Journal of Data Science xx (xx), 1–11
July 2022
DOI: 0000

Evaluating Perceptual Judgements on 3D Printed Bar Charts

Tyler Wiederich [1,*] and Susan VanderPlas[1]
[1]Department of Statistics, University of Nebraska-Lincoln, United States of America

Abstract

Graphical design principles typically recommend minimizing the dimensionality of a visualization - for instance, using only 2 dimensions for bar charts rather than providing a 3D rendering, because this extra complexity may result in a decrease in accuracy. This advice has been oft repeated, but the underlying experimental evidence is focused on fixed 2D projections of 3D charts. In this paper, we describe an experiment which attempts to establish whether the decrease in accuracy extends to 3D virtual renderings and 3D printed charts. We replicate the grouped bar chart comparisons in the 1984 Cleveland & McGill study, assessing the accuracy of numerical estimates using different types of 3D and 2D renderings.

Keywords *graphics; 3D bar charts; 3D printing.*

1 Introduction

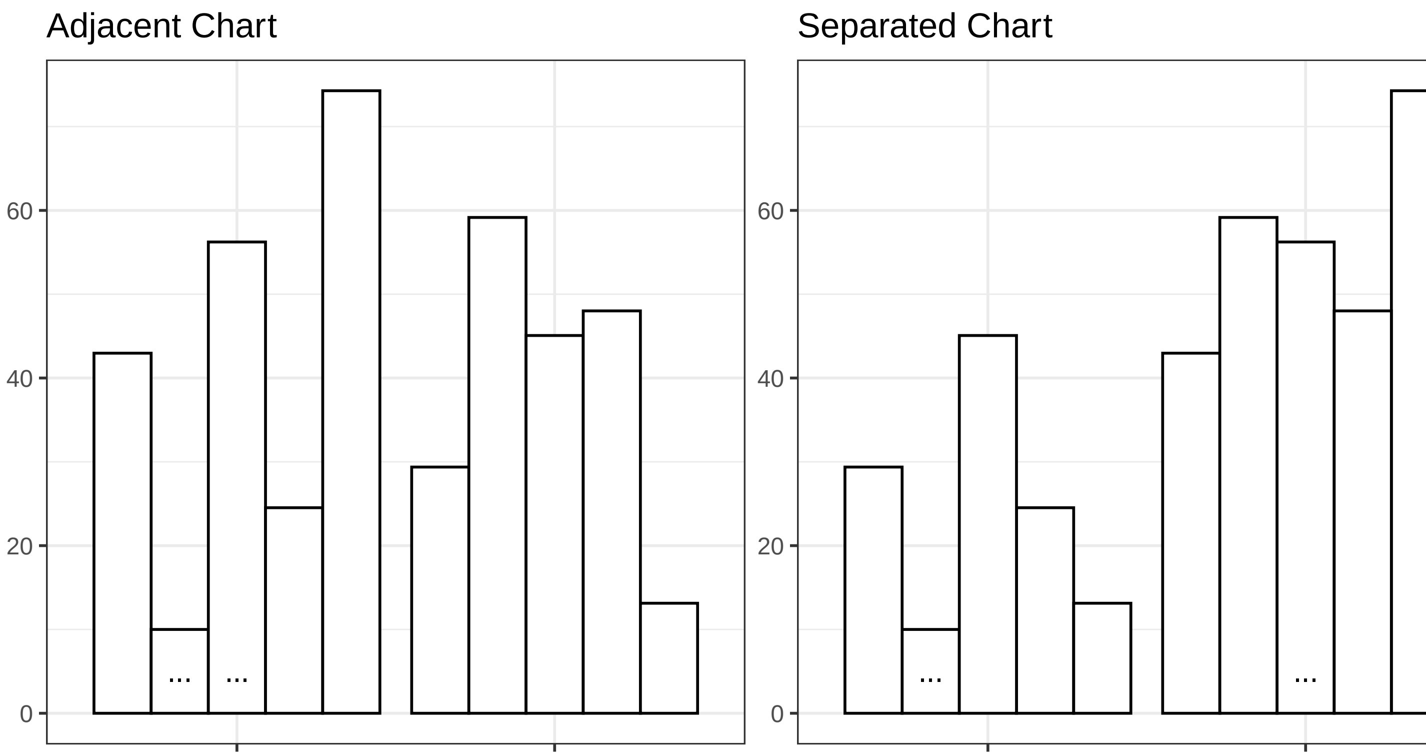
Good communication requires both that the information be transmitted correctly and that the intended recipient be able to decode and understand the transmitted information accurately. In order to communicate effectively, we must use graphical forms that accurately convey information relevant to the task in question. In many cases, this means we must understand how accurately people can read quantitative information off of charts. While accuracy is not the only quantity of interest in graphical investigations (Hullman et al., 2019), it is an important factor in assessing the utility of many different data graphics.

The accuracy of graphical forms has been studied for almost a century (von Huhn, 1927; Eells, 1926; Croxton and Stein, 1932; Croxton and Stryker, 1927), as new ways of representing information evolve, we must revisit old studies to determine whether these representations have the same limitations as previous versions. This is particularly true in areas like graphics which are affected by the immense technological innovation in hardware and software which has taken place since the early 1990s.

1.1 Elementary Graphical Tasks

Cleveland and McGill (1984) established the comparative accuracy of different "elementary perceptual tasks" (EPTs). Elementary Perceptual Tasks, according to these experiments, include assessing graphical elements such as position along a common scale, length, angle, and volume, and estimating the corresponding numerical value of these representations. The study relied entirely on estimation accuracy, which may not always be relevant when extracting information from graphs. For example, estimation is less relevant when ordering values by size. As a result of the Cleveland and McGill (1984) study, it is possible to assemble an ordering of perceptual accuracy for the elements of length, position, and angle. Heer and Bostock (2010) replicated

*Tyler Wiederich Email: twiederich2@huskers.unl.edu or susan.vanderplas@unl.edu.



(a) Adjacent

(b) Separated

Figure 1: @clevelandGraphical1984 used two different types of grouped bar charts, which are replicated here.

some parts of Cleveland and McGill (1984) in an online setting using Mechanical Turk, largely validating the results of the original study while demonstrating the utility of the Mechanical Turk platform for graphical testing.

The first experiment in Cleveland and McGill (1984) (the position-length experiment), used five types of bar charts: two types of grouped bar charts and three types of stacked bar charts. Each chart had two bars marked for comparison; participants were asked to determine which bar was smaller and give the perceived ratio of the smaller bar to the larger bar. **?** show the two types of grouped bar charts. We are primarily interested in the grouped bar charts (in part because 3D printing is not yet inexpensive enough to make moderate-scale stacked bar chart experiments viable), which consisted of two comparison bars which were either adjacent or in separate groups. These grouped bar charts will be referenced as adjacent and separated graph types in this paper, respectively.

1.2 3D Graphical Perception

Chart perception is often affected by the visual system's implicit assumption that visual stimuli are three-dimensional; after all, most of the visual input we process does come from a three-dimensional world, but charts are artificial and largely exist in two dimensions. This occasionally causes problems: the line-width illusion, for instance, has been attributed to implicit 3D perception of 2D stimuli and can affect perception of error bands, candlestick plots, and Sankey diagrams (VanderPlas and Hofmann, 2015; Day and Stecher, 1991; Hofmann and Vendettuoli, 2013).

The use of 3D graphics have been explored in multiple studies. Fisher et al. (1997) explored user preference for 2D or 3D charts and found that subjects tended to prefer simpler 2D graphs when tasked with extracting information. Barfield and Robless (1989) compared 2D and 3D graphs presented on paper and on computers, showing that the accuracy of subject answers depended on their skill level: novice subjects were more accurate with 2D paper graphs, while experienced managers were more accurate with 3D computer graphs. For both experience levels, participants were more confident in their answers when using 2D graphs. There are instances where 2D graphs perform better than 3D graphs, but there are times where 3D graphs may better encode information - for instance, when X and Y are used to represent spatial dimensions, it may be preferable to use a 3D chart to convey numerical information instead of using color, which is perceived much less accurately. Brath (2014) highlights the intrinsic attributes of 3D graphs and the benefits when used appropriately with other 3D elements such as lighting and correct portrayal of data attributes.

There is thus good reason to be wary of the use of three dimensions where only two are necessary to convey data. However, the situation is different now than it was in 1984 when Cleveland & McGill published their seminal work; it has even changed since Heer's replication study (2010) in 2010.
Digital graphics has developed quickly, along with the hardware necessary to support these software developments. As a result, we have much more natural rendering of 3D objects virtually, and we can also print graphics in three dimensions, moving artificial charts into a more natural, physical setting. As a result, it is reasonable to reconsider the use of 3D charts, not only because of new technological developments, but also because these charts provide the opportunity to make data graphics accessible to those with limited or absent vision.

In this paper, we discuss a study designed to examine Cleveland & McGill's experiments on grouped bar charts using modern graphics in two and three dimensions. This study lays the groundwork for additional empirical studies on the use of 3D graphics, rendered and 3D printed, for visualizing complex data. In order to explore perception of fully 3D graphics, it is prudent to start with the simplest possible 3D graphic: one in which the third dimension is not necessary, so that we can easily compare to two-dimensional representations without loss of information. In the next section we provide details about design and execution of our experiment, including the process of replicating stimuli from **?** in order to create 2D, 3D projected, 3D rendered, and 3D printed bar graphs. The next section presents the results, and we conclude the paper by discussing this experiment in the context of existing work on the perception of 2D and 3D graphical elements.

2 Methods

Our study is designed to replicate and expand upon the position-length experiment from Cleveland and McGill as closely as possible. In this section, we will discuss the replication process and the design of our modified version of this experiment.

2.1 Replicating Cleveland and McGill

The first step of replicating the position-length experiment was to determine the heights of the bars that participants use for comparisons. These values for the bar heights are linear on a log

Figure 2: Two dimensional, two-dimensional digital rendering, and 3D-printed charts used in this study.

scale and are given by

$$s_i = 10 \cdot 10^{(i-1)/12}, \qquad i = 1, ..., 10$$

Each graph presents two bars from the values given above where the participants are asked to judge the ratio of the smaller bar to the larger bar. The ratio of bars used by Cleveland and McGill were 17.8, 26.1, 38.3, 46.4 (twice), 56.2, 68.1 (twice), and 82.5 (twice). The exact numeric comparisons were not disclosed, but the comparison values used in our study were subjected to the constraints of having the same ratio values and that no value was used more than twice.

Each graph is presented so that there are ten bars where only two of the bars are marked for identification. Cleveland and McGill did not specify the random process for the heights of the eight other bars, so we used a scaled Beta distribution with parameters that limit excessive noise around the bars used for comparisons. Code to reproduce the data generation process, data underlying the plots used in this study, the rendered plots and STL files, anonymized user data, and analysis code can found be at `https://github.com/TWiedRW/2023-JDS-3dcharts`.

2.2 Stimuli Construction

The graphs share a common layout across all formats, where two groupings of five bars are identified by "A" and "B", respectively, and circles and triangles are used to identify the bars participants should compare. Example graphs are shown in Figure 2. There are some graphical elements that cannot be easily portrayed via 3D printing. For this reason, all graph types do not have axes, grid lines, or floating titles.

The ggplot2 (Wickham, 2016) package was utilized to create the 2D bar charts. The scale axis was removed, leaving only the bars and a bar grouping identifier. The bars used for comparisons had the identifying mark at a height of 5 out of 100 for the 2D plots, and the 3D plots had the identifying marks on top of the bars.

The 3D renderings and 3D printed charts were both created using OpenSCAD (Kintel, 2023), which creates STL files from markup describing the object's geometric composition. Charts were composed of a platform, raised labels for the A and B groups of bars, circle and triangle markers indicating the bars of interest, and the bars themselves; values for the bar heights were inserted into the markup using R (R Core Team, 2023). In addition, an ID code was engraved into the bottom of the platform to uniquely identify each object; this allows the researchers to ensure that the 3D printed charts are correctly allocated to stimulus sets.

Digital renderings of the generated STL files were created using Murdoch and Adler (2023), which integrates into Chang et al. (2023) using the Mozilla Foundation extension. Rendered

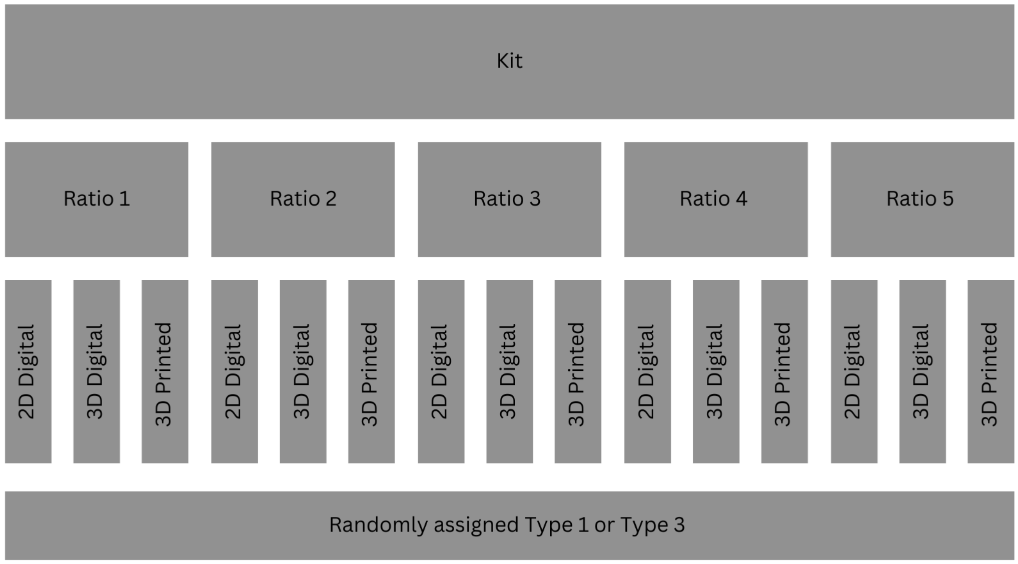


Figure 3: A graphical representation of the study design. Only five of the seven ratios were used in each kit where the ratios are randomly selected. A total of 21 kits were created to include all combinations of the five ratios.

3D charts were initially angled corresponding to the default 3D bar charts present in Microsoft Excel, but WebGL's interactivity allows the user to rotate, scale, and otherwise interact with the chart to change the angle. 3D renderings were colored to correspond to the 3D printed chart filament color. The default `rgl` lighting was replaced with three lights located in fixed positions around the rendered figure. The lights were positioned so that one was behind the rendered figure, another in front of the figure, and one light below the figure. 3D charts were printed with colored filament corresponding to a specific ratio comparison; this allowed researchers to visually assess kits to ensure that they contained five unique ratios. Colors corresponding to each ratio were assigned randomly to ensure that chart color provided no useful information about the ratio value. As printed, the base of the chart was 13cm x 3cm x 1cm, with the highest bar rising 9.5cm above the chart base. Raised letters and shapes were 2mm above the base or bar, respectively.

2.3 Experiment Design

Participants were provided a kit with five 3D-printed charts, comprising five of the seven unique ratio comparisons; we then used the kit ID to ensure that participants saw computer-rendered charts with ratios corresponding to those in the kit of physical charts. This ensured that the experimental design was balanced across chart type and randomized with respect to the type of comparison (adjacent or separated). We printed 21 kits containing five 3d-printed charts each.

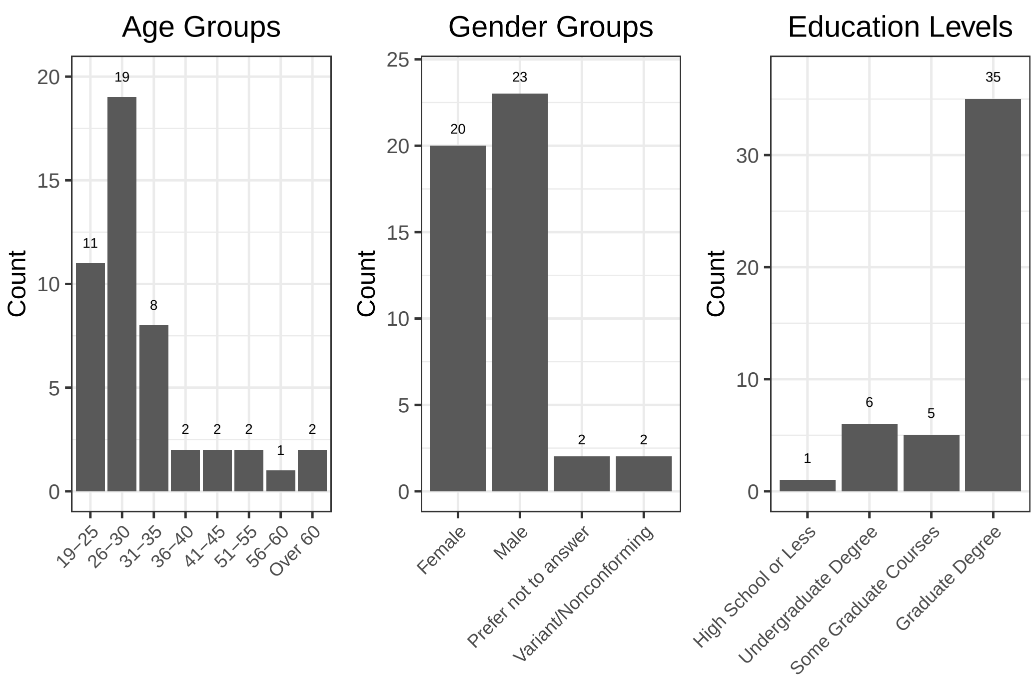


Figure 4: Demographic breakdown of participants in the study. All subjects were recruited from faculty and students in the statistics department at University of Nebraska-Lincoln.

2.4 Participant Recruitment

One interesting facet of Cleveland and McGill (1984) is the participant recruitment methodology: "For each experiment the subjects fell into two categories: (1) a group of females, mostly housewives, without substantial technical experience; (2) a mixture of males and females with substantial technical training and working in technical jobs. Most of the subjects in the position-length experiment participated in the position-angle experiment; in all cases repeat subjects judged the position-angle graphs first." It would seem likely that the authors recruited individuals within their respective departments as well as their wives. In the spirit of replicating the study, members of the UNL Statistics department and their spouses, partners, and roommates were asked to participate in our study; this replicates the spirit of the original study without the implicit assumptions that graduate students and professors are (1) largely male, (2) heterosexual, and (3) have unemployed partners.

A total of 48 participants completed the study; demographics are shown in 4.

2.5 Data Collection

A Shiny applet was used for data collection, along with the provided kit of 3D printed charts. Participants provided informed consent through the applet, and then were asked for demographic information (age, gender, education level). Then, participants were shown a "practice" page which allowed them to experiment with the data collection interface and practice estimating the ratio between the bars, as shown in Figure 5

Directly before the experiment started, participants were asked to provide the kit ID, along with directions indicating that if the instructions indicated that participants should use a 3D chart for a task, the participant should select a chart from the kit, enter the ID code of that

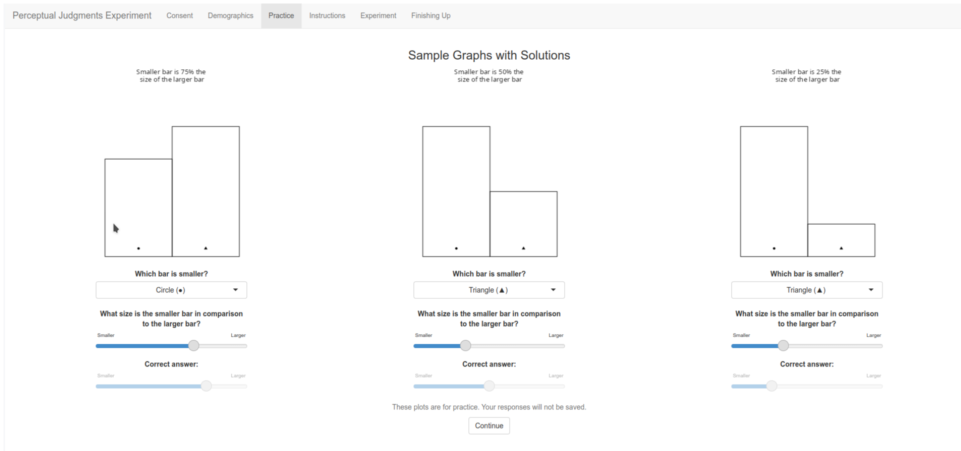


Figure 5: Screenshot of Shiny application practice screen. Three 2D bar charts with different ratios were provided, along with sliders indicating the correct proportion. Participants could practice with the sliders and preview the questions that would be asked as part of the task.

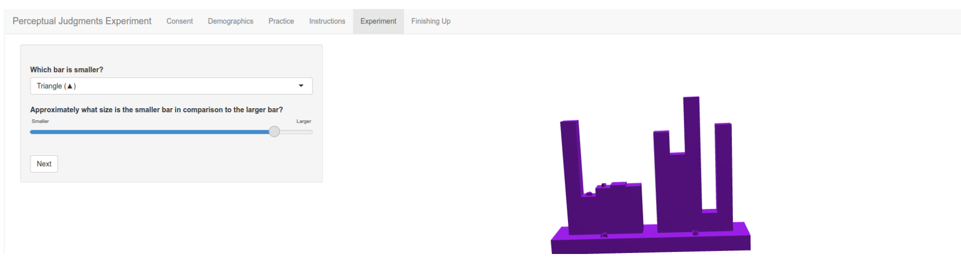


Figure 6: Screenshot of the applet collecting data for a 3D rendered chart task. Participants were asked to select which bar (circle or triangle) was smaller, and then to estimate the ratio of the smaller bar to the larger bar.

chart from the bottom of the object, and complete the requested task. Participants were also instructed to make quick judgments for each graph and not to measure or estimate ratios using physical objects.

Each graph (or prompt, in the case of 3D printed charts) in the applet had two corresponding questions for participants to answer: first, participants were to identify the smaller bar by shape, and then, participants were to estimate the ratio of the size of the smaller bar to the size of the larger bar, as shown in Figure 6.

3 Results

All responses that incorrectly identified the smaller bar were removed from the study before analysis.

3.1 Midmeans of Log Absolute Errors

Cleveland and McGill used

$$\log_2(|\text{Judged Percent} - \text{True Percent}| + 1/8)$$

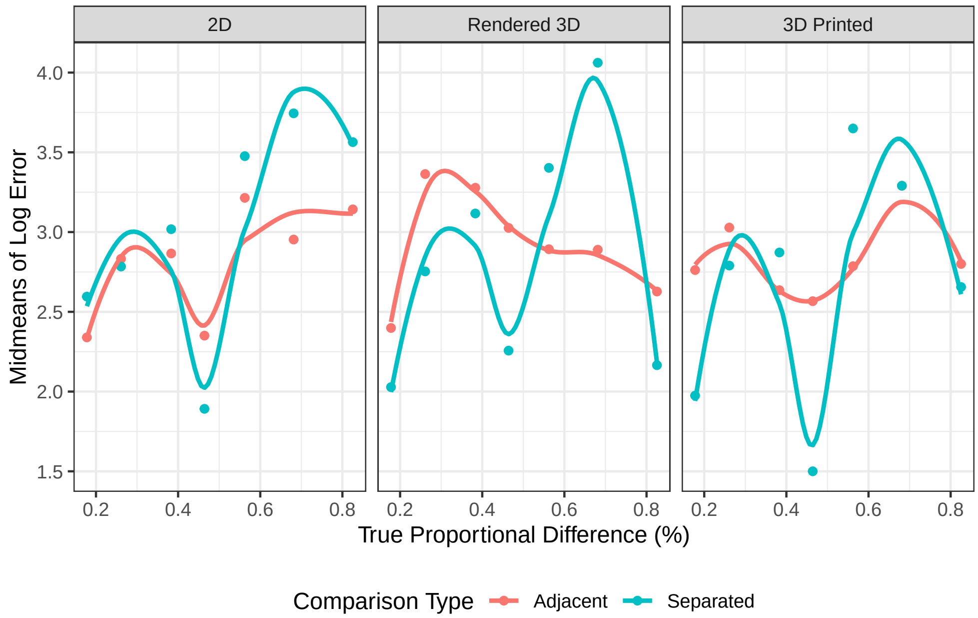


Figure 7: Midmeans of log absolute errors for the true ratio of bars. Each overlaying line are computed with loess.

to measure accuracy of their participant's responses. In their study, log base 2 seemed appropriate due to "average relative errors changing by factors less than 10." They also added 1/8 to prevent distortions when the errors were close to zero. Heer and Bostock (2010) followed the same analysis method, replicating many (but not all) of the results presented in the original paper.

Figure 7 shows the midmeans of the log absolute errors compared to the true ratio of the bars for each graph type and comparison type. The results tend to indicate that the log absolute errors increase for greater differences between the smaller and larger bars, but are consistent across the graph types.

This is somewhat different than the results in Cleveland and McGill (1984) and Heer and Bostock (2010); in both cases the midmean log absolute errors increased until about 55% of the true proportional difference and then decreased. It is possible that this difference is due to the fact that we did not require an explicit numerical estimate of the ratio but instead asked participants to indicate the ratio on a slider (which is essentially a number line). This should reduce the cognitive load required to transition from spatial comparison to numerical comparison and then to compute the proportion, but may also have impacted the results.

3.2 Linear Mixed Effects Model

In addition to replicating the (primarily graphical) analysis of participant errors, we also took a more statistical approach and fitted a linear mixed effects model that accounts for participant variation as well as the effect of comparison type, graph type, and ratio. This allows us to test for significant differences (though given the midmean log absolute error plots, we do not expect to find any) as well as to quantify effect sizes for future studies. The formal statistical model is as follows:

$$y_{ijklm} = \mu + S_i + R_j + G(R)_{(k)j} + T_l + \epsilon_{ijklm}$$

where
- $y_{ijklm} = \log_2(|\text{Judged Percent} - \text{True Percent}| + 1/8)$
- $S_i \sim N(0, \sigma_S^2)$ is the effect of the i^{th} subject
- R_j is the effect of the j^{th} ratio
- $G(R)_{(k)j}$ is the effect of the k^{th} graph type nested in the j^{th} ratio
- T_l is the effect of the l^{th} comparison type
- $\epsilon_{ijklm} \sim N(0, \sigma_\epsilon^2)$ is the random error

No differences were detected for the true ratio of bars (p-value = .7881), whether the bars were adjacent or separated (p-value = .3375), or for the plot nested within the true ratio (p-value = .6868). While Heer & Bostock provided a zip file containing data and code for their paper, the link is no longer active, so it is not possible to fit a similar model to their data at this time.

4 Discussion and Future Work

Previous work in 3D graphics would suggest that the errors for the 3D graphs would be larger than the errors for the 2D graphs. While we did not find any significant results indicating that 3D graphs are read less accurately, there are two possibilities that might account for this discrepancy.

The first potential explanation is that this study is underpowered - the effect size is small, and our 48 participants were insufficient; the original study included 51 participants, which is slightly larger. However, we should note that the analysis methods used in Heer and Bostock (2010) and Cleveland and McGill (1984) do not include significance testing beyond the graphical display of confidence intervals.

The second possibility is more interesting: we examined 3D charts using rendered 3D graphs and 3D printed charts; both of these options allow for participants to interact with the chart, rotating it, and generally perceiving it as one might perceive any other 3D, real, object. This is a far cry from the 3D perspective charts in the original study, which have a fixed angle and perspective and are thus not equivalent to our 3D charts. Future studies should include an additional fixed 3D perspective bar chart, which will at least enable us to examine whether modern 3D rendering environments allow for more accurate conclusions than fixed 3D perspectives. Future iterations of this study will include "traditional" 3D graphs created by Microsoft Excel (that is, graphs with a fixed 3D perspective rendered in 2D). This option will allow us to examine fixed perspective 3D plots compared to 3D renderings and 2D plots; it will also enable online data collection in addition to the in-person data collection used in this experiment.

Another interesting aspect of our study is that the method used to record participant estimates is different from the method used in the original study as well as Heer & Bostock's replication study. A method similar to our slider input, marking position on a line, was used in Spence (1990), but Spence asked participants to estimate $A/(A + B)$, where we asked participants to estimate A/B; thus, our results are still not directly comparable to previous studies. We expect that the specific ratio estimated would also have an effect on observed participant errors.

The slider method for input of ratio estimates should be easier for participants, as it does not require explicit transformation to the numerical domain. What is clear is that it would be beneficial to assess the impact of measurement method on participant errors directly, so that the results of these different studies might be explained and interpreted with regard both to the stimuli used and the measurement method employed in the experiment. Future iterations of this experiment will likely address this estimation difference; such modifications in experimental design are relatively straightforward in Shiny and will provide useful insight into the design of future experiments evaluating the perception of statistical graphics.

4.1 Supplemental Material

Stimuli, code, and data for this experiment are provided at https://github.com/TWiedRW/2023-JDS-3dcharts.

References

Barfield W, Robless R (1989). The effects of two- or three-dimensional graphics on the problem-solving performance of experienced and novice decision makers. *Behaviour & Information Technology*, 8(5): 369–385.

Brath R (2014). 3D InfoVis is here to stay: Deal with it. In: *2014 IEEE VIS International Workshop on 3DVis (3DVis)*, 25–31. IEEE, Paris, France.

Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, et al. (2023). *shiny: Web Application Framework for R*. R package version 1.7.4.1.

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

Croxton FE, Stein H (1932). Graphic Comparisons by Bars, Squares, Circles, and Cubes. *Journal of the American Statistical Association*, 27(177): 54–60.

Croxton FE, Stryker RE (1927). Bar Charts Versus Circle Diagrams. *Journal of the American Statistical Association*, 22(160): 473–482.

Day RH, Stecher EJ (1991). Sine of an illusion. 20(1): 49–55.

Eells WC (1926). The Relative Merits of Circles and Bars for Representing Component Parts. *Journal of the American Statistical Association*, 21(154): 119–132.

Fisher SH, Dempsey JV, Marousky RT (1997). Data Visualization: Preference and Use of Two-Dimensional and Three-Dimensional Graphs. *Social Science Computer Review*, 15(3): 256–263.

Heer J, Bostock M (2010). Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 203–212. ACM. Tex.ids= heer2010a, heer2010crowdsourcing, heerCrowdsourcingGraphicalPerception2010, heerCrowdsourcingGraphicalPerception2010a tex.organization: ACM.

Hofmann H, Vendettuoli M (2013). Common angle plots as perception-true visualizations of categorical associations. (12): 2297–2305. Number: 12 00010.

Hullman J, Qiao X, Correll M, Kale A, Kay M (2019). In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1): 903–913.

Kintel M (2023). OpenSCAD.

Mozilla Foundation (????). WebGL: 2d and 3d graphics for the web.

Murdoch D, Adler D (2023). *rgl: 3D Visualization Using OpenGL*. Https://github.com/dmurdoch/rgl, https://dmurdoch.github.io/rgl/.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Spence I (1990). Visual psychophysics of simple graphical elements. 16: 683–692. Tex.ids= spenceVisualPsychophysicsSimple1990a number: 4 place: US publisher: American Psychological Association.

VanderPlas S, Hofmann H (2015). Signs of the sine illusion—why we need to care. 24(4): 1170–1190.

von Huhn R (1927). Further Studies in the Graphic Use of Circles and Bars. *Journal of the American Statistical Association*, 22(157): 31–36.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.