J Vis
https://doi.org/10.1007/s12650-022-00888-x

REGULAR PAPER

Yihan Hou · Haotian Zhu · Hai-Ning Liang · Lingyun Yu

A study of the effect of star glyph parameters on value estimation and comparison

Received: 14 August 2022 / Accepted: 26 September 2022

Abstract Star glyph is widely used as a typical radial plot to visualize multi-dimensional data, allowing the comparison of multiple attributes while displaying them. Though many alternative designs for star glyphs exist, there is no experimental evidence for the impact of the encoding methods in understanding and comparing multi-dimensional values. This paper reports a controlled user experiment exploring the effect of fundamental design parameters of star glyphs on efficiency and accuracy. Three design parameters (position, length, and area) were tested through four tasks (finding extremes, retrieving values, comparing values of adjacent attributes, and comparing values of non-adjacent attributes) with two dimensions (low and high). In general, the results show a significant difference in efficiency in the tasks of finding extremes, comparing values for both adjacent attributes and non-adjacent attributes for the design parameter of area encoding and length encoding. Length encoding can improve the efficiency of judgment in all comparison tasks. However, surprisingly, in the finding extremes task, the augmented points affect users' efficiency on tasks with high dimensions. In terms of accuracy, no significant difference was observed among the different design parameters in all tasks. Furthermore, we report the strategies participants used in completing the tasks, users' preference of different designs, and the level of confidence in making decisions. Based on these findings, we propose design considerations for star glyphs regarding the effect of different parameters.

Keywords Quantitative evaluation · Glyph-based techniques

1 Introduction

Star glyphs, encoding data with a radial layout, have become a ubiquitous visualization technique for showing hierarchical structures and multi-dimensional data (Draper et al. 2009; Tominski et al. 2004). In a star glyph, the coordinate axes are circularly arranged around a central point and the data points are drawn at positions on the respective axes and are connected with a line to form a contour (Chambers et al. 2018). Radar chart is a typical type of star glyph. Another common star glyph is the Nightingale rose chart, which can be regarded as a polar coordinates bar chart (Wilkinson 2012). Recently, many researchers have focused

Yihan Hou conducted this research while affiliated with Xi'an Jiaotong-Liverpool University.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12650-022-00888-x.

Y. Hou
Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

H. Zhu · H.-N. Liang · L. Yu (✉)
Xi'an Jiaotong-Liverpool University, Suzhou, China
E-mail: Lingyun.Yu@xjtlu.edu.cn

Published online: 03 November 2022

on novel glyph design and add visual elements to the basic radial structure, such as contours and reference lines. Star glyph visualizations have been widely used in social (Wu and Qu 2018), medical (Seide et al. 2021), industrial (Hongliang et al. 2008), and technical (Caporaso et al. 2020; Peng et al. 2019) fields.

Many evaluation studies have been conducted comparing star glyph visualization with other chart visualizations in understanding multivariate data. For example, Stafoggia et al. (2011) have compared radar plots with target plots and spie charts with clinical data. Opach et al. (2018) compared the performance differences between star glyph and polyline glyph for six visualization tasks, including data estimation, search, and comparison. Albo et al. (2015) explored the users' preference for three radial visualizations with different tasks and dimensions. These works focus on the overall design of the star glyph without considering the effect of individual design parameters.

The evaluation work on the individual visual parameters of star glyph is limited. Klippel et al. (2009) explored the impact of the shape of the star glyph in the classification tasks through two experiments. Fuchs et al. (2014) conducted three experiments on the contours and reference structures of the star glyph to explore their impact on detecting data similarity. However, these works focus on the impact of different designs on a specific high-level task. Skau and Kosara (2016) tested the impact of individual data encodings (such as the arc, the angle and the area) in pie and donut charts. Their results showed that angle is the least effective visual encoding for both charts. Cai et al. (2018) conducted three experiments to further study the individual visual elements of pie and donut charts in the proportion estimation. Through their experiments on the fundamental design parameters, additional visual cues (such as tickmarks, central point) are proposed to improve the accuracy of proportion judgments. These studies offer a basis for understanding how visualizations are constructed through visual elements and how these visual designs could impact user perception and cognition in low-level visualization tasks. The findings can be used to guide the design of effective visualizations in real-world applications.

Our work aims to compare the design parameters of star glyph in the value estimation and comparison tasks on efficiency, accuracy, and user experience. We conducted a within-subjects study to investigate three design parameters, four visualization tasks and two data dimensions in star glyph visualizations, including design parameters (Borgo et al. 2013) (*position*, *length*, and *area*), tasks (Amar et al. 2005) (*finding extreme*, *retrieving values*, *comparing adjacent values*, and *non-adjacent values*), and data dimensions (Saary 2008; Kandogan 2000; Dy et al. 2021) (*low dimensions* and *high dimensions*). We aim to analyze the effect of design parameters in low-level visualization tasks on *efficiency*, *accuracy*, and *user experience*.

The results of our study show that, in terms of efficiency, area encoding has advantages in all three value comparison tasks, while augmented position (*position encoding*) decreases the efficiency in the value estimation task. The auxiliary lines (*length encoding*) can improve the accuracy of the comparison tasks. In terms of user preferences, in general people prefer simple designs which do not consist of many visual elements. Area encoding is preferred subjectively because it makes people more confident in making judgments. Based on these findings, we propose design suggestions for star glyphs regarding the effect of different parameters.

2 Related work

Several works have proposed novel glyph visualizations to show multivariate data in business Liu et al. (2021), education Wu and Qu (2018), industry Deng et al. (2021), transportation Zeng et al. (2017), etc. As a typical glyph type, the study on star glyphs has a long history, with the radar chart and rose chart as the two well-known representations, matching the values of different attributions to the axes of the radial layout (Heckert et al. 2002). The evaluation works for glyph visualizations have widely investigated the impact of different glyph designs on user perception and cognition and further guide to design effective visualizations for real-world applications. Zhao et al. (2019) conducted an experiment for evaluating fuzzy clustering analysis using four multi-dimensional visualization techniques. Fuchs et al. (2015) provided a review of quantitative experiments on various data glyphs from the past 70 years. However, the study on the effect of design parameters of star glyphs is rather limited. In this section, we first introduce the previous evaluation works for star glyphs and then discuss other glyph visualizations.

Several works compared star glyphs with other visualization forms. For example, Stafoggia et al. (2011) compared three visualization designs, including target plots, radar plots, and "spie" charts. They reported user performance for understanding an eight-dimensional healthcare data among these three designs. While the effect of the order of data attributes in different designs has been noticed, especially for radar charts,

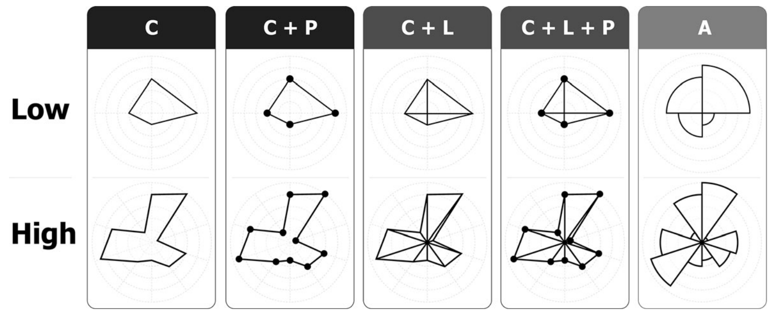


Fig. 1 Five star glyph designs in the user study: C: contour, C + P: contour and augmented position, C + L: contour and length, C + L + P: contour, length and augmented position, A: area encoding. Low: low-dimensional cases with four variables, High: high-dimensional cases with ten variables

however, the reason why the radar charts have potential flaws has not been further explored. To find out if star glyph and polyline glyph have advantages in any specific tasks, Opach et al. (2018) constructed a formal study to compare the performance differences between the two glyphs for six low-level visualization tasks (Amar et al. 2005), including estimation, searching, comparison, etc. Their results showed that the polyline glyph is more beneficial for reading specific values, while star glyph is beneficial for searching tasks. Albo et al. (2015) explored the user performance on three radial visualization schemes, Flowers, Radar, and Circle. They measured the correctness rate, completion time on the tasks, and participants' subjective opinions. Dy et al. (2021) investigated the differences in user performance on four types of chart visualizations in completing decision-making tasks based on multivariate data, including scatter plot matrices, parallel coordinates, heat maps, and radar charts. These works are similar to discussing star glyph as a visualization type, and ignore visual parameters Dy (Borgo et al. 2013) of these glyphs, such as position, length, and areas. While these works provide valuable suggestions for designing star glyphs, it is still unclear how the individual visual parameters influence user performance in low-level visualization tasks.

Klippel et al. (2009) explored how the shape of the star glyph influences users' performance in the classification task and whether the effect of shape can be counteracted by adding coding for the color of the data value line. Fuchs et al. (2014) conducted three experiments around the contours and reference structures of the star glyph to explore their impact on detecting data similarity. These two works focused on the effect of the fundamental visual elements in the visualization. However, they focus on specific high-level tasks, such as data classification and similarity detection. As discussed by Matthew et al. Brehmer and Munzner (2013), the completion of complex tasks is influenced by multiple simple tasks. In this work, we aim to explore how star glyphs are constructed through fundamental visual elements and how these visual designs influence user perception and cognition in low-level visualization tasks.

In order to study how the effect of individual visual elements is explored through low-level visualization tasks, we also reviewed evaluation works for other visualization forms. Cai et al. (2018) presented three experiments on how different visual elements influence the proportion estimation accuracy of Doughnut charts and Pie charts. They reported that marking the center of the doughnut chart or adding tick marks at intervals can improve the accuracy of proportion judgments. Skau and Kosara (2016) explored the effect of perceptual factors, such as the length, center angle, and segment area, used to encode data in pie charts and doughnut charts. Xia et al. (2021) conducted an empirical study about human perception of scatterplots, and summarized the visual factors' influences on clustering perception. Furthermore, Wei et al. (2020) presented a user study on the perceptual bias of three visual features (numerosity, correlation, and cluster separation) in scatterplots caused by geometric scaling. These studies focus on the effect of fundamental visual features and the findings can be used to inspire design considerations in various applications.

In summary, much has been done in evaluating star glyph for various visualization tasks for multivariate data, as well as the development of novel designs for star glyph. Not much work has been put into investigating the effect of individual design parameters in low-level visualization tasks. In our experiments, we refer the following works in selecting the visualization parameters (Borgo et al. 2013; Fuchs et al. 2014; Cai et al. 2018; Skau and Kosara 2016), low-level tasks (Amar et al. 2005; Opach et al. 2018), and other experimental conditions, such as dimensions (Saary 2008; Kandogan 2000; Dy et al. 2021).

3 Study design

In this section, we present the experiment to study the effect of design parameters of star glyph (position, length, and area). We aim to answer the questions as follows:

RQ1 [Efficiency] Which parameters have an impact on the efficiency of task completion?
RQ2 [Accuracy] Which parameters have an impact on the estimation and comparison accuracy of star glyph?
RQ3 [User Experience] Which designs are preferred? Which designs can make people more confident in completing tasks?

3.1 Hypotheses

RQ1 [Efficiency] We expected that all three parameters would have an impact on the efficiency. Specifically, among position, length, and area encoding designs, users would take less time in the comparison tasks using area encoding since size is an effective visual channel in general according to previous research (Li et al. 2010). Augmented points (position highlighted) would reduce the time cost in completing tasks since additional visual elements can be jointly perceived together with other visual channels (for instance, contour with highlighted position). Furthermore, we were also interested in investigating the interaction effect of different design parameters in the value estimation and comparison, for instance, position plus length. We also assumed that participants would need more time to complete high dimension questions than low dimension questions, since the increase in complexity of the visualization also increases users' cognitive load. Thus, in general, we hypothesize that design parameters and dimensions would affect the efficiency in the value estimation and comparison. Specifically, we have the following hypotheses.

H1 Different design parameters will affect the efficiency in value estimation and comparison:
H1.1 Area encoding is faster than position and length encoding in the comparison tasks.
H1.2 Length encoding improves the efficiency in the value estimation task.
H1.3 Augmented points can improve the efficiency in the value estimation task.
H1.4 People can complete tasks with low dimensions faster than with high dimensions.

RQ2 [Accuracy] We expected that design parameters would affect the accuracy of task completion. Since encoding information by location facilitates accuracy (Cleveland and McGill 1984; Healey et al. 1996), we hypothesized that highlighted position would lead to more accurate results in the task of *retrieving values*. Specifically, we have the following hypotheses.

H2 Different design parameters will affect the accuracy in value estimation and comparison:
H2.1 Length encoding improves the accuracy in the comparison tasks.
H2.2 Highlighted position improves the accuracy of the value estimation task.
H2.3 People can complete tasks with low dimensions more accurately than with high dimensions.

RQ3 [User Experience] According to previous research (Peng et al. 2004), we expected that complexity of glyph may affect users' preference and level of confidence as follows.

H3 People have higher confidence in answering questions with simple designs (contour and area) than using complex designs (contour + length, contour + highlighted position, contour + highlighted position + length).

H4 Simple designs (contour and area) are more preferred than complex designs (contour + length, contour + highlighted position, contour + highlighted position + length).

3.2 Experiment design

3.2.1 Design alternatives

After reviewing the existing literature on star glyph visualizations, we chose the three fundamental designed parameters (*position, length, and area*). Based on these, we generate five design variations of star glyphs: Contour (C), Contour with Position highlighted (C + P), Contour with Length (C + L), Contour with Length and Position highlighted (C + L + P), and Area encoding (A) (as shown in Fig. 1).

C: Baseline design. Its design parameters are *{Contour}*. The data attribute is encoded through the position of a point on the corresponding axis in the polar coordinate system. The points are connected by a contour line to form a closed shape (Fig. 1: C (Contour)).

C + P: Its design parameters are *{Contour, Augmented Position}*. The data attribute is encoded through the position of a point on the axis and the point is highlighted (Fig. 1: C + P (Contour and Augmented Position)).

C + L: Its design parameters are *{Contour, Length}*. Except for the contour, the data attribute is also encoded by the length of the line from the central point of the polar coordinate system to the position of the data point (Fig. 1: C + L (Contour and Length)). However, the position of the point is not highlighted.

C + L + P: Its design parameters are *{Contour, Length, Augmented Position}*. Except the contour, the data attribute is also encoded through the length of the line and the position of the data point is highlighted (Fig. 1: C + L + P (Contour, Length and Augmented Position)).

A: Its design parameter is *{Area}*. Multiple sectors are circularly arranged around a central point, each of which represents one dimension of data (Fig. 1: A (Area encoding)). The angles of the sectors are equal. The position of the points is not highlighted since the area is used to encode data value.

All five designs mentioned above have appeared in the literature using star glyphs and have been widely used in information visualization. However, no previous research has been done to compare them in value estimation and comparison.

3.2.2 Tasks

Four low-level visualization tasks were selected in the experiment based on the task model proposed by Amar et al. (2005). However, considering that the sorting task rarely occurs in practice, it was not selected. Instead, comparison tasks with two values are often concerned, for instance, when the star glyph is used for showing student's marks of different subjects. Thus, we add the comparison task in the experiment. Furthermore, considering that time cost to complete the task may be related to whether the two attributes are adjacent in radial layout, we divide the comparison tasks into two categories: comparing adjacent attributes and non-adjacent attributes.

T1: Finding extremes. *Which data attribute has the maximum / minimum value in the star glyph?*

T2: Retrieving values. *What is the value of a specific data attribute?*

T3: Comparing values of adjacent attributes. *Which of the two adjacent attributes has greater value?*

T4: Comparing values of non-adjacent attributes. *Which of the two non-adjacent attributes in the star glyph has greater value?*

3.2.3 Dimensionality

To investigate the performance of the design parameters at different complexity levels, we tested all design variations in both low-dimensional and high-dimensional cases. This can broaden the valid range of our experiments, making the conclusions applicable to multivariate data of different dimensions. Although there were no constraints on the number of dimensions that can be identified in star glyphs, the previous studies showed that users' performance on task completion decreased when the observed dimensions increased (Fuchs et al. 2014). For this reason, many studies on multivariate data visualization constrained the number of dimensions between 3 and 12 (Saary 2008; Kandogan 2000; Dy et al. 2021). In our experiments, we set the dimensionality as follows:

🅛 Low-dimensional data consist of *4 dimensions* which is the minimum setting that can satisfy the requirement to compare two non-adjacent attributes.

Ⓗ High-dimensional data consist of *10 dimensions*, referring to the setup used in the user study of Fuchs et al.'s work (Fuchs et al. 2014).

3.3 Participants

We invited 25 participants in the user study, including 12 males and 13 females. The user study is a within-subjects (or repeated-measures) design, which means that the same participant test all the conditions. Their age range was between 19 and 28 (mean = 22.76 and SD = 1.64). Among them, 23 participants reported familiarity with simple chart visualizations such as bar charts and pie charts. Their backgrounds were diverse, with some having a background in data visualization or in related fields, such as computer science and design area.

3.4 Procedure

The experiment was conducted online individually. We communicated with the participants through online meeting tool and observed their behavior through the shared screens. Participants were required to use a moderately sized desktop monitor and complete the tasks through a mouse and keyboard.

Before the experiment, we first introduced the purpose of the research project and asked them to fill up a questionnaire with basic personal information. The user study started with a training session. All participants were required to follow a tutorial that introduced position, length, and area encoding in the star glyphs. The main idea of the tutorial is to teach the participants how to read the alternative design of the star glyphs. Moreover, we also introduced the testing system and encouraged the participants to use the system to complete similar tasks. The data in the training session were not collected for analysis. The training session lasted about 15 min. However, the participants were asked to take as much time as needed to get familiar with the system. Only when the participants reported that they were ready, the formal user study would start. In the formal experiment, the participants were required to complete four tasks as accurately and quickly as possible. To prevent the learning effect, the first result from each set of experiments was considered as practice and was not included in the data analysis.

Figure 2 shows the interface of the experiment. The task instruction is shown at the top of the screen, and below is the question and answer area which shows a star glyph. Based on the tasks, the participants were asked to click and select a dimension with a mouse, or type an estimated value using a keyboard. Participants were allowed to modify their answers by changing and re-submitting their answers. The time consumption as well as all mouse and keyboard events were recorded for subsequent analysis. Whether or not the answer is modified is used to measure the degree of confidence.

3.5 Data

In our experiments, six datasets were used to generate glyphs. Low-dimensional datasets contain 4 data attributes and high-dimensional datasets contain 10 attributes. The values range from 1 to 10 without duplication. For example, the attribute values in the low-dimensional dataset in Fig. 1 are [2, 8, 6, 4], and the attribute values in the high-dimensional dataset are [3, 5, 6, 1, 10, 8, 2, 7, 9, 4]. We use two ways to avoid

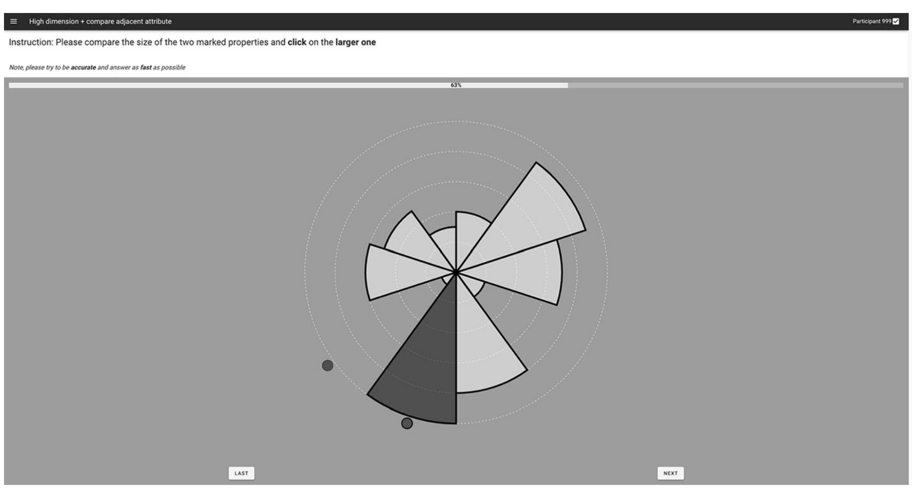


Fig. 2 Experiment interface

the learning effect. First, we randomly select one variable (for the retrieving value task) or two variables (in the comparison task), which make participants unable to predict which variable of glyph they should focus on. Second, we used a fifth-order Latin square to counterbalance the order of five star glyph designs presented to the participants.

3.6 Design

We used a repeated-measures design with the within-subject independent variable five alternative star glyph designs (C, C + P, C + L, C + L + P, A). As mentioned above, two difficult levels of tasks were used, corresponding to the low dimensions (4 dimensions) and high dimensions (10 dimensions). Each participant performed three trials for each difficulty level and task. After each task, the participants were asked complete a questionnaire for their subjective opinion. When all tasks were completed, the participants had a semi-structured interview discussing their feelings when completing the tasks. In summary, our experiments include (Table 1).

4 Results

In this section, we report statistically results on *efficiency*, *accuracy*, and *user experience*, respectively.

4.1 Measures

The experiment measured the time and correctness rates of each participant in completing the tasks. In addition, we also collected users' confidence in answering questions, their preferences for different designs, as well as the strategies used in completing tasks through questionnaires and interviews.

- **Time consumption** The time consumption is computed from the moment when the participant enters a new trial to the moment when the answer is submitted. To ensure the accuracy of the experimental results, time spent on unrelated interactions is subtracted, such as skipping to the next page.
- **Error rate** The error rate is determined based on the final results submitted by the user. If the participant modifies the answer to be correct before submission, the final result is considered correct. Correspondingly, the time spent on the revision is added to the time consumption.
- **Level of confidence** Participants' confidence levels were collected on a 5-point Likert scale, which was completed after completing all tasks for each design variation. For each design, the participants were asked to rate how confident they were in completing four tasks with it. In addition, we measure user confidence by whether they modify their answers and by whether they hesitate between different choices. This information is gathered from mouse trajectories.
- **Preference of design** Participants' preference for each design variation is measured by a 5-point Likert scale. We also collected the participants' preferences for the five designs by asking them to rank five designs in the interviews.
- **User strategy** Users' strategies for completing the four tasks using five design variations were collected during the interview phase. We further analyzed the users' mouse trajectories to verify the strategies they mentioned.

Table 1 Experiment design and total number of trails

5	Alternative designs	×
2	Dimensions (low, high)	×
3	Repetitions	×
4	Tasks	=
120	Trials per participants	×
25	Participants	=
3000	Trails in total	

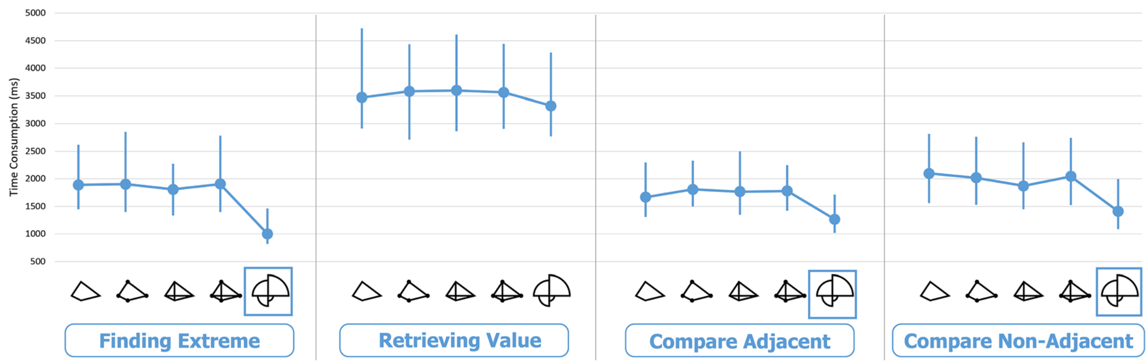


Fig. 3 Completion time for five glyph designs in four low-level visualization tasks: finding extreme, retrieving value, comparing adjacent values and non-adjacent values

4.2 Completion time

We compared the time consumption of three design parameters (position, length, and area) on four tasks (finding extremes, retrieving values, comparing values of adjacent attributes, and comparing values of non-adjacent attributes), respectively.

After the normality test, the nonparametric Friedman's test is used to show the significant effect on time consumption between different design parameters on four tasks. On average, each *task* took 2.48 s (SD = 1.49). The statistical results for different design parameters are shown below.

- **Area encoding** We found a significant effect of *area encoding* in all data comparison tasks: finding extremes (d.f. = 175, p <.001), comparing values of adjacent attributes (d.f. = 175, p <.001), and comparing values of non-adjacent attributes (d.f. = 175, p <.001). However, no significant difference is found in the value estimation task (retrieving values). Area encoding has distinct lower time consumption as shown in Fig. 3. Thus, H1.1 is confirmed.
- **Length encoding** The analysis of task completion time did not show a significant effect for Length encoding (C + L, C + L + P) in all tasks. Thus, H1.2 is rejected. However, we observed differences in the mean values for the length encoding condition, so we decided to further investigate the effect of length encoding. We re-grouped the questions in the comparison tasks: questions with small differences in values (differences of 1 to 3) and questions with large differences in values (differences of 4 to 9). In the question of big value differences (differences of 4 to 9), the analysis shows a significant difference in completion time with the length encoding designs being significantly faster (d.f. = 187, p <.001), as shown in Fig. 4.
- **Augmented position** The analysis of completion time showed a significant difference for finding extreme tasks with the Augmented position condition being significantly slower than without Augmented position (d.f. = 349, p <.05), as shown in Fig. 6. However, no difference is found in the other three tasks. Thus, H1.3 is rejected.

Besides, we find that time cost on the tasks with low dimensions (4 attributes) is significantly lower than that with high dimensions (10 attributes) in the tasks of finding extremes (d.f. = 299, $p < .001$), comparing values of adjacent attributes (d.f. = 298, $p < .005$), and comparing values of non-adjacent attributes (d.f. = 298, $p < .001$). However, no significant difference is found in retrieving values. Thus, H1.4 is partly confirmed.

4.3 Correctness rate

The results show an overall high accuracy in all tasks (mean = 99%, SD= 0.95%). No significant difference is detected between the three design parameters, which reject H2.1, H2.2 and H2.3. One possible reason is that all four tasks chosen in the experiment were low-level tasks that were relatively easy for the participants. Nevertheless, we observe by the mean value that the correctness rate of the two comparison tasks improves further when the auxiliary line (length encoding) is provided.

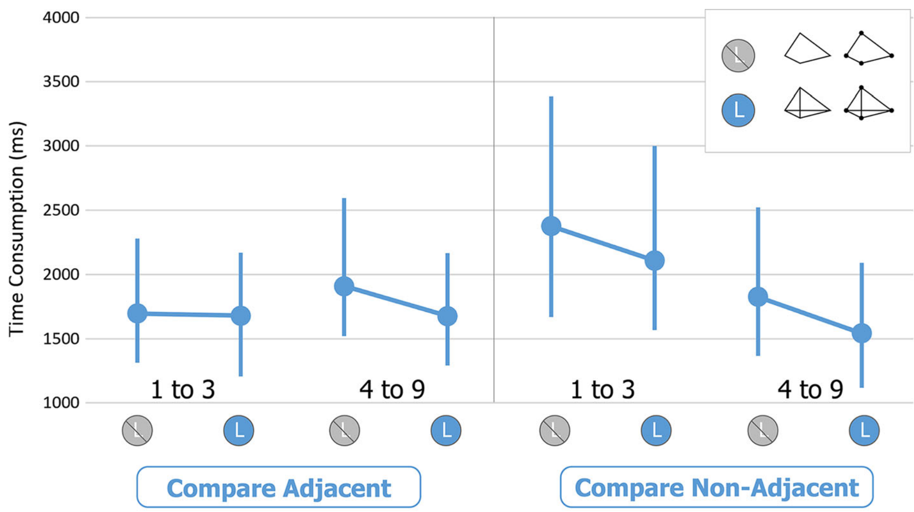


Fig. 4 Completion time for Length encoding designs in comparing adjacent values and non-adjacent values: (C, C + P), (C + L, C + L + P); The data are re-grouped by the difference between two compared attributes

4.4 Log analysis

The interaction logs reveal two patterns in user making the judgments. During the experiment, we noticed that the participants showed two patterns of behavior. Therefore, we analyzed the mouse trajectories to verify the existence of these two patterns.

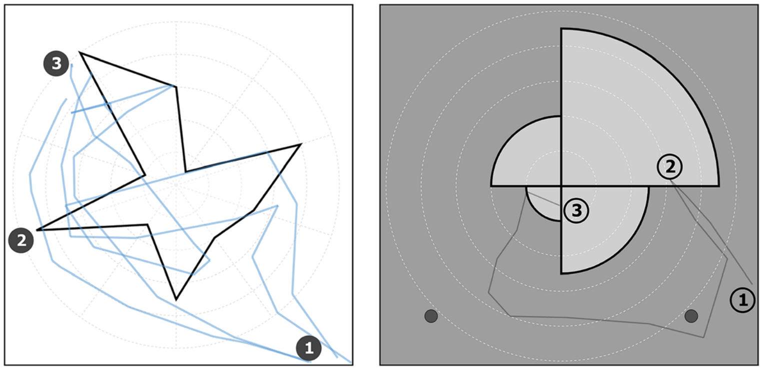
Figure 5a shows the mouse trajectories of four participants in the same trial in the task of "finding the extreme". The numbers in the figure indicate the order of mouse events. When a trial started, the cursor was always moved back to the bottom right of the interface ①. The value of the variable corresponding to ② is 9 and the value of ③ is 10. According to the task, the cursor should start from the initial point ① and move directly to the maximum value ③. However, in the mouse trajectories, the cursor moved first to the position ②, and only afterward to the position ③. This is because the participant hesitated between these two data points which have similar values (values 9 and 10).

Figure 5b shows a scenario in the task of comparing two value for adjacent attributes. The participants were asked to compare two attributes (two marked sectors whose values were 2 and 5). Similar to other conditions, before the experiment we thought that the participants would only focus on the two attributes. However, we found that the participant moved the cursor from the initial point ① to ②, before heading to the correct answer ③. Note that, the attribute under ② was not asked to be compared but the value of the attribute was the greatest value of the dataset. This indicates that the participants may be distracted by a great value in the star glyph even if it was not relevant to the task.

4.5 User feedback

We measured the rating of participants' confidence and preference by a 5-point Likert scale. Here, we use related-samples Wilcoxon signed rank test to verify the significance of the data distribution.

- **Level of confidence** Participants show a higher confidence while using area encoding design (mean = 4.6) than other designs (mean = 4.3) in the tasks of *finding extremes* (d.f. = 24, p <.005), *comparing values of adjacent attributes* (d.f. = 24, p <.005), and *comparing values of non-adjacent attributes* (d.f. = 24, p <.001), as shown in Fig. 7. Thus, H3 is partly confirmed. For the use of additional visual elements, such as auxiliary lines (length encoding) and highlighted points (position encoding), the difference in the level of user confidence was not apparent. For the length encoding designs, the star glyphs with auxiliary lines (C + L, C + L + P) (mean = 4.385) give users more confidence than without ones (mean = 4.215). For the position encoding designs, the star glyph with highlighted points (mean = 4.355) gives users more confidence than without points being highlighted (mean = 4.245).
- **Preference** Participants' preferences are shown in Fig. 8, which presents a significant preference in area encoding than other four designs (d.f. = 24, p <.001). Besides, simple designs (C, A) are more preferred



(a) the participant hesitated between two similar values **(b)** the participant was wrongly attracted by the maximum value

Fig. 5 Mouse events in the tasks of **a** finding extreme and **b** comparing adjacent values

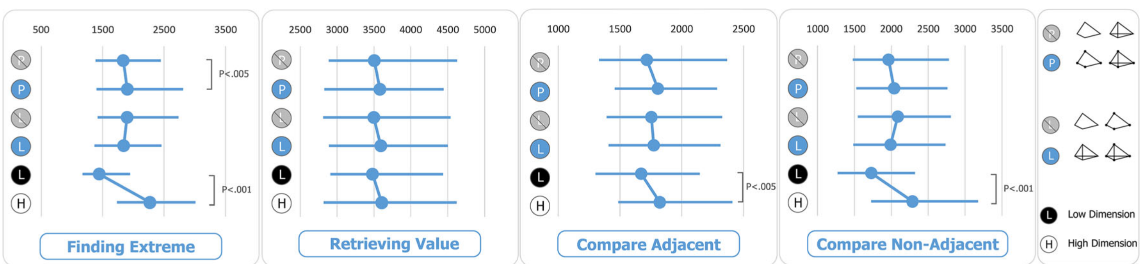


Fig. 6 Completion time comparison for position encoding: (C, C + L), P (C + P, C + L + P); length encoding: (C, C + P), L (C + L, C + L + P); and dimensions: L (Low dimensions), H (High dimensions)

than complex designs (C + P, C + L, C + L + P), with mean of 3.72 and 4.1, respectively. This confirms H4. Since there is not much difference shown between the ranks of the augmented lines (C + L, C + L + P) and the augmented points (C + P, C + L + P), it can be seen that the users' preferences are not sensitive to these parameters.

We interviewed users about their perceptions of different design variations and summarized their strategies for accomplishing different tasks.

- **In general, area encoding is preferred** The participants rated area encoding highly. In the preference ranking, 80% (20 of 25) of the participants ranked area encoding first. They gave feedback that the glyph using area encoding is the most straightforward way through which data values can be easily identified. P20 mentioned that When she used length and position encoding to find extreme values, many times she was first attracted by a prominent value and later found a larger one. However, this issue was not encountered when using area encoding. In addition, area encoding design is consistent across all internal angles, giving participants a sense of uniformity. As P10 reported that "the sense of unity makes me feel confident when answering the questions" and P24 mentioned that "The angle between each neighboring sector is the same, which makes it easy for me to compare adjacent values. Unlike in the other four designs, where I need to determine the slope of the line connecting two vertices". By checking the event logs, we also noticed their hesitation between two similar maximum values in length encoding, but not in area encoding. This comment is also supported in the participants' questionnaire.
- **Augmented position is helpful** More than half of the participants (14 of 25) explicitly stated that the highlighted points in C + P and C + L + P were very helpful to them in making judgments. P19 reported that the augmented position designs were more accurate in identifying specific values because the data points could be easily noticed. Also mentioned by P4: "The round points are easier for me to click on specific values than the sharp angles". However, some participants also claimed that the point might obstruct the view, so that the actual value cannot be easily read. "The point is not necessary since it has a certain size, which can make the valuation inaccurate", reported by P21. However, the statistical results

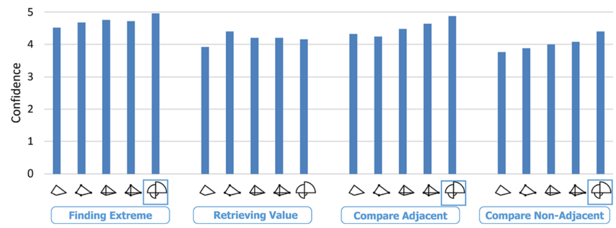


Fig. 7 Degree of confidence when using five star glyph designs to complete four low-level tasks

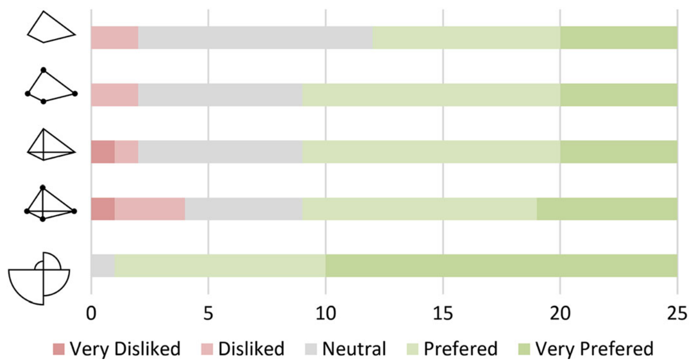


Fig. 8 Users preferences for five star glyph designs

show that the designs of augmented position neither enhance accuracy nor efficiency. This can prove that the subjective feelings of many participants deviate from the actual situation.

- **Length encoding is interfering** About half of the participants (12 in 25) reported that auxiliary lines interfered to some extent with completing their tasks. P25 did not like the design with lines: “they broke the graph into many triangles”. P13 reported that the lines made the whole picture unfocused and confusing. Contrary to the subjective perception, the statistical results showed that the length encoding did not decrease the completion time but slightly increased the efficiency. In addition to this, the length encoding improved accuracy in the comparison tasks.
- **Complex design may bring confusion** The design with both highlighted points and length (C + L + P) received polarized reviews, with ten participants complained that too many visual elements would cause confusion. Among them, one participant mentioned “I did not know what I was supposed to focus on and it made a bad user experience”. However, seven participants rated the design as their top two favorites. “It is helpful since it provides the most information”, said P1. P9 claimed that he liked C + L + P most since “There are auxiliary lines to assist the judgments, which makes this design easy to use”.

5 Discussion

In this session, we discuss the findings noticed in the study as well as the limitations of the experiment.

5.1 Findings

Through the interviews, we summarized the strategies that the participants used in the value estimation and comparison tasks with star glyphs. In general, depending on the task, the participants would first determine if it is necessary to know the exact values. For example, in the task of finding extreme values, if the difference among the values is obvious, the participants would only focus on the most outward point/longest line/biggest area in the graph. However, if there are similar values and the value difference cannot be detected at a glance, they would make use of the tick marks to estimate exact values or compare the values based on the positions of the vertex/arc. We found several characteristics of the participants when looking at the star glyph:

- **Position of the data points** In the questionnaires, the participants showed their particular attention to the position of the data points. In both position encoding (C, C + P) and length encoding (C + L, C + L + P), the majority of participants reported that they looked at the points most of the time. Even though lines were explicitly displayed in length encoding designs (C + L, C + L + P), it did not seem to have a significant effect on reading and comparing data values. P1 reported that "The line helps me better notice the position of the point". As mentioned in the previous section, more than half of the participants liked the design of the highlighted points because they thought it greatly enhanced the positions of the points. For area encoding, the participants were mainly concerned with the size of the area. The position of arcs was also noticed in the value retrieving task.
- **Shape of star glyphs** Ten participants reported their feelings about the shape of star glyphs. In the position encoding (C, C + P) and length encoding (C + L, C + L + P) designs, they mentioned that they would expect to see "sharp corners" or "spikes" in the star glyphs. Thus, if no "sharp corners" shown in the star glyph, they would understand that the dataset has no "prominent value" such that they would need to use tick marks to estimate values or compare values. Two participants reported that sometimes even if a value looks prominent, it was not the greatest value. Only one participant with a visualization background (P3) mentioned that he noticed the visualization bias "although the point goes outward and the angle looks sharp, it does not mean that the value is certainly great since the shape of the glyph depends on the adjacent values". This incorrect understanding of shape indicates the importance of the *[order]* of data attributes in star glyphs.
- **Central point** For the position encoding designs (C, C + P), the central point is not visualized. For the rest of the designs, the central points are provided but not highlighted. Six participants specifically discussed the effect of the central points in the questionnaire. Note that the central point is not the geometric center of the star glyph, but the origin of the polar coordinates. For length encoding designs (C + L, C + L + P), P19 and P22 indicated that the auxiliary line makes the location of the central point more explicit and can effectively help them estimate the distance from the data point to the center. P4, P18 and P23 suggested that the designs which do not have auxiliary lines would be better to have the center point visualized. P4 further explained, if the center point is not explicitly provided, the geometric center of the star glyph would be mistakenly regarded as the center point. It would potentially create bias since the geometric center could be far from the actual center point of the polar coordinates if several great values are adjacent.

The test results show that the efficiency for area encoding is significantly higher than the other conditions in the comparison tasks. There are two possible reasons as follows:

- **Size of the area magnifies the difference** In the area encoding designs, the data value is encoded through the radius of the sectors. Compared with length encoding designs (C + L, C + L + P), the difference between the two values looks more obvious since the size of the area perceptually magnifies the difference. Thus, many participants mentioned that they could clearly tell which attribute has greater value in the area encoding design. In addition, in the value estimation task, the statistics do not show that the accuracy of area encoding is less than in other conditions. We assume that, when people estimate the actual value of data attributes through the positions of the arcs.
- **Area encoding reduces the impact of adjacent values** In the area encoding design, each sector is independent of the other, which means that each attribute value is represented independently and does not interfere with each other. However, for the other four designs, the individual data points are connected with a line to form a contour for the glyph, which may affect the correctness of understanding the actual value of individual attributes because of the positions of the attributes in the star glyph. This point has been reflected and discussed in Sect. 4.5.

5.2 Limitations and further discussion

Due to the scope of our study and experimental setup, the conclusions drawn possess some limitations. First, in our experiments, we mainly focus on the effect of position, length, and area encodings in the star graph designs, for instance, the effect of highlighted data points and the visibility of auxiliary lines. In fact, many alternative glyph designs (Borgo et al. 2013) are presented in the previous visualization work. Thus, there are various design parameters that can be further investigated in different visualization tasks, such as contour lines, the center points. The effect of these design parameters can also be linked together in complex visualization tasks. The results and findings from such experiments on perception have great value in designing complex glyph visualizations. Second, currently we only focus on the fundamental visual elements of star glyphs without considering possible user interactions. However, user interactions play an important role in many visualization tasks, for instance, filtering, selection (Chen et al. 2020) and decision making (Zhang et al. 2021). Future research could further explore the effect of user interaction on value estimation and comparison. Third, the data used in our experiments do not contain semantics, while studies have shown that the background of the data affects the participants' focus areas (Zhao et al. 2021). Semantic enhancements can be made in the future using real-world datasets. Fourth, in this work, we focus on the effect of design parameters in a single star glyph but do not consider the use of multiple glyphs in more complex tasks in much visual analytics systems (Xia et al. 2022; Yuan et al. 2021). Thus, more work is required in these directions and we hope the results obtained from this experiment will contribute to such studies.

6 Design suggestions

The results obtained from the analysis and discussion led to the following design considerations.

- **Area encoding designs for value comparison** Area encoding star graph shows a significant advantage in efficiency in all three comparison tasks except retrieving actual values. It makes people more confident about their answers when data values need to be compared.
- **Avoid complex designs** In all tasks, the design of Contour + Length + Augmented points significantly decrease the efficiency in the value estimation and comparison tasks and is not recommended by the participants due to visual clutter.
- **Display the auxiliary circles** All participants mentioned that the auxiliary circles were necessary when making estimating exact values or comparing similar, non-adjacent values.
- **For usage scenarios that require approximate but fast readings, unnecessary augmented visual elements should be avoided** Although people showed a strong favor for position encoding designs and more than half of the participants liked the Augmented points designs (C + P, C + L + P), the efficiency in the value estimation task decreased due to the augmented points. A possible reason is that people are trying hard to read the actual value, however, for usage scenarios that only requires approximate but fast readings, the interference caused by unnecessary augmented visual element may need to be avoided.

7 Conclusion

While previous work widely discussed the design guidelines of data glyphs, there are limited studies for investigating visualization parameters for user perception of the star glyphs. In this paper, we present a within-subject experiment to compare three design parameters of star glyphs in the value estimation and comparison tasks on efficiency, accuracy, and user experience.

Our results show that area encoding is more efficient in all three comparison tasks but not in the value estimation task. The augmented positions decreased the efficiency in the data estimation task. The overall accuracy in all low-level tasks is high, and the auxiliary line (length encoding) can improve the accuracy of the comparison tasks. In addition, people prefer the area encoding design, while complex glyphs are less preferred. In general, when fast data estimation is needed, simple designs are suggested.

In summary, our work provides a fundamental study to test the three design parameters (position, length, and area) through four tasks (finding extremes, retrieving values, comparing values of adjacent attributes, and comparing values of non-adjacent attributes) with two dimensions (low and high).

Based on our findings, we provide the design suggestions for star glyphs. As such glyph visualizations are widely used to represent multivariate data, we hope that our findings and suggestions will make contributions to real-world applications of star glyph visualizations.

Acknowledgements The work was supported by NSFC (62272396) and XJTLU Research Development Funding RDF-19-02-11.

References

Albo Y, Lanir J, Bak P, Rafaeli S (2015) Off the radar: comparative evaluation of radial visualization solutions for composite indicators. IEEE Trans Vis Comput Graph 22(1):569–578
Amar R, Eagan J, Stasko J (2005) Low-level components of analytic activity in information visualization. IEEE Symp Inf Vis 2005:111–117
Borgo R, Kehrer J, Chung DH, Maguire E, Laramee RS, Hauser H, Ward M, Chen M (2013) Glyph-based visualization: foundations, design guidelines, techniques and applications. Eurographics (State of the Art Reports), pp 39–63
Brehmer M, Munzner T (2013) A multi-level typology of abstract visualization tasks. IEEE Trans Vis Comput Graph 19(12):2376–2385
Cai X, Efstathiou K, Xie X, Wu Y, Shi Y, Yu L (2018) A study of the effect of doughnut chart parameters on proportion estimation accuracy. Comput Graph Forum 37:300–312
Caporaso T, Grazioso S, Di Gironimo G, Lanzotti A (2020) Biomechanical indices represented on radar chart for assessment of performance and infringements in elite race-walkers. Sports Eng 23(1):1–8
Chambers JM, Cleveland WS, Kleiner B, Tukey PA (2018) Graphical methods for data analysis. Chapman and Hall, CRC
Chen Z, Zeng W, Yang Z, Yu L, Fu C-W, Qu H (2020) Lassonet: deep lasso-selection of 3D point clouds. IEEE Trans Vis Comput Graph 26(1):195–204. https://doi.org/10.1109/TVCG.2019.2934332
Cleveland WS, McGill R (1984) Graphical perception: theory, experimentation, and application to the development of graphical methods. J Am Stat Assoc 79(387):531–554
Deng Z, Weng D, Xie X, Bao J, Zheng Y, Xu M, Chen W, Wu Y (2021) Compass: towards better causal analysis of urban time series. IEEE Trans Vis Comput Graph 28(1):1051–1061
Draper GM, Livnat Y, Riesenfeld RF (2009) A survey of radial methods for information visualization. IEEE Trans Vis Comput Graph 15(5):759–776
Dy B, Nazim I, Poorthuis A, Joyce SC (2021) Improving visualisation design for effective multi-objective decision making. IEEE Trans Vis Comput Graph. https://doi.org/10.1109/TVCG.2021.3065126
Fuchs J (2015) Glyph design for temporal and multi-dimensional data: design considerations and evaluation. PhD thesis
Fuchs J, Isenberg P, Bezerianos A, Fischer F, Bertini E (2014) The influence of contour on similarity perception of star glyphs. IEEE Trans Vis Comput Graph 20(12):2251–2260
Healey CG, Booth KS, Enns JT (1996) High-speed visual estimation using preattentive processing. ACM Trans Comput Hum Interact 3(2):107–135
Heckert NA, Filliben JJ, Croarkin CM, Hembree B, Guthrie WF, Tobias P, Prinz J, et al (2002) Handbook 151: nist/sematech e-handbook of statistical methods
Hongliang L, Anxin L, Bin Z, Tiefu Z, Xin Z (2008) A fuzzy comprehensive evaluation method of maintenance quality based on improved radar chart. Int Colloq Comput Commun Control Manag 1:638–642
Kandogan E (2000) Star coordinates: a multi-dimensional visualization technique with uniform treatment of dimensions. IEEE Inf Vis Symp 650:9–12
Klippel A, Hardisty F, Li R, Weaver C (2009) Colour-enhanced star plot glyphs: can salient shape characteristics be overcome? Cartogr Int J Geogr Inf Geovisualization 44(3):217–231
Klippel A, Hardisty F, Weaver C (2009) Star plots: how shape characteristics influence classification tasks. Cartogr Geogr Inf Sci 36(2):149–163
Li J, Martens J-B, van Wijk JJ (2010) A model of symbol size discrimination in scatterplots. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2553–2562
Liu S, Peng D, Zhu H, Wen X, Zhang X, Zhou Z, Zhu M (2021) Muluba: multi-level visual analytics of user behaviors for improving online shopping advertising. J Vis 24(6):1287–1301
Opach T, Popelka S, Dolezalova J, Rød JK (2018) Star and polyline glyphs in a grid plot and on a map display: which perform better? Cartography and Geographic Information Science 45(5):400–419
Peng W, Li Y, Fang Y, Wu Y, Li Q (2019) Radar chart for estimation performance evaluation. IEEE Access 7:113880–113888
Peng W, Ward MO, Rundensteiner EA (2004) Clutter reduction in multi-dimensional data visualization using dimension reordering. In IEEE Symposium on Information Visualization, IEEE, pp 89–96
Saary MJ (2008) Radar plots: a useful way for presenting multivariate health care data. J Clin Epidemiol 61(4):311–317
Seide SE, Jensen K, Kieser M (2021) Utilizing radar graphs in the visualization of simulation and estimation results in network meta-analysis. Res Synth Methods 12(1):96–105
Skau D, Kosara R (2016) Arcs, angles, or areas: individual data encodings in pie and donut charts. Comput Graph Forum 35(3):121–130
Stafoggia M, Lallo A, Fusco D, Barone AP, D'Ovidio M, Sorge C, Perucci CA (2011) Spie charts, target plots, and radar plots for displaying comparative outcomes of health care. J Clin Epidemiol 64(7):770–778
Tominski C, Abello J, Schumann H (2004) Axes-based visualizations with radial layouts. In Proceedings of the ACM symposium on Applied computing, pp 1242–1247
Wei Y, Mei H, Zhao Y, Zhou S, Lin B, Jiang H, Chen W (2020) Evaluating perceptual bias during geometric scaling of scatterplots. IEEE Trans Vis Comput Graph 26(1):321–331. https://doi.org/10.1109/TVCG.2019.2934208
Wilkinson L (2012) The grammar of graphics. Handbook of computational statistics. Springer, pp 375–414
Wu A, Qu H (2018) Multimodal analysis of video collections: visual exploration of presentation techniques in ted talks. IEEE Trans Vis Comput Graph 26(7):2429–2442
Xia J, Lin W, Jiang G, Wang Y, Chen W, Schreck T (2021) Visual clustering factors in scatterplots. IEEE Comput Graph Appl 41(5):79–89. https://doi.org/10.1109/MCG.2021.3098804
Xia J, Zhang Y, Song J, Chen Y, Wang Y, Liu S (2022) Revisiting dimensionality reduction techniques for visual cluster analysis: an empirical study. IEEE Trans Vis Comput Graph 28(1):529–539. https://doi.org/10.1109/TVCG.2021.3114694
Yuan J, Chen C, Yang W, Liu M, Xia J, Liu S (2021) A survey of visual analytics techniques for machine learning. Comput Vis Media 7(1):3–36. https://doi.org/10.1007/s41095-020-0191-7

Zeng W, Fu C-W, Müller Arisona S, Schubiger S, Burkhard R, Ma K-L (2017) Visualizing the relationship between human mobility and points-of-interest. IEEE Trans Intell Transp Syst 18(8):2271–2284. https://doi.org/10.1109/TITS.2016.2639320

Zhang C, Zeng W, Liu L (2021) Urbanvr: an immersive analytics system for context-aware urban design. Comput Graph 99:128–138. https://doi.org/10.1016/j.cag.2021.07.006

Zhao Y, Luo F, Chen M, Wang Y, Xia J, Zhou F, Wang Y, Chen Y, Chen W (2019) Evaluating multi-dimensional visualizations for understanding fuzzy clusters. IEEE Trans Vis Comput Graph 25(1):12–21. https://doi.org/10.1109/TVCG.2018.2865020

Zhao Y, Shi J, Liu J, Zhao J, Zhou F, Zhang W, Chen K, Zhao X, Zhu C, Chen W (2021) Evaluating effects of background stories on graph perception. IEEE Trans Vis Comput Graph. https://doi.org/10.1109/TVCG.2021.3107297