Data and Task Based Effectiveness of Basic Visualizations

Bahador Saket
Georgia Tech
Atlanta, USA
saket@gatech.edu

Alex Endert
Georgia Tech
Atlanta, USA
endert@gatech.edu

Çağatay Demiralp
IBM Research
NY, USA
cagatay.demiralp@us.ibm.com

ABSTRACT
Visualizations of tabular data are widely used; understanding their effectiveness in different task and data contexts is fundamental to scaling their impact. However, little is known about how basic tabular data visualizations perform across varying data analysis tasks and data attribute types. In this paper, we report results from a crowdsourced experiment to evaluate the effectiveness of five visualization types—Table, Line Chart, Bar Chart, Scatterplot, and Pie Chart—across ten common data analysis tasks and three data attribute types using two real world datasets. We found the effectiveness of these visualization types significantly varies across task and data attribute types, suggesting that visualization design would benefit from considering context-dependent effectiveness. Based on our findings, we derive recommendations on which visualizations to choose based on different task and data contexts.

INTRODUCTION
The demand for data visualization has significantly grown in recent years with the increasing availability of digitized data across everyday domains [44]. Visualizations aim to enhance understanding of underlying data by leveraging human visual perception, evolved for fast pattern detection and recognition. Understanding the effectiveness of a given visualization in achieving this goal is a fundamental pursuit in data visualization research and has important implications in practice.

A large body of prior research (e.g., [9, 11, 13, 54, 43, 56, 63, 27, 39, 30, 35, 61, 60]) has studied how various properties of data visualizations, from low-level visual encoding choices to Gestalt characteristics, impact the effectiveness of visualizations. Guidelines and insights [12, 43, 45, 46, 53] motivated by these earlier studies have tremendous impact on data visualization today. Nonetheless, our operational understanding of how the effectiveness of individual visual properties determine the graphical perception of a complete visual chart is limited, despite some theoretical work [37, 48].

Every submission will be assigned their own unique DOI string to be included here.

This necessitates studying the effectiveness of visualization types in a top-down approach rather than studies focusing on graphical perception of individual visual elements. While earlier work has evaluated visualization types for their effectiveness, they were conducted under conditions that were inconsistent across studies, with varying sample sizes, and a limited number of tasks. Research indicates, however, the effectiveness of a visualization depends on several factors including task/question at the hand [1], and data attributes and datasets visualized [8, 51]. For example, while one chart might be suitable for answering a specific type of question (e.g., to check whether there is a correlation between two data attributes), it might not be appropriate for other types (e.g., to find a data point with the highest value). Yet, we know little about how some of the basic data visualizations perform under common visual analysis tasks and data attribute types.

In response, we conduct a crowdsourced study to evaluate the effectiveness of five basic visualization types (Table, Line Chart, Bar Chart, Scatterplot, and Pie Chart) across 10 different visual analysis tasks [1], three different data attributes (Nominal, Ordinal, Numerical), and two different datasets (Cars and Movies). Our results indicate that the effectiveness of these visualization types often significantly varies across tasks. For example, while pie charts are one of the most effective visualizations for finding the extremum value, they are less effective for finding correlation between two data attributes. Additionally, we found that data attribute types used for creating visualizations have significant impact on the effectiveness of those visualizations. For example, line charts are very effective for performing clustering tasks when one of the axis is either Nominal or Ordinal and another axis is Numerical. However, their effectiveness decrease significantly when data attribute for both axes are Numerical.

Based on our analysis of the user task performances and visualization preferences, we provide several recommendations on which visualization types to use. There is a renewed interest [64, 5, 65] in visualization recommendation systems that aim to shift some of the burden of visualization design and exploration decisions from users to algorithms. We discuss how empirical perception in general and our results in particular can be applied in improving visualization recommendation systems moving forward.

RELATED WORK

We draw from earlier work studying the impact of visual encoding decisions on decoding of the data presented in visualizations and the general effectiveness of visualizations for different design choices and task purposes. Below, we discuss some of the most relevant studies.

Effectiveness of Visual Encodings

One of the main components of information visualizations is data representation. The fundamental focus of data representation is mapping from data values to graphical representations [13, 11]. Visualization designers use elementary graphical units called visual encodings to map data to graphical representation [11]. Consider a case in which we visualize two numerical values using two bars with different lengths. Here, length is the primary visual encoding variable used to map the data values.

A good deal of prior work has studied how different choices of visual encodings influence visualization effectiveness. Bertin recognized that different visual variables have different effectiveness levels (or capacities) for encoding types of data [3]. Through human-subject experiments, researchers have investigated the effects of visual encoding on the ability to read and make judgments about data represented in visualizations [12, 28, 35, 43, 54, 56, 63]. Consequently, prior research has provided rankings of visual variables by user performance for nominal, ordinal or numerical data [12, 43, 45, 46, 53]. Researchers have also investigated how design parameters beyond visual encoding variables such as aspect ratio [9, 27, 60], size [30, 39, 10], chart variation [62, 36], and axis labeling [61] impact the effectiveness of visualizations.

Effectiveness of Visualizations

Although prior research has proposed models of graph comprehension [37, 24, 48, 55], little is known about how visual encoding or design parameters interact with each other or different data and task contexts in forming the overall performance of a given visualization. Earlier work [19, 15, 57, 23, 14, 26, 14, 34, 16] has also studied the effectiveness of visualization types with their common design configurations for a select number of tasks.

Eells [19] investigated effectiveness of proportional comparison (percentage estimation) task in divided (stacked) bar charts and pie charts. Eells asked participants to estimate the proportions in pie charts and bar charts. He found pie charts to be as fast as and more accurate than bar charts for proportional comparison task. He also found that as the number of components increases, divided bar charts become less accurate but pie charts become more (maximum five components were considered). In a follow up study with a different setting, Croxton and Stryker [15] also tested the effectiveness of divided bar charts and pie charts using a proportional comparison task. They also found pie charts to be more accurate than divided bar charts in most cases, but contrary to Eells' study, not all.

Spence et al. [57] studied the effectiveness of bar charts, tables and pie charts. They found that when participants were asked to compare combinations of proportions, the pie charts outperformed bar charts. Their results also show that for tasks where participants were asked to retrieve the exact value of proportions, tables outperform pie charts and bar charts. In another study comparing the effectiveness of bar charts and line charts, Zacks and Tversky [66] indicated that when participants were shown these two types of visualizations and asked to describe the data, they constantly used bar charts to reference the compare values (e.g., A is 10% greater than B). Whereas with line charts, participants described trends.

Harrison et al. [26] measured the effectiveness of different visualizations for explaining correlation, finding that parallel coordinates and scatterplots are best at showing correlation. They also found that stacked bar charts outperform stacked area and stacked line. In a follow up study, Kay and Heer reanalyzed [33] the data collected by Harrison et al. [26]. The top ranking visualization, scatterplot, remained the same.

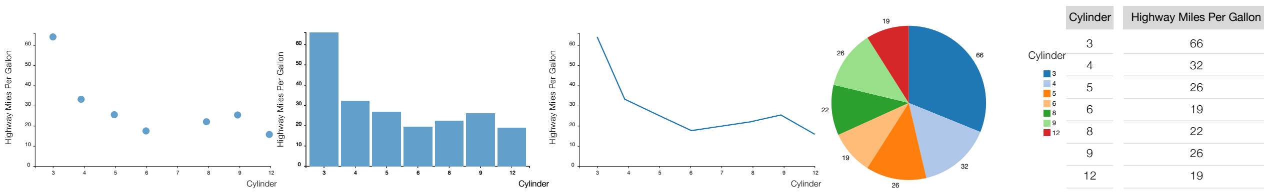
While these independent studies provide helpful generic guidelines, they were conducted under different—often, inconsistent—conditions, varying sample sizes, datasets, and for a disperse set of tasks. For example, some of the studies measured the effectiveness of charts by asking participants to estimate the proportion of part to whole [19, 57]

Also, these earlier studies have conducted experiments typically using atomic generic tasks such as comparison of data values or estimation of proportions. However, many visual analysis tasks (e.g., filtering, finding clusters) require integration of results from multiple atomic tasks, limiting the applicability of earlier findings [1, 2].

Importance of Tasks and Datasets

While performing analytical activities, users usually have a set of tasks/goals to perform [1, 6]. These tasks range from broader, "high-level" goals to more specific, "low-level" inquiries. Several studies have investigated spectrum of possible tasks that can be performed using different visualizations and summarized them as task taxonomies [1, 7, 40, 52]. Earlier research has also applied these taxonomies to evaluate different visualizations [18, 41, 38]. Previous work evidence that the effectiveness of visualizations might change with different task types [32, 18, 50]. Prior research also advocates the use of different datasets and data attribute types in evaluating the effectiveness of visualizations [8, 51].

To the best of our knowledge, the present work is the first systematic study to evaluate the effectiveness of the most common visualization types across a spectrum of tasks relevant to visual analysis using different datasets and data types.



(a) Scatterplot (b) Bar Chart (c) Line Chart (d) Pie Chart (e) Table

Figure 1. Five different types of visualizations used in this study. In this case, each visualization shows average highway miles per gallon (a numerical data attribute) for cars with different number of cylinders (an ordinal data attribute).

RESEARCH GOALS AND EXPERIMENT

Our goal is to better inform visualization design by understanding and quantifying how the effectiveness of tabular data visualizations changes with typical data analysis tasks and data attribute types. In particular, we seek to answer the following questions:

- **Q1:** How well are different tasks supported by each visualization? For example, what tasks can be performed well using a scatterplot?
- **Q2:** How do data attribute types impact the effectiveness of visualizations for each task? For example, is performing a correlation task using a scatterplot where the x axis is nominal and y axis is numerical different from the case where both axes are numerical?
- **Q3:** Is there a correlation between user preferences and their performance time and accuracy when using different visualizations for each task?

To address these questions, we conducted a crowdsourced experiment to measure accuracy, time, and user preferences in performing 10 visual analysis tasks using five visualization types. To facilitate the applicability of results, we used data sampled from two real world datasets with three different data attribute types.

We discuss the details of the experimental design along with the rationale behind our choices regarding the design variables in the following sections.

Visualization Types

When deciding which visualization types to include in our experiment, we balanced the familiarity of the visualizations considered with the comprehensiveness of the experiment. On the one hand, we would like to have more generalizable results, which suggested considering a broad set of visualization techniques in our experiment. At the same time, we would like our study to have the members of general public as our participants: this would suggest to include a set of visualization techniques which are understandable by all participants.

Previous work studied which visualization types non-experts are familiar with. Borner et al. [4] assessed understanding of the members of general public with different visualization techniques. They showed 20 different visualization techniques to 273 participants and asked participants to name and explain how to read different techniques. Their results indicate that most of the participants had a difficult time naming and interpreting many of the techniques except those that they have seen often in books, at work, on the Internet, and in the news (e.g., bar charts, pie charts, line charts, tables, geographical maps, and scatterplots). Lee et al. [42] also show that bar charts, line charts, scatterplots, and pie charts are frequently shown up in news outlet and visualization tools.

We extracted a list of visualization techniques supported by some of the well-known visualization tools (e.g., Microsoft Excel, Tableau, Spotfire, QlickView, Adobe Analytics, IBM Watson Analytics). While all these tools support a variety of techniques, few of techniques are supported by all of them.

Building on previous work [4] and investigations on visualization techniques supported by different visualization tools, we decided to include five well-recognized visualization techniques in our study. In this study, we include Bar Chart, Line Chart, Scatterplot, Table, and Pie Chart (see Figure 1).

Datasets and Data Attribute Types

To create visualizations for our experiment, we selected datasets that would be familiar to participants. This is particularly important since we did not want user performance to be affected by unfamiliarity of participants with the datasets.

We first selected five different datasets including Cereals [49], Cars [31], Movies [17], Summer Olympics Medalists [17], and University Professors [49]. We then printed a part of each dataset on paper and showed them to six participants (4 male, 2 female). We asked participants *"Please look at each of these datasets and their data attributes. Which datasets do you feel you are more familiar with?"* Cars and Movies datasets were the ones that the majority (five out of six participants) of participants selected. Both datasets include data attributes of Nominal, Ordinal, and Numerical types. The Cars dataset [31] provides details for 407 new cars and trucks for the year 2004. This dataset contains 18 data attributes describing each car. The Movies dataset [17] provides details for 335 movies released from 2007 to 2012, and contains 13 data attributes.

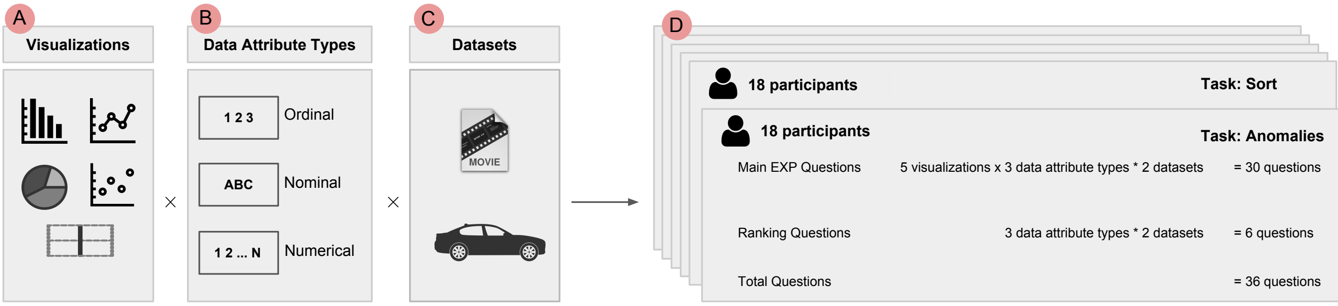


Figure 2. The figure shows our experimental design using various visualizations, data attribute types, datasets, and tasks.

Visualization Design

To generate visualizations using all three types of data attributes available in our datasets, we tested three pair-wise combinations. In particular, we used Nominal∗Numerical, Ordinal∗Numerical, Numerical∗Numerical. We did not include Nominal∗Nominal because it is not possible to represent this combination using all five visualizations considered in this study. We also did not include Ordinal∗Ordinal and Nominal∗Ordinal because these combinations are identical to Ordinal∗Numerical and Nominal∗Numerical.

To create Scatterplots, Bar Charts, and Line Charts, we used the same length, font size, and color to draw their $x-y$ axes. In addition, all the visual elements (e.g., bars in a bar chart) used in the three charts had the same blue color (see Figure 1-a, b, c).

Unlike other visualizations, pie charts do not have any axis to read the values from. That is, to create Pie Charts we had to make design decisions on how to show values of two data attributes used to generate them. The main design decision that we had to make for Pie Charts was whether to include legends. Instead of having legends, we could potentially add labels on the top of slices of Pie Charts. We tried to put the labels on the top of slices but this caused visual clutter, particularly in cases where the labels were long. Additionally, using legends for Pie Charts is a common practice in majority of commercial visualization dashboards [58, 59]. We decided to not show any value on the top of the slices of Pie Charts, instead showing the values of one data attribute using a legend and another one beside the slices (see Figure 1-d).

For Tables, we separated different rows of the table using light gray lines. We used a darker background color to make the labels (two data attributes used for creating the table) distinguishable (see Figure 1-e).

Tasks

We selected the tasks for our experiments based on three considerations. First, tasks should be drawn from those commonly encountered while analyzing tabular data. Second, the tasks should be present in existing task taxonomies and often used in other studies to evaluate visualizations. Third, the tasks should be performed in a reasonable amount of time.

Previously, Amar et al. [1] proposed a set of ten low-level analysis tasks that describe users' activities while using visualization tools to understand their data. First, these tasks are real world tasks because users came up with them while exploring five different datasets with different visualization tools. Second, different studies used these tasks to evaluate effectiveness of visualizations. Finally, these task are low-level tasks that can be performed in a reasonable amount of time. In fact, higher-level tasks are composed of these low-level tasks. With this in mind, we used the low-level taxonomy by Amar et al. [1], described below.

Retrieve Value. For this task, we asked participants to identify values of attributes for given data points. For example, *what is the value of horsepower for the cars?*

Filter. For given concrete conditions on data attribute values, we asked participants to find data points satisfying those conditions. For example, *which car types have city miles per gallon ranging from 25 to 56?*

Compute Derived Value. For a given set of data points, we asked participants to compute an aggregate value of those data points. For example, *what is the sum of the budget for the action and the sci-fi movies?*

Find Extremum. For this task, we asked participants to find data points having an extreme value of an data attribute. For example, *what is the car with highest cylinders?*

Sort. For a given set of data points, we asked participants to rank them according to a specific ordinal metric. For example, *which of the following options contains the correct sequence of movie genres, if you were to put them in order from largest average gross value to lowest?*

Determine Range. For a given set of data points and an attribute of interest, we asked participants to find the span of values within the set. For example, *what is the range of car prices?*

Characterize Distribution. For a given set of data points and an attribute of interest, we asked participants to identify the distribution of that attribute values over the set. For example, *what percentage of the movie genres have a average gross value higher than 10 million?*

Find Anomalies. We asked participants to identify any anomalies within a given set of data points with respect to a given relationship or expectation. We crafted these anomalies manually so that, once noticed, it would be straightforward to verify that the observed value was inconsistent with what would normally be present in the data (e.g., movies with zero or negative length would be considered abnormal). For example, *which genre of movies appear to have abnormal length?*

Find Clusters. For a given a set of data points, we asked participants to find clusters of similar data attribute values. For example, *how many different genres are shown in the chart below?*

Find Correlation. For a given set of two data attributes, we asked participants to determine if there is a correlation between them. To verify the responses to correlate tasks, we computed Pearson's correlation coefficient (r) to ensure that there was an strong correlation ($r \leq -0.7$ or $r \geq 0.7$) between the two data attributes. For example, *is there a strong correlation between average budget and movie rating?*

USER EXPERIMENT

In this section, we explain the details of the experiment. Figure 2 shows our experimental setting and Figure 3 indicates our experimental procedure (various stages of the experiment).

Experimental Platform & Participants

We conducted our experiment by posting it as a job, Human Intelligence Task (HIT), on Amazon's Mechanical Turk (MTurk). Earlier work demonstrates the viability of graphical perception experiments run on MTurk by reproducing in-lab study results [29].

To be able to participate in our study, MTurk workers (who perform tasks posted on MTurk), had to have an approval rate of 95% and at least 100 approved HITs as a quality check. We implemented our experiment as a web application hosted on a server external to MTurk. Participants accessed the experiment through a URL link posted on the MTurk site. Each worker could participate in our study only once. The study took about 40 minutes to complete and we compensated the workers who participated $4.

In order to determine the minimum number of participants needed for our study, we first conducted a pilot study with 50 participants on Amazon's Mechanical Turk. Based on the data collected from our pilot study, we conducted a statistical power analysis to ensure that our experiment included enough participants to reliably detect meaningful performance differences across independent variables of the experiment. Our power analysis based on the results of the pilot study indicated that at least 160 participants would be required to detect a large effect.

After determining the number of subjects required to participate in our study, we recruited 180 (105 Male, 75 Female) workers. The age of our workers ranged from 25–40 years. All workers participated in our experiment were based in the United States and have used visualizations before. 107 of the participants had experience creating visualizations using Microsoft Excel. Five of the participants also had experience in creating visualizations using Tableau software.

Procedure

Before starting the main experiment, participants were briefed about the purpose of the study and their rights. At this stage, the participants were also asked to answer to some demographic questions (e.g., age, sex, and prior experience in creating visualizations). Participants were then asked to perform 5 trial questions as quickly and accurately as possible. During this session, after answering each question participants received feedback that showed the correctness of their answers. To prevent the participants from skipping the training questions, participants were not able to move to the next training question unless they answered the question correctly.

During the main experiment 180 participants were randomly assigned to 10 tasks (18 participants per task). So, each participant performed questions designed for one type of task. For each type of task, we had 30 questions (3 *Data Attribute Type Combinations* $\times$ 5 $Visualizations \times$ 2 *datasets*). Questions were presented in a random order to prevent participants from extrapolating new judgments from previous ones. After performing 30 main questions, the participants were asked to perform 6 additional ranking questions (3 *Data Attribute Type Combinations* $\times$ 2 *datasets*). In each ranking question the participants were asked to rank the five different visualizations in the order of their preference for performing this task. Before finishing the experiment, we asked participants to *"Please enter the criteria you used for ranking the charts along with any other additional comments you have about the experiment in general"*. This was to allow the participants to convey their feedback and ideas and in order to solicit potentially unexpected insights. A screenshot of the software for the experiment is shown in Figure 4.

Data Analysis

To address our research questions (**Q1, Q2, Q3**), we needed to test how the different visualizations and differences in data attribute types affected the user performance time and accuracy for different tasks.

To analyze the differences among the various visualizations, we first calculated separate mean performance values for all questions. That is, we averaged outcome values of questions for each visualization and task. We then checked whether the collected data met the assumptions of appropriate statistical tests. The assumption of normality was satisfied for parametric testing,

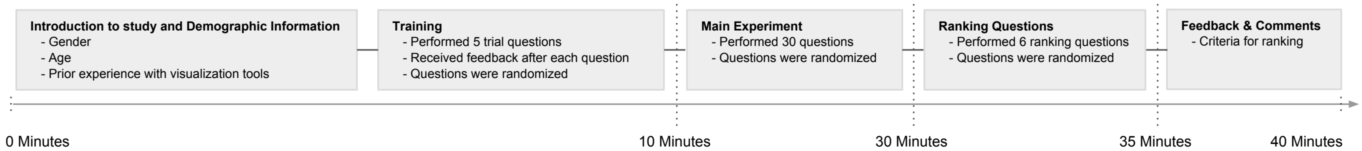


Figure 3. Experimental Procedure indicates different steps with main highlights that participants go through during the experiment. The timeline indicates an approximate time taken by the users to complete each step.

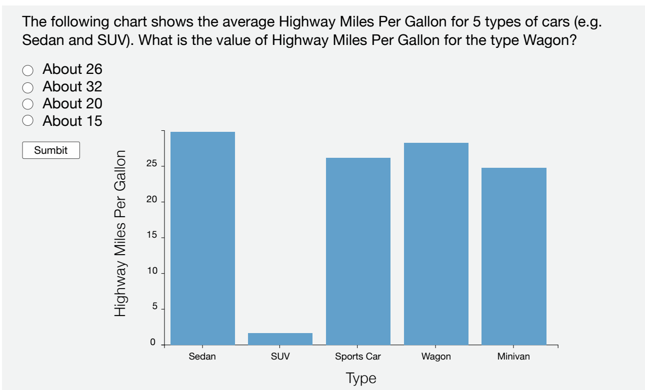


Figure 4. A screenshot of one of the trials used in this experiment. This specific trial asks users to retrieve the value of highway miles per gallon for the car type Wagon. The task used in this example is a Retrieve Value task. Data attribute types used to create the visualization are Nominal ($x - axis$) and Numerical ($y - axis$).

but Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for time. To address this issue, we report test results with corrected degrees of freedom using Greenhouse-Geisser estimates for $\epsilon < 0.75$ and otherwise with Huynh-Feldt correction.

We conducted repeated-measures analysis of variance (ANOVA) for each task independently to test for differences among the various visualizations, data attribute types, datasets and their interactions with one another. While Visualization and Data Attribute Type had significant effects on both accuracy and time, Dataset had no significant effect on accuracy or time.

We make all the relevant materials for our analysis publicly available[1], including the web application used to run the experiment along with its source code, the anonymized data collected from the subjects, and the statistical test results. Now we discuss our results. We first give an overview of our analysis of the results and then discuss them in detail for each task. Throughout the following sections, accuracy refers to values in percentages (%) and time refers to values in seconds.

SUMMARY OF RESULTS

Figure 5 summarizes performance time, accuracy, and user preferences of five charts for different tasks. Results, aggregated over tasks, data types, and datasets, show that Bar Chart is the fastest and the most accurate visualization type. This result is inline with prior work on graphical perception showing that people can decode values encoded with length faster than other encodings such as angle or volume [11, 55, 62].

Conversely, Line Chart has the lowest aggregate accuracy and speed. However, Line Chart is significantly more accurate than other charts for Correlation and Distribution tasks. This finding concurs with with earlier research reporting the effectiveness of line charts for trend finding tasks (e.g., [66]). Nonetheless, the overall low performance of Line Chart is surprising and, for some tasks, can be attributed to the fact that the x-axis values ("ticks") were drawn at intervals. This makes it difficult to precisely identify the value for a specific data point.

While Pie Chart is comparably as accurate and fast as Bar Chart and Table for Retrieve, Range, Order, Filter, Extremum, Derived and Cluster tasks, it is less accurate for Correlation, Anomalies and Distribution tasks. Pie Chart is the fastest visualization for performing Cluster task. High performance of Pie Chart for these tasks can be attributed to its relative effectiveness in conveying part-whole relations and facilitating proportional judgments, particularly when the number of data points visualized is small [19, 57]. Pie Chart may have been further helped by having colored slices with text labels showing the data values.

Overall, Scatterplot performs reasonably well in terms of both accuracy and time. For the majority of tasks Scatterplot is among the most effective top three visualizations, and it was never the least accurate or slowest visualization for any of the tasks. One reason for this could be that people are very accurate and fast in perceiving "position" [11].

Table, Bar Chart, and Pie Chart are the fastest and the most accurate visualizations for performing Retrieve tasks. Successful performance of Retrieve tasks highly depends on readers ability to accurately identify the value for a certain data point. As Ehrenberg [20] points out, tables are well-suited for retrieving the numerical value of a data point when a relatively small number of data points are displayed. While performing Retrieve task using a Table is fast and accurate, performing Order, Extremum and Distribution tasks are relatively slow.

Overall, Bar Chart and Table are the two visualization types highly preferred by participants across most of the

[1] https://github.com/AnonymousSubmission007/CHI

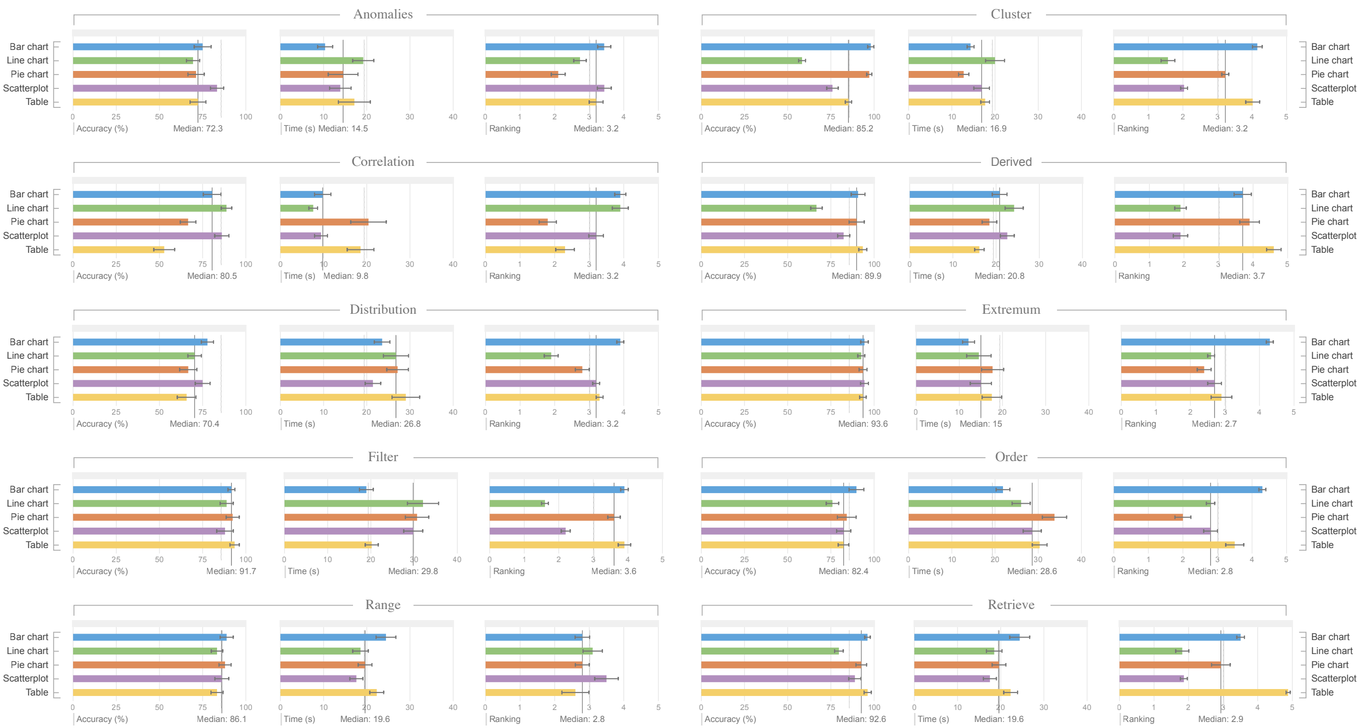


Figure 5. This figure shows performance results for different tasks. Performance results for each task are shown using three sub-charts. Mean accuracy results are shown on the left (mean accuracy is measured in percentage), mean time results are shown in the middle, and user preferences/rankings are shown at the right (1 shows least preferred and 5 shows the most preferred). Medians for each task is indicated using the vertical line. Dashed lines show the median across all tasks. Error bars represent standard error.

tasks. Bar Chart is always among two top-performing visualizations for almost all tasks, so this makes sense that people prefer using Bar Chart over other visualizations. Surprisingly, while performing some of the tasks (e.g., Distribution, Anomalies) using Table is relatively slow and less accurate, participants still prefer Table for performing these tasks. Familiarity of people with tables and easiness in understanding tables could have helped people to prefer using tables over other visualizations.

To determine whether performance time and accuracy are related to user preferences, we calculated the correlation between performance time, accuracy, and user preference. We found a positive correlation between accuracy and user preference (Pearson's $r_{(5)} = 0.68, p < 0.05$), indicating people have a preference for visualizations that allow them to accurately complete a task. We also found a weak negative correlation between performance time and user preferences (Pearson's $r_{(5)} = -0.43, p < 0.05$).

RESULTS

We provide detailed analysis of the results, breaking them down by time, accuracy, and user preferences for each task.

Find Anomalies. We found a significant effect of Visualization ($F_{(3.4,4915.1)} = 3.03, p < 0.05$, $\eta_p^2 = 0.15$) on accuracy. Overall, Line Chart, Table, and Pie Chart had lower accuracy compared to Bar Chart and Scatterplot. Although difference in accuracy was significant only among two of the visualizations. In particular, results of Bonferroni-corrected post-hoc comparisons showed that Line Chart was significantly less accurate than Scatterplot. We also found a significant main effect of Visualization ($F_{(2.8,1163.9)} = 3.2, p < 0.05$, $\eta_p^2 = 0.16$) on time. Bar Chart was the fastest visualization for performing this type of tasks. Posthoc comparisons indicate that Bar Chart was significantly faster than Line Chart.

To explore reasons behind Line Chart's low performance rate, we looked at the differences between various combinations of data attribute types in this visualization. This allowed us to check if a specific combination of data attribute types let to the low performance of line charts. We found significant main effects of Data Attribute Type on both accuracy ($F_{(1.9,4937.1)} = 10.7, p < 0.001$, $\eta_p^2 = 0.39$) and time ($F_{(1.5,1486.1)} = 5.14, p < 0.05$, $\eta_p^2 = 0.23$). Surprisingly Line Chart performed well in Nominal$*$Numerical condition in terms of both accuracy and time. What really paralyzed Line Chart's overall accuracy is their low accuracy rate for Numerical$*$Numerical and Ordinal$*$Numerical conditions. The low performance rate of Line Chart for these two conditions can possibly be explained by the fact that the x-axis values ("ticks") were drawn at intervals. This makes it difficult to precisely identify the value of Ordinal or Numerical data points.

We found a significant effect of Visualization ($F_{(3.1,45.56)} = 5.9, p < 0.05$, $\eta_p^2 = 0.26$) on user preference. For the Anomalies task type, results of pairwise comparisons show that user preference in performing Anomalies tasks using Bar Chart and Scatterplot were significantly higher than other visualizations. More specifically, user preference for using Pie Chart and Line Chart are significantly lower than other visualizations.

Find Clusters. Results indicate that there was a significant effect of Visualization ($F_{(2.6,45065.1)} = 60.7, p < 0.01$, $\eta_p^2 = 0.78$) on accuracy. Results of Bonferroni-corrected posthoc comparisons show that Pie Chart and Bar Chart were significantly more accurate than other visualizations. In terms of time, we found a significant effect of Visualization on time ($F_{(2.3,1485.1)} = 3.9, p < 0.05$, $\eta_p^2 = 0.18$). Pie Chart was the fastest visualization for performing Cluster tasks. While Pie Chart was significantly faster than Table, there was no significant difference between Pie Chart and the other three visualizations. We believe that uniquely coloring different slices of pie charts improved the performance of Pie Chart.

We also found a significant effect of Data Attribute Type on both accuracy ($F_{(1.9,59727.1)} = 177.5, p < 0.001$, $\eta_p^2 = 0.91$) and time ($F_{(1.7,98983.1)} = 47.5, p < 0.001$, $\eta_p^2 = 0.71$). For Nominal$*$Numerical condition, accuracy of all visualizations was relatively high (above 94%) and we found no significant difference among visualizations. Accuracy of Line Chart, Scatterplot, and Table decreased significantly in Numerical$*$Numerical condition. However, performance accuracy and time for Bar Chart and Pie Chart were more consistent in all data attribute type combinations. This indicates bar charts and pie charts are a more consistent visualization for Cluster tasks.

Results indicate a significant main effect of Visualization ($F_{(2.9,188.56)} = 30.2, p < 0.001$, $\eta_p^2 = 0.64$) on user preferences. User preferences in using Bar Chart and Table were significantly higher than other visualizations. While user preferences in using Bar Chart can be explained by its high accuracy and speed, it is surprising that Table was also highly preferred by users for Cluster tasks. Table had the lowest performance time and its accuracy was significantly lower than Pie Chart and Bar Chart.

Find Correlation. We found a significant main effect of Visualization ($F_{(2.5,20528.2)} = 12.1, p < 0.001$, $\eta_p^2 = 0.41$) on accuracy. Pairwise comparison show that Line Chart and Scatterplot were significantly more accurate than Pie Chart, Bar Chart and Table. Bar Chart was also significantly more accurate than Pie Chart and Table. In terms of time, there is a significant effect of Visualization ($F_{(2.2,4367.2)} = 12.4, p < 0.001$, $\eta_p^2 = 0.3$) on time. We found that Line Chart, Bar Chart and Scatterplot were significantly faster than Pie Chart and Table. In fact, our results validates the findings of the previous work that showed the effectiveness of Scatterplots and Line charts for Correlation tasks [26, 47].

Our results also indicate a significant main effect of Data Attribute Type on both accuracy ($F_{(1,28444.4)} = 17.5, p < 0.001$, $\eta_p^2 = 0.5$) and time ($F_{(1,2619.0)} = 8.2, p < 0.05$, $\eta_p^2 = 0.2$). While Table had high accuracy for the Ordinal$*$Numerical condition, its accuracy significantly decreased for the Numerical$*$Numerical condition. In terms of accuracy, Line Chart and Scatterplot performed well across all conditions. In terms of time, Bar Chart, Line Chart, and Scatterplot were faster than Pie Chart and Table in both Numerical$*$Numerical and Ordinal$*$Numerical conditions.

We found a significant main effect of Visualization ($F_{(3.6,75.2)} = 13.6, p < 0.001$, $\eta_p^2 = 0.44$) on user preference. User preference in performing Correlations tasks using Bar Chart, Line Chart and Scatterplot were significantly higher than that of Pie Chart and Table. Positive correlation between accuracy and user preference can be seen here again.

Compute Derived Value. We found a significant main effect of Visualization ($F_{(2.7,18234.2)} = 16.2, p < 0.001$, $\eta_p^2 = 0.49$) on accuracy. Table, Bar Chart, and Pie Chart had the highest accuracy for this type of task. Accuracy of Line Chart was significantly lower than rest of the four chart types. On the other hand, there was no significant difference among Bar Chart, Scatterplot, Pie Chart, and Table. We also found a significant main effect of Visualization ($F_{(2.8,1540.0)} = 6.8, p < 0.05$, $\eta_p^2 = 0.29$) on Time.

Table not only had the highest accuracy, but also the fastest speed for Derived tasks. There was, however, no significant difference in time among Table, Pie Chart, and Bar Chart. Notice that Bar Chart was as effective as Pie Chart and Table even though, Bar Chart does not show the exact values of data points.

We found a significant main effect of Visualization ($F_{(3.1,187.8)} = 35.3, p < 0.001$, $\eta_p^2 = 0.67$) on user preference. Participants preference for using Table, Pie Chart, and Bar Chart is significantly higher than Scatterplot and Line Chart.

Characterize Distribution. The different visualization types are not significantly different from each other in terms of accuracy. Scatterplot and Bar Chart have the highest accuracy for this type of tasks. We found a significant effect of Visualization ($F_{(4,1226.2)} = 4.7, p < 0.05$, $\eta_p^2 = 0.21$) on time, and our results indicate that Scatterplot and Bar Chart are significantly faster than Pie Chart and Table for Distribution tasks. There is not a significant difference between Line Chart, Bar Chart and Scatterplot.

We also found a significant main effect of Data Attribute Type on time ($F_{(1.8,1776.4)} = 7.5, p < 0.05$, $\eta_p^2 = 0.29$). Interestingly, performance time for Scatterplot, Line Chart and Bar Chart are consistent across all three data attribute type combinations. However, this was not the case for Table and Pie Chart. Table and Pie Chart were significantly slower for Numerical∗Numerical condition. The low speed of Table is likely due to the fact that counting a large number of data points in a Table could take time. Additionally, low speed of Pie Chart might be due to small slices of the Pie Chart which are hard to identify.

Participants preferred Bar Chart, Scatterplot, and Table significantly more than Pie Chart and Line Chart. It is surprising that even though Table was not faster than the other four visualizations, participants highly preferred using them.

Find Extremum. While there is not a significant effect of Visualization on accuracy, we found a significant effect of Visualization on time ($F_{(3.9,779.1)} = 7.1, p < 0.001$, $\eta_p^2 = 0.28$). In terms of time, Bar Chart is the fastest and Pie Chart is the slowest visualization for this type of tasks. In fact, Bar Chart is significantly faster than Table and Pie Chart. Previous work also recommends using Bar Chart in cases where readers are looking for a maximum or minimum values [21].

We also found a significant main effect of Data Attribute Type on time ($F_{(1.8,2026.4)} = 21.9, p < 0.001$, $\eta_p^2 = 0.56$). While performance time is not significantly different across Bar Chart, Line Chart, and Scatterplot in all three conditions, Pie Chart and Table performance time decreased significantly for Numerical∗Numerical condition.

There is a significant main effect of Visualization on user preference ($F_{(2.8,89.4)} = 8.2, p < 0.001$, $\eta_p^2 = 0.31$). For Extremum tasks, participant preference in using bar charts is significantly higher than all other visualizations.

Filter. There is a significant main effect of Visualization on time ($F_{(2.8,5830.4)} = 10.3, p < 0.001$, $\eta_p^2 = 0.37$). Results indicate that Bar Chart and Table are significantly faster than other visualizations. We did not find a significant difference among Scatterplot, Line Chart, and Pie Chart for time. Moreover, we did not find a significant main effect of Data Attribute Type on either accuracy or time. This means all visualizations were consistent in different conditions.

We found a significant main effect of Visualization on user preference ($F_{(2.2,210.5)} = 42.2, p < 0.001$, $\eta_p^2 = 0.72$). In particular, participant preference towards using Table is significantly higher than other visualizations for Filter tasks. After Table, Bar Chart and Pie Charts are the two visualizations that are highly preferred by participants for performing Filter tasks.

Sort. We found a significant main effect of Visualization on accuracy ($F_{(4,.03)} = 2.6, p < 0.05$, $\eta_p^2 = 0.17$). Bar Chart has the highest accuracy for Sort tasks. In addition, Bar Chart is significantly more accurate than Line Chart. We did not find a significant difference among Bar Chart, Pie Chart, Scatterplot, and Table. There is also a significant main effect of Visualization on Time ($F_{(2.3,3661.5)} = 8.2, p < 0.05$, $\eta_p^2 = 0.32$). Bar Chart is significantly faster than other visualizations. We did not find a significant difference among Line Chart, Scatterplot, Pie Chart and Table in terms of time.

We also found a significant main effect of Data Attribute Type on accuracy ($F_{(1.9,995.4)} = 26.6, p < 0.001$, $\eta_p^2 = 0.61$) and time ($F_{(1.4,3793.4)} = 8.07, p < 0.05$, $\eta_p^2 = 0.32$). Scatterplot, Line Chart, and Table visualizations have lower accuracy in Numerical∗Numerical condition compared to other two conditions. Interestingly, accuracy of Bar Chart and Pie Chart are not changed significantly across data attribute types conditions. This indicates stability of these two visualizations in performing Sort tasks across various conditions. For Nominal∗Numerical and Ordinal∗Numerical conditions, Bar Chart and Line Chart, and Scatterplot were significantly faster than other visualizations.

We found a significant main effect of Visualization on user preference ($F_{(3.0,103.3)} = 11.8, p < 0.001$, $\eta_p^2 = 0.44$). For Sort tasks, users preferred Bar Chart significantly more than other visualizations. Moreover, our results indicate that user preference in using Pie Chart is significantly lower than other visualizations. There was not a significant different in user preference for Line Chart and Scatterplot.

Determine Range. We did not find a significant main effect of visualization on either time or accuracy. Although the difference is not significant, Table and Line Chart have the lowest accuracy for Range tasks. Line Chart, Scatterplot, and Pie Chart are faster than Bar Chart and Table, but the difference is not significant among visualizations. We found a significant main effect of Data Attribute Type on accuracy ($F_{(1.8,12462.4)} = 10.1, p < 0.001$, $\eta_p^2 = 0.35$). For Nominal∗Numerical and Ordinal∗Numerical conditions, accuracy of Pie Chart and Line Chart are as high as other visualizations. However, accuracy of these two visualization types decreased significantly for Numerical∗Numerical condition. We also did not find a significant main effect of Visualization on user preference for Range tasks. This indicates that user preference was not significantly different among all visualizations.

Retrieve Value. We found a significant main effect of Visualization ($F_{(2.9,7114.1)} = 7.7, p < 0.001$, $\eta_p^2 = 0.32$) on accuracy. Overall, Bar Chart, Table and Pie Chart were significantly more accurate than Line Chart. The difference between accuracy in Scatterplot and Line Chart was not significant. We also found a significant main effect of Visualization ($F_{(2.2,1516.9)} = 4.8, p < 0.05$,

$\eta_p^2 = 0.2$) on time. Table is the fastest visualization for performing this type of tasks. Post-hoc comparisons indicate that Table is significantly faster than Line Chart and Scatterplot for Retrieve tasks.

We did not find a significant main effect of Data Attribute Type either on time or accuracy. This means that they were consistent across different data attribute type combinations.

We found a significant main effect of Visualization on user preference ($F_{(1.5,417.2)} = 47.1, p < 0.001$, $\eta_p^2 = 0.73$). User preference for performing Retrieve tasks using Table is significantly higher than other visualizations. After Table, Bar Chart is the second most visualization type highly preferred by users to perform this type of tasks. Table is significantly more preferred than Bar Chart. Moreover, user preference in using Bar Chart is significantly higher than Pie Chart. Line Chart and Scatterplot got the lowest user preference.

DISCUSSION

In this section, we reflect on the results of our work more broadly with respect to information visualization. We discuss several aspects of our work, implications of the study findings, and provide a series of guidelines on choosing visualizations based on different tasks and data attribute types.

No One Size Fits All

Depending on the task at the hand, various visualizations perform differently (**Q1**). That is, we do not advocate generalizing the performance of a specific visualization on a particular task to every task. For example, throughout the history of the graphical perception research, pie charts have been a subject of passionate arguments [19, 15, 57] for and against their use. Although the current common wisdom among visualization researchers is to avoid them, pie charts continue to be popular in everyday visualizations. Results of our study present a more nuanced view of pie charts. We found that pie charts can be as effective as other visualizations for task types such as Cluster, Extremum, Filter, Retrieve, and Range. On the other hand, our results suggest that pie charts perform poorly in Correlation and Distribution tasks.

Our results also show the performance of visualizations changes with data attribute types (**Q2**). For example, for Distribution tasks, pie charts and tables have a weaker performance in Numerical$*$Numerical conditions compared to Nominal$*$Numerical and Nominal$*$Numerical conditions. This again suggests that generalizing performance of a specific visualization across both dataset types and tasks is unclear at this time.

User Preferences

How do user preferences relate to user performance? Our results show user preferences correlate with user accuracy and speed in completing tasks. Before completing the study, we asked participants to explain the criteria they used for ranking the visualizations. Some of participants explicitly mentioned perceived accuracy of the charts as one of the factors that influenced their decision while ranking visualizations. Below are examples from subject comments.

> Participant: *"I ranked based on the one that showed me the information accurately (in my perception)."*

> Participant: *"Just by how accurate I felt my own answer was, and how easy it was to derive the answer from the graphs."*

Neither accuracy nor speed appear to be the only criteria by which participants describe their individual rankings. Additionally, perceived accuracy does not always match with task accuracy. We noticed that for some task types such as Distribution and Cluster, preference for using tables and bar charts is significantly higher than other visualizations, even though these two visualizations are not the most effective ones for these type of tasks. Interestingly, some of the participants took into account their familiarity with visualizations as one of the factors for preferring some visualization over others.

> Participant: *"I just went with the ones I felt were familiar to me."*

> Participant: *"I deal with bars and tables everyday. I know how to read them."*

We anticipate that users' familiarity with visualizations has positive correlation with their preference in using those visualizations, but this remains to be formally studied.

Which Visualization Type to Use?

Based on our findings in the study, we derive the following guidelines on choosing visualizations under various task contexts.

G1. For all tasks, bar charts and scatterplots are good defaults. Not only do these two visualizations have a high overall performance across all tasks, but they also have the lowest variation of performance across three different data attribute type combinations. This indicates that bar charts are very robust across various tasks and data attribute types.

G2. Avoid line charts for tasks that require readers to precisely identify the value of a specific data point. The low performance of line charts for some tasks such as Derived and Cluster might be attributed to the fact that the x-axis values (i.e., the "ticks") were drawn at uniform intervals. This makes it difficult to precisely identify the value of a specific data point.

G3. Use scatterplots and bar charts for finding anomalies. Results of our study indicate that scatterplots and bar charts have high accuracy, speed, and are highly preferred by users for this type of task.

G4. Use pie charts and bar charts for finding clusters. Our results indicate that these two visualizations performed well for this type of task.

G5. Use line charts and scatterplots for finding correlations. We found that line charts and scatterplots have high performance for finding correlations.

G6. Use bar charts for finding extremums. Based on our results bar charts are the fastest, most accurate, and highly preferred visualization for finding extremum tasks. Previous work also suggests using bar charts for finding maximum/minimum values [21].

G7. Use bar charts for sorting tasks. Considering speed, accuracy, and user preference, bar charts have a higher performance compared to other visualizations for sorting tasks.

G8. Use tables for retrieving values. Tables have the highest performance time and accuracy for this type of task. Additionally, user preference in using tables is significantly higher than other visualizations.

LIMITATIONS AND FUTURE WORK

Our experimental results should be interpreted in the context of the specified visualizations, data attribute types, tasks, and datasets. Due to practical limitations of conducting the study (e.g., length and complexity of the experiment), we did not test for all possible types of visualizations or tasks. However, we tested the most common low-level analytical tasks [1, 2] that users encounter during visual data exploration.

An important avenue for continued research is creating a recommendation engine that incorporates the findings of this study. In particular, we plan to design an engine that takes into account performance time, accuracy, and user preferences when suggesting visualizations. Once we have such an engine, we can use it to develop new visualization recommendation systems based on user-specified tasks and data attributes. One relevant application area of such a recommendation engine can be natural language interfaces for data visualization [22]. In such interfaces people tend to specify tasks as a part of their questions (e.g., "Is there a correlation between price and width of cars in this dataset?"). The engine can be used to leverage this knowledge to suggest more effective visualizations for the given query. Another category of systems that might benefit from such an engine are behavior-driven recommendation systems [25]. Such systems often extract analytical tasks from user interactions. These extracted tasks can be used as input to the engine for these systems to suggest more effective visualizations. All these applications would also benefit from models trained on the results of the current study (or any other empirical performance data) that can predict the task performance of "unseen" visualizations.

To facilitate future research by the community, we have made the results and source code of our experiment publicly available at https://github.com/AnonymousSubmission007/CHI.

CONCLUSION

In this work, we report the results of a study that gathers user performance and preference for performing ten common data analysis tasks using five basic visualization types, Table, Line Chart, Bar Chart, Scatterplot, and Pie Chart, across three data attribute types. We use data sampled from two real world datasets to further support the ecological validity of results. We find that the effectiveness of the visualization types considered significantly changes from one task or data type to another. We compile our findings into a set of recommendations to inform data visualization in practice.

Prior work on graphical perception has served visualization practitioners well by providing guidelines and heuristics on visual encoding and design choices. However, no measure of effectiveness is free from consideration of task and data. To further increase the impact of graphical perception research in visualization practice, we must tightly couple perceptual considerations into everyday visualization tools. This starts with systematically gathering empirical evidence on how visualizations perform across contexts and building models based on the evidence. Our work contributes towards the former.

REFERENCES

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

2. Robert Amar and John Stasko. 2004. BEST PAPER: A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations. In *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS '04)*. IEEE Computer Society, Washington, DC, USA, 143–150. DOI: http://dx.doi.org/10.1109/INFOVIS.2004.10

3. Jacques Bertin. 1983. *Semiology of graphics*. University of Wisconsin Press.

4. Katy Börner, Adam Maltese, Russell Nelson Balliet, and Joe Heimlich. 2015. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization* (2015), 1473871615594652.

5. Fatma Bouali, Abdelheq Guettala, and Gilles Venturini. 2015. VizAssist: an interactive user assistant for visual data mining. *The Visual Computer* (2015), 1–17.

6. Matthew Brehmer and Tamara Munzner. 2013. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2376–2385.

7. Stuart Card, JD Mackinlay, and B Shneiderman. 2009. Information visualization. *Human-computer interaction: design issues, solutions, and applications* 181 (2009).

8. Sheelagh Carpendale. 2008. Evaluating Information Visualizations. In *Information Visualization*, Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North (Eds.). Springer-Verlag, Berlin, Heidelberg, 19–45.

9. William S. Cleveland. 1993. *Visualizing Data*. Hobart Press.

10. William S Cleveland, Persi Diaconis, and Robert McGill. 1982. *Variables on scatterplots look more highly correlated when the scales are increased.* Technical Report. DTIC Document.

11. William S Cleveland and Robert McGill. 1984a. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association* 79, 387 (1984), 531–554.

12. William S. Cleveland and Robert McGill. 1984b. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554.

13. William S Cleveland and Robert McGill. 1985. Graphical perception and graphical methods for analyzing scientific data. *Science* 229, 4716 (1985), 828–833.

14. Michael Correll and Michael Gleicher. 2014. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2142–2151.

15. Frederick E Croxton and Roy E Stryker. 1927. Bar charts versus circle diagrams. *J. Amer. Statist. Assoc.* 22, 160 (1927), 473–482.

16. Michael Dambacher, Peter Haffke, Daniel Groß, and Ronald Hübner. 2016. Graphs versus numbers: How information format affects risk aversion in gambling. *Judgment and Decision Making* 11, 3 (2016), 223.

17. Tablea Datasets. 2015. https://public.tableau.com/s/resources. (2015). https://public.tableau.com/s/resources

18. Steven M. Drucker, Danyel Fisher, Ramik Sadana, Jessica Herron, and m.c. schraefel. 2013. TouchViz: A Case Study Comparing Two Interfaces for Data Analytics on Tablets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2301–2310. DOI: http://dx.doi.org/10.1145/2470654.2481318

19. Walter Crosby Eells. 1926. The relative merits of circles and bars for representing component parts. *J. Amer. Statist. Assoc.* 21, 154 (1926), 119–132.

20. ASCASC Ehrenberg Ehrenberg. 1975. *Data Reduction: Analysing and interpreting statistical data*. John Wiley and Sons, London.

21. Stephen Few. 2006. *Information dashboard design*. O'Reilly.

22. Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G. Karahalios. 2015. DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 489–500. DOI: http://dx.doi.org/10.1145/2807442.2807478

23. Rocio Garcia-Retamero and Mirta Galesic. 2010. Who proficts from visual aids: Overcoming challenges in people's understanding of risks. *Social science & medicine* 70, 7 (2010), 1019–1025.

24. Douglas J Gillan and Robert Lewis. 1994. A componential model of human interaction with graphs: 1. Linear regression modeling. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 36, 3 (1994), 419–440.

25. David Gotz and Zhen Wen. 2009. Behavior-driven Visualization Recommendation. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI '09)*. ACM, New York, NY, USA, 315–324. DOI: http://dx.doi.org/10.1145/1502650.1502695

26. Lane Harrison, Fumeng Yang, Steven Franconeri, and Ronald Chang. 2014. Ranking visualizations of correlation using weber's law. *Visualization and Computer Graphics, IEEE Transactions on* 20, 12 (2014), 1943–1952.

27. Jeffrey Heer and Maneesh Agrawala. 2006. Multi-Scale Banking to 45 Degrees. *IEEE Trans. Visualization & Comp. Graphics* 12 (2006), 701–708.

28. Jeffrey Heer and Michael Bostock. 2010a. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *ACM Human Factors in Computing Systems (CHI)*.

29. Jeffrey Heer and Michael Bostock. 2010b. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 203–212.

30. Jeffrey Heer, Nicholas Kong, and Maneesh Agrawala. 2009. Sizing the Horizon: The Effects of Chart Size and Layering on the Graphical Perception of Time Series Visualizations. In *ACM Human Factors in Computing Systems (CHI)*.

31. Harold V Henderson and Paul F Velleman. 1981. Building multiple regression models interactively. *Biometrics* (1981), 391–411.

32. Harald Ibrekk and M Granger Morgan. 1987. Graphical communication of uncertain quantities to nontechnical people. *Risk analysis* 7, 4 (1987), 519–529.

33. Matthew Kay and Jeffrey Heer. 2016. Beyond Weber's Law: A Second Look at Ranking Visualizations of Correlation. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)* (2016).

34. Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. 2016. When (ish) is My Bus? User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5092–5103.

35. Nicholas Kong, Jeffrey Heer, and Maneesh Agrawala. 2010. Perceptual Guidelines for Creating Rectangular Treemaps. *IEEE Trans. Visualization & Comp. Graphics* 16, 6 (2010), 990–998.

36. Robert Kosara and Drew Skau. 2016. Judgment Error in Pie Chart Variations. In *Proceedings of the Eurographics/IEEE VGTC Symposium on Visualization*. Wiley Online Library, 91–95.

37. Stephen M Kosslyn. 1989. Understanding charts and graphs. *Applied cognitive psychology* 3, 3 (1989), 185–225.

38. Bum Chul Kwon and Bongshin Lee. 2016. A Comparative Evaluation on Online Learning Approaches Using Parallel Coordinate Visualization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 993–997. DOI: http://dx.doi.org/10.1145/2858036.2858101

39. Heidi Lam, Tamara Munzner, and Robert Kincaid. 2007. Overview Use in Multiple Visual Information Resolution Interfaces. *IEEE Trans. Visualization & Comp. Graphics* 13, 6 (2007), 1278–1285.

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

41. Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. 2006b. Task Taxonomy for Graph Visualization. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV '06)*. ACM, New York, NY, USA, 1–5. DOI: http://dx.doi.org/10.1145/1168149.1168168

42. S. Lee, S. H. Kim, and B. C. Kwon. 2016. VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics* PP, 99 (2016), 1–1. DOI: http://dx.doi.org/10.1109/TVCG.2016.2598920

43. Stephan Lewandowsky and Ian Spence. 1989. Discriminating strata in scatterplots. *Journal of American Statistical Association* 84, 407 (1989), 682–688.

44. Peter Lyman and Hal Varian. 2004. How much information 2003? (2004). http://www.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf

45. A.M. MacEachren. 1995. *How Maps Work: Representation, Visualization, and Design*. Guilford Press.

46. Jock D. Mackinlay. 1986. Automating the Design of Graphical Presentations of Relational Information. *ACM Trans. Graph.* 5, 2 (1986), 110–141.

47. Anshul Vikram Pandey, Josua Krause, Cristian Felix, Jeremy Boy, and Enrico Bertini. 2016. Towards Understanding Human Similarity Perception in the Analysis of Large Sets of Scatter Plots. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3659–3669. DOI: http://dx.doi.org/10.1145/2858036.2858155

48. Steven Pinker. 1990. A theory of graph comprehension. *Artificial intelligence and the future of testing* (1990), 73–126.

49. UCI Machine Learning Repository. 2016. https://archive.ics.uci.edu/ml/datasets.html. (2016). https://archive.ics.uci.edu/ml/datasets.html

50. Bahador Saket, Paolo Simonetto, Stephen Kobourov, and Kai Borner. 2014. Node, node-link, and node-link-group diagrams: An evaluation. *Visualization and Computer Graphics, IEEE Transactions on* 20, 12 (2014), 2231–2240.

51. Beatriz Sousa Santos. 2008. Evaluating Visualization Techniques and Tools: What Are the Main Issues. In *the 2008 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods For information Visualization (BELIV'08)*.

52. Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*. IEEE, 336–343.

53. B. Shortridge. 1982. Stimulus processing models from psychology: can we use them in cartography? *The American Cartographer* 9 (1982), 155–167.

54. David Simkin and Reid Hastie. 1987a. An Information-Processing Analysis of Graph Perception. *Journal of American Statistical Association* 82, 398 (1987), 454–465.

55. David Simkin and Reid Hastie. 1987b. An information-processing analysis of graph perception. *J. Amer. Statist. Assoc.* 82, 398 (1987), 454–465.

56. Ian Spence and Stephan Lewandowsky. 1991a. Displaying Proportions and Percentages. *Applied Cognitive Psychology* 5 (1991), 61–77.

57. Ian Spence and Stephan Lewandowsky. 1991b. Displaying proportions and percentages. *Applied Cognitive Psychology* 5, 1 (1991), 61–77.

58. SpotFire. 2015. http://www.spotfire.com. (2015). http://www.tableau.com/

59. Tableau. 2015. Tableau Software, http://www.tableau.com/. (2015). http://www.tableau.com/

60. Justin Talbot, John Gerth, and Pat Hanrahan. 2011. Arc Length-based Aspect Ratio Selection. *IEEE Trans. Visualization & Comp. Graphics* (2011).

61. Justin Talbot, Sharon Lin, and Pat Hanrahan. 2010. An Extension of Wilkinson's Algorithm for Positioning Tick Labels on Axes. *IEEE Trans. Visualization & Comp. Graphics* (2010).

62. J. Talbot, V. Setlur, and A. Anand. 2014. Four Experiments on the Perception of Bar Charts. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec 2014), 2152–2160. DOI: http://dx.doi.org/10.1109/TVCG.2014.2346320

63. Lothar Tremmel. 1995. The Visual Separability of Plotting Symbols in Scatterplots. *Journal of Computational and Graphical Statistics* 4, 2 (1995), 101–112.

64. Manasi Vartak, Samuel Madden, Aditya Parameswaran, and Neoklis Polyzotis. 2014. SEEDB: automatically generating query visualizations. *Proceedings of the VLDB Endowment* 7, 13 (2014), 1581–1584.

65. Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. 2015. Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)* (2015). http://idl.cs.washington.edu/papers/voyager

66. Jeff Zacks and Barbara Tversky. 1999. Bars and lines: A study of graphic communication. *Memory & Cognition* 27, 6 (1999), 1073–1079.