Eye tracking for visualization evaluation: Reading values on linear versus radial graphs

Joseph Goldberg and Jonathan Helfman

Information Visualization 2011 10: 182 originally published online 18 July 2011
DOI: 10.1177/1473871611406623

Info Vis

Research Paper

Information Visualization
10(3) 182–195

DOI: 10.1177/1473871611406623
ivi.sagepub.com
SAGE

Eye tracking for visualization evaluation: Reading values on linear versus radial graphs

Joseph Goldberg[1] and Jonathan Helfman[2]

Abstract

An eye tracking methodology can help uncover subtle cognitive processing stages that are otherwise difficult to observe in visualization evaluation studies. Pros and cons of eye tracking methods are discussed here, including common analysis metrics. One example metric is the initial time at which all elements of a visualization that are required to complete a task have been viewed. An illustrative eye tracking study was conducted to compare how radial and linear graphs support value lookup tasks for both one and two data-dimensions. Linear and radial versions of bar, line, area, and scatter graphs were presented to 32 participants, who each completed a counterbalanced series of tasks. Tasks were completed more quickly on linear graphs than on those with a radial layout. Scanpath analysis revealed that a three-stage processing model was supported: (1) find desired data dimension, (2) find its datapoint, and (3) map the datapoint to its value. Mapping a datapoint to its value was slower on radial than linear graphs, probably because eyes need to follow a circular, as opposed to a linear path. Finding a datapoint within a dimension was harder using line and area graphs than bar and scatter graphs, possibly due to visual confusion of the line representing a data series. Although few errors were made, eye tracking was also used here to classify error strategies. As a result of these analyses, guidelines are proposed for the design of radial and linear graphs.

Keywords

evaluation, information visualization, eye tracking, information graphics

Introduction

Eye tracking and information graphics

Information graphics are visualizations that convey information about data trends and distributions, such as stock or temperature graphs. Although checking a stock price or comparing daily temperatures seem to be simple tasks, they can require substantial cognitive and perceptual resources. We often take for granted the subtleties involved in selecting and designing appropriate information graphics to support value lookup and comparison for a set of data. Despite earlier research and guidelines pertaining to graph design, there is still a limited understanding of how individuals read and compare data on information graphics. As an initial step in improving this understanding, this study used an eye tracking methodology to investigate individuals' scanning strategies in reading specific types of graphs.

Eye tracking as a research methodology

Eye tracking has been historically under-utilized as a methodology for understanding how individuals read and process data on information graphics. While completion time and accuracy on specific tasks may indicate that differences or problems exist, a deeper understanding of visual scanning strategies on information graphics may help to determine specific guidelines for designing graphs and selecting graph types for particular datasets and tasks. An eye tracking methodology can help to uncover these visual scanning

[1]Applications User Experience, Oracle, Redwood Shores, CA, USA.
[2]Measurement Research Laboratory, Agilent Technologies, Santa Clara, CA, USA.

Corresponding author:
Joseph Goldberg, Applications User Experience, Oracle, Redwood Shores, CA, USA
Email: joe.goldberg@oracle.com

strategies, providing rich information beyond that available from response time and accuracy-based methodologies.

Researchers have been developing and comparing cognitive models for graph usage for at least two decades. Top–down models break tasks into theoretical sub-processes that are sensitive to variations in tasks and individuals. Bottom–up models start by comparing observed behaviors, ultimately drawing implications to higher cognitive processes. Lohse[1] developed a computational, top–down predictive model for the time required to complete common graphical tasks such as reading data, comparing data, and understanding trends. The model considered eye movements, memory, and other cognitive decision events when using bar graphs, line graphs, and tables.

Both graph design factors and individual differences (e.g., such as graph reading experience or cognitive load) have been investigated. Cognitive load or mental effort may impact the effectiveness of graph visualizations.[2] Measures that incorporate cognitive load are potentially sensitive to situations in which readers of graphs perform identically, but incur greatly varying cognitive effort. Shah[3] examined the pattern and duration of individual's scanpaths on several line graphs qualitatively, in order to validate a conceptual model of graph comprehension. The majority of eye movement time was spent relating information from the data lines to labels and other referents, rather than viewing the lines themselves. Experience in viewing these graphs was an important individual difference, in that less experienced graph readers rely upon several time-consuming information retrieval and comparison sub-stages. Carpenter and Shah[4] also modeled graph comprehension as a series of internal sub-processes. They included encoding, pattern interpretation, and integration stages. Using eye tracking methods, they drew implications for improving the format of graphs to make them easier to interpret. When reading a radar versus a multi-dimensional Starchart visualization of environmental changes over time, individual differences may not be a significant predictor of overall visualization effectiveness.[5] However, there is evidence that those with slower perceptual speed preferred the radar over the more complex visualization.[5] Two-dimensional graphs usually invoke faster response time than three-dimensional graphs, partly due to the presentation of less chart junk, and a greater ratio of data ink to surface area.[6,7]

Graph research has also investigated how individuals explore node-link, or "network," graph diagrams, to improve layout effectiveness. While these graphs have a more flexible layout than visualizations such as bar or radar graphs, they still require users to explore edges to seek out answers to questions. When extracting information from network graphs, important factors include path length, path continuity, number of branch nodes, and number of edge crossings.[8] There is also eye tracking evidence that individuals exploring social relationships follow edges on their way to target nodes, and tend to avoid areas of dense edges and nodes.[9]

Advantages and disadvantages of eye tracking

Eye tracking has several unique advantages as an evaluation method, but the analysis of eye tracking data is still tedious and potentially subject to error and misinterpretation.[10–12] Advantages of eye tracking methods include:

Accessing pre-attentive behavior: Eye movements may appear to be very stochastic when observed directly, but are controlled by higher level cognitive processes.[5] Individuals are not aware of their lower level eye fixations, but are often aware of their high-level control strategies. In this sense, recording an individual's stream of fixations while conducting tasks is accessing pre-attentive behaviors that, when evaluated statistically across many conditions, can provide meaningful strategy information.

Understanding sequential strategies: Whereas traditional evaluation studies have relied upon response time measures, sequences of fixated areas can help inform the strategy used between the appearance of a task, and its completion.[13] Understanding differences in these sequential strategies between various design alternatives is valuable for improving designs to help maximize efficiency.

Error analysis: Comparing differences in sequential strategies between error and non-error trials provides diagnostic information to a designer that is hard to obtain from response time methods. Errors that occur within a fraction of a second can be investigated by studying landing locations of individual fixations.

Eye tracking methods, however, must be used with caution. Problematic issues in eye tracking include:

Defining fixations: Gaze samples typically collected at 60–120 Hz are algorithmically filtered to fixations that are approximately 3 Hz. There are various temporal and/or spatial dispersion algorithms used, that are often unreported.[14] It is not yet well-understood how the use of different dispersion algorithms might affect results of analysis on the resulting fixations.

Specifying areas of interest: In order to generate metrics of eye tracking performance, distinct areas on a viewed scene or image, known as areas of interest (AOI), can be defined after obtaining a dataset. Metrics such as time to first fixation in AOI, AOI-order, percentage time viewing AOI, and others are quickly generated by most eye tracking software. Specifying these AOIs, however, is not always straightforward. Figure 1 shows both linear (A) and radial (B) bar graphs with AOIs defined for bar heights, regions, and labels. There are no standard practices pertaining to the minimum size of AOIs, padding surrounding an AOI, consistency among AOI areas, or specifying AOI hierarchies. In Figure 1, (darker) circle diameters are as wide as possible in the linear graph, so as not to overlap with adjacent circles. In radial graphs, however, these circle diameters are constrained by the separation of rings denoting graph values, and must be smaller near the graph center to avoid overlap. Visual scenes are also not static – AOIs come and go frequently, as when an observer opens a menu, which may both hide and reveal AOIs, or expands a panel in place, which may trigger a new layout and subsequent repositioning of AOIs.

Defining metrics: Many metrics may be defined from eye tracking data, but selection of appropriate metrics depends on task questions. Sources are available that define and compare potential metrics on parameters such as efficiency and effectiveness.[15,16]

Gaze location error: Despite claimed accuracies of 0.5 arc degrees or less by eye tracker manufacturers, there are many sources of gaze location error. Error can accumulate from poor calibration,[10] observer attentional dissociation,[12] and differences in retinal acuity among individual eyes.[15,11] A location error of 1 cm is commonly used with reference to eye tracking on computer displays, with an eye-screen distance of 50 cm.

Scanpath interruptions: The sequence of fixations made by an observer's eyes and recorded by an eye tracker may be interrupted for reasons such as blinking or looking off-screen. Because eye tracking software solutions usually link the last recorded fixation with the next available fixation, interruptions can create artificially long saccades. Best practice guidelines are needed to handle these interruptions.[11]

Studies that are concerned with sequential visual strategies for completion of tasks can benefit from eye tracking methods. However, careful attention must be paid to defining appropriate eye tracking metrics. Even then, it can be difficult to compare results and conclusions across eye tracking studies, due to differences in defining methods and metrics.

Linear versus radial information graphics

This research includes both linear and radial graphs. Linear graphs encode values along perpendicular dimensions. Popular linear graph types include bar, line, area, and scatter graphs (Figure 2, A–D).

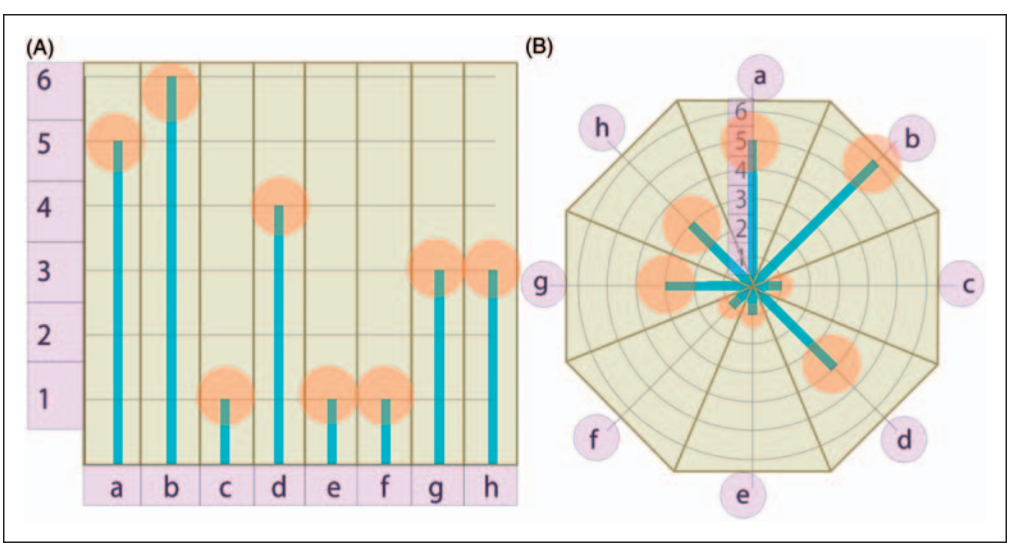

Figure 1. Linear (A) and radial (B) bar graphs, showing AOI placement for background regions, dimension labels, value labels, and data values.

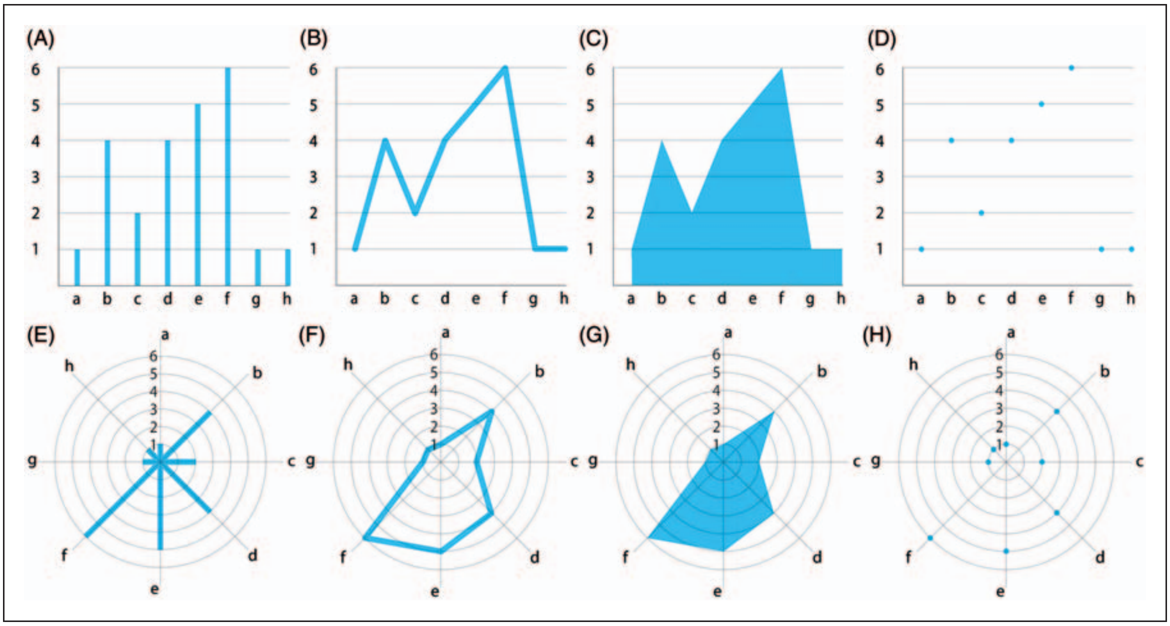

Figure 2. Eight graph types, plotting identical 8-point datasets: linear bar (A), line (B), area (C), and scatter (D), radial bar (E), line (F), area (G), and scatter (H).

Each of these plots the same dataset; each conveys quantitative information along the y-axis and categorical information (a–h here) along the x-axis. (Note that conveying categorical information on a line or area graph is not typically recommended, as misinterpretation could occur in showing a non-existent quantitative trend.)

Radial graphs are circular data visualizations that encode quantitative values along axes that radiate from the center outward. Figure 2 (E–H) shows the same dataset plotted as radial bar, line (a.k.a. "radar"), area, and scatter graphs. Although used in business domains such as human resource management and power generation monitoring, radar graphs are not generally considered to be as effective as bar graphs because it can be difficult to read values arranged in a circle effectively.[17] Radar are, however, effective for assessing the symmetry of values across a quantitative dimension, rather than comparing their magnitudes where a bar graph would be more effective.[14] They can also highlight extreme values along these dimensions.[18]

Processing stages in reading graphs

Information graphics support at least three different levels of perception and interpretation: (1) lookup of values, much like a tabular representation, (2) comparison of values, and (3) appreciation of higher level patterns, such as trends and distributions.[3,17] A comprehensive understanding of each of these levels is quite complex. The focus in this article is on how eye tracking methods can inform the easiest of these to define and compare: value lookup. Consider the same dataset plotted in linear and radial bar graphs (Figure 3A and B). A sample task on these is to *find the value of dimension* "d". As illustrated by the numbered fixation sequences in Figure 3, the cognitive stages for this task are expected to be:

(1) Find dimension →
(2) Find associated datapoint →
(3) Get datapoint value

These stages are similar to those expressed by other researchers.[1] Differences in difficulty and temporal duration can be expected for each of these three stages, based in part, on graph clutter and clarity.[7]

Graph reading tasks may also commonly require multiple completions of the above three-stage model. Consider the task, "What is the reported difference in values between dimensions 'b' and 'e'?" Figure 3 also shows both linear (3C) and radial (3D) area graphs, with the hypothetical stages required to complete this task. The first dimension must be located (Stage 1), its datapoint found (2), and its value noted (3). Then, the second dimension is located (4), datapoint found (5), and value noted (6). The cognitive subtraction can then be conducted and reported. Task difficulty

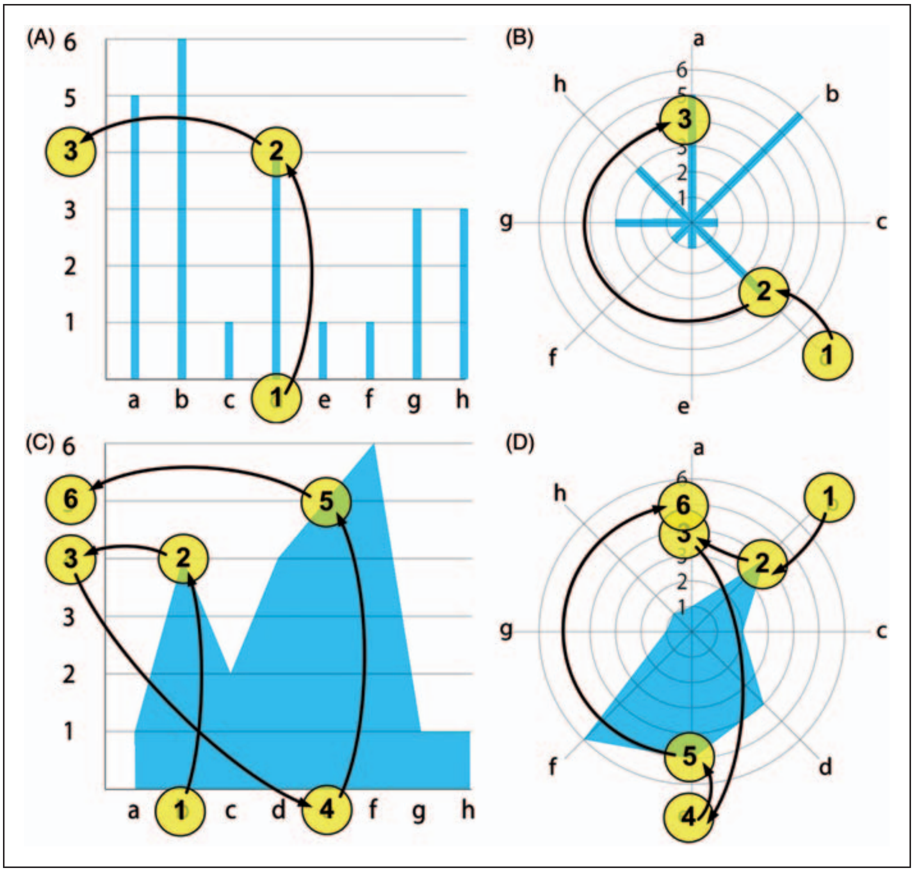

Figure 3. Four types of graphs, plotting the same dataset, illustrating strategy differences in reading and comparing values. Dimension ''d'' value lookup on linear bar graph (A), and on radial bar graph (B). Reporting the difference in value between dimensions ''b'' and ''e'' on linear area graph (C) and radial area graph (D).

increases if the separation of dimensions is hard to see due to line crossings, and with physical separation of data from dimension labels. The difficulty also increases with dimensions that are hard to identify, and with individual values that are masked or obscured. In addition, eye movements will need to follow a circular path in radial graphs, which may be harder to coordinate than simpler horizontal or vertical visual movements.[14,19] This predicts that mapping datapoints to their values in radial graphs should be harder for longer circular traverses (toward the outer rings) than for shorter circular traverses (inner rings). In addition, rings that are too close together may cause a user to "jump" to a neighboring ring, thereby misreading a value.

While simple comparisons are essentially two-value lookup tasks, harder comparison tasks may require comparison and lookup of many more values and dimensions, with concomitant demands placed on perception and memory. Harder tasks also yield much more variance in task solution scanning strategies, increasing the difficulty of data analysis.

Objective

Providing effective graph visualizations for a dataset and task is critical to successful transmission of complex information. Business tasks such as reading and comparing sales information between graph categories are frequently performed tasks. Despite potential shortcomings, eye tracking can provide useful information for effective incorporation of information graphics into a user interface. This study generated information to support the development of qualitative scanning strategies, using quantitative and objective data generation and analysis. We conducted a study to illustrate the value of eye tracking in the evaluation of visualizations, in general, and to specifically compare strategies for value lookup in linear and radial graphs. The study compared individuals' ability to read plotted data in bar, line, area, and scatter graphs presented in both linear and radial configurations. At a high level, we illustrate the value of eye tracking methods for generating guidelines for presenting visualizations. Our specific study objective was to determine whether differences in measured performance among graph types can be attributed to differences in recorded eye tracking-based metrics.

Methods

Participants

Thirty-two colleagues (18 males, 14 females) from Oracle Corporation participated in this study, without remuneration. Each was college educated, and each was experienced at reading data series in business graphs. These participants had a mean (median) age of 36.5 (37.0), with an overall range of 24–61 years. Each was screened for acceptable color vision using six Ishihara color plates on an Optec 2000P vision tester.

Apparatus

A Tobii T60 binocular eye tracking system, running Tobii Studio version 2.1.14 software collected eye tracking data.[20] This system provides a 17-inch diagonally measured 1280×1024 display ($\sim$98 dpi). Analysis was conducted using a combination of Tobii Studio, Microsoft Excel, and Minitab software. Low-level fixations were defined using 50-pixel minimum radius combined with a 100-ms minimum temporal threshold.

Graph stimuli

A single graph containing a single data series was presented on each trial of this experiment. Each graph displayed ordinate values, labeled 1–6, and abscissa values, labeled a–h. Graphs plotted one of two, 8-point datasets that were previously randomly generated using discrete values in the range of 1–6. Each dataset had mean = median = 3.0, and each had variance of 3.0–4.0. Graph styles included bar, line, area, and scatter, in both linear and radial graph types, as previously shown in Figure 2. Each graph was defined within a 20×20 cm area, using equal-width data lines and points. To minimize misinterpretation of which ink was used to show data, all datapoints, lines, and areas were drawn in blue, while the graphs' axes, grid lines, and labels, were drawn in gray.

Tasks

A sales scenario was used, with sales values presented by region in each graph. Tasks required participants to look up values in either one or two dimensions as follows:

- Value lookup in single dimension (eight tasks): participants stated the sales value associated with a given region, e.g. *What sales value was reported by Region e?*
- Add or subtract values from two dimensions (two tasks): participants looked up the values associated with two dimensions, then added or subtracted them, e.g. (1) *What was the total sales reported from regions c and f?* and (2) *What was the reported difference in sales between regions b and e?*

Procedure

Each participant was tested individually over a 30-min period. The eye tracker height and distance (50–60 cm) was adjusted to the participant, and a 5-point calibration was completed. The following scenario was read aloud: *"You are a sales analyst, looking at current sales numbers for your global company. The following graphs represent sales revenues within eight global regions, labeled a-h. Sales results are integers, labeled 1-6 on the graphs"*. Five practice tasks were completed, followed by 16 non-practice trials.

Each trial started with the presentation of a task, which the participant read aloud. Following any clarifying questions, the participant was shown the graph. Elapsed time for each task was determined when the participant verbally responded with the information requested by the task. The experimenter then presented the next task. No input device was used by the participant.

Experimental design

The data presented in this article represent a portion of the data collected from a larger study. The larger study presented four graph types (bar, line, area, scatter) $\times$ two graph styles (linear, radial) $\times$ two task types (value lookup, value comparison) to each participant. Each participant viewed two replicates of each graph style-type presented in one of 16 defined orders, for a total of 32 trials. The counterbalanced experimental design also included eight replicates of each task type assigned to each graph.

Due to space constraints, this article includes only value lookup tasks from the larger study. Although quite simple, these lookup tasks provided a wealth of eye tracking data to inform the stage results presented below. The value comparison tasks promoted many different scanning strategies, which are beyond the scope of the stage analyses presented here.

Results

AOI definitions

AOIs were defined for linear and radial graphs in order to capture fixations relative to the present tasks. Thirty AOIs were assigned within several groups, for both linear and radial graphs: value labels (1–6), dimension labels (a–h), dimension regions (a–h), and data vertices (a–h). Figure 1 showed examples of AOI assignments for linear and radial bar graphs. AOI sizes could not be controlled *between* linear and radial graphs, for three reasons: (1) less space was available for data vertices toward the center of radial graphs to avoid overlap, so a smaller diameter AOI was used for the innermost ring, than all other rings; (2) regions mapped all available space for a dimension, which was not equal between the graph types; (3) labels for values and dimensions differed somewhat in size between the two graph types.

Metrics

Several metrics were defined and used in this study, in order to extract information related to visual scanning strategies.

Response time (seconds): After the participant read and understood the task question, the response time was initiated with the appearance of a graph. The response time clock stopped with the participant's verbal answer.

Required AOIs: Each task-graph combination had a set of AOIs that had to be viewed in order to complete the task. As previously introduced, the value lookup tasks required participants to find a (1) dimension label, (2) datapoint, and (3) data value.

First fixation time (seconds): The initial time that a fixation was made within a specified AOI.

Minimum time (seconds): The (theoretical) minimum time on a task was defined at the instant a participant had completed initial fixations within all of the required AOIs on a task-graph combination. This is the theoretical earliest time at which a correct response could be made on a trial.

Processing stage durations (seconds): The minimum time is comprised of stages, each of which has starting/ending times, a duration, and an order. These are determined by the pattern of first fixations within the required AOIs.

Validation time (seconds): The difference between the response time and the minimum time is the validation time, a period where the participant often repeatedly looked back and forth before responding. Longer validation times are associated with greater uncertainty.

Univariate, mutifactor ANOVAs on the time-related metrics treated graph type, graph style, and task as main effects. Additional covariates in the model included trial order, dataset, and participant. Tukey's pairwise comparisons (using $p < 0.05$ family error) helped to explain significant factor effects with more than two levels.

Single-dimension value lookup tasks

Minimum and response times. Radial graph response times were about 1 s slower than linear graph response times (3.4 *vs.* 2.5 seconds, $F_{1,237} = 61$, $p < 0.001$), but no significant difference in times was noted between the four graph styles ($p > 0.10$). The minimum time showed the same advantage for linear over radial graph lookup time (2.1 *vs.* 1.5 seconds, $F_{1,243} = 32$, $p < 0.001$). Figure 4 shows both response and minimum times as a function of graph type and style. Participants generally needed 0.6–1.0 seconds for validation, before actually responding; radial graphs required an additional 0.3 seconds validation time compared to linear graphs, a significant difference ($F_{1,243} = 12.5$, $p < 0.001$).

As determined by eye tracking data, the minimum time was influenced by the graph type ($F_{1,215} = 41$, $p < 0.001$), lookup dimension of the task ($F_{7,215} = 8.6$, $p < 0.001$), and the dimension × graph style interaction ($F_{21,215} = 1.7$, $p < 0.05$). Dimensions "d" and "e", overall, produced slower minimum times than the other dimensions (Figure 5; Tukey, $p < 0.05$). The times were slightly faster on either end of the eight dimensions in linear graphs, and increased for dimensions "d" and "e" in radial graphs. It is likely that participants initiated scanning in the upper half of the radial graphs, and on either end of the linear graphs.

Processing stages. Analysis of eye movements showed that most of the scanpaths from the 32 participants followed the three-stage pattern: (1) FindDimension, (2) FindDatapoint, and (3) GetDatapoint Value. If a scanpath skipped a stage, moving directly from a dimension to its peripherally viewed value, the skipped stage was assigned a 0-s duration. Figure 6 shows the mean first fixation times for these stages, by graph type and style. An ANOVA corroborated that the stage times for radial graphs were slower than those for the linear graphs ($F_{1,734} = 114$, $p < 0.001$), and that the three stage times significantly differed ($F_{2,734} = 139$, $p < 0.001$; Tukey, $p < 0.05$). Time differences between the four graph styles were not significant ($p > 0.1$). There was greater slowing in first fixation times for radial than linear graphs as processing stages moved forward (graph type × stage interaction, $F_{2,734} = 15$, $p < 0.001$). Further analysis revealed that the time difference between (2) FindDatapoint and (3) GetValue stages was significant for radial graphs, but not for linear graphs (Tukey, $p < 0.05$). Overall, linear graphs provided faster first fixation times than radial graphs for each of the stages; mapping a datapoint to its value was especially slow for radial graphs, compared to linear graphs.

As illustrated in Figure 7, durations within processing stages generally decreased as processing progressed ($F_{2,734} = 57$, $p < 0.001$). Radial graphs also resulted in significantly longer stage duration times than linear graphs ($F_{1,734} = 56$, $p < 0.001$), but differences among the four graph styles were not significant ($p > 0.1$). A significant graph style × stage interaction ($F_{6,734} = 2.3$, $p < 0.05$; Tukey, $p < 0.05$) revealed that: (1) there was no difference in processing duration within each stage among the four graph styles, (2) finding a datapoint (Stage 2) was faster for scatter and bar graphs, than for area and line graphs, and (3) while finding a datapoint (Stage 2) was relatively slow for area graphs, mapping a datapoint to its value (Stage 3) was completed as fast as for other graph styles.

Add/subtract values in two dimensions tasks

Minimum and response times. Each of these tasks was more complex than the single-dimension

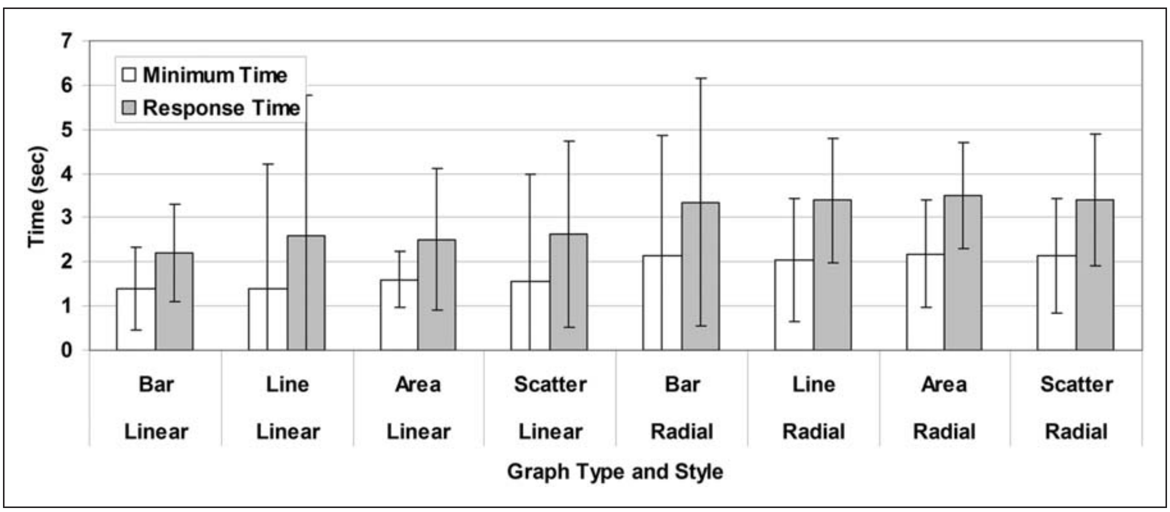

Figure 4. Mean minimum and response times by graph type and style for value lookup tasks. Error bars indicate ±1 standard deviation (SD).

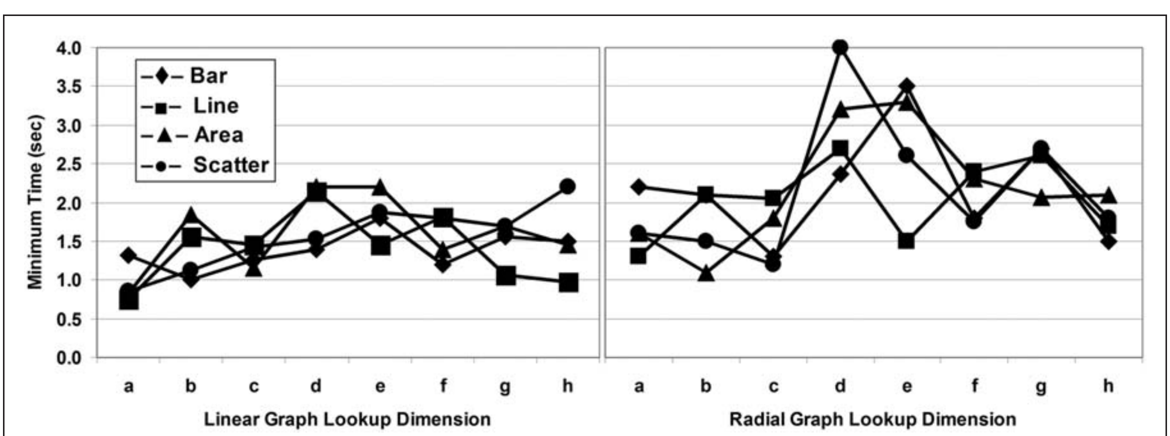

Figure 5. Comparison of mean minimum time by required dimension and graph type. Linear graphs are shown on the left and radial graphs on the right.

value lookup tasks, requiring looking up values on two dimensions, performing an arithmetic operation: either adding ("Task 1") or subtracting ("Task 2"), and then reporting the results. Of the 64 presented trials across participants, 1 trial was eliminated due to insufficient eye tracking data, and 3 were eliminated due to response errors. Minimum times were defined as soon as fixations were made in all six required AOIs, regardless of order. Both minimum and response times, shown in Figure 8, were significantly slower for radial graphs than for linear graphs (response times: 6.5 *vs.* 4.9 seconds, $F_{1,45} = 11.2$, $p < 0.01$; minimum times: 4.6 *vs.* 3.9 seconds, $F_{1,141} = 23$, $p < 0.001$).

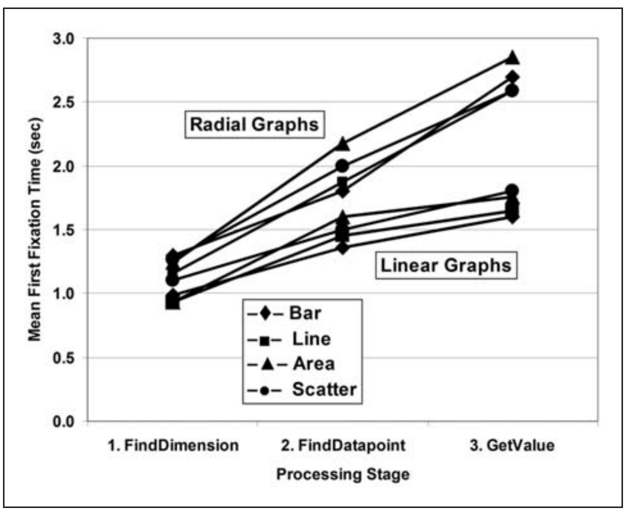

Figure 6. Mean first fixation times as a function of graph type, style, and processing stage. Radial graphs produced the upper four lines, and linear graphs produced the lower four lines.

A significant difference between these two tasks was noted in the number of fixations required at the minimum time. Task 1 (addition tasks) and Task 2 (subtraction tasks) had an average of 13.1 and 16.2 fixations, respectively ($F_{1,44} = 5.9$, $p < 0.05$). The number of fixations did not depend on the graph type or style ($p > 0.05$), therefore it is possible that tasks requiring values to be added (e.g. Task 1) may require fewer fixations than those requiring a subtraction (e.g. Task 2). The subtraction operation depends on the order of its operands, whereas addition does not. Additional fixations may be generated during subtraction, when deciding which of two values is larger.

The order of initial fixations on dimensions could potentially vary in these tasks, because information from more than one dimension was required for each task question. The mean first fixation times from Task 1 showed that dimension "c" was always fixated before dimension "f"; means from Task 2 showed that dimension "b" was fixated prior to dimension "e".

Processing stages. Nearly all scanpaths from these two-dimension tasks each flowed *twice* through the same three processing stages used by the single-dimension value lookup tasks. An additional cognitive stage could also be added for summing or subtracting the two values, once they were found. Minimum times were significantly influenced by processing stage (Figure 8; $F_{2,141} = 22$, $p < 0.001$), with Stage 1 times significantly faster than Stage 3 times (Tukey, $p < 0.05$). The task × graph type interaction was significant ($F_{1,141} = 13$, $p < 0.001$), due to slower linear than

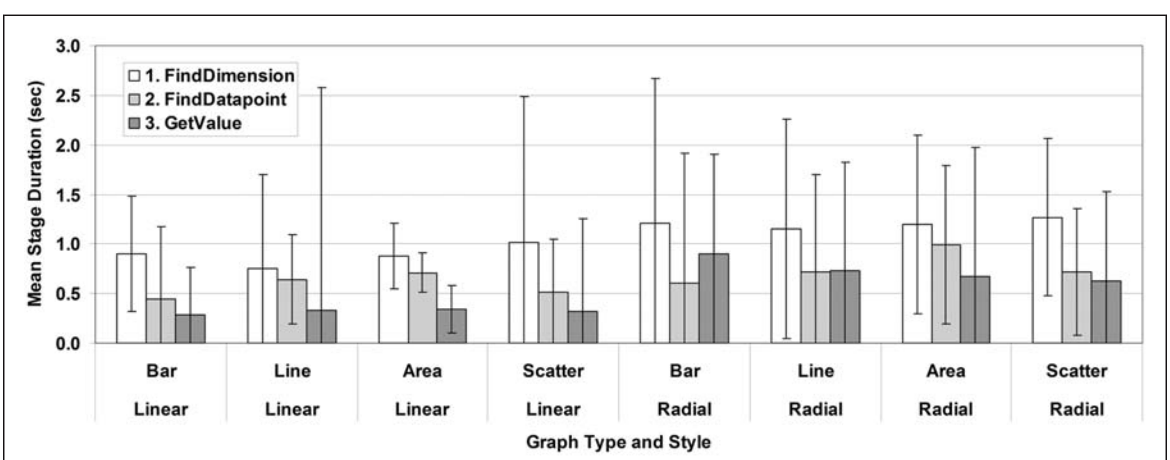

Figure 7. Mean durations of processing stages in single-dimension lookup tasks, by graph type and graph style. Error bars indicate ± 1 SD.

radial graph times for Task 1 (addition), and the reverse for Task 2 (subtraction). A significant task × processing stage interaction was significant ($F_{3,141} = 3.0$, $p < 0.05$), due to slower Stage 3 times for Task 2 (subtraction) than for Task 1 (addition).

Durations of the scanpath-determined processing stages were longer using radial than using linear graphs (Figure 9; $F_{1,141} = 6.3$, $p < 0.05$). Durations also differed among the three processing stages ($F_{2,141} = 27$, $p < 0.001$), with Stage 1 requiring more time than Stage 3 (Tukey, $p < 0.05$). A significant stage × task interaction (Figure 10; $F_{2,141} = 6.5$, $p < 0.01$) was caused by longer Stage 3 durations for Task 2 (subtraction) than for Task 1 (addition; Tukey, $p < 0.05$).

Error analysis. Classification of incorrect scanning strategies that lead to errors can be conducted, based upon eye tracking scanpaths. These strategies can be studied easily in the present case because only three errors were made. Scaling up to greater numbers of error strategies can be extremely tedious, requiring pattern analysis, or classification tools to define groupings.[8]

Classification of the three errors in this study is given in Table 1. Two of the errors were due to following the incorrect ring on a radial bar graph. The third error was due to finding an incorrect datapoint after identifying the correct dimension in a linear bar graph. Each of these errors has potential design implications for the graphs, discussed further below.

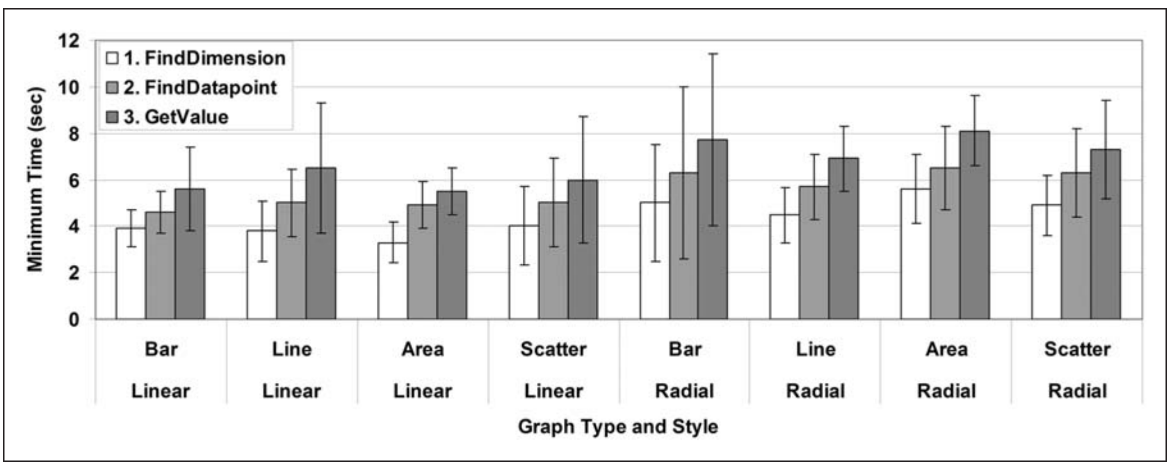

Figure 8. Mean minimum times for processing stages for two-dimension add/subtract tasks, by graph type and style. Error bars indicate ± 1 SD.

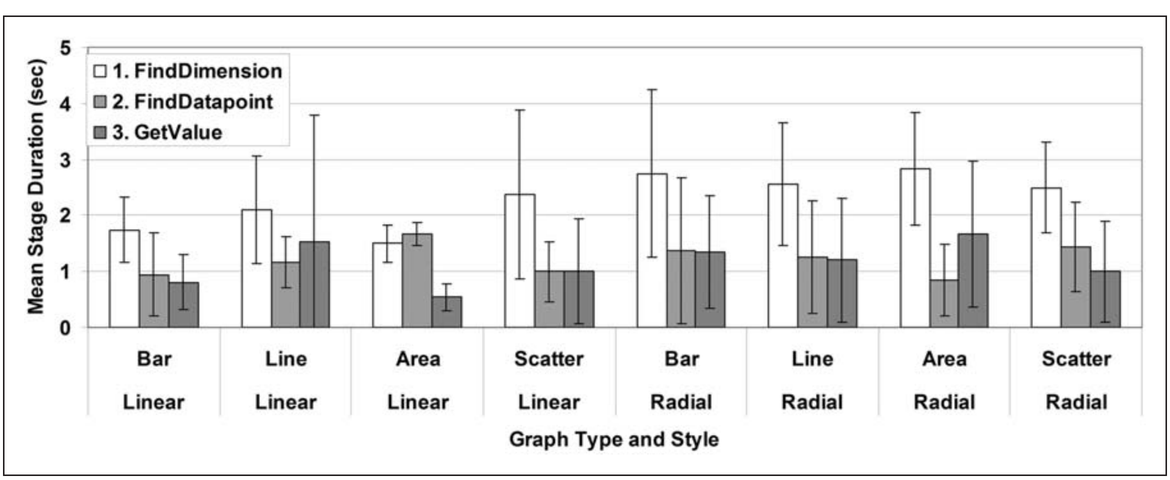

Figure 9. Durations of processing stages, by stage, graph type, and graph style, in two-dimension add/subtract tasks. Error bars indicate ± 1 SD.

Results summary

Single-dimension lookup. Both response and scanpath-computed minimum times using radial graphs were slower than those using linear graphs. No response errors were made, and most minimum time was due to validation and error-checking prior to responding. Graph style did not directly influence either minimum or response times in these tasks. The dimension influenced value lookup, however, implying that participants started scanning at the top of radial graphs and at either end of linear graphs. Participants followed a three-stage, (1) FindDimension, (2) FindDatapoint, (3) GetValue, lookup model in these tasks. Initial fixation times increased from Stages 1 to 3 for all graphs, but this increase was greater for radial graphs than for linear graphs. In particular, mapping a datapoint to its value (Stage 3) required substantially more time to achieve (0.75 *vs.* 0.25 s) for radial than for linear graphs. Considering the graph styles, the decrease in duration of Stage 2 from Stage 1 was greater for scatter and bar graphs, than for line and area graphs. Once a dimension was located, its datapoint could be found more quickly for scatter and bar graphs, than for line and area graphs. All graph styles resulted in equally rapid mapping of datapoints to their values.

Two-dimension lookup. Six target AOIs were fixated in any order to complete these tasks, amounting to two repetitions of the prior three-stage model. Minimum times and stage duration times were computed by adding the first fixation times from relevant AOIs. Participants required substantially more time to complete these tasks using radial than linear graphs (6.5 *vs.* 4.9 s), but these times were not influenced by graph style. Neither the minimum times nor the number of fixations were directly influenced by graph type or style, but the subtraction task (Task 2) required three additional fixations than the addition task (Task 1). This result was due to longer duration spent in Stage 3 for the subtraction task (Task 2) than for the addition task (Task 1). Overall, the radial graphs yielded faster minimum times for the subtraction task, but the linear graphs supported faster minimum times for the addition task. After classifying three failed scanning strategies, two errors were made while using radial bar graphs, visually following the wrong concentric ring. The third error was made while using a linear bar graph, incorrectly associating a dimension with the wrong datapoint.

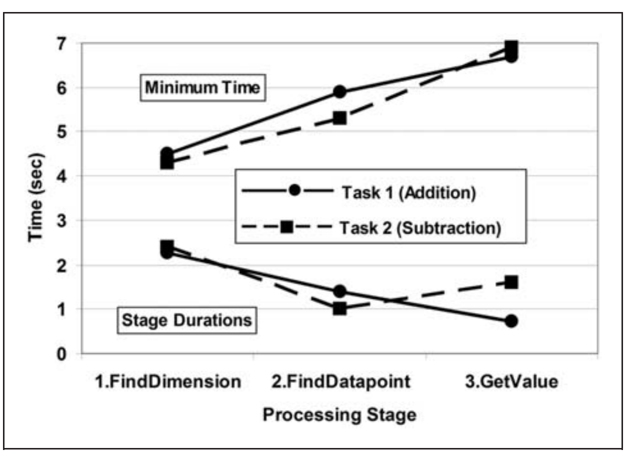

Figure 10. Processing stage mean durations and minimum times. Minimum times are plotted in upper two series, and stage durations in lower two series. Significant interaction was noted between the two duration series.

Discussion

Eye tracking has been under-utilized in the analysis of visualizations, such as the information graphics presented here. Although reduction and analysis of results can be tedious, this study demonstrates the potential richness of this methodology for understanding visual scanning behaviors while using visualizations. Design improvements and guidelines can be developed by considering scanpaths, theoretical minimum completion times, and error strategies.

Reading information graphics may seem to be an activity requiring little cognitive resources, but it can

Table 1. Classification of errors made in two-dimension tasks

Tasks	Graph	Processing stage	Error description
1 (Addition)	Radial bar	3. GetValue	Followed wrong ring when mapping 2nd datapoint to value; circular scanning error
2 (Subtraction)	Linear line	2. FindDatapoint	Mapped dimension to wrong datapoint; horizontal scanning error
2 (Subtraction)	Radial bar	3. GetValue	Followed wrong ring when mapping 2nd datapoint to value; circular scanning error

be a complex activity, requiring the extraction and comparison of multiple pieces of information. While multiple stages of processing in reading graphs have been previously theorized,[1,4] this study explicitly defined and compared these stages across a number of graph types.

Eye tracking supported an understanding of the cognitive stages required during information lookup that is difficult to capture using more traditional methods. Most eye movements are pre-attentive, and can reveal scanning trends that are not subject to one's awareness. Precise metrics can be defined from the timing and duration of eye tracking-derived processing stages, allowing statistical comparison among graph types and styles. Eye tracking analysis can also delineate stages in the use of other visualizations, such as treemaps or social network graphs.[8,9] Eye tracking is a rich information source that can support categorization of search strategies while using a visualization.

Eye tracking practice and analysis methods are subject to many potential sources of error. Calibration can drift over time, fixations may not locate directly on a gaze location, and scanpaths may be interrupted. AOIs defined during data analysis may not all be the same size, due to constraints present in stimuli. Frequent, somewhat random, eye movements are typically made, for which patterns are difficult to uncover. Defining metrics from fixations, scanpaths, and AOIs can also be an extremely tedious process. This study attempted to overcome these and other challenges by using counterbalanced experimental design, careful calibration, precise definition of metrics, statistical analysis incorporating covariate factors, and consistent fonts, colors, and line widths in the stimulus images.

Definition of AOI size and shape in eye tracking studies is an important topic for future research. If context is not considered, larger AOIs should capture greater numbers of fixations about a visual target, and should therefore result in faster times to first fixation on that target. Although AOIs in this study were of vastly different shapes, the greatest impact on fixation times should be the diameter of the AOI in the direction of the scanpath at the time of first entry. In this study, scanpath direction was generally left or right for linear graphs, and clockwise or counter clockwise for radial graphs. The AOI diameters in these directions were actually fairly similar the graph types, especially toward the inner concentric rings of the radial graphs. Further research should establish whether scanning direction, AOI size, AOI shape, label size, or other factors influence the time to first fixation within these areas.

There are several explanations why tasks completed with radial graphs were completed more slowly than the same tasks completed with linear graphs. First, the type of lookup tasks that were used in this study are not typically recommended for radial graphs, due to the non-data ink and complexity of radial graphs compared to linear graphs.[7] Radial graphs are most useful for finding extreme values[18] and for determining whether values are symmetric across several dimensions.[14] The design of this study was more focused on careful experimental controls, than on ensuring that each visualization was appropriate for each task. Second, although participants in this study were all familiar with radial graphs, they had more experience with linear graphs, possibly resulting in stronger expectations about the placement of data labels and values.

Strong, repeatable evidence was found for a three-stage process in looking up data values on graphs. The data dimension must be located, its datapoint found, then the datapoint is mapped to a value. This process held for all four styles of both radial and linear graphs, and also held (using two iterations) for two-dimension lookup tasks. Finding the right dimension(s) took the longest, and mapping the datapoint to its value was typically the fastest stage. Finding the datapoint value took longer when the dimension was located farther from the value axis. This stage was slower for those graphs with dimensions toward the right side of linear graphs, or toward the bottom of radial graphs. Radial graphs also have the potentially error-prone property that the reader's eyes must follow one of several concentric rings to map a datapoint to its value.

Error analysis can illuminate alternative strategies, or may point to unclear or poorly comprehended elements of a visualization. Only three errors were made in this study, too few to seriously impact design. However, given sufficient errors, categorizing these can be a valuable use of eye tracking data. Two categories of errors were noted here, based on processing stage analysis across both graph types: mismatching a datapoint to its value, and mismatching a dimension to its datapoint. Both errors that stemmed from using the radial graph were caused by mistakes in mapping specific datapoints to their values. Each of these required traversing a circular ring around the graph to find a value label. Providing rings that have greater separation, or unique line codings could aid the mapping of datapoints to their values. Providing multiple axes with value labels can increase clutter, but may also shorten the scanning distance from datapoint to value label. The sole error using a linear graph was due to mismatching a dimension to its datapoint, in a line graph. This could occur if categories (or values) along the abscissa are too dense, or in cases where a line series contains many datapoints.

Finding a datapoint in single-dimension lookup tasks was easier for scatter and bar graphs than for line and area graphs. A possible explanation may be that line and area graphs have additional diagonal edges and non-data ink within the central area of the graphs that could provide visual distractions. The ink used to draw the bars in bar graphs is also extra and does not actually show data, but the ink along each bar's length may be used as a visual aid to help the eyes track the distance between dimension labels to datapoints. Unlike bar graphs, the extra ink in line and area graphs does not connect dimension labels to datapoints.[7] In fact, the extra ink in line and area graphs cuts across the distance between dimension labels to datapoints at many different angles that are a function of the difference between consecutive data values and the distances between dimensions (on the horizontal axis for linear graphs or the angular distance for radial graphs). Area graphs, in particular, also convey emergent shapes, which could further confuse the observer into perceiving false angular lines.

Visualization tasks requiring subtraction of values may require more fixations than tasks requiring addition of values. Although replication of this result with other tasks is needed, a possible implication is that tasks requiring order dependent operations (e.g., subtraction) generate longer scanpaths than tasks in which operands can be processed in any order (e.g., addition). Order dependency may promote additional comparison by requiring an individual to identify the largest of two or more operands.

Conclusion

The efficiency of information graphics may be potentially enhanced by considering ways to support each of the stages within the value lookup model. Information visualizations such as the graphs presented here allow the viewer to comprehend larger trends, beyond single data dimensions and values. Linear graphs can better support the initial, dimension-finding stage than radial graphs, because their horizontally or vertically aligned dimension/value labels better support an organized search than radially aligned labels. In fact, participants started searching both clockwise and counter clockwise for radial dimensions, unlike the rightward searching in linear graphs.

Although realistic expectations must be held concerning eye tracking's shortcomings, this methodology is effective for evaluating specific design issues of many types of data visualizations. Analysis of pre-attentive, scanpath, and fixation-based metrics can point to design and navigation issues, and their potential improvements. This study illustrated the utility of eye tracking by comparing eight different information graphics to determine their relative ability to support value lookup tasks. These same methods can be employed in the analysis of information lookup and comparison tasks for other visualizations.

Although guidelines for graph design as an outcome of this study are neither surprising nor novel, the present empirical results help to establish a scientific foundation for design heuristics. Guidelines from the present results include:

- Tasks requiring value lookup on one or two dimensions should generally use linear, rather than radial graphs. Visual scanning can be conducted more quickly along the vertical and horizontal axes of these graphs. Circular scanning along rings is error-prone and not reliable.
- The number of concentric rings in a radial graph should be defined by the minimum number required to support relevant tasks. Graphs with rings that are too close invite errors when looking up values.
- Concentric rings in radial graphs should be distinguishable by varying major and minor value line thickness, or colors. Of course, increasing the saliency of these rings must be balanced with their potential to also increase graph clutter.
- Value labels in radial graphs could be repeated on multiple axes, to minimize visual scanning errors. Minimizing extensive traversal distance should lessen the chance of mismatching a value label to a datapoint.
- The density of dimensions (or labels) along a category axis should be limited, to avoid mismatching a dimension label to its respective datapoint.
- Plotting very small values in a radial graph should be avoided, because confusion among dimensions increases toward the origin of these graphs.

Acknowledgement

The authors sincerely and greatly appreciate the efforts of Sally Tang, for both testing participants and reducing large quantities of eye tracking data.

References

1. Lohse J. A cognitive model for the perception and understanding of graphs. *CHI '91 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New Orleans, Louisiana), ACM Press: New York, 1991; 137–144.
2. Huang W, Eades P and Hong S. Measuring effectiveness of graph visualizations: a cognitive load perspective. *Inform Visual* 2009; 8: 139–152.
3. Shah, P. A model of the cognitive and perceptual processes in graphical display comprehension. *Proceedings of American Association for Artificial Intelligence Spring Symposium, AAAI Technical Report FS-97–03*. Stanford University, CA, 1997.

4. Carpenter PA and Shah P. A model of the perceptual and conceptual processes in graph comprehension. *J Exp Psychol Appl* 1998; 4(2): 75–100.
5. Conati C and Maclaren H. Exploring the role of individual differences in information visualization. *AVI '08, Proceedings of the Working Conference on Advanced Visual Interfaces* (Napoli, Italy), ACM Press: New York, 2008; 199–206.
6. Stewart B, Cipolla J and Best L. Extraneous information and graph comprehension: Implications for effective design choices. *Campus-Wide Inform Syst* 2009; 26(3): 191–200.
7. Tufte ER. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983.
8. Ware C, Purchase H, Colpoys L and McGill M. Cognitive measurements of graph aesthetics. *Inform Visual* 2002; 1(2): 103–110.
9. Huang W and Eades P. How people read graphs. *Proceedings of Asia Pacific Symposium on Information Visualization 2005*, 27–29 January (Sydney, Australia), 2005, 51–58.
10. Bojko A and Adamczyk KA. More than just eye candy. *User Experience* 2010; 9(3): 4–8.
11. Goldberg JH and Helfman JI. Comparing information graphics: a critical look at eye tracking. *Proceedings BELIV'10 Workshop, Beyond Time and Errors in the Evaluation of Visualizations* (Atlanta, USA), ACM Press: New York, 2010; 71–78.
12. Goldberg JH and Wichansky AM. Eye tracking in usability evaluation: a practitioner's guide. In: Hyona J, Radach R and Deubel H (eds) *The Mind's Eye: Cognitive and Applied Aspects of Eye Movements*. Amsterdam: Elsevier Science Publishers, 2003, pp.493–516.
13. Goldberg JH and Helfman JI. Scanpath clustering and aggregation. *Proceedings of ACM/ETRA 2010, Symposium on Eye-Tracking Research & Applications* (Austin, USA), ACM Press: New York, 2010; 227–234.
14. Few S. *Keep Radar Graphs Below the Radar – Far Below.* 2005. http://www.perceptualedge.com/articles/dmreview/radar_graphs.pdf.
15. Duchowski AT. *Eye Tracking Methodology*, 2nd edn. London: Springer-Verlag, 2007.
16. Goldberg JH and Kotval XP. Computer interface evaluation using eye movements: methods and constructs. *Int J Ind Ergon* 1999; 24: 631–645.
17. Few S. *Information Dashboard Design*. Sebastapol, CA: O'Reilly Press, 2006.
18. Harris RL. *Information Graphics: A Comprehensive Illustrated Reference*. New York: Oxford University Press, 1996, pp.320–321.
19. Carpenter RHS. *Movement of the Eyes*. London: Pion Press, 1988.
20. Tobii Technology, http://www.tobii.com/