have not seen or touched before. We focused on generating text description that conveys chart types, mainly their visuals (i.e., how it looks) through encoding information (e.g., how a chart type accommodates different data types) that allows BLV users to use as a "template" to construct a mental model of the given chart type before filling in the detailed data [14]. We surveyed theories from the field of education on what makes an explanation effective when communicating unfamiliar concepts. Based on the theories, we identified several strategies to communicate unfamiliar chart types to BLV users. Then we devised a prototype system to generate the explanation automatically. We conducted an exploratory study to observe how BLV participants interpreted chart types from the explanations.

3.1 Factors that May Affect Understanding Chart Type Explanation

3.1.1 Dimensions of Explaining Strategies. We surveyed seminal works in education to define the space of explanation strategies. In the process, one of the co-first authors, who has an advanced degree in education, reviewed the theories of some of the most cited authors in the field. We then studied how each of these theories can be applied to 1) the domain of data visualization, 2) the setting where a learner reads a text explanation with no human instructor, and 3) the BLV users as learners. Based on these criteria, we identified three orthogonal dimensions of explanation strategies.

D1. Comparing with Prior Knowledge vs. No Comparison Learners can better learn new concepts by exploiting their prior knowledge. This approach is grounded on Ausubel's concept of the comparative organizer, which states learners can organize the new information by assessing the similarity and differences with existing knowledge to promote better learning [3]. This concept is also based on Piaget's theory of schema, a structure through which individuals interpret the world. When a schema encounters new information, it tries to reach equilibration between itself and the new information by the process called assimilation and accommodation [25]. Pointing out the relationship between a new piece of information and the information already in the schema can facilitate this process.

In the context of our work, we assume that BLV users possess a schema of basic chart types. When BLV users learn a new chart type, they must understand how it fits into the concept of a chart type (assimilation) as well as how it is different from other chart types that they already know (accommodation). This process can be expedited by providing an explanation that compares the new chart type with a more familiar chart type.

D2. Declarative vs. Procedural Knowledge A piece of knowledge can be classified as either declarative or procedural. Declarative knowledge is factual knowledge expressed with language, and procedural knowledge is behavioral knowledge that can be acted out unconsciously [6, 28, 32]. In education, traditional teaching methods like lectures are considered to teach mostly declarative knowledge. On the other hand, active learning, a student-centered way to attain knowledge through engagement, is effective for teaching procedural knowledge, especially problem-solving skills [22].

In the context of our work, a *declarative* explanation communicates the visuals of a chart type by simply conveying how it looks like. Then it explains what data each visual element represents. Most alternative texts and captions generated by captioning technologies [44] used in practice follow this strategy. On the other hand, a *procedural* explanation achieves the same goal by teaching how to create the chart from given data. For procedural explanation, we also implement principles of active learning to provide learners with opportunities to practice the procedures they are learning, which has been shown to enhance visualization learning [39]. For example, we prompt the learner to draw the chart in the air with their fingers.

D3. Abstract vs. Concrete This dimension concerns the level of abstraction of the learning material. For example, one approach to learning math is to gain knowledge of the rules first before working out the specific problems. Another is to look at the solutions to

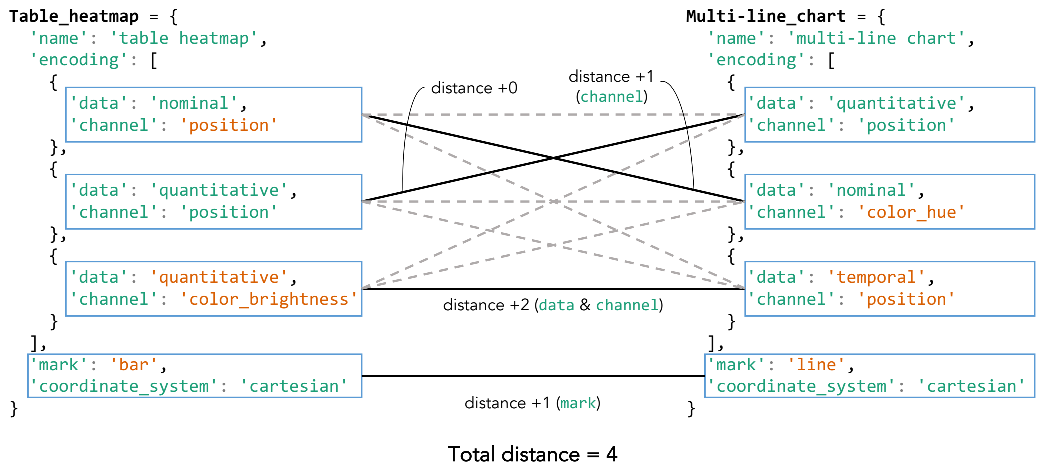

Figure 2: A demonstration of calculating the distance between table heatmap and multi-line chart. On the left is the specification of a table heatmap, and on the right is the specification of a multi-line chart. Both chart types have three `encodings`. First, we perform minimum weight bipartite matching from `encodings` of table heatmap to `encodings` of multi-line chart. The bold lines indicate the resulting edges. The mismatches between elements that are contributing to the distance are colored in red. The second `encoding` of table heatmap is matched to the first `encoding` of multi-line chart. They are identical, so no distance is added. However, the first `encoding` of table heatmap is matched to the second `encoding` of multi-line chart, and their `channel` is different. Distance 1 is thus added. Similarly, the third `encoding` of table heatmap is matched to the third `encoding` of multi-line chart. Distance of 2 is added because both `data` and `channel` are different. Lastly, `mark` and `coordinate system` are compared, and distance 1 is added for each mismatch. This results in a total distance of 4.

the problems before learning the high-level rule that governs them. Both types of learning are advocated by the literature. Abstract knowledge can help learn more specific concepts [3]. Once an abstract concept has been established in the learner, more specific knowledge that falls under that concept will activate the learner's cognitive structure and will be better incorporated. On the contrary, sufficient evidence also points to the effectiveness of example-based learning. Observing worked examples can help the learning of cognitive skills [2].

In the context of our work, an *abstract* explanation describes the chart type without mentioning any one instance of that chart type. A *concrete* explanation, on the other hand, explains the chart type by referring to a real-world instance of that chart type. For example, an abstract explanation of a bar chart says, "a bar's length represents a number," while a concrete explanation says, "a bar's length represents the price of fruit." For all of the concrete explanations, we chose to use an example about fruits (e.g., price, color, or type of fruit) to ensure that the example is familiar to all the participants.

Defining Terminology In this paper, we use **dimension** to refer to the three dimensions defined above. Each dimension has two **conditions**, and each condition corresponds to one explanation **strategy**. We use **condition** and **strategy** interchangeably. We italicize references to explanation strategies.

3.1.2 Individual User Traits. Individual traits impact visualization literacy for sighted people [19, 47]. We investigate how these factors can affect BLV users in understanding chart type explanations and if they should be considered in creating these explanations. We focus on the locus of control and spatial ability, which are known to influence visualization comprehension [11, 36]. To measure them, we used the questionnaires from prior works [26, 35]. The validity of these self-measured matrices is demonstrated through many research (e.g., [7, 8, 35]).

The questionnaire for spatial ability measures three factors, including object-manipulation spatial ability (OMSA), spatial navigational ability (SNA), and visual memory (VM). VM measures the ability to recognize and recall other people's appearance. Since they did not apply to BLV people, we removed the questions. For other questions, we modified and removed some of the items based on their applicability to BLV people. For example, Q8 "I can easily identify a three-dimensional shape drawn on paper" is changed to "I can easily identify a three-dimensional shape drawn on paper if it is accessible." We removed Q9 "If I see photographs of a building taken from different perspectives, I can visualize this three-dimensional structure in my mind." The questionnaire for measuring locus of control is used without any changes. For all the questions, we used a 1 to 5 Likert scale to allow participants to give a neutral answer to induce a higher accuracy and reliability of the collected data [27].

We also conjecture that participants' age, gender, and whether they have functional vision may influence their understanding of chart type and incorporate it into our analysis.

3.2 Stimuli: Explanation Generating System

We first defined what a chart type is to identify what information needs to be communicated in explaining it. Among prior work (Section 2.2), we adopt the systematic view of visualizations and define a chart type by its specification that contains its mark, data type, channel, encoding, and coordinate system [23, 43]. Table 1 shows some examples of what each element could be. Additionally, an encoding is defined as a pair of one data type and one channel. We

No comparison – Declarative – Abstract

```
A chart has chart.coordinate_system. There are chart.marks, which encodes
information through their chart.channels.
FOR e IN chart.encodings:
    The e.channel of the chart.mark represents e.data.
    e.channel.change represents e.data.change.
```

No comparison – **Procedural** – Abstract

```
Imagine that you are drawing the chart in the air with your fingers. Let me explain how to
create a chart. Assume you have a list of data, each consisting of for chart.data. First,
draw chart.coordinate_system. You will then draw chart.marks to visualize data.
FOR e IN chart.encodings:
    e.data in the data is mapped to the e.channel of the chart.marks.
    e.data.change is mapped to e.channel.change.
If you are done, you have finished drawing a chart.
```

No comparison – Declarative – **Concrete**

```
A chart has chart.coordinate_system. There are chart.marks, which encodes
information through their chart.channels. For example, consider a chart that shows
information about fruits in a grocery store.
FOR e IN chart.encodings:
    e.data ← POOL( [ "the type of fruit", "the color of fruit", "price of fruit",
                     "number of fruit sold", "weight of fruit", "size of fruit", "year" ] )
    The e.channel of the chart.mark represents e.data.
    e.channel.change represents e.data.change.
```

Comparing with prior knowledge – Declarative – Abstract

```
A chart1 is similar to a chart2, with minor differences.
A chart1 has chart1.coordinate_system,
[ which is the same for chart2 | whereas a chart2 has chart2.coordinate_system ].
A chart1 shows data with chart1.marks.
[ Similarly, a chart2 has chart2.marks. | However, a chart2 shows data with chart2.marks. ]

# Reorders encodings so that chart1.encodings[i] and chart2.encodings[i] are
# connected with an edge as a result of minimum weight bipartite matching
Minimum_Weight_Bipartite_Matching(chart1, chart2)

FOR (e1, e2) IN (chart1.encodings, chart2.encodings):
    In a chart2, the e2.channel of the chart2.mark represents e2.data.
    [ This is the same for a chart1.
     | However, in a chart1, the e1.channel of the chart1.mark represents e1.data.
     | This is not present in a chart1. ]
    e1.channel.change represents e1.data.change.

IF LENGTH(chart1.encodings) > LENGTH(chart2.encodings):
    Furthermore, a chart1 has additional properties.
    FOR e1 IN REMAINING(chart1.encodings):
        The e1.channel of the chart1.mark represents chart1.data.
        e1.channel.change represents e1.data.change.
```

Figure 3: Pseudocodes for generating explanation using different strategies. Among the $2 \times 2 \times 2 = 8$ possible explanations, we show 4 of them here. The top-left pseudocode uses *no comparison, declarative,* and *abstract* strategy. The other three pseudocodes show how the explanation changes as one of the strategies changes.

chose this view of visualizations as it is widely used in many visualization creation tools such as ggplot2 [42] and Tableau (formerly Polaris) [33]. This approach allows us to create a system that can automatically generate explanations from existing visualizations on the web, which aligns with one of the goals of our work.

We curated a set of 50 chart types in the following way. First, we included basic charts known to be familiar to BLV users [40] (Fig 6). These include pie charts, bar charts, line charts, and scatterplots. These charts can be used in *comparison with prior knowledge* explanation strategy and are compared to a more unfamiliar chart type. Second, we surveyed chart types on the web. We searched on Google with keywords "visualization types" and "chart types." We also examined the example galleries of Vega-Lite, Highcharts, and D3, resulting in 50 types. Note that we only consider the basic instance of a chart type. To elaborate, we exclude visualizations that encode a variable with more than one channel. For example, some bar charts have bars with different colors because they encode a nominal variable with both color and position. We do not include such duplicate encoding in the specification and just consider it a normal bar chart. Figure 2 shows what the specifications for a table heatmap and a multi-line chart look like.

In addition, to generate explanations that compare one chart type to another, we designed an algorithm that can calculate the degree of similarity between two chart types. The algorithm takes the specifications of the two charts as its arguments and outputs a distance between the two. A distance of 0 means that the two chart types are identical, and a higher distance means they are increasingly different.

Figure 2 shows an example of how the distance is calculated. The algorithm first compares the two `encoding`s. We formulate this task as a minimum weight bipartite matching problem. A node represents an `encoding` which is a pair of a `data` type and a `channel`. Two nodes that do not belong to the same chart type can have an edge between them. The weight of the edge is the Hamming distance between two nodes: 0 if both the `data` type and the `channel` are the same, 1 if only one of them is the same, and 2 if neither are the same. A solution to the minimum weight bipartite matching is then found, and all of its edges' weights are added, plus 2 for each node without any edge (e.g., when the number of nodes of two chart types is different). Finally, 1 is added to the distance for each mismatch in `mark` and `coordinate system`.

Then, we organize a total of 2x2x2=8 different methods of explanations from the 3 dimensions of explanation strategies. We create a template for each strategy so that the corresponding explanation can be automatically generated from a chart type specification. Exclusively for *comparing with prior knowledge* explanations, the system takes an additional chart type specification as input and based on the bipartite matching described above, outputs an explanation that uses the comparison. Figure 3 shows the pseudocode used to generate 4 out of 8 possible combinations of strategies. Below are two example explanations of the nightingale chart that our system generated.

- Nightingale chart: *No comparison & Procedural & Abstract* *Imagine that you are drawing the chart in the air with your fingers. Let me explain how to create a nightingale chart. Assume you have a list of data, each consisting of a category, a number, and time. First, think of a circle. You will then draw pie slices to visualize data. Now, a category is mapped to the color of the pie slices. Each category is mapped to a different color. Next, a number is mapped to the length of the pie slices. A higher number is mapped to a longer pie slice, and a smaller number is mapped to a shorter pie slice. Next, time is* *mapped to the clock position of the pie slices. A later time is mapped to the clockwise direction of its previous time. If you are done, you have finished drawing a nightingale chart.*

- Nightingale chart: *Comparing with prior knowledge & Declarative & Concrete*
  A nightingale chart is similar to a pie chart, with minor differences. A pie chart has a circular shape, which is the same as a nightingale chart. A pie chart shows data with pie slices. Similarly, a nightingale chart has pie slices. Consider a nightingale chart and a pie chart that shows information about fruits in a grocery store. In a pie chart, the color of the pie slice represents the type of fruit. This is the same for a nightingale chart. Each color represents a different type of fruit. Then, in a pie chart, the central angle of the pie slice represents the price of the fruit. However, in a nightingale chart, the length of the pie slice represents the number of fruits sold. A longer pie slice represents more fruits sold than a shorter pie slice. Furthermore, a nightingale chart has an additional property. The clock position of the pie slice represents the year. A pie slice in the clockwise direction of another represents the next year.

3.3 Study Method: Interview

We used mailing lists we received from organizations for BLV people (e.g., the National Federation of the Blind). The recruitment criteria were participants with 1) age over 18, 2) legal blind status, 3) daily screen reader use, and 4) living in the United States. A total of 24 BLV people consented to participate and provided demographic information. 13 were female, 10 were male, and one person preferred not to state their gender. The participants' average age was 38 (SD=10). 14 had no functional vision, and 10 had a functional vision. 14 were blind from birth, and 10 became blind after birth. We also collected information on their assistive technology use and experience with data and visualizations. The full information is in the supplementary material. We compensated a $25 gift card per hour for their participation.

We scheduled an online Zoom interview with each participant. During the interview, participants interacted with the study interface we created. The interviews were recorded and were used by the research team members approved by the IRB. The study took 61 minutes on average (SD=19.4). The first few sessions were treated as pilot studies where we reflected on their input to iteratively polish the design of the experiment and the interface.

We chose 12 chart types that we thought most participants would be unfamiliar with (Table 2). We defined complexity as the number of variables that can be mapped to a chart type through different channels. We chose 6 from low-complexity chart types (at most 2 variables) and the other 6 from high-complexity chart types (at least 3 variables) and set them as the stimuli pool. At the start of the interview, we ask several questions through the interface to check their prior knowledge of chart types. To be specific, we ask them whether they can explain what a chart type looks like for all the chart types in the stimuli pool. If their answer is yes, that chart type is removed from the stimuli pool. We select four chart types from the ones that the participant does not know and use them as stimuli. For each stimulus, we assign explanation strategies such that each participant sees exactly 2 of each strategy for all three dimensions (Figure 4).

Complexity	Chart type	Prior knowledge	Distance
4	Colored bubble chart	Multi-line chart	4
3	Nightingale chart	Pie chart	3
3	Marimekko chart	Bar chart	4
3	Stacked bar chart	Bar chart	2
3	Table heatmap	Multi-line chart	4
2	Radar chart	Line chart	3
2	Radial bar chart	Bar chart	2
2	Treemap	Bar chart	2
2	Histogram	Bar chart	1
2	Donut chart	Pie chart	1
1	Violin plot	Area chart, Line chart	2, 3
1	Bee swarm chart	Scatterplot, Line chart	3, 4

Table 2: Chart types used in the interview.

This means that two of the stimuli are explained by *comparing with prior knowledge* strategy. To select the chart type to use as prior knowledge for comparison to the stimulus, we create a list of chart types by sorting by their distance to the stimulus in ascending order and filtering for those with lower complexities. The violin plot and bee swarm chart are exceptions because they have a complexity of 1, the lowest possible complexity. They are compared to chart types that have higher complexity than them. We then check whether the participant knows the chosen chart type by asking them if they can explain how it looks. If the participant knows the chart type, we use it to generate the explanation. Otherwise, we choose the next chart type in the list and repeat the process of asking the participant. For instance, to explain a violin plot *with comparison*, we compare it to an area chart which is the closest chart type to a violin plot. However, if the participant says they do not know what an area chart is, then a line chart is used, which is the next closest chart type. This is repeated until we find a chart type that the participant knows. However, if the distance between the chart type and the stimulus exceeds 4, the stimulus changes to another chart type, and the process is repeated. After the 4 stimuli and 2 chart types to be used for comparison are all determined, the interface automatically generates one explanation for each stimulus, using the templates that we defined earlier.

The participants then perform 4 tasks. In each task, participants are asked to listen to an explanation of a certain chart type. They could listen to the explanation with a screen reader at most three times. Then, they are asked to draw the general appearance of the chart type. They were not allowed to read the explanation again once they started drawing. We recommended they draw with a thick pen on a piece of paper but encouraged them to use whatever tools they are comfortable with. We took a screenshot of their drawings while they showed them via webcam. Then, the participants performed a question answering (QA) task where they answered 3 questions that assessed their comprehension of the chart type. These questions are adapted from VLAT [18], a popular visualization literacy test. The QA task first describes which specific data is shown in the chart. Each question asks the participant to perform an analysis task. For example, here is the prompt for the QA task of radar charts:

Assume you are looking at a radar chart that shows a student's test scores by subject.

(1) *Describe how you would learn from the chart her scores on biology.*

Score	Rubric for Drawing Task	Rubric for Question Answering Task
0	The participant could not draw anything, or the drawing had no correct visual element.	The participant did not answer, or the answer did not identify any concepts.
1	The drawing had one correct visual element (e.g., axes, mark, channel).	The answer identified one concept correctly (e.g., channel, mark, data).
2	The drawing had two correct visual elements.	The answer identified at least two concepts correctly but no channel-data pair was correctly identified.
3	The drawing had more than two correct visual elements.	The answer identified a correct channel-data pair but had missing concepts.
4	The drawing had all the visual elements but had a slight error.	The answer identified all concepts correctly but described a wrong task.
5	The drawing had all the visual elements and resembled the actual chart.	The answer identified all concepts correctly and the visual explanation made sense.

Table 3: The rubric used to quantitatively assess the artifacts generated by the participants during the experiment.

(2) *Describe how you would learn from the chart the subject where she got the highest score.*
(3) *Describe how you would learn from the chart whether she is well-rounded or is specialized in one subject.*

Participants answered how to perform the given task from the visualization. We chose this format because we wanted to measure the participants’ comprehension of not only the chart’s appearance but also its relationship to the underlying data. The scope of the task goes beyond just retrieving a single value and encompasses more complex tasks, such as summarizing data distribution and finding a correlation between two variables.

Before reading the generated explanations, participants went through a practice task. The practice task asks them to read a very simple explanation of a bar chart and draw and solve questions like in a real task. The practice task shows them example answers to set the right expectations about the tasks and gives them a chance to get used to the interface.

We conducted a short interview with the participants after the experiment. We wanted to see how the participants perceived the effectiveness of each strategy. We also collected their opinions on the drawing and the QA task. Some of the questions that we asked include: How difficult was the task of drawing charts on paper? Which explanation did you find the easiest and why? Which explanation did you like? What do you usually do when you encounter a chart that you cannot understand? What information did you think was missing in the explanations?

Finally, after the interview, participants filled out a survey that measured their spatial ability and the locus of control on a self-report questionnaire, as described in Section 3.1.2.

3.4 Analysis approach & Evaluation

Since our goal is to understand qualitatively the requirements for creating the most effective chart type explanation, we adopted a qualitative content analysis method. During the interviews and after transcription, each researcher took note of any interesting signals and insights of each researcher’s findings. Each researcher analyzed what participants said and how they performed the drawing and QA tasks. Then the researchers reviewed each other’s notes and extracted common qualitative themes. We then classified these themes according to each dimension of the explanation strategy.

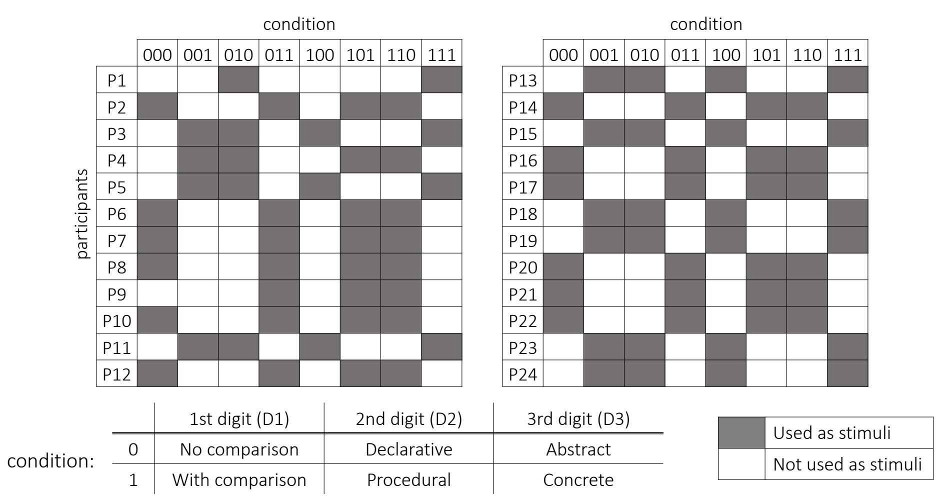

Figure 4: Conditions assigned to each participant. A participant was given two of each strategy for each dimension.

In addition to the qualitative analysis, we sought to have some quantitative measures that can signal participants' performance. Since there is no prior work on performance measurements for BLV users, we attempted to devise them in the context of our study. To measure how they constructed the images of visualizations in their head, we evaluated the participants' drawings and answers on a 0 to 5-point scale using the rubric we created (Table 3, left). For a full score, a drawing must contain accurate representations of the axes and marks, and the marks should show variance along their channels. For example, a bar chart must show horizontal and vertical axes and bars that have different lengths. Often, the drawings satisfied all of these criteria but were visually different from the chart type that was explained to them. In such cases, we deducted one point from the full score and analyzed the explanations to investigate what was missing and how to augment them.

The responses to the QA task were similarly evaluated on a 0 to 5-point scale. We first collaboratively created answers to the questions. From those answers, we identified key concepts, including mark, channel, and data type, that are necessary to answer the question. For a full score, an answer must address all the concepts correctly and carry the given tasks out accurately using the concepts (Table 3, right). We deducted a point when participants were able to identify the correct concepts but could not connect them to carry out the given analytical tasks. When participants did not use the specific terms used in the explanations, if their explanations demonstrated comprehension of a concept, we counted it as correctly addressed. The three questions in a single QA task were added up so that the scores ranged from 0 to 15.

We also consider individual traits as potential factors that influence their comprehension of the explanations. The locus of control and spatial ability was calculated as a number between 1 and 5, from the participants' responses to the self-diagnosis questionnaires after the interview. Age, gender, and whether or not participants had functional vision were collected from the demographic survey before the interview.

We used R's lme4 package to perform linear mixed effects analysis of how explanation strategy and individual traits influenced the task scores. Each strategy dimension, stimuli, and demographic factor were set as fixed effects, and participant id was set as a random effect. We also looked at the interaction between stimuli and condition to see if the effectiveness of each explanation strategy was different by chart type. We used normal approximation to calculate p-values of fixed effects, with t-scores produced by lme4. We then analyzed the post-task interview to learn about the perceived effectiveness of each strategy and the participants' needs. Lastly, we analyzed if each participant performed better on strategies that they thought to be effective. We backed our findings with qualitative evidence.

4 RESULTS

4.1 Findings on strategies

We present our findings on the perceived effectiveness and the effectiveness of the explanations by each strategy dimension. We first report our observations from the interviews and present statistics derived from our evaluation of the participants' drawings and their answers to the QA task. The performances of a participant on these two tasks were weakly correlated (R^2=0.26).

Each interview took an hour on average. We were not able to collect drawings from 2 participants (P13 & P19) since P13 did not prepare any drawing tools and P19 went through a technical issue with the camera. We excluded the two participants from the analysis of drawing tasks.

4.1.1 D1. Comparing with Prior Knowledge vs. No Comparison. Contrary to our expectations and the theories, most participants found explanations with *no comparison*s to be more effective than explanations *with comparison*. P13 shared that "it might seem easy but comparing two charts makes me more confused." P23 echoed, "it could make it a little hard to decipher." Some implied that simultaneously comparing two charts imposed cognitive load. P7 mentioned "I find it confusing because I have to think about two things at once, I think about the original and the new modified work." P14 also shared "I liked it when it was only one chart because otherwise, it got too many things in the mix. It just got too overwhelming for me." Participants felt that explaining the differences and similarities altogether was superfluous. P20 shared "The focus is on what a radar chart is. But I feel like that's being overshadowed by what a radar chart is not."

Sometimes, the familiar chart type that is used for comparison had a strong priming effect and gave participants a wrong impression about the unfamiliar chart. For instance, when a nightingale chart was compared to a pie chart, the participants assumed that how color is represented in both charts is the same. This is because both chart types use color to represent a category. Therefore, instead of realizing that different colors can exist in a single pie slice, many assumed that each pie slice in a nightingale chart would have one color as in a pie chart. Participants' answers to QA tasks (e.g., P18 "I would find the longest slice and find out what category its color corresponded to", P14 "The slices that were the same color the most often...") signal this misconception. Similar phenomena were observed when explaining a bee swarm chart. When it was compared to a scatterplot, many participants drew both a horizontal and a vertical axis like a scatterplot, rather than just one (Figure 5).

Another reason for *comparison with prior knowledge* being not much effective is that the participants can possess inaccurate prior knowledge. We asked each participant if they knew a chart type at the start of the experiment. Specifically, we asked whether they have seen, touched, or learned how it looks. Even though we strongly encouraged them to answer "no" if they were unsure, some participants who answered "yes" showed a lack of understanding of the chart type in the following tasks. Also, some participants misunderstood the names of the chart types, as there are no fixed nomenclatures for them. P24 said "I remember, like Venn diagrams, line graphs, box, and whisker plots...maybe it was called a different name. Maybe like radial (bar chart) is supposed to be pie chart, so sometimes maybe it's introduced under a different name."

Nonetheless, some participants perceived the *comparison* to be more effective. P23 shared, "in some ways, it was helpful because for the ones I didn't recognize it gave me like a basis to go off of." P18 said, "Comparison is better as I can use background knowledge." P20, though skeptical about detailed comparison, said, "the first comparison, I think, is necessary because it relates to the information you already may have." This aligns with P1's opinion that

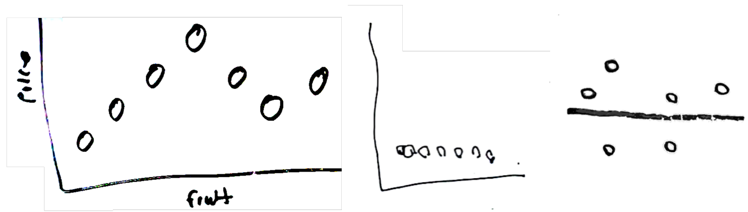

Figure 5: Bee swarm charts drawn by P6 (left) and P8 (middle) illustrate misconception due to the priming of scatterplot. P2's drawing (right) does not exhibit the sign of misconception.

"explaining all the similarities first and then differences later might be easier to understand."

The effectiveness of *comparison*s depended on the complexity of the chart type being explained. The more complex the chart type, the less helpful the comparison was as perceived by participants. This could be because high-complexity charts tend to get compared with other high-complexity charts, making keeping track of all the elements difficult. For example, a colored bubble chart is of complexity 4, the highest in the stimuli pool. It is compared to a multi-line chart, which is of complexity 3, meaning that the participants had to work with a total of 7 data types and channels. Therefore, explanations with comparison (M=6.27, SD=2.73) resulted in significantly lower QA scores than without (M=13.95, SD=2.18, t=2.2, p<.05). On the other hand, a violin chart (complexity 1) is compared to an area chart or a line chart (complexity 2). This produces a much shorter explanation that is easier to process by the participants. Thus, comparisons were more effective in terms of the QA tasks (M=6.67, SD=1.94) than without (M=2.06, SD=2.67). This interaction between D1 (*with comparison* vs. *no comparison*) and complexity is also evidenced by our quantitative analysis (t=-2.0, p<.05). Treemap is another example where drawing scores were higher with comparison (M=3.03, SD=0.75) than without (M=0.90, SD=0.53, t=-2.3, p<.05). Treemap was compared to a bar chart, both of complexity 2, which is relatively low.

The comparison was also more effective when the two chart types being compared were more similar than different. For example, many participants thought the comparison with a pie chart helped explain a donut chart (P19 "Donut (chart) was the easiest because I know the pie (chart)."), but not for explaining a marimekko chart with a bar chart (P20). This is because the distance between a donut chart and a pie chart is 1, whereas the distance between a marimekko chart and a bar chart is 4.

4.1.2 D2. Declarative vs. Procedural. For this dimension, many participants said they do not prefer one over the other. When prompted, P7 mentioned, "I think I don't care." In some cases, participants did not recognize or recall the difference between the two strategies at all. Only a few participants had a clear preference for one over the other. Some favored *declarative* explanations because they helped visualize the chart better. P15 shared, "procedural explanations have too many words and give me cognitive problems". P10 echoed "I prefer the one with (declarative because) I can visualize it better."

Those who favored *procedural* explanations emphasized the sequential nature of the instructions that they could follow step-by-step. Prompting active learning also helped the perceived effectiveness of *procedural* explanations. P22 said "Imagine you're drawing a circle in the air with your finger. That made it a lot easier." P10 physically drew on the air while listening to the explanation. P14, along the same line, shared, "when it could tell me how to draw it, I found that I was trying to draw it." P8 suggested a combination of both strategies by saying, "I want it to explain how it looks like first and then explain how to draw it step by step".

While no significant difference was observed in the number of participants who preferred one option over the other, *declarative* explanations (M=8.98, SD=1.06) yielded significantly higher QA scores (t=-2.3, p<.05) than did *procedural* explanations (M=7.19, SD=1.05) across all chart types. In drawing tasks, we did not observe any differences between the two strategies (t=-1.1, p=.27).

While there was a signal that *declarative* explanations can be more effective, *procedural* explanations were more effective than *declarative* explanations on some chart types. Participants performed reliably better on the drawing task of a violin plot with *procedural* explanations (M=2.55, SD=0.50) than with *declarative* explanations (M=0.16, SD=0.71, t=-2.8, p<.01). This may be due to its lowest complexity in the stimuli pool, meaning the rules for drawing it is relatively simple. Thus, the *procedural* explanation for violin plots did not put too much cognitive load on the participants while providing the benefits of step-by-step instructions.

4.1.3 D3. Abstract vs. Concrete. The majority of participants answered that they prefer *concrete* explanations. For P10, "examples give a better understanding of the values." P10 even demanded more concreteness by saying "I want more specific examples, like what each bar actually represents". P18 shared that explanations with examples were easier to understand, saying "marimekko chart was the easiest because it was quite specific." The example of fruits also helped participants visualize the chart type. In his drawing of a bee swarm chart, P6 labeled the axes "fruit" and "price", and said "I like examples because it is like the real world". Similarly, P23 liked *concrete* explanations "because they gave me something concrete to picture and to categorize."

On the contrary, some participants thought *abstract* explanations were more effective. For example, P12 said "(abstract) description seem pretty straightforward." P14 said, "if the example was more complex, it almost made the whole experience a little more overwhelming for me." When listening to a concrete explanation of histogram, P7 expressed his wish to get rid of the examples and think more abstractly, saying "Isolating from examples can be better (because) length and position is confusing." P16 was ambivalent and said "Examples did not matter too much."

4.2 The Effect of Individual User Traits

Both the locus of control and spatial ability had a significant influence on the participants' task performance, especially on the QA task. Participants with a more external locus of control (t=2.3, p<.01) and those with a higher spatial ability (t=2.2, p<.01) had higher QA scores on average. This was consistent with prior findings with sighted users [47] and shows that similar conclusions can be made on the visualization skills of BLV people. However, the locus of control (t=1.3, p=.20) and the spatial ability (t=0.8, p=.44) did not reliably impact their performance on the drawing task.

We found that *comparison with prior knowledge* was more effective in QA tasks for participants with high spatial ability than for those with the low spatial ability (t=2.5, p<.05). This may indicate that spatial ability is closely related to the BLV user's ability to compare and contrast two chart types in their minds. In contrast, explanations with *no comparisons* did not show a difference between participants with high and low spatial ability.

Participants who had a functional vision (M=3.26, SD=0.34) scored better on the drawing task than participants without functional vision (M=2.75, SD=0.32, t=2.1, p<.05). Age (r^2<.1) and gender (t=1.6, p>.1) did not correlate with performance on both tasks.

4.3 Other Insights

We also report additional findings that can help us better design an explanation system for unfamiliar charts.

Discrepancy between Effectiveness and Perceived Effectiveness In most cases, participants performed better on the drawing and QA tasks with explanation strategies that they perceived to be more effective. However, this was not always the case. Participants sometimes performed much better on explanation strategies that they did not perceive to be effective. P4 said that *comparing with prior knowledge* and *no comparison* were similarly effective, but struggled more on QA tasks with explanations with comparisons. P2 favored *procedural* explanations over *declarative*, saying that "really specific step-by-step explanation, like first draw this then draw that, was super helpful," but performed noticeably worse on QA tasks with *procedural* explanations. P23 pointed out that "explanations with examples were better because they were concrete" but showed better performance with *abstract* explanations.

Thinking Color as Texture Participants were more comfortable thinking of texture in place of color in the explanations. Our stimuli pool contained many chart types (e.g., marimekko chart, stacked bar chart, nightingale chart, etc.) that use color as one of their channels. Early in the study, P3 showed difficulty grasping how color could be used to represent data and also how to express color in their drawings. P3 suggested that "(color) could be replaced with different texture," which we adopted since the texture is a modality that BLV users are more familiar with than color. With explicit instruction to think of color as texture, participants more comfortably expressed color with different shapes (e.g., P17 "I tried to make like a different shape," P12 "I tried to add a different...dots or lines," Figure 6), or verbally explained how different textures would fit in their drawings.

Difficult Terms Around Circular Chart Types Some terminologies related to circular (i.e., polar) coordinate systems were not intuitively understood by the participants. Radial bar charts, radar charts, and nightingale charts all suffered from this problem. Our explanations adopted more colloquial terms like "circular", "clockwise", "width", "pie slice", and "donut slice" instead of more technical terms like "polar", "radial", "angular", "circular sector", and "annulus sector", respectively. However, the colloquial terms that we chose were insufficient. Many participants pointed out the difficulty of understanding how these terminologies translate into a visual figure. P9 expressed confusion by saying "I wonder what they mean by the position of the pie slice." P14 shared, "(What) does that mean when you make a bigger central angle? I wasn't quite sure what that meant."

Difficulty with Stacking Marks Many participants failed to visualize the stacking of marks. Stacking is visually connecting more than one mark along an axis. A stacked bar chart, for example, uses stacking to connect bars of different colors. Participants could not visualize the stacked bars for a stacked bar chart (Figure 7), though the explanation explicitly said that the "bars in the same horizontal position are vertically stacked on top of each other." They commonly drew a grouped bar chart, where bars are stacked horizontally instead of vertically (e.g., P23 "Bars are horizontally stacked on top of each other"). P20 interpreted the stacking as overlapping the bars and drew a bar inside another. Many other participants drew it just like a bar chart, but with different colors assigned to each bar (e.g., P4, P12).

5 DISCUSSION

In this study, we observed how various types of strategies impact participants' understanding of visualizations that are not familiar to them. The perceived effectiveness of each explanation strategy varied by individuals. The participants' level of comprehension mostly aligned with the perceived effectiveness of the given explanation strategy, but in some cases, they were at odds with one another. The declarative explanation strategy (D2) was more effective overall than the procedural explanation strategy when measured with QA tasks. For the accuracy of the drawing, there was no significant

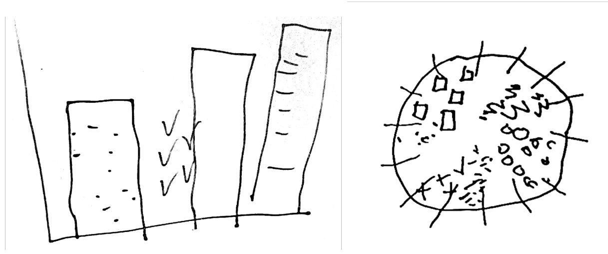

Figure 6: Examples of expressing color as texture. Stacked bar chart by P12 (left) and donut chart by P17 (right).

effect of each explanation strategy. Individual factors, including spatial ability, locus of control, and having a functional vision, affected the participants' overall performance.

5.1 Design Considerations for Future Systems

The result points to the feasibility of creating an automated system that can generate chart type explanations from a specification that describes its mark, channel, and data. Based on the findings, we list the implications for making more effective explanations and identify design considerations for future systems.

Spectrum Between the Abstract and the Concrete Some participants wanted an even more concrete explanation than what we provided. Their demands suggest that Dimension 3 (Abstract vs. Concrete) is not binary, but instead has multiple levels between the two extremes. The system must be able to generate different explanations for each level to satisfy the users' needs. For example, an explanation of a bar chart that is more concrete than the ones used in our study would provide more information, such as how many bars there are, which fruit each bar represents, and the exact length of each bar, one by one.

Personalized Approach on Providing Explanation The perceived effectiveness of each strategy differed by each participant. This implies that participants' preferences differ from person to person. Therefore, it makes sense to allow BLV users to select the strategy that they want to see in the explanations, if possible. Users would be able to set their preferred strategy as the system's default explanation strategy and also switch between different explanation strategies when they want to.

On the other hand, it could be better for some choices to be controlled by the system. For example, D2 (*declarative* vs. *procedural*) was the least appreciated dimension, with some participants not even noticing the difference between the two strategies. Surprisingly though, it was the most significant predictor of the participant's performance on the tasks. Furthermore, the fact that a participant preferred one explanation strategy over another did not necessarily mean that they performed better on those tasks. This discrepancy between preference and performance suggests that it is ineffective to grant users too much freedom in selecting what explanation strategy to use. An alternative approach would be providing a list of recommended strategies to use while still considering the user's preferences.

Finding Clear Terminology The lexicon for generating explanations must not be too limited. Some chart types, to accurately describe their peculiar geometry, require technical terms in their explanation. The user might not understand them, so alternative vocabularies to express abstract concepts in layman's terms are necessary. This results in a trade-off between the accuracy of the explanation and its difficulty. In such situations, deciding which terminology best supports comprehension would rely upon the user. BLV users with mathematical backgrounds may prefer technical terms such as "polar coordinates", while others may prefer simpler terms such as "circle", which are easier to grasp but less precise. Understanding the nature of such a trade-off is required to generate an effective explanation.

Chart Type Learning Order For leveraging prior knowledge to be effective, the user must know a chart type that is very similar to the given unfamiliar chart type. Based on this observation, it is possible to create a sequence of chart types that start from one of the user's prior knowledge to the target unfamiliar visualization, where adjacent chart types have high similarity. This sequence can act as a curriculum for the user to learn sequentially. For example, explaining a violin plot could be preceded by explaining a histogram. How histogram looks could be explained by comparing it to a bar chart, which is one of the most basic chart types. Then, a violin plot could be explained by comparing it to a histogram since they have many similarities. Sequential learning like in this case, can be more effective than suddenly trying to learn something that is very dissimilar to one's prior knowledge.

Preventing and Resolving Misconception Since there is no trivial way for BLV users to verify whether their mental model of the visualization is correct or not, it can be very hard to resolve misconceptions. Participants of our study showed various misconceptions about the chart types that were explained to them. Therefore, explanations must focus on preventing these misconceptions. For instance, when a user recalls a chart type that they already know, it can prime the user favorably as much as it can unfavorably. Thus, the explanation must focus on toning down this priming effect by emphasizing the differences between the two chart types more than the similarities. Documenting commonly misunderstood concepts (e.g., stacking) and explaining them in more detail when necessary could also help.

5.1.1 *Envisioning an application.*

We can envision a real-world application based on the design considerations we listed above. This application can be an Internet browser (e.g., Chrome) extension that detects digital visualization on the current web page. The application can then extract its specification to know what chart type it is. If it is not a basic chart type that many users are familiar

Figure 7: Stacked bar charts drawn by P23 (left) and P20 (right). P23 stacked the bars in parallel, just like a grouped bar chart, instead of serially stacking bars. P20 overlayed a bar in front of another. Both drawings show misconceptions about stacking.

with, it can generate an explanation for the chart type using our templates to automatically insert the explanation in the alternative text of the visualization. The user will be able to select the desirable strategy from multiple options according to their preference. The application can also collect preference data from multiple users to learn which strategies are preferred the most for a particular chart type and can recommend the corresponding strategy.

Our findings can also inform the future design of chart question answering (QA) systems, which allow users to pose natural language queries. Questions like "What is a treemap?", which is highly likely to be asked if the alt text mentions a chart being a treemap, can be answered using an approach similar to ours. More complex questions like "How do treemap and bar chart differ?" and "Tell me how to draw it" can also be answered by our system. Our findings on which terminology to use in the explanation can also help QA systems formulate their answers.

5.2 Evaluating Visualization Perception of BLV Users

Before designing the study, we brainstormed many ideas on how to capture BLV participants' visual understanding of visualizations. Inspired by prior work that demonstrated BLV individuals actively visualize the visualizations while listening to the alternative text [14], we devised a way to prompt participants to draw the visualizations they constructed mentally. For example, in our study, we asked participants to prepare sheets of letter-sized paper and a thick sharpie and to use the full paper using the corner of the paper as a reference for their drawing. By prompting participants to draw, not only can we measure how they construct the image, but also we can get interesting insights into their misconceptions that are revealed through the drawings. However, there is no evaluative metric for assessing one's understanding of chart type from their drawings, which is an exciting topic for future work.

We believe that this method has some potential. Many participants shared that they enjoyed the chart drawing part of the interview. P2 said "This was kind of fun." and P20 "I didn't struggle with the drawing piece as much as I thought I would." P17 optimistically said "it was a little bit challenging, but I think that I felt (I was) able to do it." Besides enjoyment, the activity of drawing was not easy, especially for those without functional vision. Mentally keeping track of their drawings was challenging. Participants struggled with drawing circles in particular. When drawing a colored bubble chart, P7 showed concern by saying "(drawing circles) is probably gonna be terrible." P23 said "I definitely felt like the lines or bars weren't always meeting the axes, or it wasn't always like a perfect shape."

Some participants desired to have assistive drawing tools, such as tactile boards, to complete this task. Further research can leverage assistive technology to design a better measure.

6 LIMITATION & FUTURE WORKS

We found several areas of improvement in the explanations. For example, some participants wanted to know when and why a chart type is commonly used. For example, a scatterplot is mainly used to find the correlation between two variables, and a violin plot to understand the distribution of data. This information can hint at the chart's visual appearance, as suggested by the positive correlation between QA score and drawing score. Future researchers could explore ways to automatically extract the information on usage from visualizations.

Many more chart types exist than the 50 that we curated. For example, our study did not concern chart types that show network data with nodes and edges, nor the three-dimensional (3D) chart types. While they are not as common, it would be essential to investigate those complex types. As mentioned in the results, our method of measuring prior knowledge of the user was sometimes unreliable. A more precise technique to prompt and codify prior knowledge should be studied.

The complexity measure we defined did not always reflect the actual difficulty experienced by the participants. For instance, the colored bubble chart, which was ranked the highest on our complexity scale, was one of the most easily understood chart types. On the contrary, visualizations with fewer variables, such as histogram and violin plots, were among the most low-scoring chart types. We consequently did not explicitly analyze data with this variable. However, future research can define a complexity metric that better models the users' cognitive load.

Our similarity metric between chart types takes a theoretical approach grounded in existing visualization frameworks rather than an empirical approach. Because the metric is intended for the specific purpose of aiding chart type explanation for BLV people, evaluating the metric via separate tests was difficult. We encourage the visualization community to explore how such a metric could be further developed and assessed.

We focus on operationalizing different explanation strategies suggested by education theories and not on creating fluent natural language text. Future work should apply recent NLP advancements.

7 CONCLUSION

Our work investigates how text explanations can help BLV users' understanding of unfamiliar chart types. We identify 3 dimensions of explanation strategies and create a prototype system that can automatically generate explanations using one of 8 strategies from a chart type specification. We interview BLV participants where they read the generated explanations of unfamiliar chart types, then draw out the chart, and answer the questions. We analyze the findings using both qualitative and quantitative methods. We find that declarative explanation is overall more effective for conveying the visuals of a chart type. The perceived effectiveness of strategies varies by participant, but does not always align with the actual effectiveness. The spatial ability, locus of control, and having functional vision also influence the understanding of explanations. Also, we report valuable qualitative findings on the actual needs and wants of BLV users that should be addressed to make accessible explanations. Based on our results, we present considerations for authoring effective chart type explanations and point to the possibility of a system that can automatically generate explanations that effectively convey the visuals of a chart type to BLV users. We hope our work provide way for future research on the problem of communicating chart type information to BLV users to make visualizations more accessible.

REFERENCES

[1] Robert Amar, James Eagan, and John Stasko. 2005. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE, 111–117.

[2] John R Anderson and Jon M Fincham. 1994. Acquisition of procedural skills from examples. *Journal of experimental psychology: learning, memory, and cognition* 20, 6 (1994), 1322.

[3] David P Ausubel. 1960. The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of educational psychology* 51, 5 (1960), 267.

[4] Cristian Bernareggi, Dragan Ahmetovic, and Sergio Mascetti. 2019. μGraph: Haptic Exploration and Editing of 3D Chemical Diagrams. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility.* 312–317.

[5] Craig Brown and Amy Hurst. 2012. VizTouch: automatically generated tactile visualizations of coordinate spaces. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction.* 131–138.

[6] Albert T Corbett and John R Anderson. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4 (1994), 253–278.

[7] Ashley R Craig, John A Franklin, and Gavin Andrews. 1984. A scale to measure locus of control of behaviour. *British Journal of Medical Psychology* 57, 2 (1984), 173–180.

[8] Patricia C Duttweiler. 1984. The internal control index: A newly developed measure of locus of control. *Educational and psychological measurement* 44, 2 (1984), 209–221.

[9] Christin Engel, Emma Franziska Müller, and Gerhard Weber. 2019. SVGPlott: an accessible tool to generate highly adaptable, accessible audio-tactile charts for and from blind and visually impaired people. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments.* 186–195.

[10] Christin Engel and Gerhard Weber. 2017. Analysis of tactile chart design. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments.* 197–200.

[11] Tera M Green and Brian Fisher. 2012. Impact of personality factors on interface interaction and the development of user profiles: Next steps in the personal equation of interaction. *Information Visualization* 11, 3 (2012), 205–221.

[12] Enamul Hoque, Parsa Kavehzadeh, and Ahmed Masry. 2022. Chart Question Answering: State of the Art and Future Directions. *arXiv preprint arXiv:2205.03966* (2022).

[13] Shakila Cherise S Joyner, Amalia Riegelhuth, Kathleen Garrity, Yea-Seul Kim, and Nam Wook Kim. 2022. Visualization Accessibility in the Wild: Challenges Faced by Visualization Designers. In *CHI Conference on Human Factors in Computing Systems.* 1–19.

[14] Crescentia Jung, Shubham Mehta, Atharva Kulkarni, Yuhang Zhao, and Yea-Seul Kim. 2021. Communicating Visualizations without Visuals: Investigation of Visualization Alternative Text for People with Visual Impairments. *IEEE transactions on visualization and computer graphics* 28, 1 (2021), 1095–1105.

[15] Nam Wook Kim, Shakila Cherise Joyner, Amalia Riegelhuth, and Y Kim. 2021. Accessible visualization: Design space, opportunities, and challenges. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 173–188.

[16] Gregory Kramer, Bruce Walker, Terri Bonebright, Perry Cook, John H Flowers, Nadine Miner, and John Neuhoff. 2010. Sonification report: Status of the field and research agenda. (2010).

[17] Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. 2006. Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization.* 1–5.

[18] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2016. Vlat: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 551–560.

[19] Zhengliang Liu, R. Jordan Crouser, and Alvitta Ottley. 2020. Survey on Individual Differences in Visualization. *Computer Graphics Forum* 39, 3 (2020), 693–712. https://doi.org/10.1111/cgf.14033

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

[21] Kim Marriott, Bongshin Lee, Matthew Butler, Ed Cutrell, Kirsten Ellis, Cagatay Goncu, Marti Hearst, Kathleen McCoy, and Danielle Albers Szafir. 2021. Inclusive data visualization for people with disabilities: a call to action. *Interactions* 28, 3 (2021), 47–51.

[22] Joel Michael. 2006. Where's the evidence that active learning works? *Advances in physiology education* (2006).

[23] Tamara Munzner. 2014. *Visualization analysis and design.* CRC press.

[24] John G Neuhoff. 2019. Is sonification doomed to fail? Georgia Institute of Technology.

[25] Jean Piaget. 2013. *Origin of Intelligence in the Child: Selected Works vol 3.* Routledge.

[26] Rémi Piatek and Pia Pinger. 2010. Maintaining (locus of) control? Assessing the impact of locus of control on education decisions and wages. *Assessing the impact of locus of control on education decisions and wages* (2010), 10–093.

[27] Melanie A Revilla, Willem E Saris, and Jon A Krosnick. 2014. Choosing the number of categories in agree–disagree scales. *Sociological methods & research* 43, 1 (2014), 73–97.

[28] Henry L Roediger. 1990. Implicit memory: Retention without remembering. *American psychologist* 45, 9 (1990), 1043.

[29] Alper Sarikaya and Michael Gleicher. 2017. Scatterplots: Tasks, data, and designs. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 402–412.

[30] Ather Sharif, Sanjana Shivani Chintalapati, Jacob O Wobbrock, and Katharina Reinecke. 2021. Understanding screen-reader users' experiences with online data visualizations. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility.* 1–16.

[31] Linda Sheppard and Frances K Aldrich. 2001. Tactile graphics in school education: perspectives from teachers. *British Journal of Visual Impairment* 19, 3 (2001), 93–97.

[32] Larry R Squire. 1987. Memory and brain. (1987).

[33] Chris Stolte, Diane Tang, and Pat Hanrahan. 2002. Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics* 8, 1 (2002), 52–65.

[34] Ryo Suzuki, Abigale Stangl, Mark D Gross, and Tom Yeh. 2017. Fluxmarker: Enhancing tactile graphics with dynamic tactile markers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility.* 190–199.

[35] Melih Turgut. 2015. Development of the spatial ability self-report scale (SASRS): reliability and validity studies. *Quality & Quantity* 49, 5 (2015), 1997–2014.

[36] Maria C Velez, Deborah Silver, and Marilyn Tremaine. 2005. Understanding visualization through spatial ability differences. In *VIS 05. IEEE Visualization, 2005.* IEEE, 511–518.

[37] Rafael Veras and Christopher Collins. 2019. *Discriminability Tests for Visualization Effectiveness and Scalability.* Technical Report arXiv:1907.11358. arXiv. http://arxiv.org/abs/1907.11358 arXiv:1907.11358 [cs] type: article.

[38] B. N. Walker and L. M. Mauney. 2010. Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. *ACM Transactions on Accessible Computing* 2, 3 (2010), 1–16. https://doi.org/10.1145/1714458.1714459

[39] Jagoda Walny, Samuel Huron, Charles Perin, Tiffany Wun, Richard Pusch, and Sheelagh Carpendale. 2017. Active reading of visualizations. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 770–780.

[40] Ruobin Wang, Crescentia Jung, and Y Kim. 2022. Seeing Through Sounds: Mapping Auditory Dimensions to Data and Charts for People with Visual Impairments. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 71–83.

[41] Tetsuya Watanabe, Toshimitsu Yamaguchi, and Masaki Nakagawa. 2012. Development of software for automatic creation of embossed graphs. In *International Conference on Computers for Handicapped Persons.* Springer, 174–181.

[42] Hadley Wickham. 2010. A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics* 19, 1 (Jan. 2010), 3–28. https://doi.org/10.1198/jcgs.2009.07098

[43] Leland Wilkinson and Graham Wills. 2005. *The grammar of graphics* (2nd ed.). Springer, New York.

[44] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning.* PMLR, 2048–2057.

[45] Yalong Yang, Kim Marriott, Matthew Butler, Cagatay Goncu, and Leona Holloway. 2020. Tactile presentation of network data: Text, matrix or diagram?. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–12.

[46] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. 2008. Data sonification for users with visual impairment: a case study with georeferenced data. *ACM Transactions on Computer-Human Interaction (TOCHI)* 15, 1 (2008), 1–28.

[47] Caroline Ziemkiewicz, Alvitta Ottley, R. Jordan Crouser, Krysta Chauncey, Sara L. Su, and Remco Chang. 2012. Understanding Visualization by Understanding Individual Users. *IEEE Computer Graphics and Applications* 32, 6 (Nov. 2012), 88–94. https://doi.org/10.1109/MCG.2012.120 Conference Name: IEEE Computer Graphics and Applications.