MEASURING THE EFFECTS OF CHART EMBELLISHMENTS TO BETTER UNDERSTAND OUR PERCEPTION OF CHARTS

by

Drew West Skau

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2016

Approved by:

____
Dr. Robert Kosara

____
Dr. Celine Latulipe

____
Dr. Wenwen Dou

____
Dr. Zachary Wartell

____
Christopher Beorkrem

ABSTRACT

DREW WEST SKAU. Measuring the effects of chart embellishments to better understand our perception of charts. (Under the direction of DR. ROBERT KOSARA)

News organizations, non-profits, and even government agencies use information graphics to advertise and communicate their messages. Data visualizations are used heavily in these graphics, but they also often incorporate unusual design elements to help catch viewers' eyes. In the struggle to rise to the top of the crowd, the data visualizations in infographics are often embellished with additions and modifications to the raw chart. The general consensus is that these embellishments can make charts less effective at communicating information, but most of them have never been tested to see if this is true. This work examines the factors in bar, pie, and donut charts that affect our perception of the charts.

I approach this in two different ways, both using a series of surveys on Mechanical Turk. The work on pie charts examines the individual contribution of arc-length, angle, and area variables so that embellishments embellishments may be evaluated based on their use of visual variables. The bar chart work examines some of the most common embellishments designers make to bar charts. This approach allows the isolated study of embellishments to determine which hinder or contribute the most to our perception of charts. I conclude with concrete recommendations based on the findings of the studies. My results show that conventional wisdom about how these charts are perceived is not always correct, and some types of embellishments are harmful while others have virtually no effect.

ACKNOWLEDGMENTS

This entire research process can be traced back to one interaction with a fellow student, Roman Pyzh. After attending the first session of a computer science class that did not interest me, Roman recommended a class taught by Dr. Robert Kosara. A year later, I was enrolled in the PhD program with Robert as my advisor. Robert has been and continues to be a magnificent advisor, providing guidance and leadership well beyond anything I could have hoped for. His community presence, masterful writing, and perspective on the field have all been invaluable through this process.

Along the way, a number of other people and organizations have contributed to my work. I'm grateful to the founders and employees of Visually Inc. Along with providing me employment, my time there revealed the need for evaluating the effectiveness of embellished charts. I'm also thankful to Dr. Lane Harrison for his excellent co-authorship and the wonderful Experimentr framework. I'm also appreciative of the funding Tableau Software provided for several of my studies.

Above all else, I'm grateful to my family for the support they have given me. My parents, John Skau and Judith West, raised me with a diversity of experience and respect for knowledge to which I owe all of my accomplishments. My brother, Erik Skau, has given me a sense of brotherly competition that will forever help me push myself to new levels. And my wife, Alexandra Skau has provided an incredible combination of emotional support, assistance, and encouragement.

I would like to thank you all for the contributions you have made to me and to my research throughout this process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

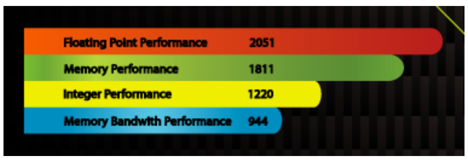
CHAPTER 1: INTRODUCTION

As data visualizations have increased in ubiquity, data visualization designers have pushed the boundaries of recommended charting techniques, adding visual embellishments to catch the eye. These embellishments are most pronounced in information graphics (infographics), but also exist in interactive web and mobile applications, and even printed media. Advertisers, news organizations, non-profits, and even government agencies have begun to use infographics to advertise and communicate their messages.

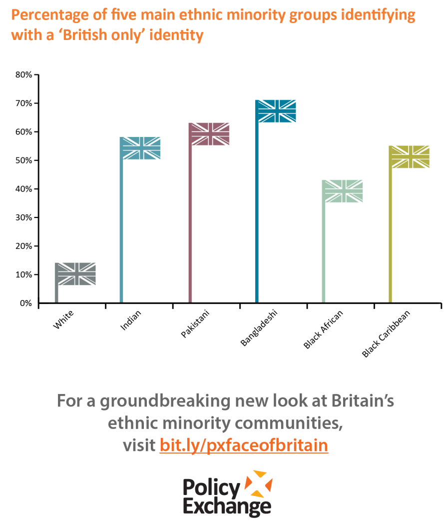
1.1 Conventional Wisdom

Data visualizations are often used heavily in these graphics, but they also usually incorporate unusual design elements to help catch viewers' eyes. In the struggle to rise to the top of the crowd, the data visualizations in infographics often are embellished with additions and modifications to the raw chart (Figure 1). Unfortunately, the foundations of knowledge on top of which designers have built these chart embellishments are shaky and incomplete. The perceptual effects that charts rely on are not fully understood, so modifications to charts are made without an awareness of the effects the changes will have on the chart's ability to communicate.

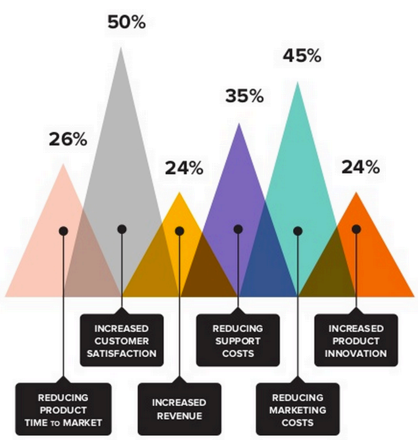
As designers have done increasingly creative things to charts, the potential for data distortion has also grown, however conventional wisdom only discourages modifica-



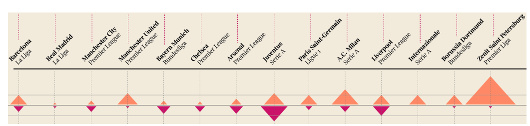
(a) Rounded ends bar chart.



(b) Capped bar chart.



(c) Overlapping triangles bar chart.



(d) Quadratically increasing area bar chart.

Figure 1: A sampling of charts used in infographics, taken from examples found on Visually [40].

tions, and does not offer any insight into what modifications might be okay, and which may be harmful. This leads to a mismatch between what the community creates and the knowledge that can guide and evaluate those creations.

Regardless of infographics or chart embellishments, pie, donut, and bar charts are all common chart types. The creation of these charts has been made more accessible through increasing support by office suite software, design suites, and software that manages and produces data like CRM systems, project tracking platforms, and analytics platforms. The frequency of use of these charts justifies deeper study into the mechanisms that make them work.

The data visualization community has recently experienced a large growth, with an abundance of new books, however the expertise visible to people outside the field and present in most books in the field primarily stems from work by a few leaders, primarily conducted decades ago. The spread of knowledge through these books has outpaced the community's ability to verify the older work that it is based on, leading to a shaky foundation of conventional wisdom that is often untested and occasionally untrue. Conventional wisdom says that many chart embellishments harm the communication of the data by damaging, impeding, or distracting from the pure visual representations of the data (often referred to as *chart junk* [38]). Despite this, infographics and the field of data visualization in general continue to manipulate charts to include embellishments.

As designers continue to push the boundaries of accepted chart practices, it is up to the data visualization research community to evaluate those embellishments. This dissertation explores the visual variables within pie, donut, and bar charts that contribute to our perception of the data the charts represent and evaluates the impact of common embellishments on communication accuracy. Together, these strategies build a foundation of knowledge about how charts work, and what embellishments can be used safely without harming communication accuracy.

1.2 Overview

In this document, the work is split into two primary chapters, bar charts (Chapter 2), and pie and donut charts (Chapter 3).

First, for bar charts, I abstracted modified charts into simplified categories of common embellishments. The abstracted versions are designed to examine the overall shape of embellished bars (Figure 2). I conducted a user study to compare judgement accuracy for embellished vs. baseline charts. The results of this study were published in *An Evaluation of The Impact of Visual Embellishments In Bar Charts* at EuroVis 2015 [34] (Section 2.2). I combined this with a second study testing full color embellishments, and determined that the primary impact from embellishments is from the outer boundaries of the bars, not internal color changes or visual cues (Section 2.3).

Pie and donut charts necessitated starting with a different strategy. Instead of testing common embellishment types, I started by isolating the individual visual variables that a chart uses. This is done by designing a version of the chart with only one visual variable remaining. Once isolated, I examined the levels of communication accuracy that each variable can produce. By decomposing the chart embellishments based on the visual variables they impact, the repercussions of individual embellishments can be evaluated. The results suggest that some embellishments are better than others, that pie and donut charts are equally good for communication accuracy, and allow me to offer advice on which embellishments are appropriate to use in which situations.

This work was published as a paper (*Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts* [35]) at EuroVis 2016, and consisted of two user studies (Sections 3.2 and 3.3). Alongside that paper, I published another study that looked at the impact of specific embellishments (*Judgment Error in Pie Chart Variations* [20]). This study used the same strategy as my bar chart work to look at the impacts of simplified categories of specific embellishment types (Section 3.4). These two papers together worked to call into question over 90 years of conventional wisdom saying that angle was the primary visual cue for reading pie and donut charts.

These two methodologies for evaluating the chart embellishments, and the perceptual significance of individual visual variables within a chart is a general contribution of my work. When both methodologies are used to study a chart type, the findings from each can be used to help test the findings of the other, and work toward a more general understanding of how the chart type is read by our perceptual system.

This document concludes with a chapter summarizing the contributions of the work, and a list of rules for designing bar, pie, and donut charts (Chapters 4 and 5).

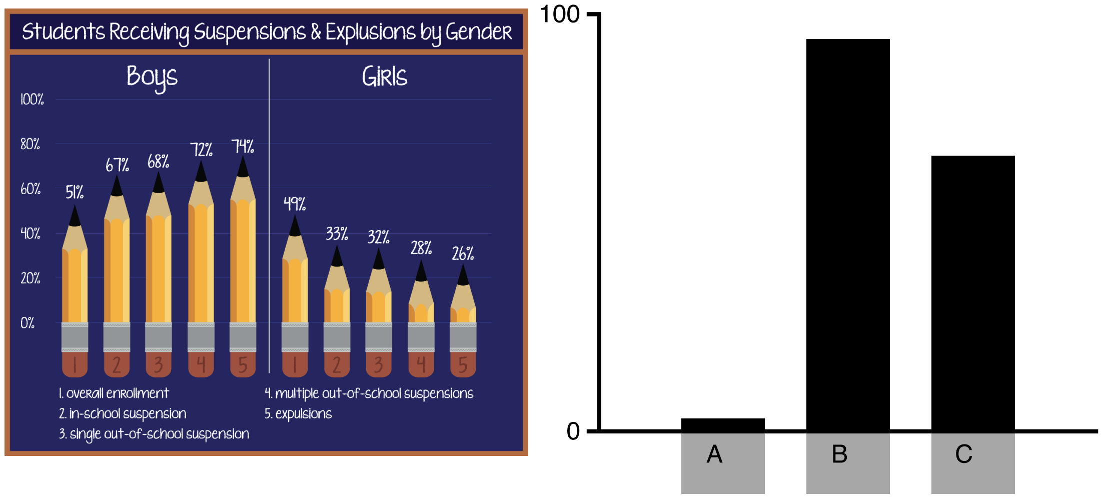
CHAPTER 2: BAR CHARTS

Bar charts are a great tool for showing categorical or time series data. The first bar chart is generally attributed to William Playfair, published in *The Commercial and Political Atlas* [29] in 1801. Bar charts are ideal for displaying small discrete sets of continuous values, making them one of the most commonly used data visualizations. In infographic circles, their prevalence has helped lead to a common perception of standardized bar charts as plain and dull. This has prompted exceptional levels of creativity with bar chart aesthetics. Designers frequently change the shape of bars, use recognizable objects as bars, or turn the whole bar chart into part of an image.

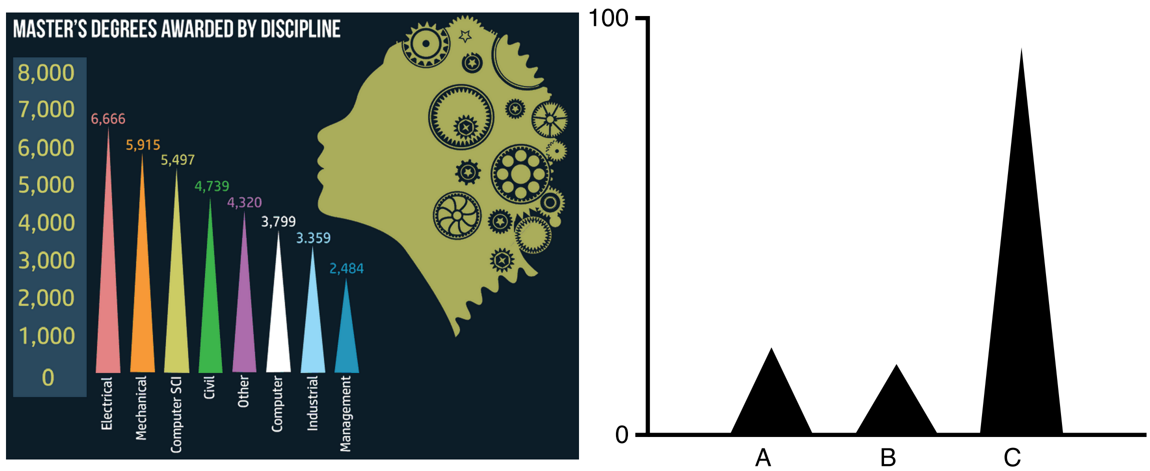
This is done to provide context on the topic of the chart and make it more memorable, and undoubtedly also to make the chart more visually attractive. For an information graphic to actually be informative, however, the reader has to be able to read the charts. How much common chart embellishments decrease a chart's readability needs to be studied so that designers can make informed decisions about how to balance attractiveness and distortion of the data that is displayed.

Common embellishment shapes include design elements that the information visualization community generally recommends against: non-rectangular bars (in particular triangles), rounded tops, the use of color and images in bars, etc. (Figure 2).

Color, imagery, and altered chart geometries introduce a number of complex visual features into the chart. They are usually considered clutter or *chart junk*. The goal



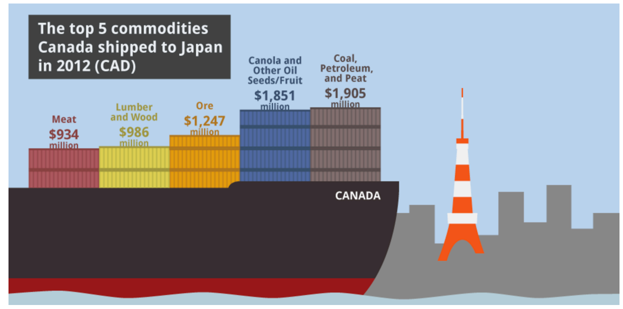
(a) Bars extend below zero line.



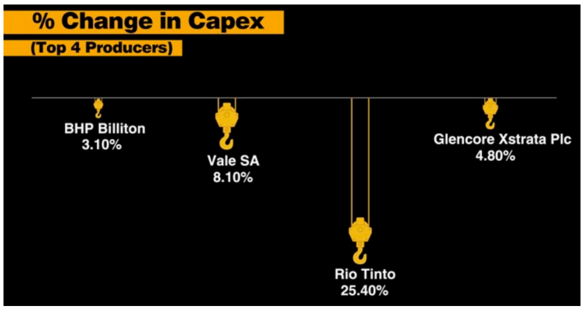
(b) Triangle bar chart.

Figure 2: Two examples of embellished charts and abstracted versions of the embellishments.

(a) Miscellaneous images as bars.



(b) A ship with stacked containers as bars.



(c) Crane hooks as bars.

(d) Ice cream stacked in cups as bars.

Figure 3: Examples of bar charts embellished with images taken from various infographics found on Visually [40].

of this work is not to assess the potential positive effects of this (such as memory or attention), but strictly to look at their impact on the accuracy with which the charts can still be read. Does accuracy suffer? Even if not, does it take longer to read the charts?

My work on bar charts consists of two studies. The first study reduces common embellishments to geometric forms, testing how accurately each form can be read (Section 2.2). The second study delves into more complex embellishments to determine if color changes internal to bars have a significant impact on bar readability (Section 2.3).

2.1 Related Work

Tools like Datavisual [42], RAW [13], Infoactive [6], Visage [26], Plotly [30], or Lyra [32] make it possible for non-technical designers to build visualizations. These tools often allow the user to export files that are then easy to edit further in design programs like Adobe Illustrator.

Despite their increasing prevalence in infographics, chart embellishments have typically been viewed as decreasing a chart's ability to communicate the data involved. Edward Tufte has been one of the major drivers behind this point of view, coining the term "*chart junk*" to describe embellishments, and calling them *non-data-ink* or *redundant data-ink* [38]. Despite a general lack of empirical evidence supporting this negative view on chart embellishments, it has driven a culture of design simplification and sterilization in the data visualization community. Tufte's writings encourage this approach of reducing the data/ink ratio. This may be a good rule of thumb for visualizations, but not for infographics. It fails to address the impact of design modifications on the real salient features of a chart. By this measure, bars in a bar chart would be better if they were reduced to a single-pixel wide line. There is a balance to be struck between simplicity and practicality, and work by others has begun to investigate this. Some have begun to investigate memorability. However, being memorable and comprehensible are not the only goals for a chart; it is also important to communicate data accurately.

In the midst of the controversy, websites like Visually [40], Visualizing.org [39], Dadaviz.com [12], or ILoveCharts [18] have brought data visualizations merged with design into the public spotlight. They are proof of the increasing occurrences of design merged with data visualization, and their showcases are rife with embellished charts. This points to a stark divide between what is done in practice, and what is encouraged by the theoretical side of the community.

An evaluation of embellished charts would provide tangible proof of the impact of designers' creativity. There are well established ways to evaluate charts. Cleveland and McGill performed seminal work in 1984 with their studies exploring differences between bar and pie charts for different tasks [8]. Specifically, the bar chart portion of their study selects a confounding factor (e.g. distance between bars) and examines how that factor has an impact on the chart's performance for a given task (comparison between bars). They used the results of this to determine best practices when selecting which chart type to use for a given task. Their study has been replicated by Heer et al. using Amazon's Mechanical Turk [17], an internet based platform for sourcing work for any scale of human intelligence tasks. Other groups have proven the platform for studies on perception [21] and visual decisions [19]. The scalability, price, and relatively automated process that Mechanical Turk provides make it an attractive platform for running browser-based user studies.

The work that comes the closest to evaluating the communication abilities of embellished charts mostly addresses memorability. Research by Bateman et al. has suggested that people can still interpret embellished charts accurately [1], and that embellishments may actually improve memorability [4]. Other research by Borgo et al. suggests that charts embellished with semantically meaningful objects can have an impact on working memory, long term memory, visual search, and concept grasping [3]. This study, however, also focuses on memorability, not on accuracy of communication.

There is a body of work that specifically investigates the perception of bar charts. Talbot et al. have taken Cleveland and McGill's study and delved deeper into the questions of how different bar chart configurations impact accuracy [37].

In another study that looked at the impact of similar design changes on perception accuracy, Zacks et al. compared three versions of charts, a single thin line for each bar, a thicker bar, or a projection of a 3D bar [43]. They found that while 3D perspective depth cues lowered accuracy, distractions from neighboring elements were a more damaging source of inaccuracies. It has been found that the pictorial embellishments in ISOTYPE charts did not have a negative impact on accuracy or reading time [16]. Correll et al. explored the effectiveness of alternate error bar designs on communication of confidence intervals on bar charts [9]. Newman et al. have shown that predictions of averages across a bar chart always end up weighted lower than they should be as a result of the alignment of the bars to the bottom end of the scale [28]. Elzer et al. constructed a model of perceptual task effort aimed at improving communication of the message a bar chart is intended to convey [15]. Their work shows that having an understanding of the perceptual effects going on in a bar chart is critical to making it communicate accurately. All of this work points to an incomplete understanding of the perceptual issues involved in bar charts.

2.2 Simplified Bar Chart Embellishments

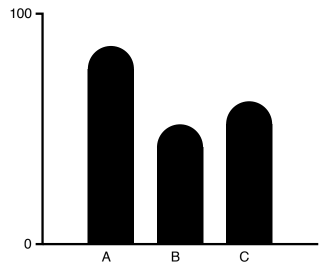
The effects of bar chart embellishments on communication accuracy are incompletely understood, and this means there is potential for designers to use or create harmful embellishments without knowing it. In this study, I identify common embellishment types (Figure 4) and test them against a baseline bar chart. This allows me to measure the impact any bar chart embellishment types have on the chart's ability to communicate data.

This section gives an overview of the embellishment design space before delving into the study materials, procedure, and results. I conclude with a discussion of the implications of my findings, some advice for designers, and a summary of the research.

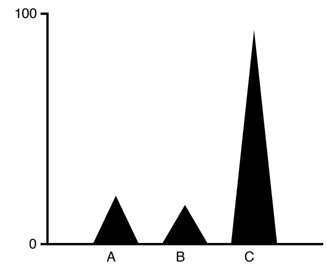
2.2.1 Design Space Overview

The first step to evaluating chart embellishments is defining what they are. I define embellishments as deviations from a baseline chart. In the case of a bar chart, deviations include changes to the shape of the bars, added components, or bars with an altered set of data encodings.

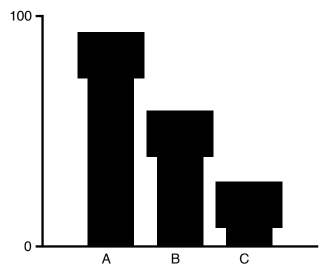
The second step is classifying the different types of embellishments. There is a wide range of chart embellishment styles and types, and different embellishments likely have different impacts on the chart. For example, a curved or pointed end to a bar means there is no strong line at the end for the viewer to extend to the value axis. At the same time, a triangular bar has the angle of the triangle edges that also communicate the data. Bars with quadratically increasing area, or bars that overlap mean the bar's area does not accurately encode the data, so the area and



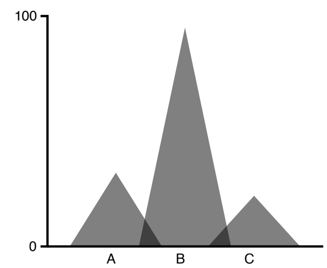
(a) Rounded tops bar chart.



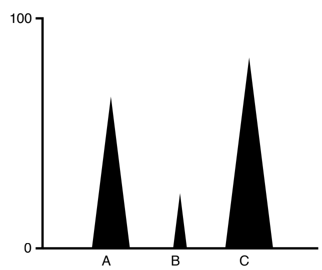
(b) Triangle bar chart.



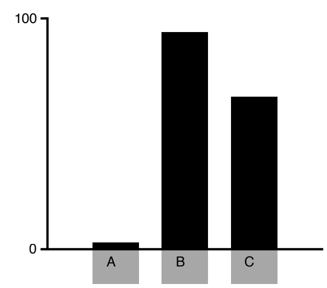
(c) Capped bar chart.



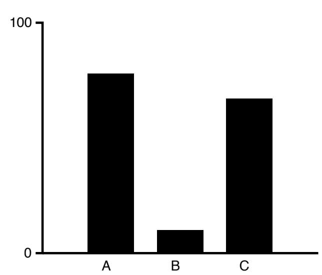
(d) Overlapping triangles bar chart.



(e) Quadratically increasing area bar chart.



(f) Bars extend below zero line.



(g) Baseline bar chart.

Figure 4: Embellished bar charts simplified for use in the study.

bar height provide conflicting signals. Bars that extend below the zero point to allow labeling may make comparisons between bars more difficult as the overall height is not proportional to the data. Caps on bars can introduce ambiguity in what signifies the top of the bar (the middle of the cap? the bottom edge? the top edge?). These possibilities provide motivation for studying the effect of embellishments on graphical perception.

Through an informal survey of infographics found on Visually [40], I have identified a set of bar chart embellishments that occur frequently. This set serves as a starting point for establishing the impact of embellishments on graphical perception.

1. Rounded corner charts (Figures 1a and 4a) do not have a strong line at the end for the viewer to mentally extend to the value axis. Sometimes these are rounded due to being a portion of an illustration, or as part of a pseudo-3D effect.

2. Triangle charts (Figures 2b and 4b) have the same issues as the rounded corners charts, however they also lack any vertical edges to help judge height. These also add a data encoding, as the angle of the end point changes based on the height (albeit, not with linear proportionality).

3. Capped bars (Figures 1b and 4c) come in many forms, but in all cases, they have a wider end, or an end with a stronger color contrast. The change in the visual weight of the bars is the primary change from the baseline chart.

4. Overlapping triangle charts (Figures 1c and 4d) generally have some level of transparency so the overlapping regions are visible. The overlapping technique

used with these is used almost exclusively with triangular bar charts. This is likely because of the added visual complexity that comes from overlapping angles rather than overlapping rectangular regions, and because overlapping triangles are easier to tell apart than overlapping bars.

5. Quadratically increasing charts (Figures 1d and 4e) have shapes ranging from simple rectangles or triangles to an illustrated figure like the Shrinking Family Doctor [38]. Still, the most common version of these is triangles.

6. Bars that begin before the origin point (Figures 2a and 4f) are often illustrations of real world objects, or the bar labels are an extension of the bar itself. This embellishment makes the bar read as being longer than it actually is, possibly causing the value the bar represents to be misinterpreted.

This set of six embellishments was identified as worth exploring for two reasons: they occur frequently in infographics, and they can be created using standard tools. In addition, all chosen embellishments have some impact on the salient features of the chart in a way that may change our ability to interpret the data. Some of them are isolated versions of a larger set of chart modifications that are done for a certain effect. Some of them are abstracted versions of embellishments that occur frequently in different forms. For example, the capped bar embellishments include a variety of different caps, however the version I chose is abstracted to test the general principle of having more visual weight at the top of the bar than at the bottom. Triangular charts are extremely prevalent in infographics, at least partly due to easy creation with tools like Microsoft Excel or Piktochart. Rounded bar ends are also common in infographics, sometimes as a part of other effects like pseudo 3D cylinders, or often just as rounded ends.

2.2.2 Study

The primary goal that motivates measuring the performance of visual communication of charts is communication accuracy. This step validates how well secondary goals like communication speed or information retention can be achieved based on the accuracy of the information that can be retained or communicated quickly. To realize this goal, I adapt the experiment design from Cleveland and McGill [8] and Heer and Bostock's replication of their experiment [17], making necessary changes to study the impact of embellishments on communication accuracy.

This study relies on comparisons to a baseline chart style. This chart style uses a grayscale color theme, a clear to read and familiar font (Helvetica), and standardized labels and axes. An example of the baseline style for a bar chart can be found in Figure 4g. Based on the previous section, Table 1 shows the details of my hypotheses for each bar chart embellishment's performance as compared to the baseline. Overall, I expected most embellishments to perform roughly equivalently to or worse than the baseline, with a few exceptions for the bars with end caps. I reasoned that the end of the bar farthest from the zero line was the most important visual portion. Most of the embellishments either damage an encoding, or reduce the prominence of the end of the bar farthest from the zero line. Only end caps make that portion of a bar more pronounced, so it is possible they will improve the communication of the chart's data.

Table 1: Hypotheses for each chart embellishment's performance as compared to the baseline chart (Figure 4g).

Modification	Figure	Absolute Judgement	Relative Comparison
Rounded tops	4a	equal	equal
Triangles	4b	worse	worse
End caps	4c	better	better
Overlapping triangles	4d	worse	equal
Disproportionate area	4e	worse	equal
Extend below zero	4f	worse	worse

2.2.2.1 Materials

Rather than testing the "in the wild" versions of these charts that often use bright colors, fancy fonts, and all manner of labeling schemes, I recreated versions of the charts that deviate from the baseline in a controlled manner, and by only one criterion at a time (Figure 4). All charts use the same font and labeling technique as the baseline chart.

The charts display three values each, labeled A, B, and C for reference in the questions. I constrain the bars to three to contain the scope of the study to issues caused directly by the embellishments, avoiding compound issues that may appear when two compared bars have many bars between them (as found by Cleveland and McGill [8] and Talbot et al. [37]). The minimum value on the y-axis always starts at 0, while the maximum y-axis value is capped at 100. To prevent overly simple judgements where the bars line up with the bottom or top of the y axis, the bar values are generated randomly in the range from 3 to 97.

2.2.2.2 Procedure

The study began with an introduction page and a short demographic form. This was followed by a page discussing the procedure for answering questions and a block of one type of question (Figure 5). At the end of the first block of questions, there was a short break page reminding the user of the procedure for answering questions and a block of a second type of question. The order of the two blocks of question types was randomized to ensure that there were no learning effects from one question type that would influence the results of the other. The first embellishment type in each block was rotated per participant, also to ensure that there were no learning effects from one embellishment type that would influence the results of the others. The order for the subsequent embellishment types was randomized per participant to ensure that transition order between embellishments had no influence on the results.

For some charts, it may not be immediately obvious that they should be read as bar charts. In these cases, seeing the baseline bar chart could tip viewers off to the pattern of them all being bar charts. In the wild, these charts are often encountered without any instructions or context to suggest they are bar charts (Figure 1). There are several things I did to ensure viewers come to their own conclusions on how to read the charts.

First, there was no specific mention of bar charts, column charts, bars, or columns throughout the study or participant materials leading up to the study. The charts were merely referred to as 'charts', and the bars were referred to by their label.

Second, the order in which the variations were presented to the participants was

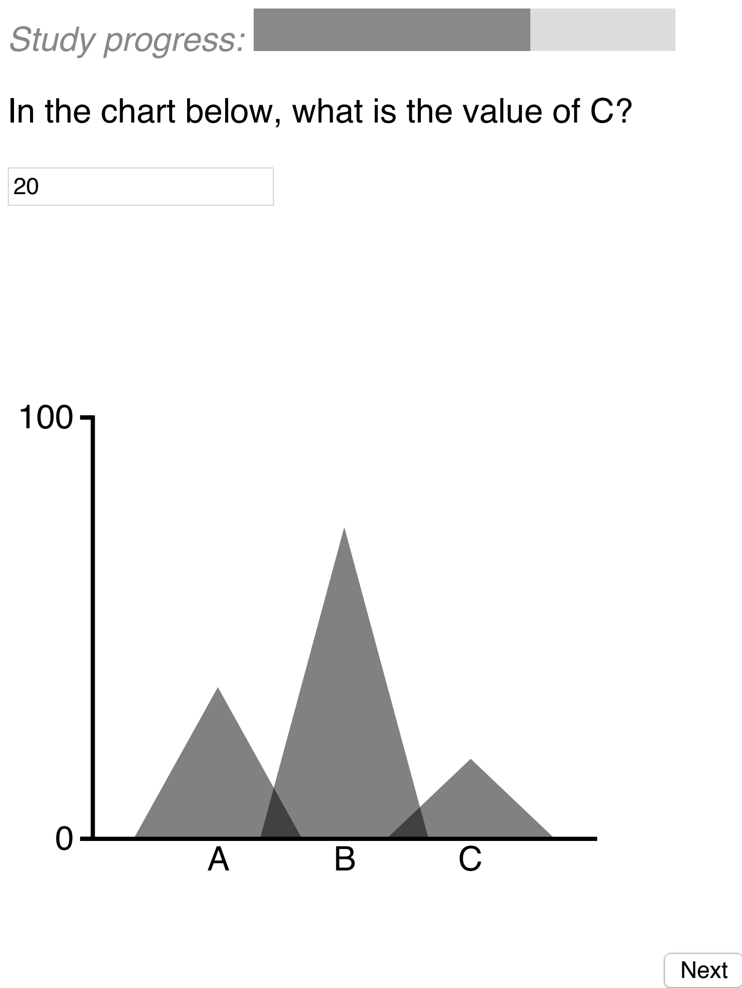


Figure 5: An example of what a survey participant saw when answering questions. This screen uses the absolute value question and the baseline bar chart. The participant has entered “20” into the text entry.

controlled. The first embellishment type a participant saw was cycled per participant ensuring equal coverage per type, using a Latin Square. The remaining rows of the square were shuffled randomly to distribute the transitions between chart types.

2.2.2.3 Question Types

The two different question types were designed to test the main tasks associated with bar charts: comparison between bars, and reading the value of a single bar.

A. *In the chart below, what is the value of A?*

B. *In the chart below, what percentage is B of A?*

These questions do not address a user's comprehension of what the chart communicates. I have intentionally left off units and only assigned values to the charts, so that I can focus only on the accuracy of communication and the perception of the charts. It is possible that these embellishments could impact higher level comprehension based tasks, however that research is beyond the scope of this study.

The questions asking about the absolute value of a single bar (question type A) are simpler to answer, so participants were shown five of these questions for each of the seven embellishment types for a total of thirty-five questions in this section. The questions asking the participant to make a comparison between two bars on the chart (question type B) are more difficult to answer and have more permutations of bar position along the x-axis, so participants were shown eight of these questions for each of the seven embellishment types for a total of fifty-six questions. In addition, the y-axis is not necessary for this task, and could actually allow the participants to

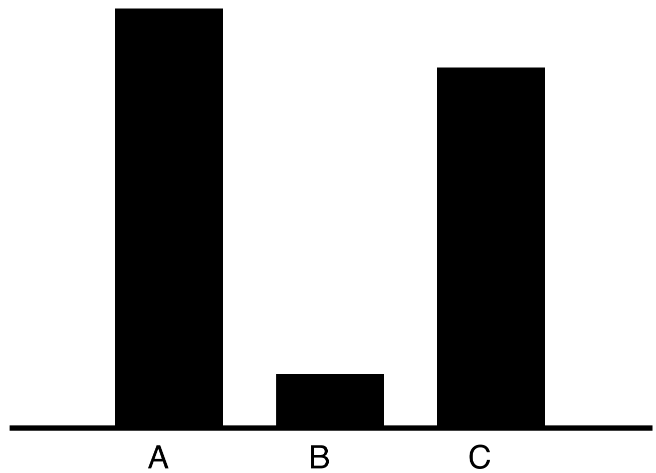


Figure 6: Chart used in the relative-question condition did not include a y-axis, to ensure that participants compared the two bars based on their perceived differences rather than numerical estimation.

"cheat" by first judging the absolute height of each bar and mathematically computing the percentage. To prevent this, the y-axis was removed from the charts for these questions (Figure 6).

2.2.2.4 Results

I recruited 100 participants through Amazon's Mechanical Turk for this experiment. Mechanical Turk refers to a task someone does on their platform as a HIT or Human Intelligence Task. Of the original 100 HITs, three were rejected and re-run by different participants because over one quarter of their answers were incorrect by more than 30%, indicating they were not paying attention or did not fully understand the questions. This resulted in a total of 103 participants with 100 being paid US $2.00 for their participation. It took approximately two days to gather all responses. Average completion time for all participants was 19 minutes and 11 seconds. My study used a total of 6 embellishments (and 1 control, the baseline), and two question types (absolute/relative), yielding 7 main conditions and 2 sections. Each participant answered questions using all seven chart types, resulting in 35 absolute and 56 relative judgements per participant.

Because the experiment is adapted from previous studies in graphical perception, I follow previous methods for computing errors and confidence intervals [8, 17]. Error was computed using the midmean of log-absolute error (MLAE), and 95% confidence intervals via bootstrapping [8]. To mititgate the effect of outliers, 6 participants were removed because their average error exceeded 172% (maximum error for the remaining 94 participants was 45%). This result is in line with Heer and Bostock's crowdsourced graphical perception experiments [17], and Mason and Suri's study examining the data quality on Mechanical Turk [25].

Consistent with the previous studies mentioned above, the resulting errors in each question type and embellishment group were non-normally distributed. To test the effect of embellishments against the baseline, I compared the error of each to the baseline condition through six Mann-Whitney-Wilcoxon tests. I use a Bonferonni correction to address the problem of multiple comparisons, resulting in an $\alpha = 0.0083$ required for rejecting the null hypothesis. All tests results and parameters are reported in Tables 2, 3, and 4. To aid visual comparisons, I provide plots of the means and their 95% confidence intervals in Figures 7 and 8.

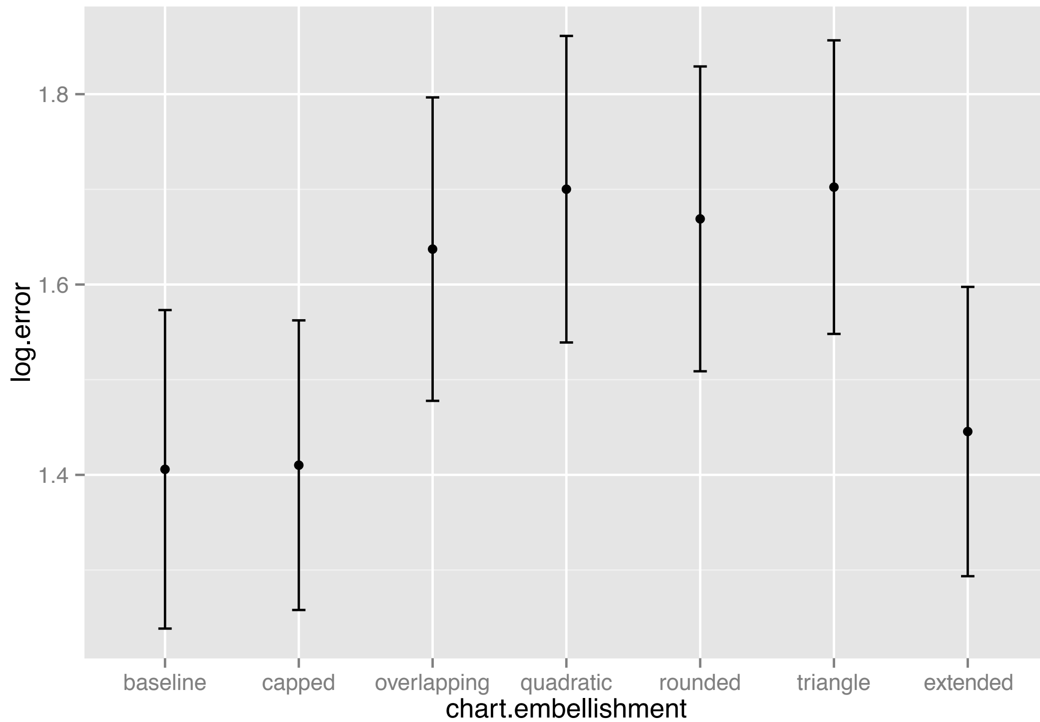


Figure 7: Log error for the absolute value question by embellishment type. Only quadratic is significantly worse than baseline, as indicated by ANOVA.

Table 2: Hypotheses for each chart embellishment's performance as compared to the baseline chart (Figure 4g). Results are based on differences in average log-error, with statistically significant differences indicated by * with $\alpha = 0.0083$.

		Absolute Judgement		Relative Comparison	
Modification	Figure	Hypothesis	Result	Hypothesis	Result
Rounded tops	4a	equal	worse	equal	worse*
Triangles	4b	worse	worse	worse	worse*
End caps	4c	better	equal	better	worse*
Overlapping triangles	4d	worse	worse	equal	worse*
Disproportionate area	4e	worse	worse*	equal	worse*
Extend below zero	4f	worse	equal	worse	equal

Table 3: *Absolute judgements: summary statistics and Mann-Whitney-Wilcoxon Tests Results comparing embellishment types to the baseline. Significant values denoted by * with* $\alpha = 0.0083$.

Embellishment	Mean	SD	p-value
baseline	1.41	1.85	
capped	1.41	1.68	0.500
overlapping	1.64	1.76	0.048
quadratic	1.70	1.78	*0.007
rounded	1.67	1.77	0.009
triangle	1.70	1.68	0.012
extended	1.45	1.68	0.562

2.2.2.5 Absolute Judgements

For absolute judgments, which involved estimating the value of a given bar, only quadratic bars ($m = 1.70, sd = 1.78$) performed significantly worse than the baseline ($m = 1.41, sd = 1.85$). One possible explanation is because the area change in the quadratic bars exaggerates the overall size change and may make it harder to estimate the height of the bar against the value axis. Other embellishments performed similarly to quadratic in terms of error, including overlapping ($m = 1.64, sd = 1.76$), rounded ($m = 1.67, sd = 1.77$), and triangle ($m = 1.70, sd = 1.68$). Notably, both extended ($m = 1.45, sd = 1.68$) and capped ($m = 1.41, sd = 1.68$) are similar to the baseline in terms of error.

2.2.2.6 Relative Judgements

The trends in the relative judgement data were strikingly different than the absolute judgement data. Relative judgments involved estimating the percentage of a given bar to another, and only the extended embellishments ($m = 1.59, sd = 1.66$) performed similarly to the baseline ($m = 1.43, sd = 1.85$). In all other cases the baseline

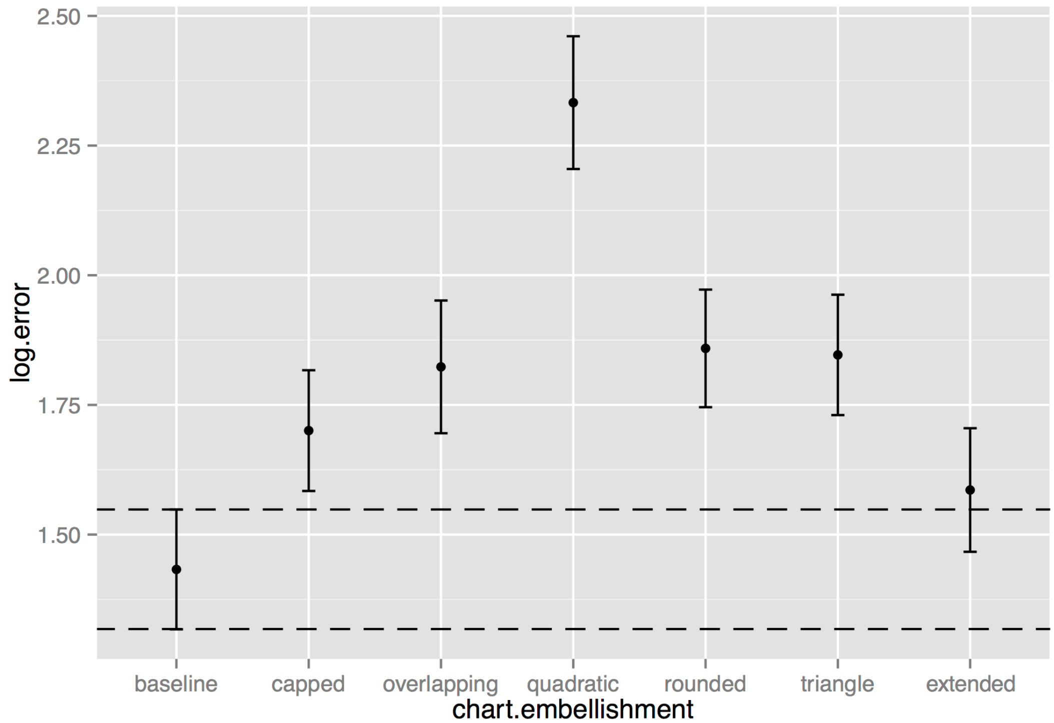


Figure 8: Log error for relative judgments (comparing two bars) by embellishment type. Extended is no worse than baseline, while the other embellishment types are significantly worse.

condition performed significantly better. Specifically, capped ($m = 1.70, sd = 1.68$), overlapping ($m = 1.82, sd = 1.76$), rounded ($m = 1.86, sd = 1.77$), and triangle ($m = 1.85, sd = 1.68$) performed similarly, while quadratic produced much higher error overall ($m = 2.33, sd = 1.78$).

2.2.3 Discussion

The results of this experiment confirm that common embellishments can significantly impact the perceptual performance of bar charts, and that these impacts differ substantially based on the task (i.e., absolute versus relative judgements).

Specifically, none of the embellishments tested in this experiment performed better at communication of the data than the baseline standardized chart. One notable

Table 4: *Relative judgements: summary statistics and Mann-Whitney-Wilcoxon Tests Results comparing embellishment types to the baseline. Significant values are denoted by * with* $\alpha = 0.0083$.

Embellishment	Mean	SD	p-value
baseline	1.43	1.85	
capped	1.70	1.68	*0.001
overlapping	1.82	1.76	* < 0.001
quadratic	2.33	1.78	* < 0.001
rounded	1.86	1.77	* < 0.001
triangle	1.85	1.68	* < 0.001
extended	1.59	1.66	0.097

exception is the *capped* bar chart, a bar chart with an additional “cap” on top which is wider than the bottom portion of the bar. These performed equally well as, and with slightly lower variance than, the baseline chart for absolute judgement questions. This result suggests that users indeed rely on strong lines at the ends of bars to mentally extend the bar end to the value axis, especially when considering the comparatively poor performance of the embellishments that distort the top of the bar (rounded caps, triangles, etc.) – see Figure 9.

All adaptations except the *extended* embellishment performed significantly worse than the baseline on relative judgements. Even small changes, for example the rounded bar, produced a significantly higher error rate. This confirms the hypothesis of Tufte and others, that quadratically-scaled bars lead to large errors, even when compared to similar embellishment styles (i.e. triangles, overlapping triangles).

2.2.4 Recommendations

For designers creating charts, this study produces actionable advice. It is advisable to stay away from creating triangular bar charts. Triangular charts that overlap

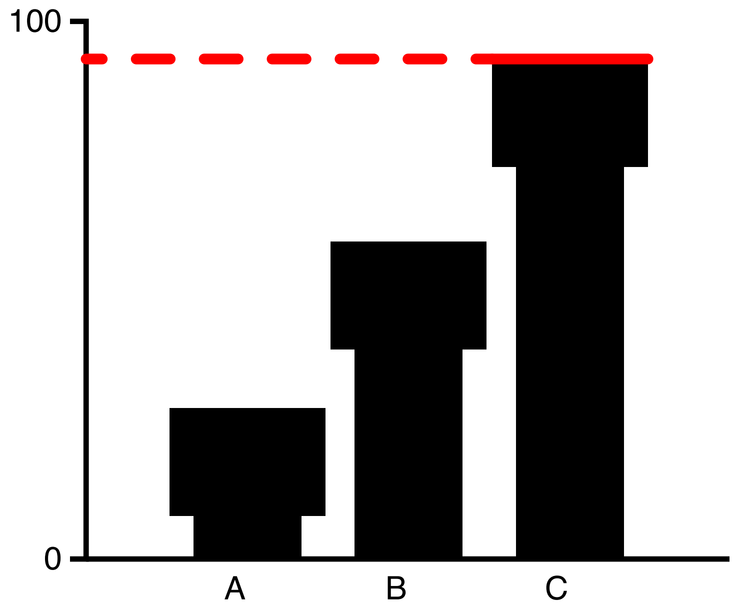


Figure 9: The possible mental operation viewers use to tell the value of an individual bar depends heavily on the shape of the top of the bar and the axis.

and that have quadratically changing areas are especially worth avoiding. While there are no guarantees about shapes other than triangles with disproportionately changing areas, these results suggest that it is inadvisable to scale chart elements on two axes simultaneously. This is in line with common wisdom. End caps with a strong horizontal top are not advisable for tasks that involve comparing bars, but are fine (and perhaps better) for absolute judgements. Bars that have a portion extending below the zero point on the value axis seem to be fine to use, assuming the portion that extends is a visibly different color from the value portion of the bar.

The results of this study qualify findings by Borkin [4] and Borgo [3] suggesting that memorability can be aided by embellishment. Changes to charts that affect the primary chart elements can reduce the communication accuracy of the chart. Increasing the memorability of a chart is certainly a worthwhile pursuit, however it must be balanced with the need to communicate information accurately in the first place.

Several promising areas for future work follow the results of this experiment. Quantifying the impact of embellishments on perceptual accuracy establishes a baseline to test the impact of other factors that are at play when designing charts. By quantifying the impact of design factors on perception, it is possible to explore and optimize compromises that may be struck between accurate communication and high-level design goals.

2.2.5 Conclusions

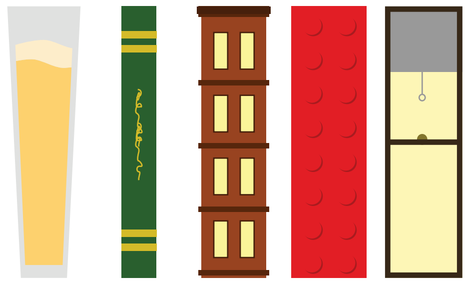
I present a crowdsourced experiment to investigate the impact of common chart embellishments on the accuracy of absolute and relative judgements in bar charts. The results of this experiment establish that bar chart embellishments do indeed have an impact on how well the data within the chart can be communicated. For nearly all tested chart embellishments, even small changes like rounding the top of a bar, led to higher error rate. However, there was one notable exception, the T-shaped *capped* bar chart.

These results advance our understanding of the intricacies of how we use the visual cues in bar charts. They identify which visual cues in bar charts have the most impact, and establish a basis for exploring the impact of low-level design elements in graphical perception.

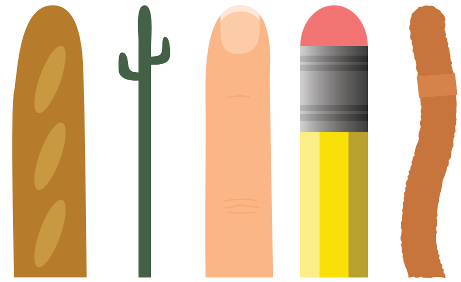
2.3 Fully Embellished Bar Charts Study

In the previous study, I determined that bar chart embellishments can have a negative impact on a chart's ability to communicate data. However, the embellishments I tested were generalized types of embellishments with a limited set of aesthetic distractions. In reality, embellished charts often have many more variables impacting their communication accuracy. In this study I took a handful of embellishment types and created several fully embellished charts for each. By comparing these fully embellished charts against the generalized form, I identified whether or not the added aesthetic distractions had any further impact on the charts' ability to communicate.

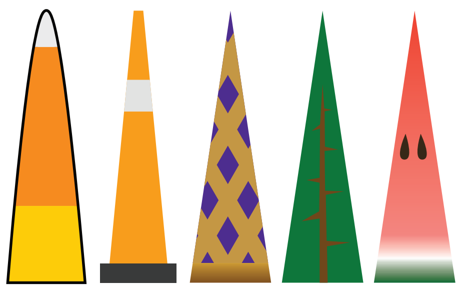
This section begins by outlining my hypotheses. I describe the materials and



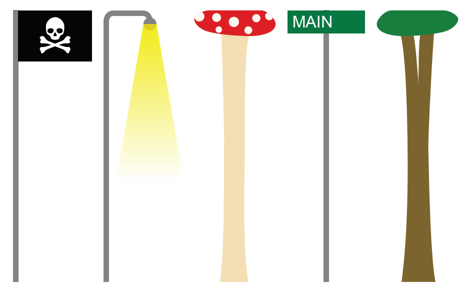
(a) Images used as rectangular bars.



(b) Images used as bars with rounded tops.



(c) Images used as triangular bars.



(d) Images used as T-shaped bars.

Figure 10: Do elaborate pictorial embellishments of bar charts lead to reduced precision when reading them? I tested four classes of infographic-style embellished bars with five designs each.

procedure used for the study, and present the results. I conclude with a discussion of the results and recommendations for designers using and creating embellished charts.

2.3.1 Embellishments and Hypotheses

The previous study outlines a set of bar chart shape embellishments that occur frequently in the infographic design space. Their designs varied shape, but did not use color or texture. In this study, I build on their designs by adding pictorial content to the basic bars. This impacts their appearance in a number of ways. First, the shapes may distort the purely geometrical base shape of the bar (Figure 10), and potentially add a third dimension. Some of the designs make shape elements more or less obvious, such as the rounded top in the baguette versus the pencil; the T-shape can be more or less pronounced, e.g. in the Jolly Roger flag versus the street lamp. Different colors also impact the perceived weight of the chart [44].

I used the previous findings on the impact of different bar shapes to narrow down the list to four with a less harmful impact:

- Rectangular bars
- Rounded corner charts
- Triangle charts
- Capped bars

In the previous study, I found that some embellishment shapes definitely harm communication accuracy – to the point that they should probably be avoided without compelling reasons. However, some embellishment shapes had a less severe impact on communication accuracy, and their use could be justified fairly easily for some of the other benefits they may provide.

In this study, I hoped to discover if versions of embellishments complete with colors and internal structures increased or decreased the impact of the shape-based embellishments on communication accuracy.

Based on my previous results, I developed the following hypotheses:

- Any additional embellishment with color and shape will lead to higher error compared to a solid bar of the same shape.
- The higher complexity of pictorial bars will require more time to read.

- I found T-shaped bars to be no different from base bars for absolute judgments in the previous study. I hypothesize that they will not be impacted by further embellishments.

2.3.2 Study

I took the four primary embellishment shape categories and ran a within-subject study on Mechanical Turk, using an experiment design adapted from Cleveland and McGill [8]. I measured communication accuracy with the same question structure as my earlier work, with tasks for relative comparisons between bars, and absolute comparisons of one bar to the chart's y-axis.

I compared all embellishments against a baseline chart style with black rectangular bars (Figure 12). All charts used the same axes as in the previous study, with relative questions having no vertical axis, and absolute questions having a vertical axis ranging from 0 to 100. As in the previous study, I used Helvetica as the font for all chart labels, and standardized the axes across all chart types.

2.3.2.1 Materials

The previous study left out colors and any internal embellishments in their study, instead focusing on the impact of different bar shapes caused by embellished charts.

While my materials adopt four of the same shape classifications, I have used pictorial embellishments complete with colors. I hypothesized that colors and markings internal to specific embellishments may have a significant impact on the ability of a chart to communicate accurately. For each shape classification I used, I created five different images for a total of twenty (Figure 10). In order to protect the recognizabil-

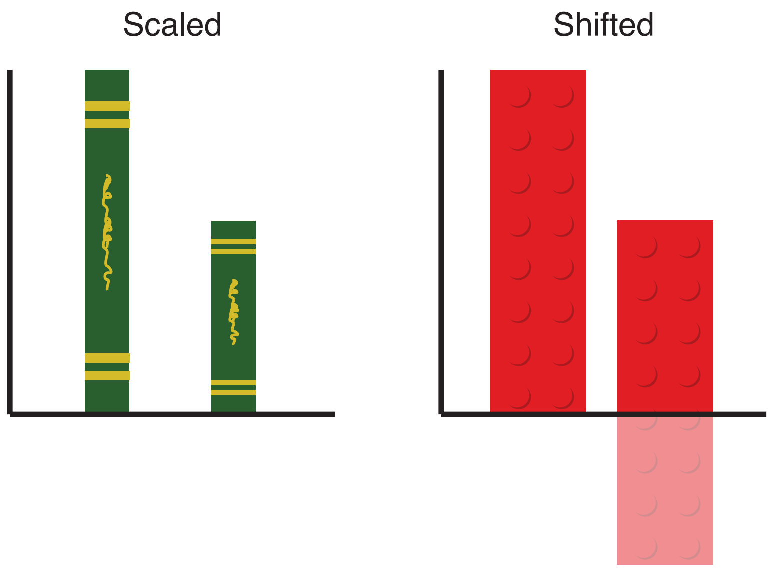


Figure 11: The two different methods for sizing bars in the chart. The lighter portion of the toy brick below the x-axis was not visible in the study. Here it illustrates the mechanism for displaying shifted bars at different heights.

ity of each image as much as possible, different images are drawn with different scaling methods. Some images can be vertically stretched and still look very much like the objects they represent, however, for some images this reduces their recognizability.

For example, books come in many different shapes and sizes, so a short, squat book is a reasonable image. A street sign, however has text which should stay near a certain aspect ratio, so the sign should be drawn at different heights by shifting down the signpost and cropping the bottom rather than stretching the entire object. To account for this in the bar charts, I sized each image with one of the two methods shown in figure 11.

I created five images for each of the four embellishment shapes, plus one baseline bar chart as a control condition.

The study consisted of two sections, one asking relative questions, the other abso-

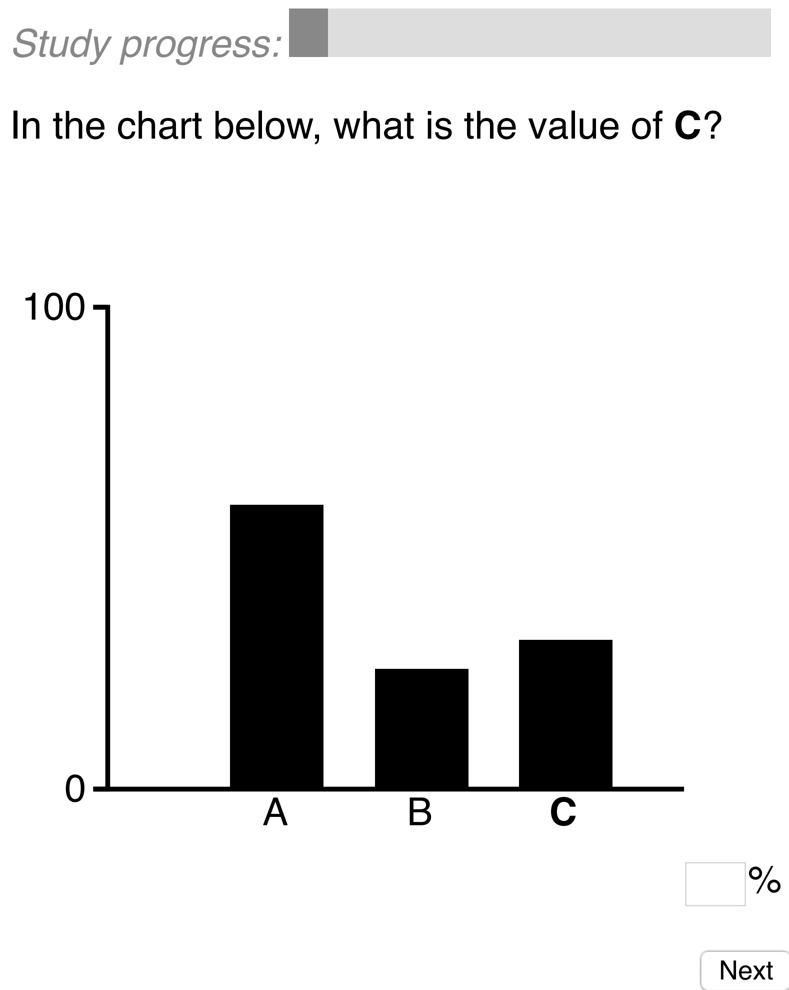


Figure 12: A screenshot of the study, showing an absolute question with the baseline chart.

lute. In each section, I used each individual image type twice, and also included the baseline bar chart four times. This yielded $4 \cdot 5 \cdot 2 + 4 = 44$ questions for each section, and 88 total.

- Baseline
- Rectangular (Figure 10a)
  - Pint glass – scaled
  - Book – scaled

  - Building – shifted
  - Toy block – shifted
  - Window – scaled
- Rounded tops (Figure 10b)
  - Baguette – shifted
  - Cactus – shifted
  - Finger – shifted
  - Pencil – shifted
  - Worm – shifted
- Triangular (Figure 10c)
  - Candy corn – scaled
  - Construction cone – scaled
  - Blueberry pie – scaled
  - Coniferous tree – scaled
  - Watermelon – scaled
- T-shaped (Figure 10d)
  - Jolly Roger flag – shifted
  - Lamp post – shifted
  - Mushroom – shifted

– Street sign – shifted

– Deciduous tree – shifted

2.3.2.2 Procedure

Participants began with a page introducing the study and a short demographic form. The next page provided instructions on how to answer the study questions, and briefly discussed the two part structure of the study. The first block of questions ended with an intermission page, giving participants an opportunity for a break before proceeding with the second half of the study.

I randomized the order of the blocks of question types, as well as the order of the embellishment shapes and images within each block to ensure there were no learning effects from one question, embellishment shape, or image to another. Like the previous study, I used a Latin Square to select the first embellishment shape a participant saw, ensuring equal coverage of initial embellishment shapes across participants.

I also avoided referring to the charts in the materials as bar charts, instead calling them simply charts, and referring to individual bars just by their labels. This allowed participants to come to their own conclusions on how to interpret the charts. This mimics the experience of most infographics, with no instructions provided to assist with the interpretation of the graphics presented.

2.3.2.3 Question Types

I adopted the question types used in my previous embellishment study and tested the accuracy of comparisons between bars and reading single bar values.

A. *In the chart below, what is the value of A?*

B. *In the chart below, what percentage is B of A?*

I chose to not address the subject matter of each chart's images with the questions. My goals were to test the communication accuracy, not other effects like comprehension or memorability.

Each question type was asked twice for each bar image, and four times for the baseline chart for a total of 44 questions in each section and a grand total of 88 questions in the study. I did not display a y-axis in the relative questions (question type B) to discourage participants from mathematically computing the percentage using the absolute heights of each bar.

2.3.2.4 Results

I used Amazon's Mechanical Turk to recruit 81 participants for the study. Each participant was paid US $2.50, with an average completion time of 16 minutes for an hourly rate of $9.07. Of those participants, 37 identified themselves as female and 44 identified themselves as male. There were seven in the 18-24 age range, 18 between 25-29, 33 between 30-39, 16 between 40-49, five between 50-59, and two older than 60. Education levels were also fairly wide ranging, with 30 who finished high school, 42 with a bachelor's degree, six with a master's, and three reporting as other. I eliminated two of the participant's results for having average error rates above 195%. Both eliminated participants were males with high school degrees. One was in the 18-24 age range, the other was in the 25-29 age range. This resulted in a total of 79 participants.

Despite my experiment being adapted from previous studies, I elected to abandon the log absolute error measure used by Cleveland and McGill. This measure can artificially make results look statistically significant. To report results, I use signed error (answer-correct) and absolute error (absolute value of error). Signed error is useful to gauge over- and underestimation, with positive values for overestimation and negative ones for underestimation. Absolute error is a better measure for precision, since it measures the distance from the correct value without the averaging-out effect between over- and underestimates the signed error suffers from.

2.3.2.5 Absolute Judgements

As in the previous study, I divided my analysis based on the absolute and relative questions.

39 participants answered the absolute questions first, while 40 saw them second. I observed no learning effects from seeing either question type first or second.

The answers for the absolute judgement questions, where participants were asked to estimate the value of one bar using the y-axis as a reference, generally led to underestimations (Figure 13, left column). Participants underestimated the values for all of the embellishment images, although their estimations were in line with the results from the shape-based embellishments from the previous study.

As in the previous study, there were significant differences observed in the mean absolute error between different embellishment shapes (Table 5), however these differences did not carry over to the different pictorial embellishments. There was no significant difference in mean absolute error between any of the images of a given embellishment type, and the means of each image were largely the same as the means

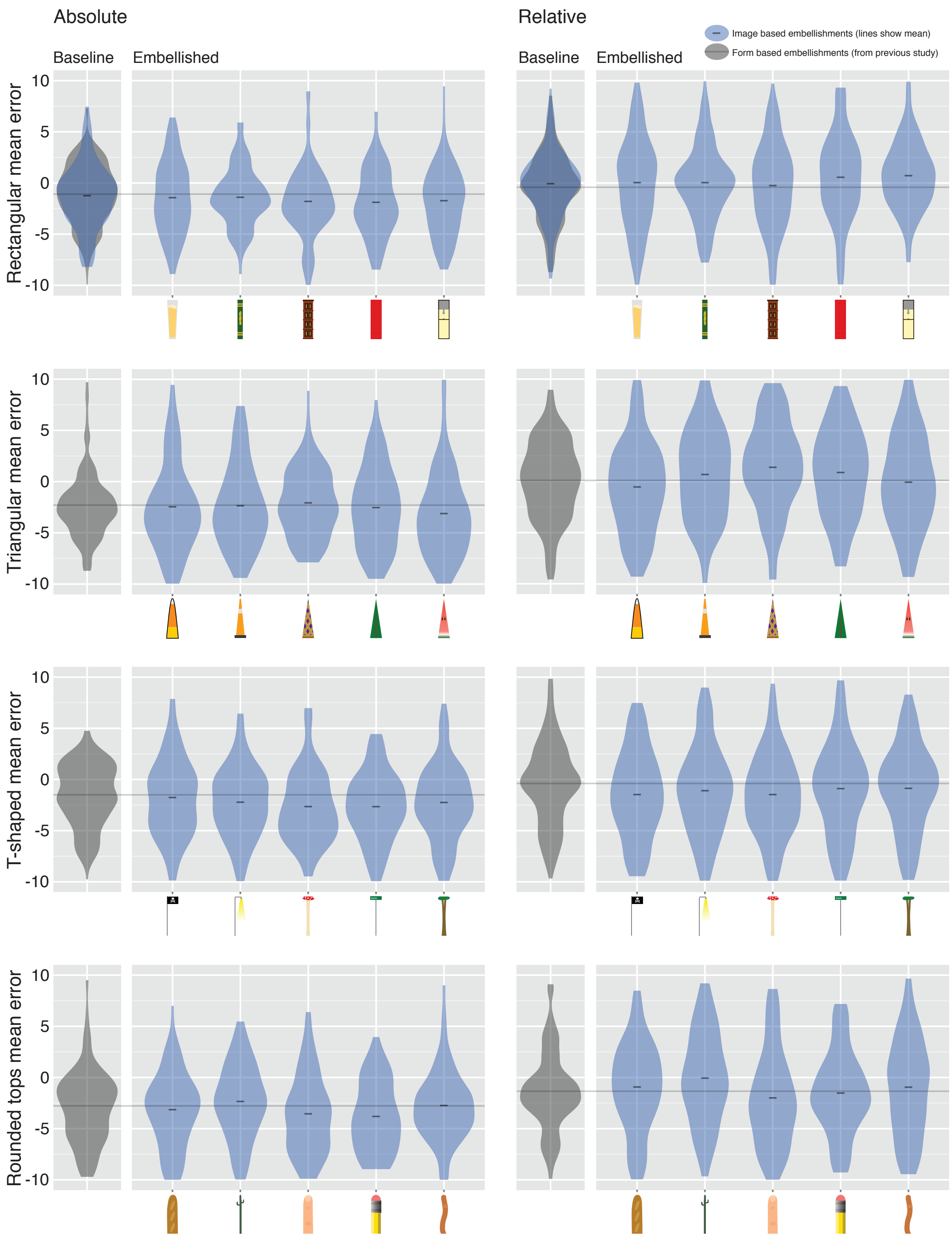


Figure 13: Violin plots of error for each embellishment, broken down by class. Blue shapes show results from the current study, gray shapes and the long gray lines provide comparisons with the previous study [34] for context.

Table 5: For absolute questions, means and 95% confidence intervals for absolute error by embellishment type (ANOVA: $F(4, 3471) = 2.861, p = 0.022$).

Embellishment	**Mean**	**95% CI**
Baseline	4.278	±0.472
Rectangular	4.678	±0.451
Triangular	6.863	±2.185
T-shaped	4.858	±0.361
Rounded tops	5.815	±0.473

for their embellishment shape.

I also found no significant difference in response times for any of the embellishment images, consistent with the previous study (ANOVA: $F(4, 3471) = 0.132, p = 0.971$). The different rendering methods (scaled vs. shifted, figure 11) for images also appeared to have no significant effect on absolute error or response time.

2.3.2.6 Relative Judgements

The answers for the relative judgement questions, where participants were asked to estimate the percentage one bar was of another, were not consistently under- or overestimated (Figure 13, right column). Mean absolute error, however, was higher than mean absolute error of the absolute judgements, in line with the results of the previous study. This confirms that this is a difficult task for people.

As with the absolute judgements and previous study, there were significant differences observed in the mean absolute error between different embellishment shapes (Table 6). Similarly, there were no significant differences in mean absolute error between embellishment images within each embellishment shape.

I did find significant effects on response times (Table 7). Baseline (unembellished) charts are the fastest, with rectangular, t-shaped, and triangular charts being slower,

Table 6: For relative questions, means and 95% confidence intervals for absolute error by embellishment type (ANOVA: $F(4, 3471) = 1.746, p = 0.137$).

Embellishment	**Mean**	**95% CI**
Baseline	5.539	±1.875
Rectangular	5.277	±0.608
Triangular	7.062	±1.364
T-shaped	6.506	±1.227
Rounded tops	6.341	±0.519

Table 7: For relative questions, means and 95% confidence intervals for response time in seconds by embellishment type (ANOVA: $F(4, 3471) = 2.569, p = 0.036$).

Embellishment	**Mean (s)**	**95% CI**
Baseline	6.51	±0.99
Rectangular	6.97	±0.75
Triangular	6.92	±1.03
T-shaped	6.81	±0.61
Rounded tops	7.03	±0.84

and bars with rounded tops the slowest. This is consistent with the previous study. Different rendering methods (scaled vs. shifted, figure 11) showed no significant effects.

2.3.2.7 Comparison with Previous Study

The gray shapes in Figure 26 provide context from the previous study. The data for rectangular bars allows direct comparison, since those were used in both studies. The means are virtually identical, and the shapes of the histograms are similar (especially for relative judgments). The long gray lines in the figure represent the means from the previous study's data for each bar shape.

While there are small differences between the pictorial embellishments in each shape class, none of them are statistically significantly different from the the solid shape. I performed ANOVAs for absolute error for each of the four classes and two question sets, and found none that were significant.

2.3.3 Discussion

While I found significant differences between the different shapes, confirming the effects found in the previous study, there appears to be no further significant effect from adding pictorial elements to them. The results by shape are in line with my previous work, and there are no differences between different designs within the classes.

As a result, I believe that the impact of pictorial elements on bar charts to be overstated. Pictorial elements have no discernible effect on the communication accuracy of basic rectangular bar charts. Other shapes are not impacted negatively beyond the effect that the shape alone already has. This suggests that the most salient features of bars in bar charts are the boundaries of the bars. This is further supported by the lack of differences in error caused by scaling vs. shifting the bars.

2.3.3.1 Recommendations

This finding leads to actionable advice for designers of bar charts in information graphics:

- Avoid bars without a strong horizontal mark indicating the top of the bar, but bars shaped like rectangles or Ts are okay.
- Ensure strong boundary contrast so the edges of the bar are clearly visible.
- Within the bounds of the bars, feel free to use any variety of colors, textures, and shapes, as this has little impact on the chart's ability to communicate

accurately.

This also provides some promising news for designers, as they should be able to reap the memory benefits found by Borkin [4] and Borgo [3] without the negative effects associated with embellished bar charts. There are minor impacts to response time for comparison tasks, however these are almost certainly trade-offs worth making. Creating eye-catching, memorable, visually interesting bar charts is possible, as long as a designer plans carefully.

2.3.3.2 Limitations and Future Work

This work provides us with a clearer understanding of some bar chart mechanisms, however there are still several gaps in our knowledge providing direction for future work. We do not know how shape embellishments impact non-vertical bar charts. Studies structured similarly to mine with different bar chart directions could shed light on this, and possibly provide overarching rules for designing embellished bar charts.

I also do not know how well differently embellished bars in a single chart can be compared. Studies looking at relative comparison tasks between differently embellished bars could help us understand the interplay between bars.

2.3.4 Conclusions

In this section I present a crowdsourced experiment to investigate the impact of pictorial chart embellishments on the accuracy of absolute and relative judgements in bar charts. The results of this experiment confirm findings from the previous study that bar chart shape embellishments do indeed have an impact on how well the data within the chart can be communicated. However, pictorial embellishments within a bar's shape have no impact on on how well the data within the chart can be communicated.

These results shed light on the visual cues our brains use to interpret bar charts. They further identify which visual components in bar charts are the most important to our perception of the data in the charts, and provide actionable advice for bar chart designers.

2.4 Bar Chart Conclusions

As designers continue to experiment with bar chart embellishments, there will undoubtedly be new types of embellishments, however this research provides some guiding principles. This type of research is critical to allowing designers the freedom to build and develop new chart designs, without fear of damaging the accuracy of the chart's communication. It also produces knowledge that can help viewers understand when a chart they are viewing may not be communicating accurately.

There are three primary takeaways from this research. Embellishments that cause conflicting encodings (area that doesn't scale proportionately with height for example), will harm the communication accuracy of the chart. In addition, this work suggests that vertical bar charts with strong horizontal tops are slightly better for absolute judgements, while bars without horizontal tops are worse for absolute and relative judgements. The testing of detailed embellishments reveals that the contents within bar boundaries have little impact, even with high levels of contrast. Instead, the bar boundaries themselves are the most critical component in the communication accuracy of the bar.

To facilitate reproducibility, I have made my materials and resulting data available on GitHub at `github.com/dwskau/bar-chart-embellishment`.

CHAPTER 3: PIE AND DONUT CHARTS

Pie and donut charts have a complex position in the visualization community. They are often chided as being bad charts, and generally, the community discourages their use. Yet, they do have merit as a chart type. They fit in a relatively small amount of space, have a pleasing circular shape that can help draw the eye to them, and they use a clearly understood metaphor to represent a part to whole relationship.

Despite these redeeming factors, the general understanding of pie charts is that they are bad, and we read them by angle. Donuts are considered worse because we can't see the central angle. Unfortunately, these "understandings" are derived from work done in 1926 [14] using self-reported findings.

William Playfair is usually credited with the invention of the pie chart, with his *Statistical Breviary* [29] published in 1801 being the first known use of this chart type. The chart took off, with Brinton complaining in 1914 about its use as a popular display for data [5]. Today, they are simple to produce using virtually all charting and data analysis tools. Elementary school students are taught how to read and draw them, and they even appear in popular culture, making them a part of the public consciousness.

Despite their prevalence, the data visualization community has been reluctant to fully embrace the pie chart. This is not without reason, as several studies have shown that bar charts are better for many tasks relating to comparison of values with part whole relationships, however pie charts are still commonly used.

The community's resistance to the pie and donut chart is perhaps partially responsible for a lack of research into the underlying mechanisms that drive how they work. My research does not attempt to establish the legitimacy of using pie and donut charts; others have done this work and have come to differing conclusions, some promoting their use, and others renouncing them. Regardless of their conclusions, pie and donut charts continue to be used to communicate data. Instead, my research attempts to understand how pie and donut charts work. I do that through three studies.

The first study looks at the individual visual variables used in pie and donut charts: arc-length, area, and angle (Section 3.2). The next study examines the ratio of inner to outer radius in donut charts (Section 3.3). These two studies were published together in *Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts* [35]. The third study tests common chart embellishments against a baseline pie chart (Section 3.4), and was published as a short paper *Judgement Error in Pie Chart Variations* [20].

3.1 Individual Data Encodings

Pie and donut charts are prevalent in all forms of communication with data, in particular when used as part of information graphics. In a random sampling of infographics on visual content website Visually [40], 36% of infographics with charts used some form of pie or donut chart. Information designers are experimenting with variations such as exploded charts, varying radius charts, icons broken into radial

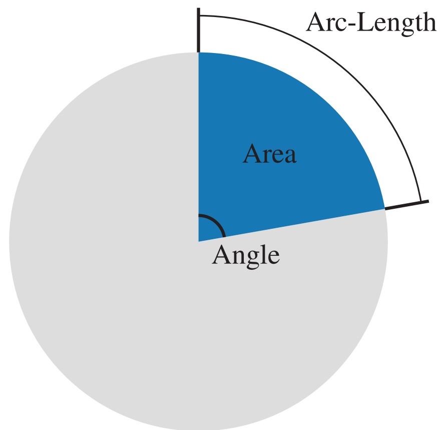


Figure 14: The three different encodings representing data in a pie or donut chart: central angle, wedge area, and arc length.

segments, nested donuts, etc. (Figure 15)

While angles are often mentioned when discussing pie and donut charts, there are three visual variables that encode data: the angle, the area of the circle wedge, and the length of the segment on the circle (Figure 14). Which of these encodings do people read, and how important is their combination? Which can be left out without doing damage to accuracy?

To answer these questions, I designed a study to separate the three visual cues and compare how well each of them would do on its own (Section 3.2). Based on this, I then designed a second study to measure the difference between pie and donut charts and the impact of the size of the donut hole (Section 3.3). Both studies point to angle being less important than arc and area.

3.1.1 Reading Accuracy with Pie Charts

Most research about pie charts looks to compare them to other chart types, primarily bar charts of varying configurations. This research has a long history, with some of the early work having taken place in the 1920s. Eells compared pie charts to stacked or "divided" bar charts and found that pie charts are more effective at helping the viewer determine the percentage of the whole [14]. In response, Croxton and Stryker performed a study to settle the beginning dispute over the chart type, but ended up with a set of recommendations that varied by the number of pie slices, the values shown, etc. [11].

Cleveland and McGill's seminal work on graphical perception [8] addresses the effectiveness of different chart types, including pie charts, for different tasks. Despite referencing Cox's call for a theory of graphical methods [10], and Kruskal's observation of the lack of theory or systematic body of experiment [24], Cleveland and McGill stop at the evaluation of the charts' suitability for tasks and do not delve into the perceptual factors of the charts themselves. Cleveland also argues [7] that pie charts are inferior for many common tasks because of *degraded pattern perception*, but does not provide a deeper explanation.

Simkin and Hastie [33] showed that pie and bar charts are equivalently suited for tasks involving estimation of the proportion of part to whole. Their work builds on Cleveland and McGill's, helping to establish the relative communication abilities for certain concepts of pie and bar charts, however it still does not look at the systems contributing to those communication abilities. Spence and Lewandowsky [36] also compared pie charts to bar charts and tables, this time using everyday tasks rather than simple magnitude judgements. They determined that pie and bar charts are definitely superior to tables, however their work still only compared chart types rather than exploring the charts themselves.

Very little work has dealt with donut charts. One study found no difference in precision between them and other charts [22], though this was just a minor part of a larger study looking at various charts. They did find that people's confidence in their answers were higher for donut than for pie charts, however.

3.1.2 Perceptual Mechanism

One model of creating visualizations based on Wilkinson's *Grammar of Graphics* [41] argues that pie charts are stacked bar charts transformed into polar coordinates. Wilkinson does not claim that this is how they are actually read, but this view would suggest that the length along the outside arc is what people are looking at, not the angle in the center (Figure 14).

Most sources do not make explicit claims as to the way we read pie charts, and when they do they do not base them on research. Brinton [5] implies that "circles with sectors" ought to be read by angle, but may mistakenly be read by area when images are inserted into the pie wedges. Bertin also claims angles as the main mechanism [2], Robbins mentions angle judgments when reading pie charts [31], and Munzner classifies pie charts as using the angle channel [27].

The only study directly addressing the perceptual mechanism underlying the reading of pie charts I am aware of is Eells' 1926 paper [14]. He lists the methods his study participants indicated as the ones they used "exclusively or predominantly": 51% reported using arc length, 25% area, 23% angle, and 1% chord length. Chord length was the mechanism claimed by earlier work, which Eells clearly disproved. Kosslyn also points out the possible systematic underestimation of area in regards

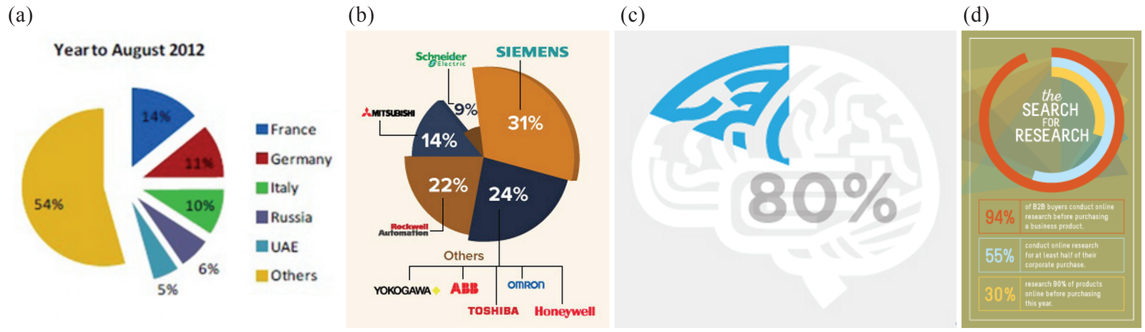


Figure 15: A sampling of pie and donut charts used in infographics, taken from examples found on Visually [40]. (a) exploded pie chart, (b) chart with varying segment radii, (c) pie chart constructed with an icon, and (d) nested donut chart.

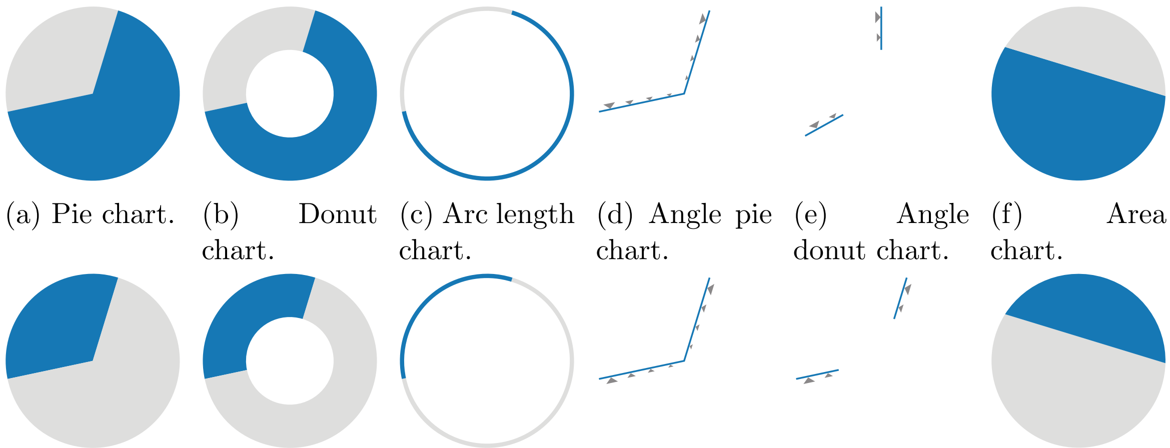
to pie charts, however the studies backing this up were not done in context of pie charts [23].

3.1.3 Use in Information Graphics

Pie and donut charts are very common in infographics, and are often modified from their canonical forms. I hope to use the results of this study to make recommendations about which of these are likely problematic, and which are probably no harder to read than regular pie charts.

Exploded pie charts (Figure 15a) don't directly violate any encodings, as all are individually present, however the arcs are no longer continuous. Varying radius pie charts (Figure 15b) alter arc length and area encodings, though angle is maintained. Charts constructed with icons (Figure 15c) often don't have accurate arc length or area encodings of the data. Nested donut charts (Figure 15d) make arc length harder to compare between the layers.

Depending on the significance of each encoding in the communication of data, these modifications to pie and donut charts may cause them to be significantly less



(a) Pie chart. (b) Donut chart. (c) Arc length chart. (d) Angle pie chart. (e) Angle donut chart. (f) Area chart.

(g) Pie chart. (h) Donut chart. (i) Arc length chart. (j) Angle pie chart. (k) Angle donut chart. (l) Area chart.

Figure 16: A sampling of charts used in the study of pie and donut chart encodings. The top row all represent 67%, while the bottom row all represent 33%.

effective. Indeed, my own work based on the results reported in section 3.4 shows that exploded pie charts lead to higher error, as do larger radii – the latter cause systematic overestimation of the value. Shapes other than circles also predictably lead to more error.

3.2 Arcs, Angles, and Areas Study

In order to test the contribution of each visual encoding, I designed new charts that would allow me to isolate each visual variable as much as possible (Figure 16). This let me test the accuracy of arc length, angle, and area independently of their counterpart encodings.

I hypothesized that the baseline charts would be more accurately interpreted than any of the individual encodings, and the baseline pie chart would be the most accurately interpreted of all chart types. Of the individual encodings, I expected the chart type displaying arc length to perform the best because it is most similar to an extremely thin donut chart. I expected the angle chart for pies to be the next best performer, and then the angle chart for donuts.

However, I did not expect the individual encodings to perform much worse than the baseline charts. The rationale for this was that people presumably use a single cue to read a chart, rather than averaging from multiple cues.

3.2.1 Materials

The design of the test charts is key to being able to independently test the data encodings. Every chart has two segments. Pie charts "in the wild" often have more divisions, but I chose to constrain the study stimuli to two parts to avoid complicating the task. In all of the charts, the blue portion is the segment referenced in the question (Figure 16). The rest of the charts are light gray, so the blue is the only color (and also darker than the gray), providing a clear focus and reducing distractions.

The study uses six different chart types (Figure 16):

- **Baseline Pie** – a standard pie chart (Figures 16a and g) using all three visual cues to represent the number.

- **Baseline Donut** – a standard donut chart (Figures 16b and h) using area and segment length to encode data. The angle is much more difficult to read due to the missing center where the lines would meet.

- **Arc** – a chart showing only arc length (Figures 16c and i), without area or angle.

- **Angle Pie** – a chart showing the angle component of a pie chart (Figures 16d and j) without a filled area or circle segment, thus removing these two cues. Little arrows point towards the part of the full circle that encodes the value from the outside.

- **Angle Donut** – a chart showing the angle component of a donut chart (Figures 16e and k), though without the lines meeting in the center that presumably allow precise judgment of the angle. Area and segment length are not represented.

- **Area Chart** – a chart using only area to represent a percentage value (Figures 16f and l). The area representing the data "fills up" proportionally as the value increases, thus removing angle cues and only providing very non-linear segment length.

Segment placement on all charts is randomized through a rotation of the entire chart to reduce the occurrence of segment edges that line up with quadrant points. This prevents participants from being able to use the natural quadrant points to gauge segment sizes.

My experiment design is adapted from Cleveland and McGill [8] and Heer and Bostock's replication of their study on Mechanical Turk [17].

Through a pilot study, I discovered some potential issues with an earlier set of chart designs. I originally used a red dot outside the chart to indicate the focus area without interfering with the chart. This produced high error rates and caused confusion for the angle charts. In the pilot, it seemed that many participants answered for the opposite side of the angle charts. For example, if I asked about a portion that was

25%, their answer would be close to 75%. I was also concerned that the dot would make it easier to mentally complete the area or arc between the angle indicators in the angle-only condition, thus skewing the results. Using only color, I was able to point out the element of interest without adding extraneous objects. The angle-only condition is the only exception, but even then I kept the additional clues outside of the indicated angle.

Producing an angle-only condition requires extra marks in order to indicate to the participant which side of the angle is being asked about. I considered changing the question language to reference the side of the angle by its relationship to 180° (greater than or less than), however this provides non-visual information about the angle, and could confound results by introducing the concept of degrees.

The angle-only condition led to more opposite answers (about 10%) than the others (about 3%). The percentage is still relatively small though, and I accounted for most of the resulting error by flipping the answers for a number of users.

3.2.2 Procedure

The study consisted of six sections:

- Introduction page and brief demographic survey
- Pre-study questions about which encoding people thought they used to read pie and donut charts
- Tutorial on how to read the more unusual chart variations
- Main part of the study asking for the values encoded in 48 different charts

- Post-study questions about encodings used, same as in the pre-study part

- Short debrief

3.2.2.1 Introduction, Pre-Study, Tutorial

The study began with an introduction page followed by a short demographic form to collect education level, gender, age range, and physical monitor size. Every page after the intro page had a next button to advance to the subsequent page, with no controls provided to go back.

The first segment of the study included six questions broken up into two groups of three, one focused on pie charts, the other on donut charts (counterbalanced so some participants saw the donut chart questions first, others the pie chart). Each group of questions began by asking the standard question for the study, "*What percentage of the whole is indicated below?*", twice. The third question in each group asked the participant which encoding they thought they were using to come up with their answer, using a diagram similar to Figure 14.

A twelve-page tutorial section followed the first segment of the study. The tutorial explained each chart type, and asked two sample questions for each, with the answer shown as a hint. Participants had to enter that number in the response field to advance to the next page.

3.2.2.2 Main Section

After the demographic form and tutorial, the main part of the study began. Each chart type was tested 8 times adding up to a total of $6 \times 8 = 48$ questions for each participant. A progress bar at the top of the page showed their progression through the study questions. After completing all questions in the body of the study, the first segment of the study asking about the individual encodings was repeated. This was done to see if participants would change their answers after having answered many more questions.

In the body of the study, charts were shown to participants in random order, however each chart type was shown eight times per participant. The data in the charts was from a pre-selected array of random integers with a possible range from 3 to 97, the same for every participant. The array was shuffled randomly for each participant, making any combination between data and chart type possible.

I asked the same question for every chart: "*What percentage of the whole is indicated below?*" Some of my chart variants made this relationship clearer than others. For example, the arc and area charts clearly have a part and a whole indicated by the blue segment and the gray segment, but the angle charts don't provide a good indicator of the whole. By keeping the question consistent and providing the brief tutorial at the beginning, I hoped to avoid confusion when participants encountered the more unusual charts.

3.2.2.3 Post-Study and Debrief

The study ended with a debriefing page explaining what the study was exploring and providing an optional free response form for feedback and comments.

For this experiment, I recruited 102 participants through Amazon's *Mechanical Turk* platform. I eliminated answers from two participants who did not complete the study. Subjects took an average of 25 minutes and 7 seconds to complete the

Table 8: Means and confidence intervals for log error by chart type (ANOVA: $F(5, 4650) = 121.955, p < 0.001$).

Chart	Mean	95% CI
Baseline Pie	1.032	±0.138
Baseline Donut	1.000	±0.137
Arc	1.294	±0.128
Angle Pie	1.967	±0.167
Angle Donut	2.279	±0.157
Area	1.306	±0.125

study from start to finish, including the introduction and demographic form and a debriefing page with optional free response feedback. They were paid US$3.00 each for their participation, resulting in an average hourly rate of US$7.20.

3.2.3 Results

For the analysis, I edited one outlier value where a participant had entered 7068 and left a note about correcting this in the feedback section (I changed it to 70). I eliminated answers from three participants based on comments they left in the feedback form, which indicated that they had misunderstood the study or made major mistakes.

Five participants had answers that were wildly inaccurate, with three of them apparently answering in degrees instead of percentages. I omitted their answers from my analysis, leaving me with 92 participants: 43 female and 49 male, with the majority in the 25–29 and 30–39 age ranges.

Just as in my pilot study, pie and donut angle charts had issues with participants answering for the opposite segment in the chart. The occurrences of this were reduced from the pilot study, however it still happened often enough to merit correction. I measured the distance between the answer given and the value represented by the two

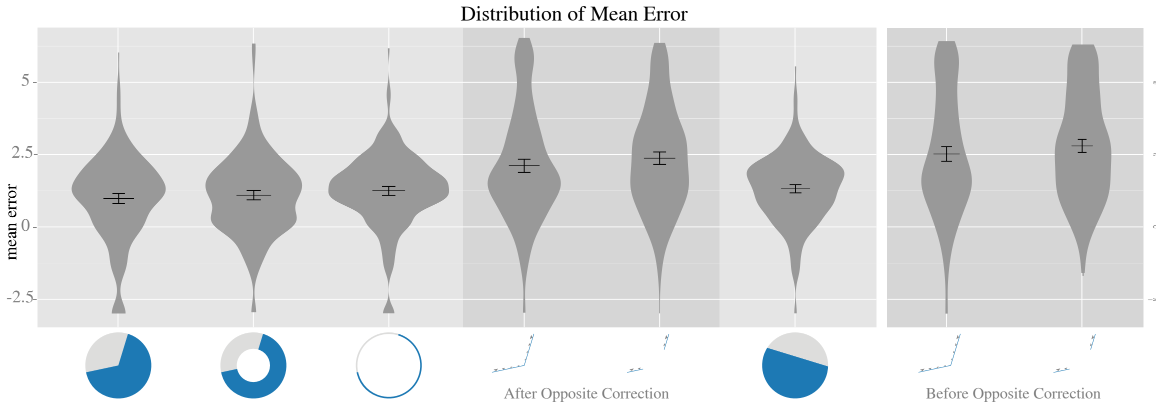


Figure 17: The distribution of amount of error per chart type after correcting for opposite answers on angle charts (uncorrected on far right). The error bars show 95% CI and the middle black lines represent the mean for each violin plot.

segments in each angle chart. When over half of a participant's answers were closer to the opposite angle, I subtracted all of their answers from 100 to get their estimate for the opposite segment. I ended up doing this for 16 participants. The discussion below is based on the corrected results. I show both for the angle charts in Figure 17 (corrected in the main part, uncorrected on the far right).

For consistency with other studies [17], I report the log absolute error:

$$log_2(|\text{judgedvalue} - \text{truevalue}| + \tfrac{1}{8})$$

3.2.3.1 Accuracy by Chart Type and Value

Means and 95% confidence intervals for log absolute error are reported in Table 8, violin plots of the same data are shown in Figure 17. I find these plots to be more informative than pure p-values, though I also report those in the table captions. Violin plots show the distribution of error better than box plots and others [9].

Error was smaller for the baseline charts, area chart, and the arc chart than the two angle-only charts. This was not what I hypothesized, and contradicts common

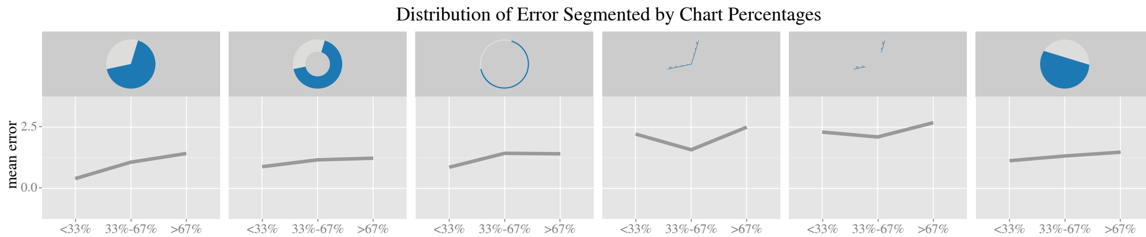


Figure 18: The distribution of error segmented by the percentage amount shown in each chart. All charts except the angle charts show an increase in error as the percentage shown in the chart increases.

wisdom that angles are critical to pie and donut chart perception.

Interestingly, the baseline donut chart had a slightly lower log error than the baseline pie chart, but well within the 95% confidence interval (virtually identical between the two).

The distribution of mean log error per participant in Figure 17 clearly shows the differences between the two angle charts and the other chart types. The relatively tall and skinny violin plots show a high degree of variance in the amount of error for the angle charts, while the other charts have relatively tight groupings, showing a consistent level of error. The arc-length chart has the tightest grouping of error, so despite a higher mean error, the error amount is more predictable.

The unusual area-only chart has very similar error to the pie and donut. This is remarkable, given how difficult it generally is to correctly estimate area, and also the chart's lack of familiarity.

I found that the size of the percentage shown in a chart also has an impact on participants' ability to interpret the chart (Figure 18). All except the two angle-only charts show more error with larger segments. The two angle-only charts have a v-shape that has lower error for the middle third percentage values.

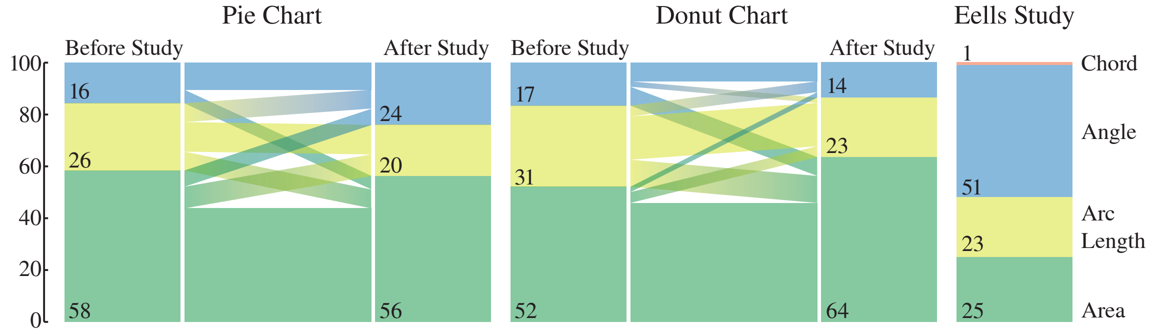


Figure 19: At the beginning and end of the study, participants were asked about the encoding they primarily used to interpret pie and donut charts. These are compared with self-reported answers from an earlier study [14].

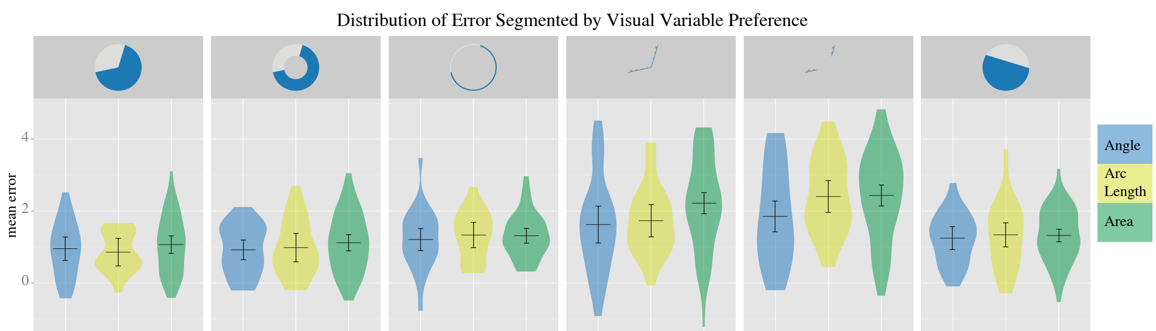


Figure 20: The distribution of error segmented by the second self-reported encoding preference for pie charts. People using angle did better in the angle-only condition than ones who reported using area or arc length. Black lines represent the means for each encoding preference per chart type, error bars show 95% confidence intervals.

3.2.3.2 Accuracy by Self-Reported Main Visual Cue

At the beginning and end of the study, participants were asked to report which encoding they were primarily using to read pie and donut charts. The exact question was, *In the previous two charts, what did you primarily use to estimate the percentage?* Interestingly, my study had far more answers for area, while Eells [14] found more people reporting the use of angle (Figure 19).

Mean log error per participant for each chart type segmented by their answers to the second self-reported encoding question suggests that there may be individual differences in people's ability to read angle, but the area and arc-length encodings help to reduce these effects (Figure 20). People who reported angle as their primary visual cue had lower mean error on the angle charts than people who reported arc-length or area, however they performed about the same for the other chart types. This suggests that people who believe they are reading angles may use them to interpret pie and donut charts, however people who believe they are reading arc-length or area are equally accurate with their preferred encodings. People who primarily use angles are able to use arc-length or area equally well for the charts where angle is not present. The mean error per person segmented by their primary visual cues are all the same, showing that arc-length and area are equivalent variablea for reading charts. In summary, angle encodings work well for some people, but arc-length and area work for all.

3.2.3.3 Demographics and Quadrant Alignment

I examined the data broken down by the demographic information provided, and found the expected effects of gender (males doing slightly better than females, also found by Eells), and age group (accuracy decreases slightly with age), but no discernible impact of highest degree completed.

The study was built to reduce the number of charts that would align with quadrants, however 300 (about 6.7%) of the charts did align on one of the quadrant edges due to random chance. I found no effect of this alignment on the results.

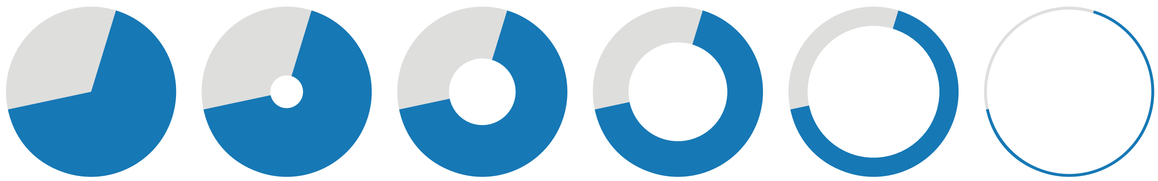
3.2.4 Discussion

My results cast doubt on the importance of angle: angle-only charts for pies and donuts performed considerably worse than the rest. The possible impact of the chart design on the angle results does make it difficult to know whether the differences in their performance derive from the chart design or the encoding itself. This suggests that angle cannot be the only way I read a pie or donut chart. At least one of the other encodings is necessary to be able to interpret the angle encoding in a chart. I found that donuts are likely no worse than pies, despite missing the center. This suggests that area and arc length can make up for the missing angle information. While arc length and area alone are better than angle alone, they are still worse than complete pie and donut charts.

Taken together, my results allow me to establish an ordering in terms of accuracy (with $\approx$ meaning "no different"):

- *baseline donut* $\approx$ *baseline pie*
- *arc* $\approx$ *area*
- *angle pie*
- *angle donut*

Encodings do not seem to combine in an additive manner. Instead, they appear to work together to substitute for the missing encoding when one is absent (as in the donut chart). Angle appears to contribute the least to the accuracy of the chart's



(a) Baseline pie chart; inner radius 0% of outer radius.

(b) Donut chart; inner radius approximately 20% of outer radius.

(c) Donut chart; inner radius approximately 40% of outer radius.

(d) Donut chart; inner radius approximately 60% of outer radius.

(e) Donut chart; inner radius approximately 80% of outer radius.

(f) Donut chart; inner radius 97% of outer radius.

Figure 21: The six inner radii tested in the second study, from a filled pie chart with no hole to a thin outline.

communication. Arc length has a greater impact on the communicative value of the chart, however it still does not match all three encodings combined.

3.3 Donut Radii Study

The results of the first study suggest that angle has a minimal contribution to our ability to perceive pie and donut charts. But within donut charts, does the size of the hole in the center make a difference? It should if angle is important, since any hole removes the most salient portion of the angle encoding: the center where the lines meet. Arc length is still present, as is area unless the donut gets extremely thin.

I therefore ran a study varying the inner radius of the charts from zero (i.e., a pie chart) to the point where only a thin outline was left. My hypothesis was that the different inner radii would show no difference in how accurately they were interpreted. Based on the previous study's results, I expected the thinnest donut chart to have somewhat worse performance because of the higher error for the pure arc length compared to the donut tested there, but was unsure at which point accuracy would start to degrade.

3.3.1 Materials

I chose a set of six inner radii to ensure good coverage of the range of possible donut designs:

- 0% – a pie chart (Figure 21a)
- 20% – a small hole in the center (Figure 21b)
- 40% – a medium hole in the center (Figure 21c)
- 60% – a thick circle outline (Figure 21d)
- 80% – a thin circle outline (Figure 21e)
- 97% – a very thin circle outline (Figure 21f)

Throughout the study, the inner radius randomly varied among the six different sizes, however each size was tested ten times per participant. Just as in the first study, I chose to limit the charts to two segments to reduce distractions and focus the participant. The blue segments indicated the portion being asked about, while the gray segments indicated the rest of the whole.

3.3.2 Procedure

The structure of this study was similar to that of the first one. It was also posted on Mechanical Turk and ran entirely in participants' web browsers.

Like the previous study, it began with an introduction page followed by a short demographic form collecting education level, gender, age range, and physical monitor size.

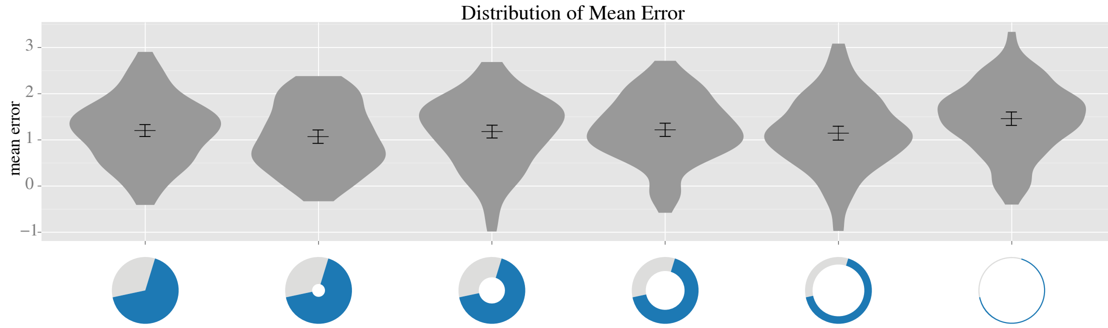


Figure 22: The distribution of amount of error per radius size. The error bars show 95% CI and the middle black lines represent the mean for each radius.

This study did not include any tutorial, however, instead jumping straight into the chart questions. Each inner radius size was tested 10 times adding up to a total of 60 chart questions for each participant. The data in the charts came from the same pre-selected array of 60 random integers as in the previous study. The array was shuffled randomly for each participant, making any combination between data and radius size possible. Every chart was rotated at a random angle to reduce quadrant effects that make values at 25%, 50%, and 75% easier to perceive.

Just as in the first study, a progress bar at the top of the page showed their progression through the study questions. This study also ended with a debriefing page explaining what the study was exploring and providing an optional free response form for feedback and comments.

3.3.3 Results

Out of 117 recruited participants, 96 fully completed the survey in an average time of 15 minutes and 42 seconds. These 96 participants were compensated $2.00 each for an average of $7.85 per hour.

One participant appeared to answer in degrees rather than percent, so I discarded their responses. Two others had average log absolute error above 3.00 (the next highest was 2.27), which is why I also omitted their data. This left me with 59 male participants and 34 female participants, with the majority in the 25–29 and 30–39 age ranges.

I again use the log absolute error to report results. Figure 22 and Table 9 show that the distribution of log absolute error values across all inner radius sizes was very similar. All donut thicknesses performed relatively similarly with the exception of the thinnest donut which performed worse.

As in the first study, demographics had an effect on this study: males perform significantly better across all inner radius sizes, and increase in error correlates with an increase in age.

3.3.4 Discussion

The results confirm that angle encoding, especially the center meeting point of the angle, is not contributing significantly to our ability to perceive pie and donut charts accurately. The lack of difference between the pie chart and all but the thinnest donut is also consistent with the first study.

The ratio of inner radius to outer radius on donut charts does not have a significant impact on the communication accuracy of a chart (with the exception of the thinnest one, which is somewhat less accurate). Although area encodings are technically preserved for all of the charts tested in this study, it is hard to imagine that the thinner donuts are being perceived using area, suggesting that arc-length may be the most important encoding in pie and donut charts.

Table 9: Means and confidence intervals for log error by inner radius size (ANOVA: $F(5, 5754) = 4.37, p < 0.001$).

Inner Radius	**Mean**	**95% CI**
0%	1.327	±0.119
20%	1.162	±0.130
40%	1.289	±0.125
60%	1.333	±0.114
80%	1.257	±0.128
97%	1.553	±0.116

3.4 Pie Chart Variations Study

Pie charts are a common feature in information graphics (infographics). Not content with regular pie charts, designers often modify them by changing their shapes, moving slices apart, or enlarging slices to emphasize them (Figure 23).

In my *Arcs, Angles, or Areas* study (Section 3.2), I found that contrary to common assumptions, central angle is likely not the primary way people read pie charts [35]. Area plays a significant role, and arc length may be involved as well (in particular when reading donut charts, which I found to perform no worse than pie charts). This leads me to predictions of the effect of pie chart design variations.

Based on these predictions, I designed a study that directly investigates four common variations of pie charts that are often used in infographics. The goal was to shed further light on the underlying mechanism that people use when reading pie charts. If they used central angle, their responses would be affected differently by these design choices than if they took area and/or arc length into account.

I also wanted to directly assess the impact of these design decisions on the readability of these charts, since they are intended to communicate data. If the ways they are rendered cause errors in the way people read them, they do not actually serve

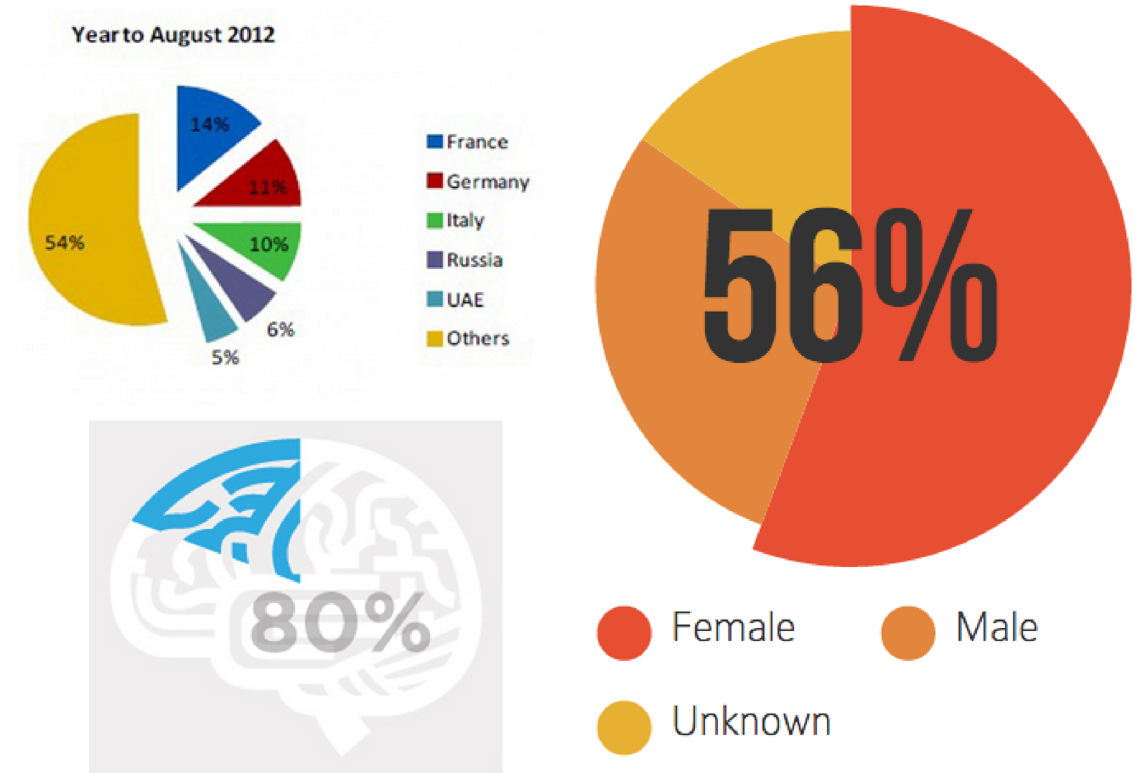


Figure 23: Examples of pie chart variations from infographic repository Visually [40]: exploded pie chart, chart with a larger slice, and chart with a non-circular shape. These directly inspired the designs of the materials for the study reported here.

their purpose. Infographic designers currently don't have much research on which they can base their designs.

Of the four pie chart variations in the study (Figure 24), three change the relationship between angle, area, and arc length. The only one that does not is the exploded pie chart.

3.4.1 Materials and Procedure

The results of my previous studies (Sections 3.2, and 3.3) gave me some insight into the effects of common pie chart variations. They were focused on decomposing pie charts into their visual cues. For this study, I modeled a number of simple charts

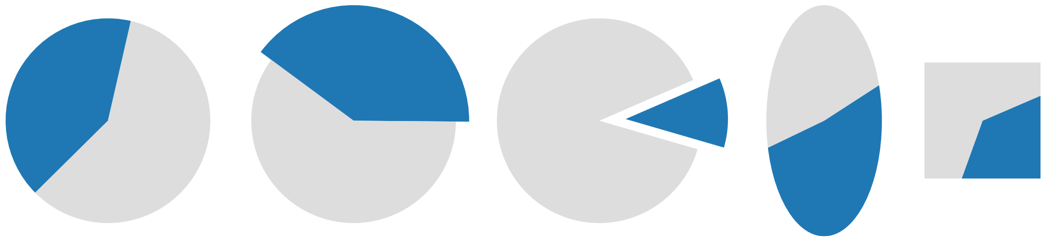


Figure 24: I tested four variations on the basic pie chart to measure their effect on error in reading. Left to right: base pie chart, chart with larger slice, exploded pie, elliptical pie, and square pie.

on the most common design choices and distortions I found in infographics.

The set of stimuli presented to study participants consisted of the following pie charts and variations (Figure 24):

- *Baseline circular pie chart* with two slices: one gray, the other blue.
- *Larger slice chart*, where the blue slice had a larger radius than the gray one, making the blue slice stick out. The larger radius led to a larger area and arc length.
- *"Exploded" pie chart* with the blue slice moved away from the center. This does not change the angle, area, or arc length.
- *Elliptical chart*, compressed horizontally into an ellipse. This was done by masking out the unneeded parts of a full circle. I used a vertical ellipse rather than a horizontal one to minimize it being read as a 3D pie chart. The ellipse strongly distorts area and arc length, but not angle.
- *Square "pie chart,"* created by cutting a square out of the basic pie chart. Just like the ellipse, this has a nonlinear effect on area and arc length.

Table 10: Means and 95% confidence intervals for log error by pie chart variation.

Chart Variation	**Mean Log Error**	**95% CI**
Baseline Pie	1.151	±0.098
Exploded	1.236	±0.101
Larger Slice	1.338	±0.096
Square	1.487	±0.097
Ellipse	1.570	±0.097

The study procedure was almost identical to the previous study. Participants were recruited on Amazon Mechanical Turk. They were first asked a set of basic demographic questions (10-year age group, gender, highest degree obtained), then they saw a brief description of the study with an image that showed them examples of all the different types of pie variations to expect in the study.

The body of the study showed them one chart at a time and asked them to estimate the percentage shown by the blue (darker) slice as a whole number. I used the same set of numbers as in the previous study; their values varied from 3% to 97%. The order of values and variations shown was randomized for each participant. Each chart was rotated at a random angle to avoid effects based on slice edges being aligned with major axes.

3.4.2 Predictions and Hypotheses

In a regular pie chart, area and arc length increase linearly with angle (both are fractions of the entire circle). When the shape is distorted or one section is larger, that relationship is more complicated. I calculated arc length and area as a function of central angle for the five variations I studied. Figure 25 shows them as a multiple of the baseline pie chart (which is identical with the exploded pie chart).

The larger slice is a simple multiple, determined by the larger radius (175 vs. 155

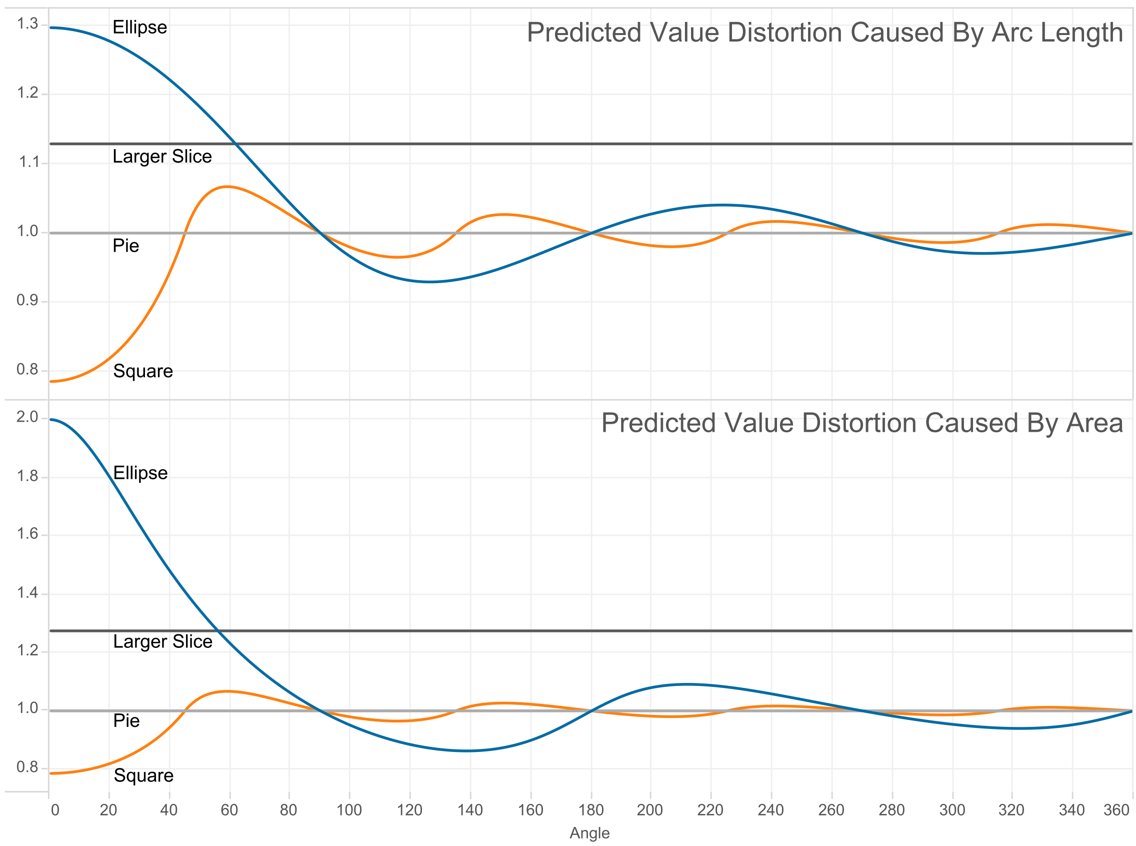


Figure 25: The effect of arc length (top) and area (bottom) as a function of central angle, by variation type. The pie chart serves as a base line, the other values are expressed as multiples.

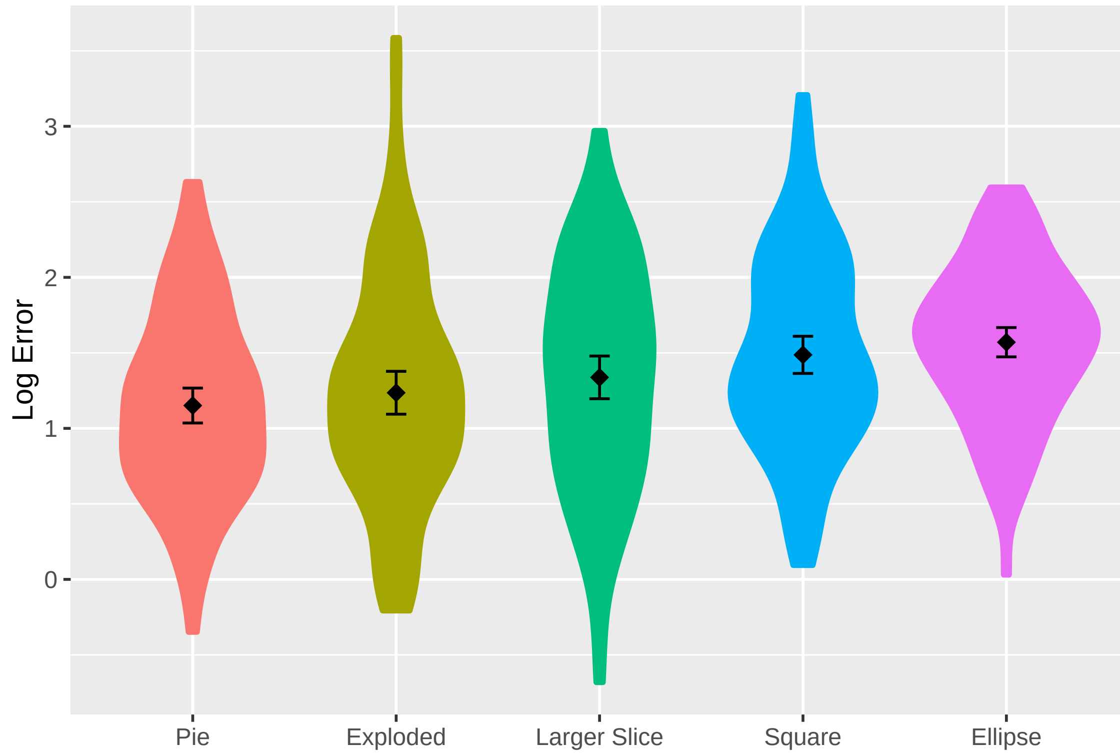


Figure 26: Mean log error (dot) and distribution of log error of responses by chart type. Error bars show 95% confidence intervals.

pixels). The square's area was determined by adding up 45° increments, and then computing the final fraction's area as a right-angle triangle. Rotation was taken into account by calculating the area for the central angle plus the rotation (measured from the positive x axis) and then subtracting the area covered by the rotation angle. Similarly, the area of the elliptical chart was determined by adding up quarters of the ellipse and then adding the final part using the formula for ellipse slice area within a quarter ellipse.

Arc length was determined in a similar way. For the square, the length of the blue border was determined by adding up 45° sections (corresponding to half the length of a side) and then adding the fraction determined by the angle within the last octant. Arc length for the ellipse cannot be determined with a simple formula, and instead was computed using numeric integration of the path integral along the ellipse.

Based on these computations, I expected the following as compared to a basic pie chart:

- A chart with a larger slice should lead to systematic overestimation of the value, since the area of the slice is larger in relation to the rest of the pie than the percentage and central angle.
- For the exploded pie chart, I did not expect a difference, since the central angle is still as readable and there is no distortion of area or arc length.
- The ellipse distorts area and arc length, and presents more complex shapes. I therefore expected it to yield considerably higher error than the circular charts.
- The square is an unusual shape to use for a pie chart, and it also leads to a complex relationship between area, arc length, and angle, resulting in more error.

3.4.3 Results

Of the 108 participants, exactly half were female. The predominant age group (43%) was 30–39 years old. High school or Bachelor's degree were reported as the highest degree by 44% each, with only 12% reporting a Master's or higher.

Participants completed the study in an average time of just over 11 minutes. They were paid $2 to participate, resulting in an extrapolated hourly rate of about $10.90.

I removed one participant's data from the analysis, since he apparently responded in degrees rather than percent. As in the previous study, I found a number of responses

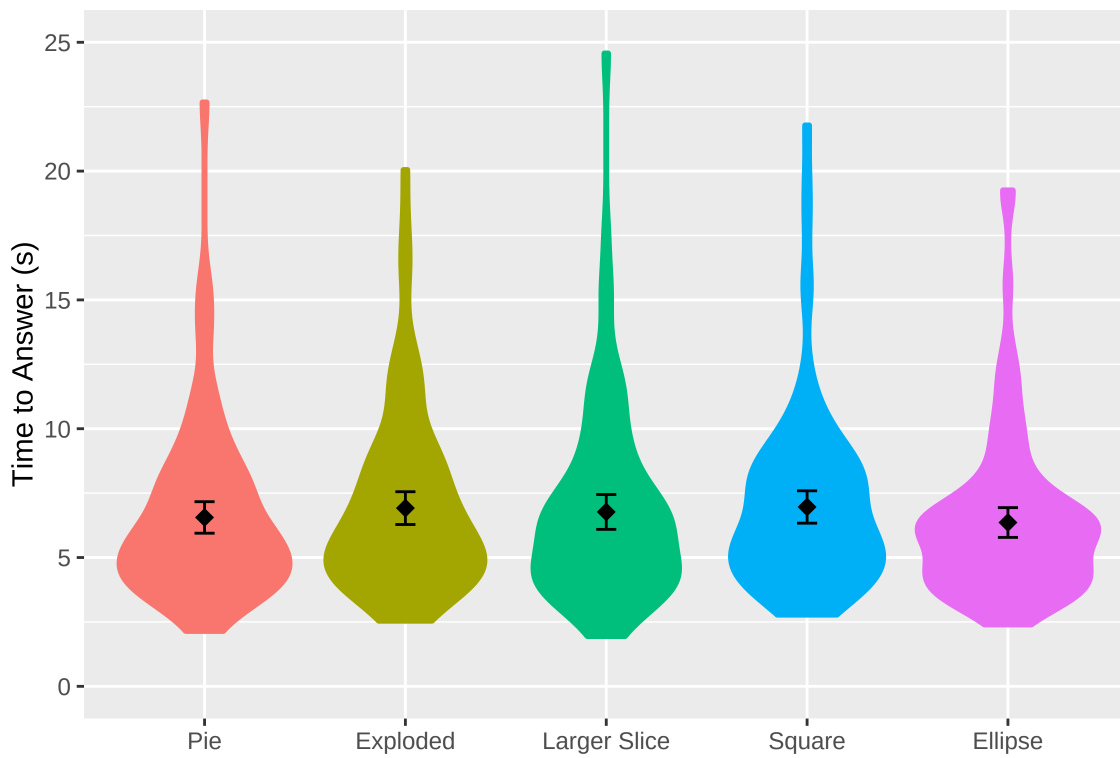


Figure 27: Response times by chart type. Error bars show 95% confidence intervals.

that appeared to be judging the wrong part of the chart. This was the case for 60 of 6420 total trials, or 0.93% of the data. In this case, I did not find any participants doing this consistently (the highest was 30% of answers), so I did not correct any of the data.

3.4.3.1 Judgment Error

The results are shown in Figure 13 and summarized in Table 10. As in the previous study, and for consistency with other work [17], I report the log absolute error, $log_2(|\text{judgedvalue} - \text{truevalue}| + \frac{1}{8})$.

There are considerable differences in error depending on the chart variation. They are visible in the distribution of error in Figure 13, and an ANOVA also shows them to be statistically significant ($F(4, 6415) = 12.071$, $p < 0.001$). The exploded pie chart

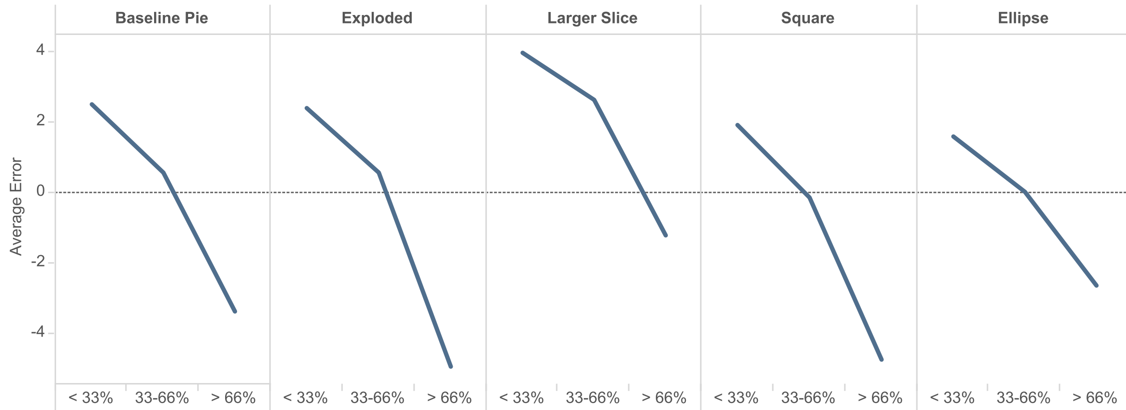


Figure 28: Direction of error for each chart variation split into thirds: value $< 33\%$, $33 - 66\%$, and $> 66\%$. All variations lead to overestimation of small values and underestimation of large ones.

has the second-lowest error, followed by the larger slice. The two charts distorting the shape had the highest errors, with the ellipse being even higher than the square.

Error varies by the angle presented, depending on the chart type (Figure 28). All charts lead to overestimation of small values and underestimation of large ones. On average, all but the larger slice chart lead to an overall underestimation of values (Figure 29). The larger slice chart leads to an overestimation by a factor of 1.6 on average. This is consistent with the idea that area plays a role (the area of the larger slice in the study materials was larger by a factor of about 1.28, which did not vary with the percentage shown).

3.4.3.2 Response Time

Time to answer does not differ between chart types (Figure 27). Only the square appears to take slightly longer to read than the others, but even that does not come near a statistically significant difference.

While I had not stated hypotheses for response time, I would have expected re-

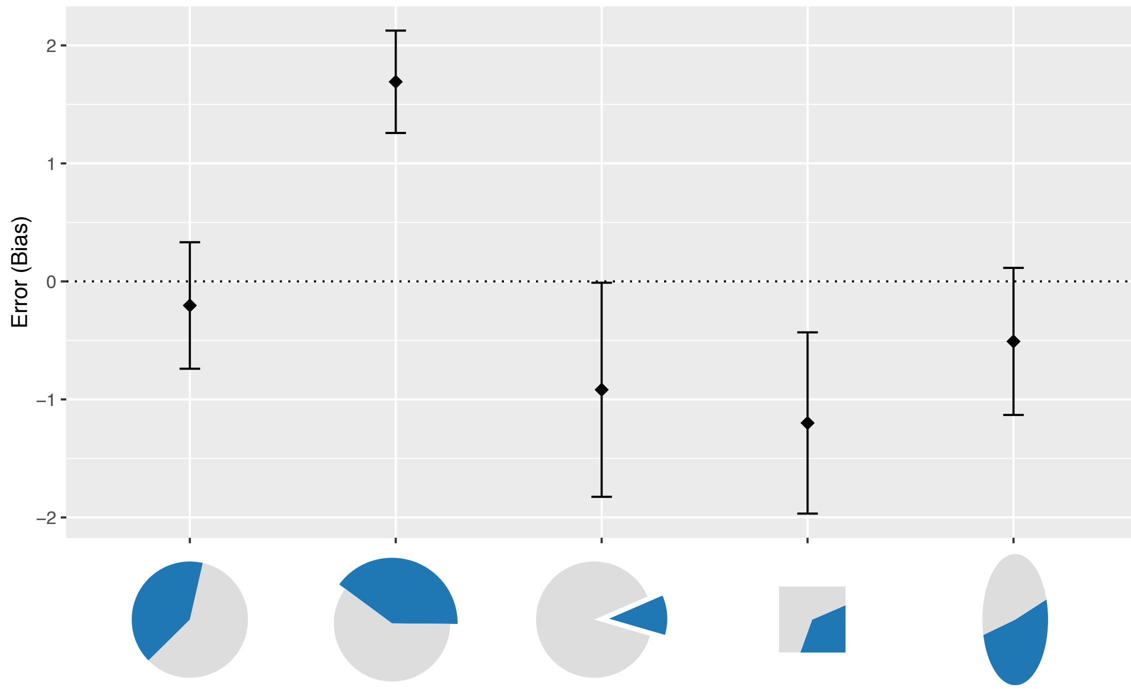


Figure 29: Signed error for each chart type with 95% confidence intervals.

sponse times to be longer for the more unusual and harder-to-read charts, like the ellipse and the square.

3.4.3.3 Comparison with Predictions

Due to the small differences in the predicted effects of arc and area in the study stimuli, I cannot decide which model is the closest fit. Using a linear model, I find multiple-R^2 values of 0.8839 for angle, 0.8731 for arc, and 0.8329 for area, respectively. Including the chart type, these increase to 0.8852 for angle, 0.8816 for arc, and 0.8693 for area. Despite these differences, all models have $p \ll 0.001$. I take this to mean that they all model the data very well and cannot be used to determine which visual cue is the most likely to be used to read the charts.

Regarding the qualitative predictions, I find that my results largely fit them:

- The larger slice leads to systematic overestimation over almost the entire range of values.
- The exploded pie chart shows higher error, which I did not expect. Perhaps the gap between the two slices adds a level of distraction that causes higher error.
- The ellipse yields much higher error than the circle, as expected.
- Likewise, the square produces larger error than the circle, just as expected. Interestingly, the ellipse actually leads to more error than the square, which I did not expect.

3.4.4 Discussion

Distorting pie charts has an effect on reading accuracy. All of my variations increased the error. Even the seemingly innocuous exploded pie chart resulted in a measurable effect.

None of my designs changed the central angle, yet they all led to considerable error. This further undermines the importance of angle as the key visual cue, since angle was always easily readable.

What was distorted in all but the exploded pie chart was area and arc length. That led to more error. The systematic overestimation in the larger slice chart condition is particularly telling: the larger slice has more area and a longer arc, but covers the same angle. That said, I was unable to find a clear link between either area, arc, or angle and participants' responses that would have allowed me to decide which of the three was the most influential. This is somewhat consistent with my previous work, which suggested a combination of the three.

The increased error in the exploded pie chart was unexpected. One explanation is that the white space between the slices made it harder to estimate the total area of the pie. Moving a slice for emphasis is common in infographics and business presentations. Based on my results, however, it should be avoided.

The pronounced effect of the square and ellipse shapes is also troubling. Shape distortions are quite common in infographics, and I found that they had the strongest effect among the variations I tested. Therefore, distorting the shape of pie charts should be avoided to preserve communication accuracy.

It is interesting to note that infographics often prominently include numbers on top of pie charts (Figure 23). This might seem to obviate the use of the chart in the first place, but the combination can be quite powerful. The pie chart, despite its distortion, still gives a rough idea of the differences between slices, especially when they are large. The number then provides the precision that the chart does not. It seems that infographics designers are intuitively aware that their design decisions are impacting the precise reading of values.

3.4.5 Limitations

The effect of the distortions tested in this study is limited. I believe that there would have been stronger effects had the ellipse been more compressed or had there been a larger size difference for the larger pie slice. I picked those values based on charts I had seen in practice, however. While there are clearly exceptions, my distortions are in line with the more common ones found in information graphics.

3.4.6 Conclusions

Together with my *Arcs, Angles, or Areas* work (Section 3.2), the results of this study call the assumption into question that pie charts are read primarily by central angle. If that were the case, error should not be different between the baseline pie and the exploded pie or the larger-slice pie. Both cause considerably higher error, though.

Design choices common in infographics cause considerable distortion. The worst offenders in the study were the ones where the shape of the "pie" was no longer a circle. I recommend that all such designs be avoided in favor of simple pie and donut charts.

Even seemingly innocuous changes, like moving a slice away from the center, can have an effect on error, however. Given the prevalence of pie charts in many different contexts, I believe that a more systematic study of these effects is called for.

In the interest of reproducibility, all study code has been made available at `github.com/dwskau/arcs-angles-area`, `github.com/dwskau/donut-radii`, and `github.com/dwskau/pie-variations`.

3.5 Conclusions

Despite their bad reputation in data visualization, pie and donut charts are commonly used in information graphics and many other areas. In these studies, I attempted to find out which visual encodings are important for reading values from these charts by splitting them into their constituent parts.

The results show that all three visual cues are important, but arc length in particular seems to provide important information. Angle is clearly not a significant bearer of information in pie charts, and in particular the central meeting point of the circle segments does not appear to be crucial. Donut charts thus appear to be no worse than pie charts for accurately communicating the data they contain.

I also note that despite the generally firmly held stance against pie charts in the data visualization community, little actual research has looked into their underlying perceptual mechanisms or the impact of design variations. More work is clearly needed, especially because of these charts' widespread use.

CHAPTER 4: CONCLUSIONS

Charts are not a new tool for the communication of data, but they are being used in new ways. In order to use charts effectively, especially charts with modifications, designers must fully understand how they function. The results for bar charts provide deeper insight into how our visual system processes these charts, and which portions of the chart may represent concepts vs. data. This insight can help designers to create more effective charts. The results for pie and donut charts have already begun to shift 90 years of thinking, and are spurring the data visualization community to re-examine other conventional wisdom.

Experience has proven that people will continue to create embellished charts, and push the aesthetic boundaries of the data visualizations they produce. Where communication is primarily visual, it is key to understand what each visual modification is doing to the ability of the chart to communicate accurately. It is possible to use knowledge about systematic under- or over-estimation caused by a chart embellishment to mislead a chart's viewers. The key to defending against this type of exploitation is to make this research known. If this knowledge is spread widely, creators can avoid accidental biasing, and, more importantly, viewers can recognize when they may be duped.

The research presented in this document provides insight into how charts work, and can provide guidelines for chart designers to more accurately communicate information with bar, pie, and donut charts. The most important factor in bar charts is the shape of the border of the bars; the contents within that border have a minimal effect on the perception of the bar's value.

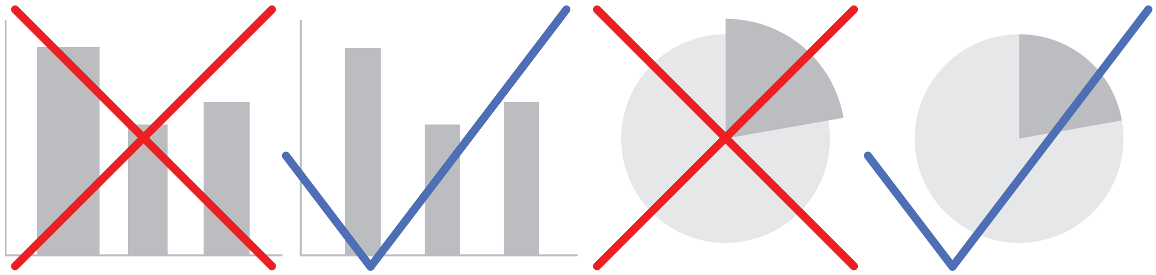
This work also provides us with an approach to research how other charts work. The process of distilling a chart down to the geometric components that represent data (visual variables), and testing these individual variables independently in custom charts can be done for any type of data visualization. This systematic geometric evaluation could lead to a more fundamental understanding of how our visual system converts geometry into values. The methods for accomplishing this are accessible to other researchers, using proven, common tools that scale well and keep financial cost minimal.

Future work using this system of geometric component isolation could tackle any number of chart types. My work on bar charts only addresses vertically oriented bars, however, horizontal bar charts are also extremely common, and are also frequently embellished. Line charts are another prime candidate for exploration using this system, however it may be more difficult to isolate their visual variables. The interaction between the representation of the data they contain and the tasks they support mean line slopes are the most salient geometric feature. In addition to common chart types, glyphs used in small multiples could be a good target for this method.

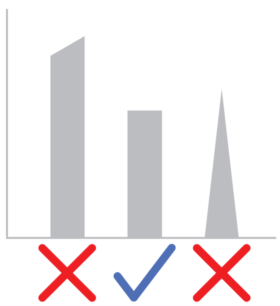
CHAPTER 5: RECOMMENDATIONS FOR BAR, PIE, AND DONUT CHARTS

The results of all this research can be summarized in a list of recommendations (illustrated in Figure 30).

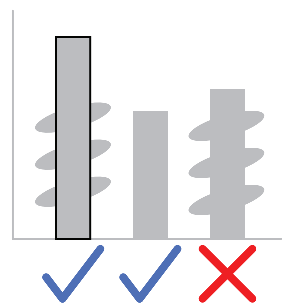
- No conflicting data encodings allowed.
- Protect the ends of vertical bars, keep the top edges perpendicular to the y-axis.
- Emphasize bar boundaries.
- Within bar boundaries, anything goes.
- Do not use overlapping bars.
- Donut charts are just as accurate as pie charts.
- Maintain the outer arcs of a pie or donut chart.
- Maintain the area of each wedge of a pie or donut chart.



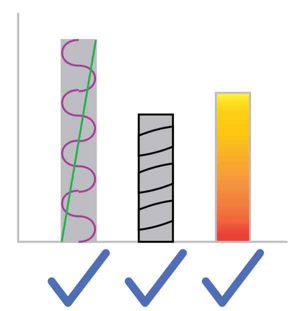
(a) No conflicting data encodings.



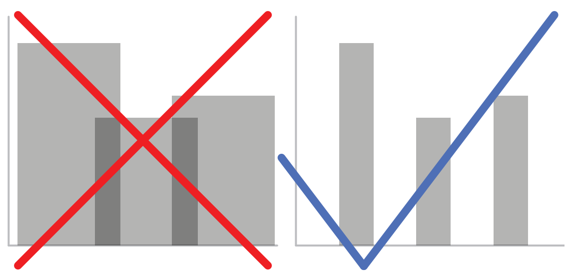
(b) Perpendicular bar ends.



(c) Maintain boundaries.

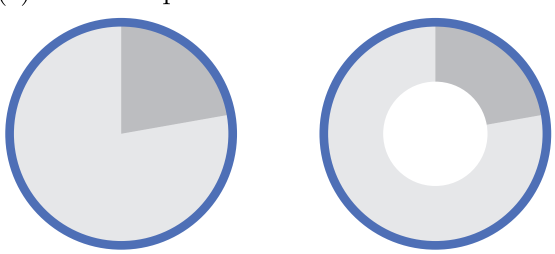


(d) Within boundaries, anything goes.

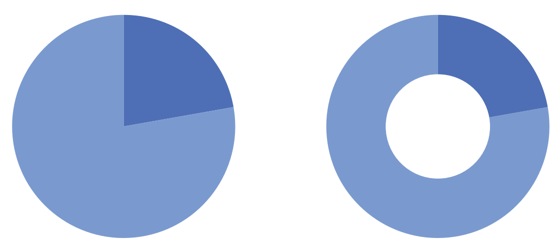


(e) No overlap.

(f) Donuts aren't so bad.



(g) Maintain the outer arcs.



(h) Maintain the wedge areas.

Figure 30: Illustrations of the recommendations that resulted from this research.

REFERENCES

[1] S. Bateman, R. L. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2573–2582. ACM, 2010.

[2] J. Bertin. *Semiology of graphics*. University of Wisconsin Press, 1983.

[3] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, L. Floridi, and M. Chen. An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2759–2768, 2012.

[4] M. M. A. Borkin, A. A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, 2013.

[5] W. C. Brinton. *Graphic Methods for Presenting Facts*. Engineering Magazine Company, 1914.

[6] T. Chiasson. Infoactive. `infoactive.co`, 2014. [Online; accessed 19-August-2014].

[7] W. S. Cleveland. *The Elements of Graphing Data*. Hobart Press, 1994.

[8] W. S. Cleveland and R. McGill. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[9] M. Correll and M. Gleicher. Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, 2014.

[10] D. Cox. Some remarks on the role in statistics of graphical methods. *Applied Statistics*, 27(1):4–9, 1978.

[11] F. E. Croxton and R. E. Stryker. Bar Charts Versus Circle Diagrams. *Journal of the American Statistical Association*, 22(160):437–482, 1927.

[12] dadaviz. Data nerds + art lovers = dadaviz. `dadaviz.com`, 2014. [Online; accessed 19-August-2014].

[13] D. Design. Raw. `raw.densitydesign.org`, 2014. [Online; accessed 19-August-2014].

[14] W. Eells. The Relative Merits of Circles and Bars for Representing Component Parts. *Journal of the American Statistical Association*, 21:119–132, 1926.

[15] S. Elzer, N. Green, S. Carberry, and J. Hoffman. A model of perceptual task effort for bar charts and its role in recognizing intention. *User Modelling and User-Adapted Interaction*, 16(1):1–30, 2006.

[16] S. Haroz, R. Kosara, and S. L. Franconeri. ISOTYPE Visualization – Working Memory, Performance, and Engagement with Pictographs. In *Proceedings CHI*, pages 1191–1200, 2015.

[17] J. Heer and M. Bostock. Crowdsourcing graphical perception: using Mechanical Turk to assess visualisation design. *ACM Human Factors in Computing Systems*, pages 203–212, 2010.

[18] ILoveCharts. I love charts. `ilovecharts.tumblr.com`, 2014. [Online; accessed 19-August-2014].

[19] S. Kim, S. Li, B. chul Kwon, and J. Yi. Investigating the Efficacy of Crowdsourcing on Evaluating Visual Decision Supporting System. *Proceedings of the Human Factors and Ergonomics Society*, 55(1):1090–1094, 2011.

[20] R. Kosara and D. Skau. Judgment error in pie chart variations. In *Short Paper Proceedings of the Eurographics/IEEE VGTC Symposium on Visualization (EuroVis)*, 2016.

[21] R. Kosara and C. Ziemkiewicz. Do Mechanical Turks dream of square pie charts? *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaLuation methods for Information Visualization*, pages 63–70, 2010.

[22] R. Kosara and C. Ziemkiewicz. Do mechanical turks dream of square pie charts? In *Proceedings BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV)*, pages 373–382, 2010.

[23] S. M. Kosslyn. *Graph Design for the Eye and Mind.* Oxford University Press, 2006.

[24] W. Kruskal. Visions of Maps and Graphs. *Proceedings of the International Symposium on Computer-Assisted Cartography, Auto-Carto II*, pages 27–36, 1975.

[25] W. Mason and S. Suri. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1):1–23, 2012.

[26] C. F. Media. Visage - beautiful, branded content. `visage.co`, 2014. [Online; accessed 19-August-2014].

[27] T. Munzner. *Visualization Analysis and Design.* A K Peters, 2014.

[28] G. E. Newman and B. B. J. Scholl. Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic bulletin & review*, 19(4):601–607, 2012.

[29] W. Playfair, H. Wainer, and I. Spence. *Playfair's commercial and political atlas and statistical breviary.* Cambridge University Press, 2005.

[30] Plotly. Plotly. `plot.ly`, 2014. [Online; accessed 19-August-2014].

[31] N. B. Robbins. *Creating More Effective Graphs.* Chart House, 2013.

[32] A. Satyanarayan and J. Heer. Lyra: An Interactive Visualization Design Environment. *Computer Graphics Forum*, 33(3):351–360, 2014.

[33] D. Simkin and R. Hastie. An Information-Processing Analysis of Graph Perception. *Journal of the American Statistical Association*, 82(398):454–465, 1987.

[34] D. Skau, L. Harrison, and R. Kosara. An evaluation of the impact of visual embellishments in bar charts. *Computer Graphics Forum (Proceedings EuroVis)*, 34(3):221–230, 2015.

[35] D. Skau and R. Kosara. Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts. *Computer Graphics Forum (Proceedings EuroVis)*, 35(3), 2016.

[36] I. Spence and S. Lewandowsky. Displaying proportions and percentages. *Applied Cognitive Psychology*, 5(1):61–77, 1991.

[37] J. Talbot, V. Setlur, and A. Anand. Four Experiments on the Perception of Bar Charts. *IEEE Transactions on Visualization and Computer Graphics*, 2014.

[38] E. Tufte and P. Graves-Morris. *The visual display of quantitative information.* CT Graphics, Cheshire, CT, 1983.

[39] Visualizing.org. visualizing.org — data visualizations, challenges, community. `visualizing.org`, 2014. [Online; accessed 19-August-2014].

[40] Visually. Visual.ly. `visual.ly`, 2015.

[41] L. Wilkinson. *The Grammar of Graphics.* Springer, 2nd edition, 2005.

[42] D. Woo. Datavisual. `datavisu.al`, 2014. [Online; accessed 19-August-2014].

[43] J. Zacks, E. Levy, B. Tversky, and D. J. Schiano. Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology: Applied*, 4(2):119, 1998.

[44] C. Ziemkiewicz and R. Kosara. Laws of Attraction: From Perceptual Forces to Conceptual Similarity. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1009–1016, 2010.