5-2023

Efficacy of Information Extraction from Bar, Line, Circular, Bubble and Radar Graphs

Hiddadura Isura Malinda Mendis Abeynayake
Hong Kong University of Science and Technology

Ravindra S. Goonetilleke
Khalifa University

Albert Wijeweera
Khalifa University

Uwe Reischl
Boise State University

Publication Information

Abeynayake, Hiddadura Isura Malinda Mendis; Goonetilleke, Ravindra S.; Wijeweera, Albert; and Reischl, Uwe. (2023). "Efficacy of Information Extraction from Bar, Line, Circular, Bubble and Radar Graphs". *Applied Ergonomics, 109*, 103996. https://doi.org/10.1016/j.apergo.2023.103996

Applied Ergonomics 109 (2023) 103996

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Efficacy of information extraction from bar, line, circular, bubble and radar graphs

Hiddadura Isura Malinda Mendis Abeynayake [a], Ravindra S. Goonetilleke [b,*], Albert Wijeweera [c], Uwe Reischl [d]

[a] Human Performance Laboratory, Department of Industrial Engineering and Decision Analytics, Hong Kong University of Science and Technology, Hong Kong
[b] Department of Industrial and Systems Engineering, Khalifa University, United Arab Emirates
[c] Department of Humanities and Social Sciences, Khalifa University, United Arab Emirates
[d] Department of Public Health and Population Science, Boise State University, USA

ARTICLE INFO

Keywords:
Data representation
Data visualization
Information extraction

ABSTRACT

With the emergence of enormous amounts of data, numerous ways to visualize such data have been used. Bar, circular, line, radar and bubble graphs that are ubiquitous were investigated for their effectiveness. Fourteen participants performed four types of evaluations: between categories (cities), within categories (transport modes within a city), all categories, and a direct reading within a category from a graph. The representations were presented in random order and participants were asked to respond to sixteen questions to the best of their ability after visually scanning the related graph. There were two trials on two separate days for each participant. Eye movements were recorded using an eye tracker. Bar and line graphs show superiority over circular and radial graphs in effectiveness, efficiency, and perceived ease of use primarily due to eye saccades. The radar graph had the worst performance. "Vibration-type" fill pattern could be improved by adding colors and symbolic fills. Design guidelines are proposed for the effective representation of data so that the presentation and communication of information are effective.

1. Introduction

Graphs, a form of representation of analog data, are a way of communicating information in day-to-day life and they are easier for us to comprehend in comparison to numbers or words (Bergauer et al., 2022; Cukier, 2010; Wickens, 2013; Gaissmaier et al., 2012). It is a way for designers to encode numerical data (Cleveland and McGill, 1985; Kelleher and Wagener, 2011; Shah and Hoeffner, 2002) and for users to visualize and decode the data easily (Bergauer et al., 2022; Durand et al., 2020). In the decoding process, vision and cognition are combined to understand the qualitative and quantitative information encoded in a graph (Chen and Jin, 2017; Ware, 2004). For this process to be effective, the graph should convey the information in a simple and usable way so that the required information can be easily extracted (Glazer, 2011).

Various types of graphs are used to explain different types of data even though the most useful type for any particular application is not well known. Bars and line graphs are frequently used for comparisons of individual values as length and position are easily perceived (Acartürk, 2014; Jeong et al., 2015; Cleveland and McGill, 1984). Most previous studies have focused on interpreting bar and line graphs and histograms. Circular, bubble and radial charts, however, are gaining popularity in the big data domain (Agarwal and Sarkar, 2022; Cairo, 2013; Draper et al., 2009; Proulx, 2011). Bubble graphs can represent 3-dimensional data while conventional graphs can represent 2-dimensional data. Radar graphs are considered to be inefficient (Few, 2009) even though they are quite popular for multi-dimensional data. In this research, we set out to find which type of graph is most appropriate for data representation in terms of search time and accuracy.

A graph is not just a figure and what comprises a graph is not primary. It has to include what users need (Carswell, 1992). The goal of every designer is to present information to their consumers or target audience in a way that allows for the greatest exchange of knowledge (Webber, 2018). Examining what a user views and processes is difficult to know. One possible way to track and understand user behavior and strategy is by using eye-gaze patterns (Delmas et al., 2021; Goldberg and Helfman, 2011; Kullmann et al., 2021; Torres et al., 2021). Eye tracking allows search time, accuracy, and other related issues to be evaluated. Thus, we used eye tracking to evaluate why some graphs are more

* Corresponding author.
E-mail address: ravindra.goonetilleke@ku.ac.ae (R.S. Goonetilleke).

https://doi.org/10.1016/j.apergo.2023.103996
Received 19 September 2022; Received in revised form 2 December 2022; Accepted 6 February 2023
Available online 17 February 2023

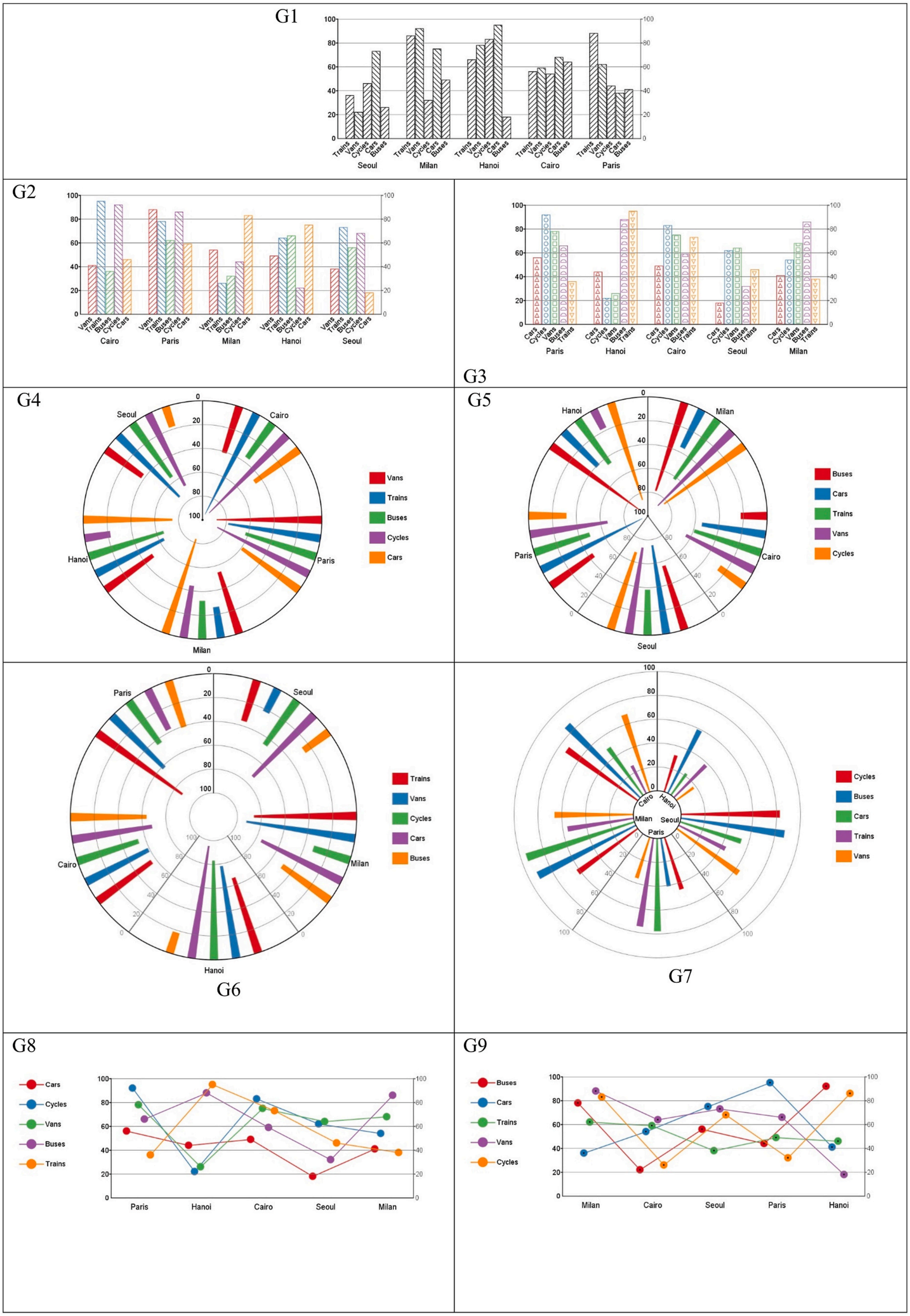

Fig. 1. The fifteen graphs that were tested.

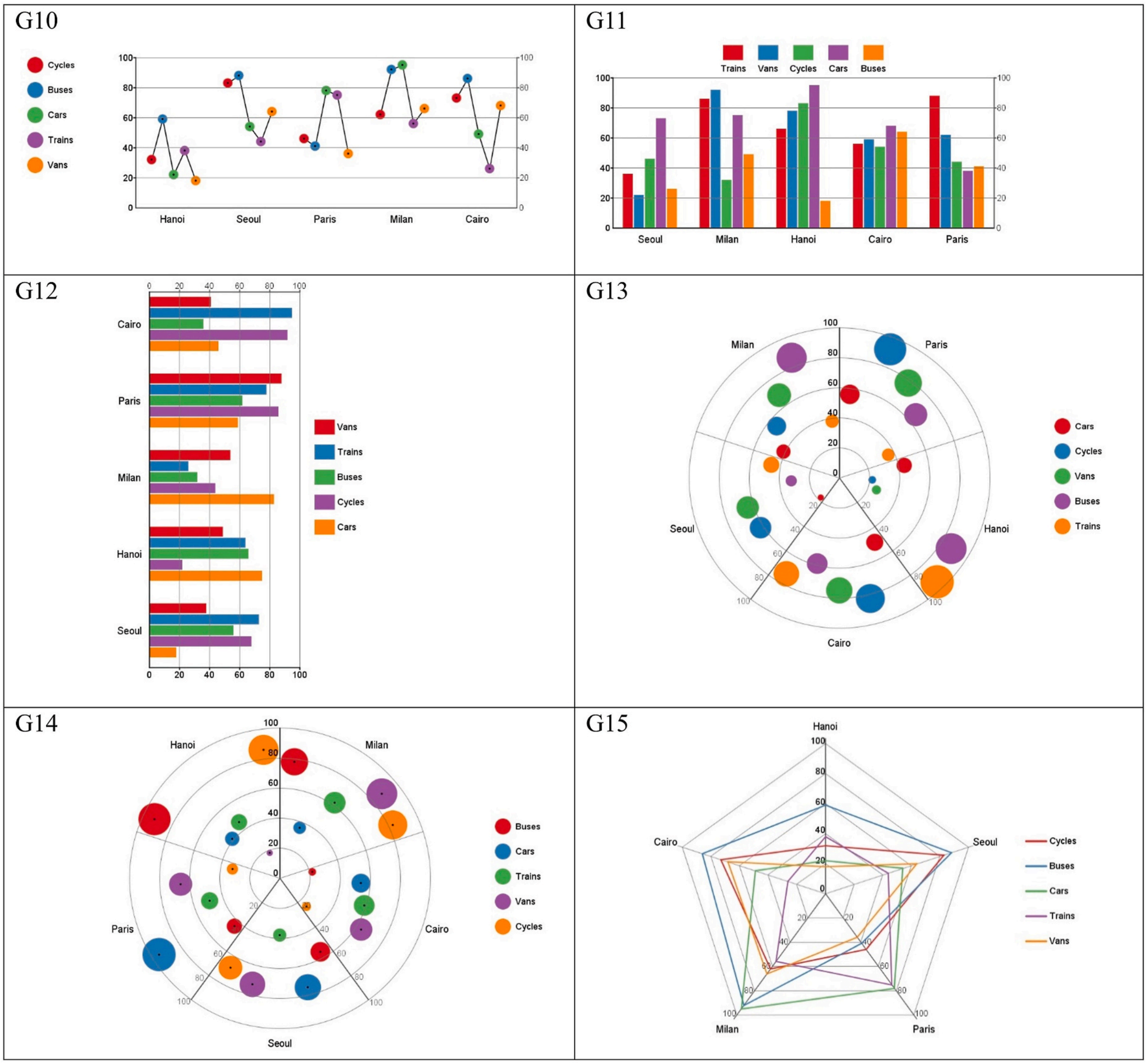

Fig. 1. (*continued*).

efficient than others.

Considering the cognitive processing that is involved, there are many principles and guidelines for the design of graphical representations (Carswell and Wickens, 1990; Few, 2012; Kosslyn, 2006; Tufte, 2001). For example, Kosslyn (2006) proposed eight principles for effective graphical displays. These are.

P1. Principle of relevance: Presenting only the required information
P2. Principle of appropriate knowledge: Users have sufficient familiarity with the graphs
P3. Principle of salience: Most important information to be visually striking so that it will draw the reader's attention.
P4. Principle of discriminability: Differences should be large enough for items to be discriminated
P5. Principle of perceptual organization: Conform to the Gestalt principles
P6. Principle of compatibility: Eliminate any form of mental transformations as a result of spatial imprecision, perceptual distortions or cultural conventions.
P7. Principle of informative changes: Variations are shown in changes in appearance.
P8. Principle of capacity limitations: Our short-term memory is limited and hence changes or trends in a variable are better identified with the slope of lines. The proximity compatibility principle (PCP) is a more specific form stating that high proximity such as by connecting points will help minimize cognitive operations to predict trends (Connor et al., 2022; Wickens et al., 2022).

The principles, P1–P8 capture the essential needs of graphs. However, putting them into practice and determining an optimal format without any contradictions for displaying analog data is not an easy task. A good example is the color and contrast of words and backgrounds. Intuition suggests that high contrasts may be better for visual search. But, Hill (1997) found that high contrast between font and background did not have better-reading performance. Thus, it is just impossible to evaluate or even predict the outcome of all possible options. Hence, in this experiment, the effectiveness of Bar, Line, Circular, Bubble and Radar graphs was evaluated based on common usage and what may be appropriate in terms of published principles and guidelines. For example, it has been reported that bar graphs are good for complex comparisons, while line graphs are good for trends (Meyer et al., 1997; Umanath and Scamell, 1988). Bar graphs are predominantly used for nominal scales while line graphs are used in situations where the variables are measured on interval or ratio scales. A continuous form of bar graph where the variations over time are visualized is known as the ThemeRiver and is popular in the big data domain (Havre et al., 2000).

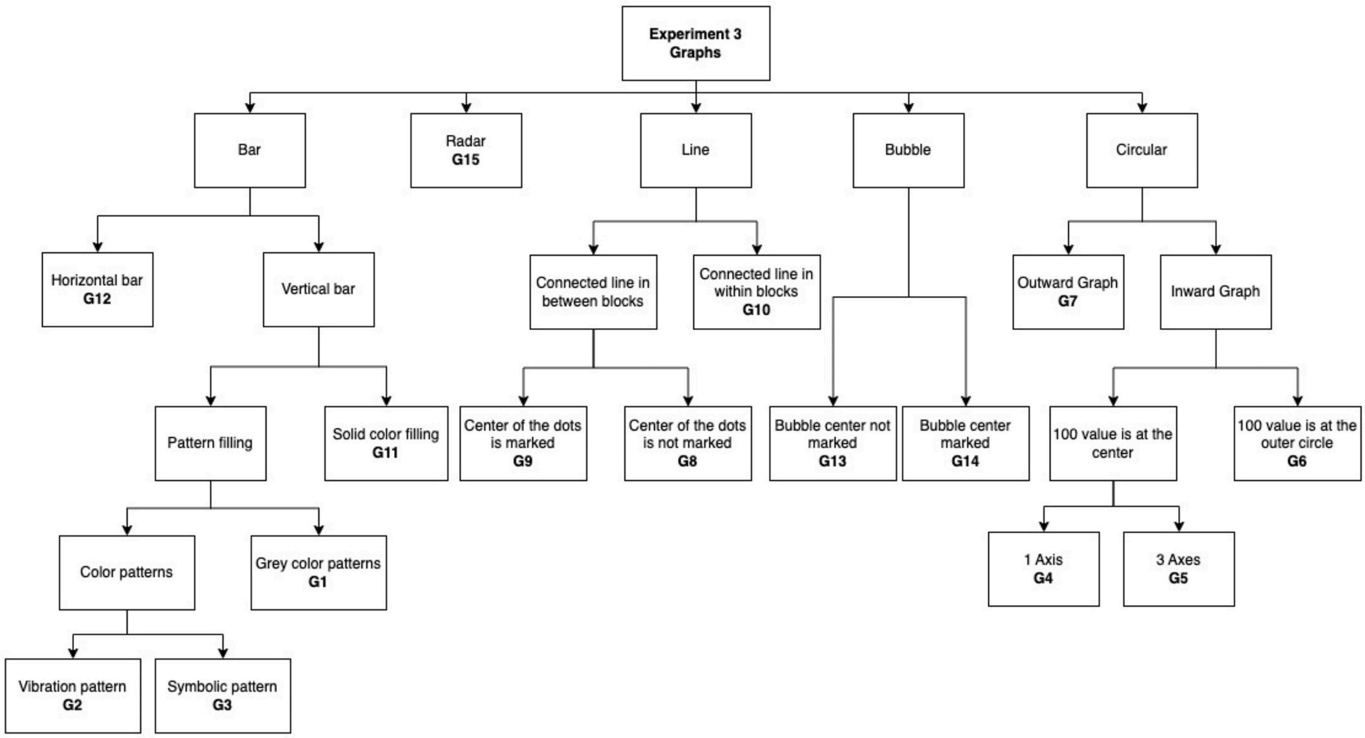

Fig. 2. The 15 graphs and their classification.

Thus, the effectiveness of a graph depends heavily on the type of data to be shown and the question that is asked (Kosslyn, 2006). But, not much has been reported in terms of the questions asked (principle of relevance). In other words, the critical question is whether data representations are for information transmission or information extraction (Durand et al., 2020). Information extraction is where the information may be "hidden" amongst what is depicted. This study attempts to explore attributes that affect information extraction and how that information is transmitted when comparisons have to be made and related to user perceptions of difficulty.

Time and accuracy are indicators of human performance. But, these two metrics alone do not tell the researchers how accuracy was affected or where the time is spent (Huang, 2007). How we "see" details in an image can directly impact a viewer's efficiency and effectiveness (Healey and Enns, 2012). Eye tracking gives insights into how a particular task is carried out and gives additional information (Delmas et al., 2021; Duchowski 2003; Holmqvist et al., 2011; Netzel et al., 2017). Moreover, extracting the fixation sequences on any object will allow improvement and redesign for more efficient use (Blascheck et al., 2017). In this respect, the fixations on the areas of interest (AOI) and the relevant transitions help understand and improve the graphs so that they are more efficient (Klein et al., 2021). The objective of this study was to evaluate the effectiveness (accuracy), efficiency (time) and perceived ease of use of a bar, circular, line, bubble and radar graphs. Only 2D graphs were used as 3D graphics are not that helpful for visualization even though they may be helpful for data analysis (Few 2012). Even if it were useful, multi-dimensional data can still be shown in 2D space with coplots and scatter plots (Cleveland, 1994).

On gender effects, Hartley and Cabanac (2014) examined 3576 single-authored science articles published in 2011 and found that men used more graphs than women even though they found in another study that there were no significant differences between men and women in the use of graphs and figures or tables in the social sciences. Due to this mixed result, in our experiment, gender was balanced and investigated.

The test graphs were subjected to a screening and improvement process before they were used in the experiment. The weaknesses of some of the graphs that were used in the pilot experiments were compensated by adding various features. The consistency among the charts was maintained by using the same pixel count for the values in the axes. Eye movements were recorded during the experiment. The fifteen graphs that were tested are shown in Fig. 1 and their classification is in Fig. 2.

The hypotheses tested were.

H1. There are no differences between males and females in terms of accurate information extraction from different types of graphs.

H2. Bar, Line, Circular, Bubble and Radar graphs are equally effective (accurate) for comparisons across different levels.

H3. Bar, Line, Circular, Bubble and Radar graphs are equally efficient (search time) for comparisons across different levels.

H4. Bar, Line, Circular, Bubble and Radar graphs have the same level of perceived ease of use.

H5. The fixation time on each type of graph is the same.

2. Method

2.1. Participants

Fourteen undergraduate and postgraduate students (seven male and seven female) of mean age of 25.9 years (SD = 4.6 years) from the Hong Kong University of Science and Technology were participants in the experiment. All participants claimed they were free of eye deformities. Each of them was paid HK$ 200 for their time. The experiment was approved by the institutional review board.

2.2. Experiment design

The experiment was a mixed design with two within-participant variables (Graph, Question) and one between-participant variable (Gender). There were 15 different types of graphs (Figs. 1 and 2) and

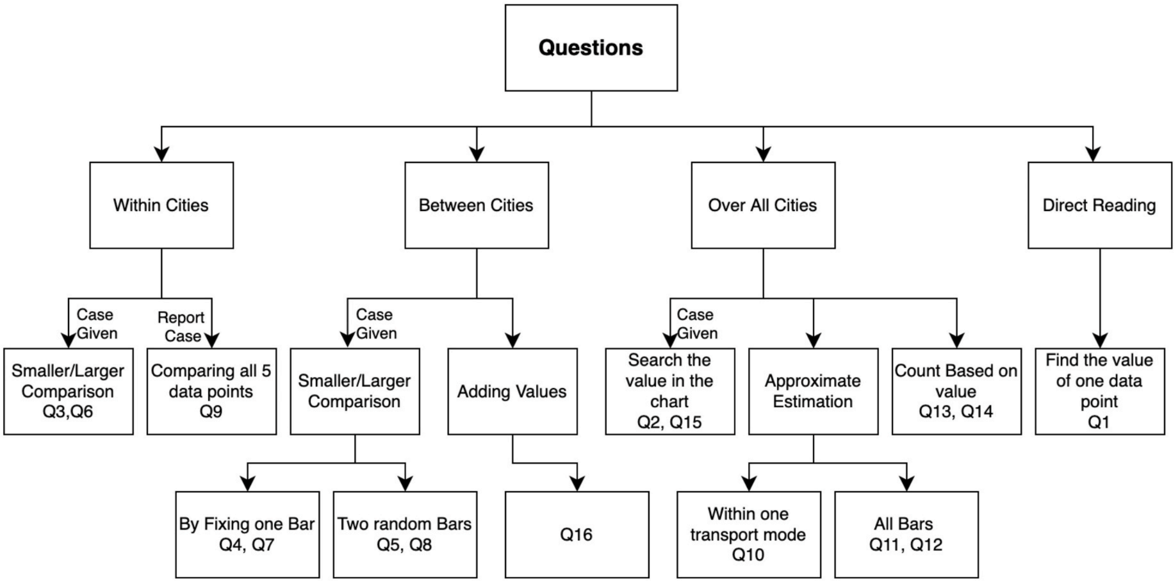

Fig. 3. Different Types of Question sets.

Table 1
A sample of the sixteen questions.

Q1	How many Trains are in Seoul?
Q2	If 73 exists in the chart, indicate
Q3	Which one is smaller: Cars in Hanoi or Cycles in Hanoi?
Q4	Which one is smaller: Trains in Seoul or Trains in Milan?
Q5	Which one is smaller: Cycles in Seoul or Buses in Cairo?
Q6	Which one is larger: Vans in Cairo or Cars in Cairo?
Q7	Which one is larger: Cycles in Hanoi or Cycles in Paris?
Q8	Which one is larger: Cycles in Seoul or Buses in Cairo?
Q9	In Cairo, which is the second highest?
Q10	Which city has the second-highest number of Cycles?
Q11	Which is the second highest?
Q12	Which is the second smallest?
Q13	How many values are smaller than 28?
Q14	How many values are larger than 89?
Q15	If 47 exists in the chart, indicate
Q16	If you add all modes of transport, which city has the larger value; Seoul or Milan?

participants had to use these graphs to respond to 16 questions that asked questions about five transport modes (trains, vans, cycles, cars and buses) in five different cities (Seoul, Milan, Hanoi, Cairo, and Paris). The questions fell into four groups: within-city comparison (or within block), between-city comparison (between blocks), comparison over all cities (over all blocks), and reading the exact value for one transport mode in one city (Fig. 3). A sample set of questions is shown in Table 1.

All graphs were created using Processing 2.2.1 open source software as it allowed to control the colors, shapes, axes, line sizes, and fonts in the graphs so that there were no salience effects. Using existing graphics software allows limited configurations and those may have driven the reader towards certain features of the graph. All 15 graphs had the same size of axis, colors and fonts with a "box" framework. The bars were marked and arranged the same way to ensure similarity among the quantities compared. Text were in mixed upper- and lowercase, cities or the main categories (Arial 20), transport mode or sub-categories (Arial 18), axis values (Arial 16) had consistent point fonts and color. All graphs had the same number of pixels for the 0 to 100 scale. Colors can be misleading especially if they are strong and adhere to certain stereotypes (Cleveland and McGill, 1985). Thus, it is important to use balanced colors. In this experiment, color codes suggested by Brewer (2020) were used. The R, G, B values of the chosen colors were Red: (228, 26, 28), Blue: (55, 126, 184), Green: (77, 175, 74), Violet: (152, 78, 163), and Orange: (255, 127, 0).

The Gestalt rules of proximity and similarity were used to separate the cities and show the similarity of the transport modes. Because absolute magnitudes were used, the "vertical" axis originated at zero (Robbins, 2005, pp. 239–241). As stated before, it is not easy to incorporate all the principles into every graph. The effect of having met some of them will be discussed after the results are presented. The details of each graph are given below (Figs. 1 and 2).

G1: *Bar Graph* with labels on the x-axis. This is a very common form of representation and the most dominating effect is that it obeys Principle 2 (P2). It had a gray "vibration" fill pattern. The guideline for hatching orientation is generally a change of around 30° (Kosslyn, 2006) between adjacent regions. In this case, the change in orientation is quite extreme and quite difficult to look at, giving rise to the vibration type of pattern (De Valois and De Valois, 1988). The sub-category labels were close to the category labels.

G2: This is similar to graph 1 except that the gray color fill in each bar is replaced with a color fill (Seva et al., 2022). It adheres to principles P1 and P4 as the differences between the modes of transportation are larger due to the addition of color.

G3: *Bar Graph* with a symbol pattern (triangles, circles, etc.) filled of different colors. The graph takes G2 to another level. The symbols represent relative quantities (King et al., 2022). This type of graph can be made interesting by adding pictures of objects (Kosslyn, 2006). Thus, the fill pattern acts as a subtle "micro-level" scale as well. That is, it is a scale-within-a-scale. If silhouettes equivalent to a certain value are used, the y-axis can be eliminated as well (Cairo, 2013). This graph incorporates principles P1, P4 and P7 as the variations among the bars are shown are clearer with the addition of the objects inside each bar.

G4: This is an "inward" radial or *circular graph.* The value 100 is at the center of the circle. It was hypothesized to have the minimum level of visual scanning as the bars are constricted towards the center of the circle. This graph violates principle P6 because the legend is "inverted" for half of the circle but discrimination among the bars is much better than the bar graph (P4).

G5: Inward radial or *circular graph* with the value 100 at the center of the circle. This graph is similar to G4 except that it had the value axis at three angles thereby enhancing the readability of the bars especially when the values are not at their maximums. Even though such a modification is an enhancement for readability, it violates the Tufte

(1990) principle of efficiency, which suggests using a data-ink ratio of close to 1.0 and eliminating all redundancy and unnecessary visual information. This graph is an improvement over G4 as perceptual distortions due to the circularity are reduced and reading the exact values is not subject to spatial imprecisions (principle P6).

G6: Inward radial or *circular graph* with the value 100 offset from the center. This feature was to reduce any potential visual clutter at the center and make the graph aesthetically more pleasing. There were 3 value-axes similar to G5. This type is a further improvement on G5 to incorporate the Gestalt rules of similarity and proximity to reduce clutter (P5).

G7: Outward radial with 3 value-axes. As the inverse of G6, it reduces clutter further as the outside of each bar extends over a greater radius.

G8: Somewhat of an unusual *line graph* as colored lines were used for connecting the categories, which is not an interval scale. Non-sequential data and categorical data that are often grouped, are not supposed to be connected by lines (Kelleher and Wagener, 2011). The reason for using this type of graph was to conform to the principle of capacity limitations (P8) (Kosslyn, 2006) and PCP (Wickens et al., 2022). The slopes of the lines (Ware, 2008) help to discern the differences, especially when the neighboring values are very similar.

G9: Similar to G8 except that a small black dot is placed at the center of the data point to identify the exact value immediately rather than having the user interpolate the center of the circle eliminating relatively difficult estimation. Spatial imprecisions in G8 as a result of the target point can be reduced by marking the center (P6)

G10: A *line graph* with a line used to connect sub-categories within a category. This again, like G8 is generally not used as a line represents a continuous variable. A small black dot was placed at the data point. This graph conforms with P6 and P8 similar to the others that have been described.

G11: Vertical *bar graph* with a horizontal spatially-compatible (Huestegge and Philipp, 2011) legend on the top of the graph. Such a horizontal legend is not common even though it conforms to P6. The prescribed legend is a vertical one (Kosslyn 2006).

G12: Horizontal *bar graph* with a spatially compatible vertical legend on the right side (P6). This graph was used primarily because people overestimate vertical line lengths (Graham, 1937), and thus, Jarvenpaa and Dickson (1988) recommend using horizontal bar graphs.

G13: *Bubble graph.* The center of the bubble represents the value, and the bubble diameter is proportional to the value as well. Using bubble size to show relative comparisons is not uncommon (Cleveland and McGill, 1984). The reason for using diameter rather than the area is primarily because the psychophysical scaling of line lengths is close to 1.0 whereas for areas it is less than 1.0 (Cairo, 2013). This graph even though not common, follows P4 and P7.

G14: Same as G13 with a small black dot for precise estimation of the center of the data point. P6 gets added here to minimize the spatial imprecision with a circle.

G15: *Radar graph.*

The broad categories of graphs were Bars (G1, G2, G3, G11, G12), Circular (G4, G5, G6, G7), Line (G8, G9, G10), Bubble (G13, G14) and Radar (G15). All graphs had only labeled values on the "Y-axis". There were no sub-divisions to reduce the information contained in the graphs. The line graphs, G8, G9 and G10 used differing colors for the different entities as colors are the most discriminable (Cleveland and McGill, 1984; Lewandowsky and Spence, 1989). In this experiment, color codes suggested by Brewer (2020) were used. Only "horizontal" inner gridlines were used to be consistent across all graphs. The inner grids were made thin and light and were discriminable with the framework and the axes. They did not obscure the information that was displayed. That is, they were "behind" bars.

A right side Y-axis was added in all bar graphs due to several reasons.

- Some circular graphs had three axes. Hence the addition of a second axis for the bar graphs may make it a somewhat fair comparison against the radial graphs.
- The points on the right tend to be underestimated with a single y-axis on the left (Kosslyn, 2006). Also, an axis on the right side may help to reduce the saccade length when reading a value from a bar on the right (for vertical bar graphs).
- Extra axes and values were in gray, to reduce the visual impact (principle of salience) on the secondary axis when compared to the main axis.

The inward-radial graph is not common in data visualization. But it has some unique characteristics to help the graph comprehension process such as when determining the maximum value or in high-value comparisons because the scan area reduces in size at the higher data values. The additional axes were to minimize confusion as the participant has to trace an arc for any one value even though this feature violates the data-ink ratio principle (Tufte, 1990).

The data depicted in all graphs were the same but were randomized so that the answer to any question was different for the differing questions. There were five similar question sets. When a question set was assigned to a participant, it was balanced across participants and graphs. In addition, the categories and sub-categories were randomized across graphs. This prevented any short-term memory effects. To make the design manageable, groups of 3 graphs were created and the 48 related questions were randomized within the group.

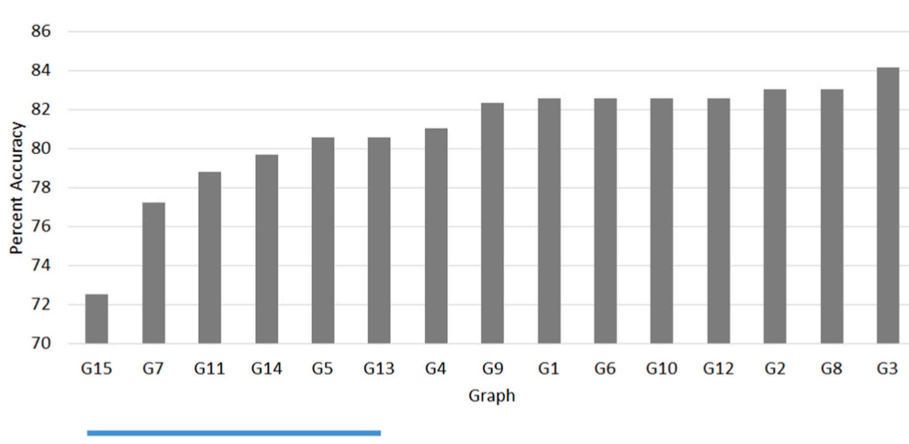

Fig. 4. Accuracy of each Graph for all Questions. Graphs that are not significantly different ($p > 0.05$) are connected with a line below the x-axis.

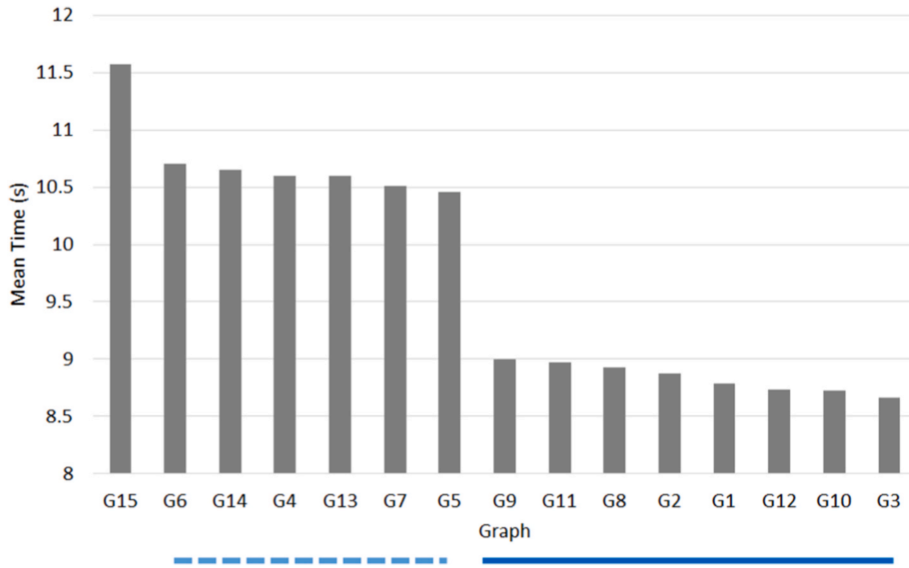

Fig. 5. Mean Time for all Questions for each Graph. Graphs that are not significantly different (p > 0.05) are connected with a line below the x-axis.

2.3. Task and procedure

Participants completed a consent form before to the experiment. There was a training session to explain the task and each participant was encouraged to ask questions before the start of the experiment. Thereafter, the TOBII system eye calibration was performed to track eye gaze. Each of the graphs was displayed on a 23-inch (58 cm) screen one at a time.

There were two trials in this experiment. The same stimuli were repeated in the second trial and there was at least a one-day gap between the two trials. Each participant had to answer 240 questions in each trial. There were 5 groups of graphs with each group having 3 graphs. The participant was given a break at the completion of each group. The participant was requested to select the most appropriate answer and verbally inform the experimenter. The experimenter recorded the answers. There was no time limit for any question. The experimenter pressed a key as soon as the participant answered the question and the Processing software calculated the search time. The perceived ease of use rating data of each graph was obtained at the end.

Eye movement data were recorded using a Tobii TX300 eye tracker (sampling rate 300 Hz) with Tobii 3.2.0 software and all questions were at the top of each graph.

3. Results

3.1. Accuracy

Time and accuracy were two dependent variables in this experiment. Accuracy was analyzed by checking the responses of the participants. The mean correct response of all participants was 80.9%. There was no significant difference in accuracy between Males and Females. Thus, hypothesis 1 that there are no differences between males and females is accepted.

The accuracy of each graph, for all questions in both trials, is shown in Fig. 4. Accuracy varied from 72.5% (G15) to 84.2% (G3). The repeated measure ANOVA showed a significant difference ($p < 0.05$) among Graphs ($F\ (14,\ 182) = 3.61$; $p < 0.001$), and Questions ($F\ (15{,}195) = 87.23$; $p < 0.001$), The Graphs × Question interaction was not significant. The main effects were compared using the Bonferroni method in a pairwise manner. G15 had significantly ($p < 0.05$) lower accuracy (72.5%) compared to the other graphs. The mean accuracy of each question ranged from 17.4% (Question 15) to 96.7% (Question 4).

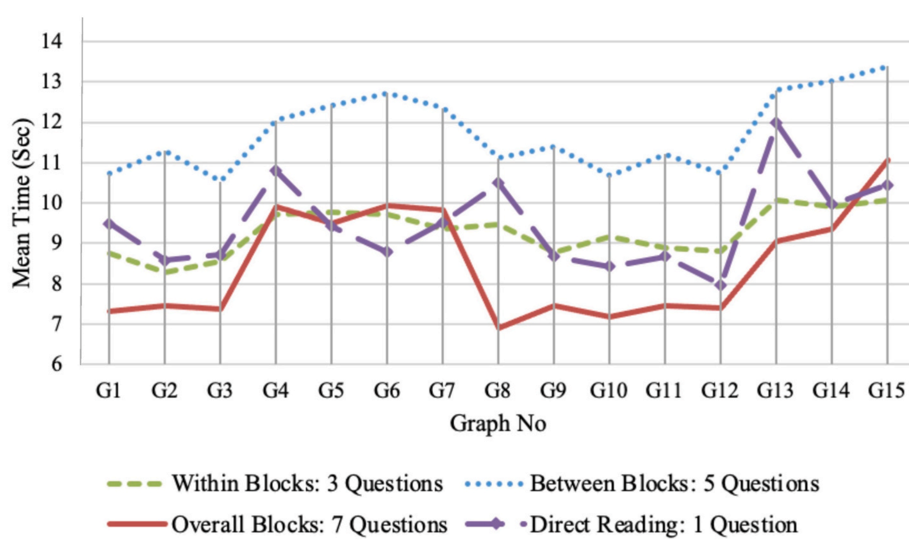

Fig. 6. Time taken for each graph for different types of question sets.

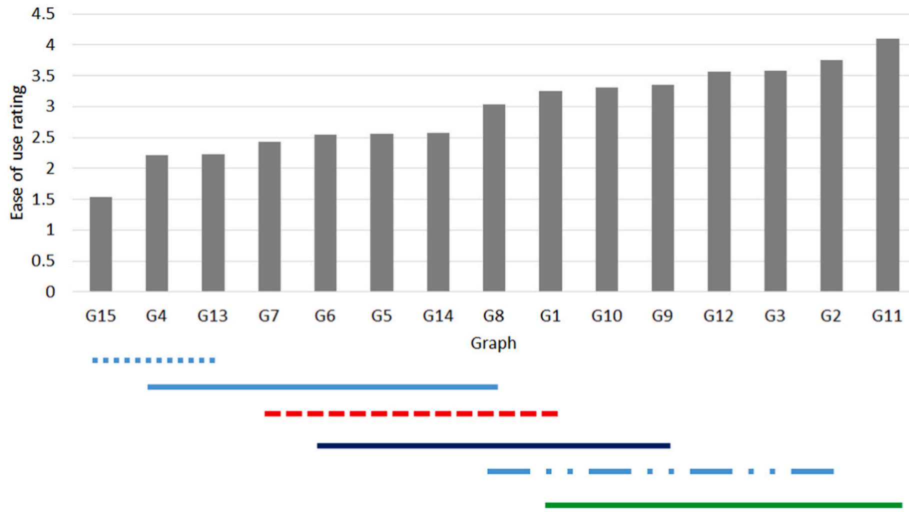

Fig. 7. Ease of Use Rating of each graph. Graphs that are not significantly different ($p > 0.05$) are connected with a line below the x-axis.

Questions 3–12 and 16 (mean in the range 90.5%–96.7%) were not significantly different in accuracy, questions 2 and 13 (mean = 71.2% and 76% respectively) had statistically similar accuracy and the worst accuracy was for question 15 (mean = 17.4%). Question 1 was slightly better than question 15 but significantly different from others with a mean accuracy of 34%. The fact that the accuracies are quite different across the different questions indicates that the level of difficulty of each question was different. Hypothesis 2 that the accuracy of the different graphs is the same is rejected.

3.2. *Search time*

The search time was chosen as the lowest time of accurate trials. The mean time of all questions for each graph is shown in Fig. 5. G3 has the fastest search time while G15 is the slowest. Thus, hypothesis 3 that all graphs have the same time is rejected. A repeated measure ANOVA showed no significant difference between genders. Graphs ($F(14, 182) = 19.0$; $p < 0.001$), Questions ($F(15,195) = 70.45$; $p < 0.001$) and Graph × Question interaction ($F(210,2730) = 1.80$; $p < 0.001$) were all statistically significant.

Based on the Bonferroni pairwise comparisons on the graph type, three distinct groups emerged.

- Bar and Linear (BL) graphs, G1, G2, G3, G8, G9, G10, G11, G12, that had the lowest time
- Circular or radial graphs: G4, G5, G6, G7, G13, G14
- Radar graph, G15, had the highest time

Due to the complexity of the interaction, the questions were grouped into four categories (Fig. 3).

- Within category (Block) Discrimination (Q3, Q6, Q9)
- Between categories (Q4, Q5, Q7, Q8, Q16)
- The complete range of data or over-all blocks (Q2, Q10, Q11, Q12, Q13, Q14, Q15)
- Point data Reading (Q1)

A F-test and a Wald test showed that there was no significant effect within each of the four groups. In other words, there were four distinct types of questions. The interaction effect with these blocks is shown in Fig. 6.

To make the results more general, the categories (or cities) will henceforth be referred to as "blocks".

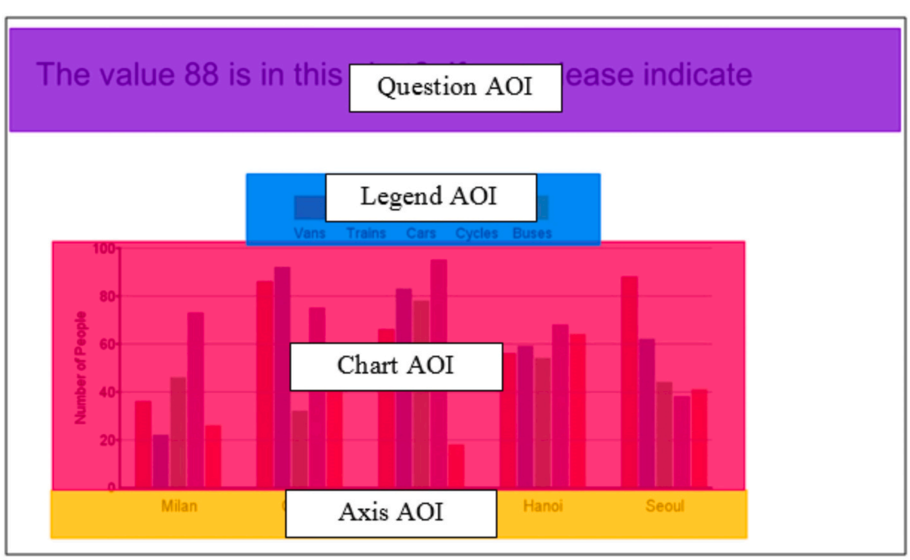

Fig. 8. Area of interests (AOIs) in the stimuli.

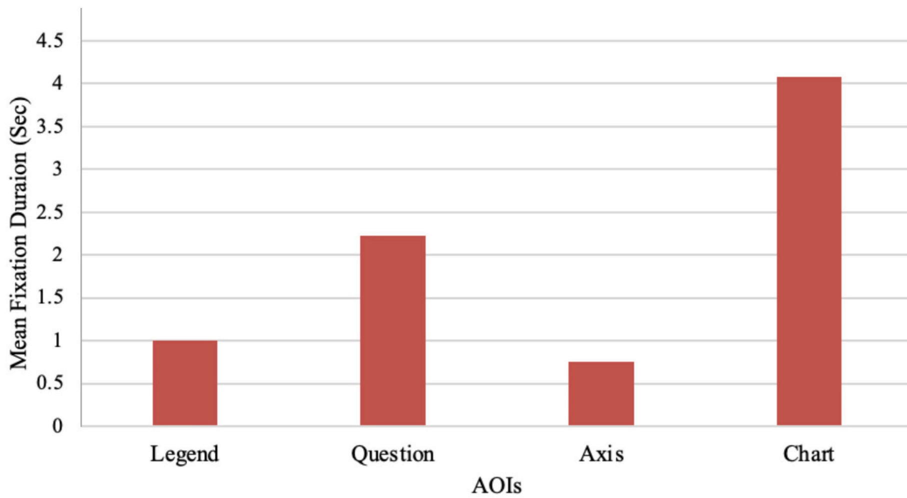

Fig. 9. Mean Fixation Durations for all graphs and questions in different AOIs.

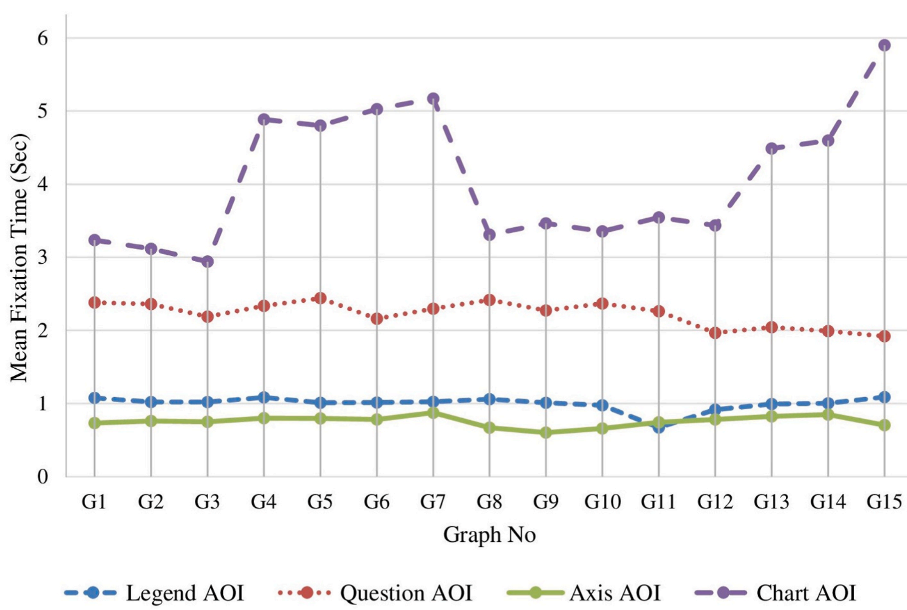

Fig. 10. Fixation Duration of each Graph Type for different AOIs.

3.3. Perceived ease of use

In this experiment, the participant had to give the overall rating for each graph at the end of the experiment. A 5-point scale (1 = very difficult, 2 = difficult, 3 = neutral, 4 = easy, 5 = very easy) was used and the participant could give a rating to one decimal (Fig. 7). The ANOVA on the rating showed the type of graph to be significant (F (14,182) = 21.4; $p < 0.001$) The highest rating was for G11 (mean = 4.1) while the lowest rating was for G15 (mean = 1.54), and these two were significantly different from many other graphs (Fig. 7). Thus, hypothesis 4 that the perceived ease of use is the same for all graphs is rejected.

Bonferroni pairwise comparisons showed that the bar and line types were the easiest for the given task.

3.4. Fixation data

Eye fixation data was obtained from Tobii studio 3.2. The "Tobii fixation filter" was used to generate the fixation plots. Fixation duration was analyzed using the stimuli presented in the experiment on the four Areas of Interest (AOIs) (Krishna et al., 2018) (Fig. 8): The Question, Chart, Legend, and Axis.

The highest fixation durations were on Chart AOI while Axis AOI had the lowest fixation time (Fig. 9). Furthermore, the fixation duration of each AOI was plotted for each graph type (Fig. 10). The legend, axis and question fixation durations are similar across all types of graphs. However, the search time within a chart is quite different. Circular, Bubble (G4, G5, G6, G7, G13 and G14) and Radar graphs (G15) have higher fixation time on chart AOI. Thus, hypothesis 5 that the fixation time on each graph is the same is rejected. Interestingly, the pattern for fixation time is very similar to the question type × graph interaction plot on time (Fig. 6).

Questions were grouped the same way as before (Fig. 3) to analyze the fixation data (Fig. 11). Chart AOI had higher fixation durations in "Over All Blocks", "Between Blocks" and "Direct Reading" questions. Question AOI has the highest fixation duration for the "Within Blocks" questions.

4. Discussion and conclusion

The main objective of this experiment was to analyze the

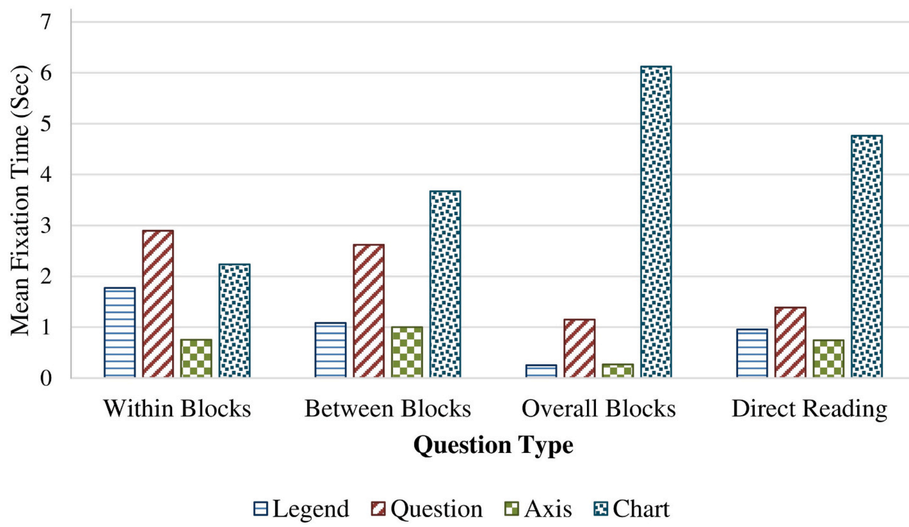

Fig. 11. Mean Fixation Duration for question type.

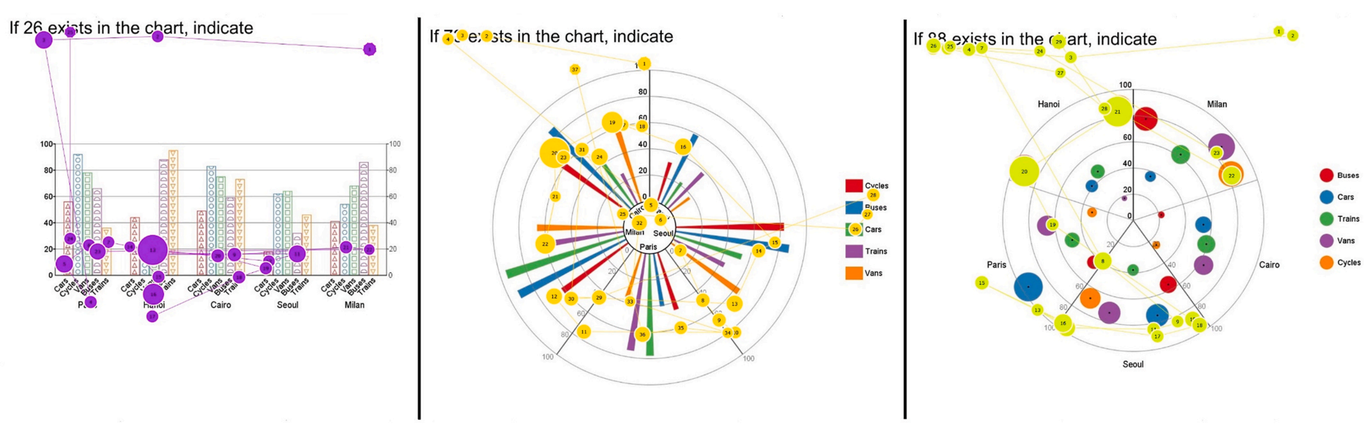

Fig. 12. Eye Fixations for linear and circular axes graphs (Fixations are from one participant for question 2).

performance in terms of effectiveness, efficiency and ease of use of bar, circular, bubble, line and radar graphs. From the analyses, it is clear that people are not as good at extracting information from circular or radar graphs compared to linear or bar graphs. Graphs, G1, G2, G3, G8, G9, G10, G11 and G12 had the lowest time for all questions (Fig. 5). They were significantly ($p < 0.05$) better than the circular (G4, G5, G6, G7, G13, G14) and radar (G15) graphs in search time. In other words, bar and line graphs are efficient compared to circular, radial or radar graphs. This result is in agreement with previous studies where it has been found that bar and line graphs are frequently used for comparisons because length and position are easily perceived (Acarturk, 2014; Cleveland and McGill, 1984). The result gives validity to the principle of appropriate knowledge (P2). Users have sufficient familiarity with the graphs and as a result, their performance is better. We had two trials to eliminate the learning effects but it seems that is not sufficient. Even though Fischer et al. (2005) reported longer decision times in horizontal graphs compared to vertical graphs, there weren't any such differences in this study. They attributed the higher time in horizontal graphs to a lack of familiarity (P2) but we think the difference in the Fischer et al. study may be a result of compatibility (P6). We made sure all the bar graphs had a similar level of compatibility.

Generally, when a task has to be performed, time and accuracy are both important as there can be a speed-accuracy tradeoff. In the experiment reported here, it is clear that no such speed-accuracy tradeoff was present as graph G3 (bar) had the fastest time and the highest accuracy while G15 (Radar) had the slowest search time and the worst accuracy very similar to the Fischer et al. (2005) study. G3 had some special characteristics relative to G1 and G2. It had colors instead of a grayscale and a very subtle micro-level scale that helped participants extract information relatively easily. In other words, P4 and P7 have been coded into G3 and that allows the information to be extracted easily. A design such as G3 may be considered a violation of Tufte's (1990) data-ink ratio principle but may be considered as adhering to the Salience principle proposed by Kosslyn (2006). This result may be a good example of where there can be conflicting principles at work and the designer ought to be able to take into account the basic characteristics depending on what the user needs to extract from the graph.

Fixations from one participant for different types of graphs are shown in Fig. 12. All three images show the fixations for question Q2. It shows how the participant looked at the linear and circular axes to read the values. Saccades are eye "jumps" from one place to another. Even though saccades may be curved for both static and dynamic images (Costela and Woods, 2019; Goettker et al., 2019; Lamasky, 1869), given the higher number of fixations and the related fixation durations in the circular and bubble graphs (G4, G5, G6, G7, G13, G14) it could be inferred that the saccade patterns are somewhat not very compatible with curved presentations of data leading to higher search time, poor accuracy and even higher perceived difficulty. This means circular saccades are not efficient. Thus the principle of compatibility (P6) should be extended, not just to account for mental transformations, but for eye movements as well. Linear saccades are more efficient than circular ones.

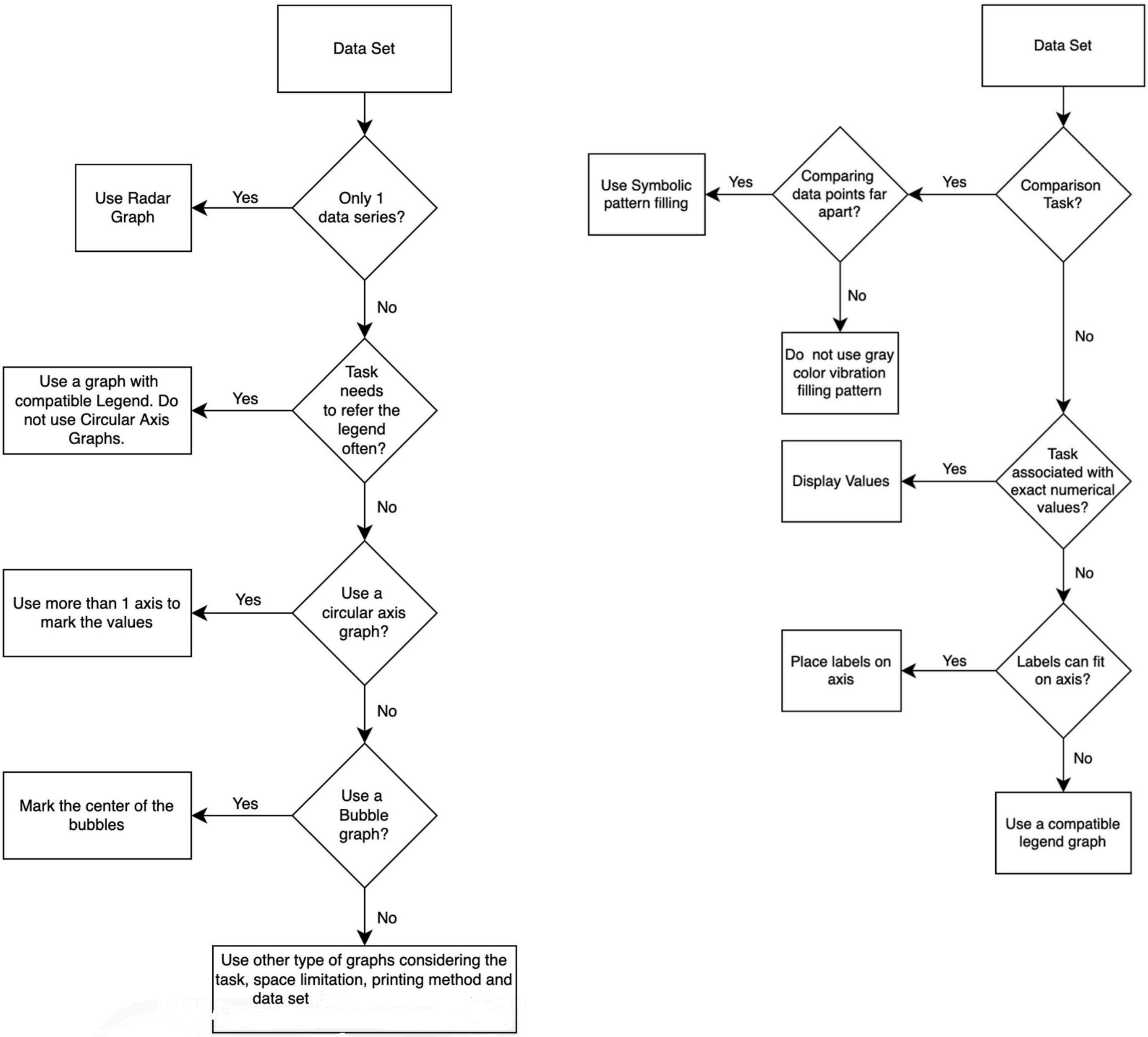

Fig. 13. Flowchart.

The Radar chart had the highest time and it was significantly different from all other graphs. The radar chart had the lowest accuracy as well and it was significantly different with all except G5, G7, G11, G13, and G14. So this is not a speed-accuracy tradeoff. Just like the bubble and circular graphs, the radar graph also has circular fixations, which are not very efficient. Besides, the crisscrossing lines on the radar chart made it difficult to read the radar graph as a radar graph is more suitable with just one data series (Schwabish, 2014). Even though the radar graph meets the proximity compatibility principle (PCP), that is not very effective when laid in a circle (P8). PCP would be useful in a linear fashion. The additional axes evenly laid around do help but the saccades take ineffective paths making the search time high and accuracy low. Schwabish (2014) proposed the three basic principles of show the data, reduce the clutter, integrate the text and graph to visualize data and recommended reducing the clutter in a spaghetti chart by having many data series be replaced with single graphs for each data series or one data series on one graph. The contrast between light and dark also helps to highlight specific trends and reduce clutter. The eye fixation durations on each graph in each AOI (Fig. 10) show that the radar graph has higher fixations in Chart AOI compared to other graphs. However, the Legend, Axis and Question AOI in a radar graph does not seem to have a significant difference when compared with the other graphs. Thus, the structure or the chart area in the Radar graph is what makes it difficult and time-consuming to extract information. The Ease of Use rating is also high for the radar chart (Fig. 6). G15 has the lowest rating and it is significantly different from others except for G4 and G13.

There are some differences among G1, G2 and G3. G1 has the vibration pattern filled in gray while G2 has them in color. G3 on the other hand has color fill and a scale within each bar. G2 follows P1 and P4 while G3 has P1, P4 and P7. Although there is no significant difference in time, accuracy or perceived ease of use of G1, G2 and G3, the results show that adding color reduces the search time and helps to improve performance (Seva et al., 2022). This is also evident from the Ease of Use rating (Fig. 6). One purpose of visualization is to show relationships and differences among variables. But, visualizations can easily mislead readers into thinking that relationships or patterns exist when in reality they may not (Azzam et al., 2013). The results show that commonly used bar graphs like G1 can be improved with color fills and symbol patterns.

Based on the results of this study and the findings, the following design guidelines can be stated adding more information to the principles of Kosslyn (2006).

1. In bar comparison tasks, differentiable filling patterns or colors in the bars can make a difference. Therefore, one should try to avoid a gray vibration filling pattern in those comparison tasks so that principle P7 is adhered to with symbols to represent relative quantities (King et al., 2022).
2. Circular graphs have poor legend spatial compatibility with data points. If there is a task that needs to refer to the legend frequently, circular graphs are not advisable because the circular saccades are not efficient and they require visual transformations. Use a graph type that has greater legend spatial compatibility so that the saccades are straight and not curved.
3. Legend compatibility in bar graphs are important. Use legends on top of the vertical bars contrary to the existing guideline of a vertical legend. This is to minimize inversion between legend and bars during viewing so that principle P6 is maintained without mental transformations.
4. When using bar graphs for complex comparison tasks, it is preferable to use a symbolic filling as in G3. This will allow easier estimation.
5. Use labels at the bottom of the axis in bar graphs instead of legends if the text fits close to the axis. This reduces the eye movement distance when identifying the bars.
6. Place values on the bars only for those tasks that require numeric values. Values on the bars are not important in comparison tasks as slopes and height differences allow reasonably good estimations.
7. Radar graphs are not suitable when there is more than one data series. It is difficult to search information when there is overlapping data on a graph. Even though a radar graph conforms to P8, that is violated when circular saccades are present.
8. If using a bubble graph, a mark at the center will help to identify the exact location of the bubble.
9. Do not include redundant legend information on a graph.

A flowchart showing the above guidelines is shown in Fig. 13. There are limitations to the study as not every parameter in a graph was investigated. For example, in bar graphs, the width of the bar was set to be the same. The effect of changing the width may need investigation as there may be an interaction of width and height in bar graphs. Also, the optimal aspect ratio for graphs was not investigated.

In conclusion, when designing and selecting graphs, one has to identify the need and the end message to deliver to the reader so that time is saved, less effort is used and accuracy is improved. Bar and line graphs in conjunction with questions related to all blocks appear to be the most efficient.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The corresponding author would like to thank the funding from FSU-2022-011 and the support of the Healthcare Engineering Innovation Center (HEIC) of Khalifa University.

References

Acartürk, C., 2014. Towards a systematic understanding of graphical cues in communication through statistical graphs. J. Vis. Lang. Comput. 76–88.

Agarwal, S., Sarkar, S., 2022. Topical analysis of migration coverage during lockdown in India by mainstream print media. PLoS One 17 (2), e0263787. https://doi.org/10.1371/journal.pone.0263787.

Azzam, T., Evergreen, S., Germuth, A.A., Kistler, S.J., 2013. Data visualization and evaluation. N. Dir. Eval. 139, 7–32.

Bergauer, L., Kataife, E.D., Mileo, F.G., Roche, T.R., Said, S., Spahn, D.R., Tscholl, D.W., Wetli, D.J., 2022. Physicians' perceptions of two ways of algorithm presentation: graphic versus text-based approach. Ergonomics 65 (10), 1326–1337. https://doi.org/10.1080/00140139.2022.2029581.

Blascheck, T., Schweizer, M., Beck, F., Ertl, T., 2017. Visual comparison of eye movement patterns. Comput. Graph. Forum 36 (3), 87–97.

Brewer, C., 2020. Color Brewer 2.0. http://colorbrewer2.org/. (Accessed 18 April 2020).

Cairo, A., 2013. The Functional Art. New Riders, Berkeley, CA.

Carswell, C., 1992. 16 reading graphs: interactions of processing requirements and stimulus structure. Percepts, Concepts and Categories - The Representation and Processing of Information Advances in Psychology 605–645.

Carswell, C.M., Wickens, C.D., 1990. The perceptual interaction of graphic attributes: configurality, stimulus homogeneity, and object integration. Percept. Psychophys. 47, 157–168.

Chen, X., Jin, R., 2017. Statistical modeling for visualization evaluation through data fusion. Appl. Ergon. 551–561.

Cleveland, W., 1994. The Elements of Graphing Data. Hobart Press, Summit, NJ.

Cleveland, W.S., McGill, R., 1984. The many faces of a scatterplot. J. Am. Stat. Assoc. 79, 807–822.

Cleveland, W., McGill, R., 1985. Graphical perception and graphical methods for analyzing scientific data. Science 828–833.

Connor, Z.A., Wickens, C.D., Patton, C.E., 2022. The proximity compatibility principle and its applications to human perception of graphic displays. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 66 (1), 1727–1731. https://doi.org/10.1177/10711813226611.

Costela, F.M., Woods, R.L., 2019. When watching video, many saccades are curved and deviate from a velocity profile model. Front. Neurosci. 12, 960. https://doi.org/10.3389/fnins.2018.00960.

Cukier, K., 2010. A special report on managing information. Economist 394 (8671), 3–18.

De Valois, R.L., De Valois, K.K., 1988. Spatial Vision. Oxford University Press, New York.

Delmas, M., Caroux, L., Lemercier, C., 2021. Searching in clutter: visual behavior and performance of expert action video game players. February 2022 Appl. Ergon. 99, 103628.

Draper, G., Livnat, Y., Riesenfeld, R., 2009. A survey of radial methods for information visualization. IEEE Trans. Visual. Comput. Graph. 15 (5), 759–776.

Duchowski, A.T., 2003. Eye Tracking Methodology - Theory and Practice. Springer.

Durand, M.A., Yen, R.W., O'Malley, J., Elwyn, G., Mancini, J., 2020. Graph literacy matters: examining the association between graph literacy, health literacy, and numeracy in a Medicaid eligible population. PLoS One. https://doi.org/10.1371/journal.pone.0241844. (Accessed 11 November 2020).

Few, S., 2009. Time series analysis. In: Now You See it: Simple Visualization Techniques for Quantitative Analysis.

Few, S., 2012. Show Me the Numbers: Designing Tables and Graphs to Enlighten, second ed. Analytics Press, Burlingame, CA.

Fischer, M., Dewulf, N., Hill, R., 2005. Designing bar graphs: orientation matters. Appl. Cognit. Psychol. 953–962.

Gaissmaier, W., Wegwarth, O., Skopec, D., Muller, A.S., Broschinski, S., Politi, M.C., 2012. Numbers can be worth a thousand pictures: individual differences in understanding graphical and numerical representations of health-related information, 2012 Health Psychol. 31 (3), 286–296. pmid:21842998.

Glazer, N., 2011. Challenges with graph interpretation: a review of the literature. Stud. Sci. Educ. 47 (2), 183–210. https://doi.org/10.1080/03057267.2011.605307.

Goldberg, J., Helfman, J., 2011. Eye Tracking for Visualization Evaluation: Reading Values on Linear versus Radial Graphs. Information Visualization, pp. 182–195.

Goettker, A., Braun, D., Gegenfurtner, K., 2019. Dynamic combination of position and motion information when tracking moving targets. J. Vis. 19, 2. https://doi.org/10.1167/19.7.2.

Graham, J.L., 1937. Illusory trends in the observation of bar graphs. J. Exp. Psychol. 20, 597–608.

Hartley, J., Cabanac, G., 2014. Do men and women differ in their use of tables and graphs in academic publications? Scientometrics 98 (2), 1161–1172, 0138–9130.

Havre, S., Hetzler, B., Nowell, L., 2000. ThemeRiver: Visualizing Theme Changes over Time, pp. 115–124. IEEE Information Visualization Symposium. Salt Lake City, Utah, Oct 9-10.

Healey, C., Enns, J., 2012. Attention and visual memory in visualization and computer graphics. IEEE Trans. Visual. Comput. Graph. 18 (7), 1170–1188.

Hill, A., 1997. Readability of screen displays with various foreground/background color combinations, font styles, and font types. Proceedings of the Eleventh National Conference on Undergraduate Research II, 742–746.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., van de Weijer, J., 2011. Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press.

Huang, W., 2007. Using Eye Tracking to Investigate Graph Layout Effects, pp. 97–100. https://doi.org/10.1109/APVIS.2007.329282, 6th International Asia-Pacific Symposium on Visualization.

Huestegge, L., Philipp, A., 2011. Effects of spatial compatibility on integration processes in graph comprehension. Atten. Percept. Psychophys. 1903–1915.

Jarvenpaa, S.L., Dickson, G.W., 1988. Graphics and managerial decision making; Research based guidelines. Commun. ACM 31, 764–774.

Jeong, M.-H., Duckham, M., Bleisch, S., 2015. Graphical aids to the estimation and discrimination of uncertain numerical data. PLoS One 10 (10), e0141271. https://doi.org/10.1371/journal.pone.0141271.

Kelleher, C., Wagener, T., 2011. Ten guidelines for effective data visualization in scientific publications. Environ. Model. Software 26, 822–827.

King, B.J., Read, G.J.M., Salmon, P.M., 2022. Clear and present danger? Applying ecological interface design to develop an aviation risk management interface. Appl. Ergon. 99, 103643 https://doi.org/10.1016/j.apergo.2021.103643.

Klein, P., Becker, S., Küchemann, S., Kuhn, J., 2021. Test of understanding graphs in kinematics: item objectives confirmed by clustering eye movement transitions. Physical Review Physics Education Research 17 (1), 013102. https://doi.org/10.1103/PhysRevPhysEducRes.17.013102.

Kosslyn, S.M., 2006. Graph Design for the Eye and Mind. Oxford University Press, New York.

Krishna, O., Helo, A., Rämä, P., Aizawa, K., 2018. Gaze distribution analysis and saliency prediction across age groups. PLoS One 13 (2), e0193149. https://doi.org/10.1371/journal.pone.0193149.

Kullmann, A., Ashmore, R.C., Braverman, A., Mazur, C., Snapp, H., Williams, E., Szczupak, M., Murphy, S., Marshall, K., Crawford, J., Balaban, C.D., Hoffer, M., Kiderman, A., 2021. Portable eye-tracking as a reliable assessment of oculomotor, cognitive and reaction time function: normative data for 18–45 year old. PLoS One. https://doi.org/10.1371/journal.pone.0260351. November 22, 2021.

Lamasky, S., 1869. Pfluger's Archiv. f. d. gesammte Physiologie II, p. 418.

Lewandowsky, S., Spence, I., 1989. Discriminating strata in scatterplots. J. Am. Stat. Assoc. 84, 682–688.

Meyer, J., Shinar, D., Leiser, D., 1997. Multiple factors that determine performance with tables and graphs. Hum. Factors 39, 268–286.

Netzel, R., Ohlhausen, B., Kurzhals, K., Woods, R., Burch, M., Weiskopf, D., 2017. User performance and reading strategies for metro maps: an eye tracking study, 2017 Spatial Cognit. Comput. 17 (1–2), 39–64.

Proulx, M.J., 2011. Individual differences and metacognitive knowledge of visual search strategy. PLoS One 6 (10), e27043. https://doi.org/10.1371/journal.pone.0027043.

Robbins, N., 2005. Creating More Effective Graphs. Wiley-Interscience, Hoboken, NJ.

Schwabish, J., 2014. An economist's guide to visualizing data. J. Econ. Perspect. 28 (1), 209–234.

Seva, R.R., Wu, J.A.G., Chinjen, K.K., Estoista, N.A.T.P., 2022. Effect of color properties in multiple time series graph comprehension. Appl. Ergon. 103, 103808 https://doi.org/10.1016/j.apergo.2022.103808.

Shah, P., Hoeffner, J., 2002. Review of Graphs comprehension research: implications for instructions. Educ. Psychol. Rev. 14 (1), 47–69.

Torres, D., Sena, W.R., Carmona, H.A., Moreira, A.A., Makse, H.A., Andrade Jr., J.S., 2021. Eye-tracking as a proxy for coherence and complexity of texts. PLoS One. https://doi.org/10.1371/journal.pone.0260236. (Accessed 13 December 2021).

Tufte, E.R., 2001. The Visual Display of Quantitative Information. Graphics Press, Chesire, CT.

Tufte, E.R., 1990. Envisioning Information. Graphics Press, Chesire, CT.

Umanath, N.S., Scamell, R.W., 1988. An experimental evaluation of the impact of data display format on recall performance. Commun. ACM 31, 562–570.

Ware, C., 2004. Information Visualization Perception for Design, second ed. Morgan Kaufman, San Francisco.

Ware, C., 2008. Visual Thinking. Morgan Kaufman, San Francisco, CA.

Webber, K.L., 2018. Let Me Paint You a Picture: Utilizing Visualizations to Make Data More Accessible. Building Capacity in Institutional Research and Decision Support in Higher Education. Springer International. N.p.

Wickens, C., 2013. Spatial displays. In: Engineering Psychology and Human Performance, fourth ed. Pearson, Columbus.

Wickens, C.D., McCarley, J.S., Gutzwiller, R., 2022. Applied Attention Theory, sixth ed. CRC Press, pp. 45–50.