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

The use of 3D data visualizations has limitations when the third dimension does not convey any additional information to the viewer. Numerous studies advocate to avoid these types of graphs whenever possible, but these studies are almost entirely focused on the 2D projections of the 3D graphs. This paper describes the partial replication of a well-known paper in data visualization and its adaptation to focus on 3D printed bar charts. While current results from our study do not show differences between 2D and 3D graphs, we will use this study to provide students of a introductory statistics course hands-on experience with the research process.

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

1   Introduction

1.1   Elementary Graphical Tasks

Cleveland and McGill (1984) published a paper that sets up the foundation for how accurately people extract quantitative information using 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. They found that judgements for comparisons using position along a common scale had smaller log errors than for judgments using length comparisons. The study relied entirely on numerical estimation, which may not always be relevant when extracting information from graphs. Heer and Bostock (2010) replicated some parts of Cleveland and McGill (1984) in an online setting using Mechanical Turk, largely replicating the results of the original study while demonstrating the utility of the Mechanical Turk platform.

Further studies have shown that 3D graphs are less accurate at portraying numeric information that 2D graphs (Barfield and Robless, 1989; Fisher et al., 1997). In certain contexts and conditions, there is some research suggesting that 3D graphs may better encode information (Brath, 2014).

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 two-dimensional stimuli and can affect perception of error bands, candlestick plots,

---

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

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 the preference of using either 2D or 3D graphs and found that subjects tended to use simpler 2D graphs when tasked with extracting information. Barfield and Robless (1989) compared 2D and 3D graphs presented on paper and on computers. Their results showed that the accuracy of subject answers depended on their skill level. Novice subjects were more accurate with 2D paper graphs and experienced managers were more accurate with 3D computer graphs. For both experience levels, participants were more confident in their answers using 2D graphs over other presentations of data. There are instances where 2D graphs perform better than 3D graphs, but there are times where 3D graphs may better encode information. Brath (2014) highlights the intrinsic attributes of 3D graphs and its 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 2014. 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 the artificial charts into a more natural setting. As a result, it is reasonable to reconsider the use of 3D charts, not only because of technological developments, but because these charts provide the opportunity to make visual graphics accessible to those with limited or absent vision.

Here, we provide the process of replication and modernization of testing perceptual judgments to 2D graphs, 3D graphs projected in 2D environments, and 3D printed bar graphs.

1.3   Selected Components from Cleveland and McGill

Cleveland and McGill provided a theory and tested for the ordering of perceptual importance for the elements of length, position, and angle. Their first experiment, referenced as the position-length experiment, used five types of bar charts. Two of these were grouped bar charts and the other three were stacked bar charts. Each chart had two bars used for comparison and participants were asked to determine which bar was smaller and give their perceived ratio of the smaller bar to the larger bar. The two grouped bar charts are for the perceptual element of position along a common scale, where one has zero distance between bars and the other has a fixed distance between bars. These grouped bar charts will be referenced as adjacent and separated graph types in this paper, respectively.

Our study replicates the procedure for the comparisons of the two grouped bar charts, but with an objective of detecting differences in accuracy between 2D graphs, 3D digital graphs, and 3D printed graphs.

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.

Figure 1: Two dimensional, two-dimensional digital rendering, and 3D-printed charts used in this study.

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 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. The code for data generation and the data for the plots we used can be found here: `https://github.com/TWiedRW/2023-JDS-3dcharts`.

The aspect ratio of the plots is approximately 4:3.3, which was determined by measuring the pixels of a figure in Cleveland and McGill's paper.

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 1. There are some graphical elements that cannot be portrayed via 3D printing. For this reason, all graph types do not have axes, gridlines, 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 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 lighting was removed and replaced with three lights located in fixed positions around rendered figure. The lights were positioned so that one was behind the rendered figure (x: 910, y: -1000, z:1000), another in front of the figure (x: 780, y: 1030, z: 1000), and one light below the figure (x: 65, y: 15, z: -100). 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 graphs sufficient to create 21 kits of five 3d-printed charts each.

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) are heterosexual, and (3) have unemployed partners.

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 3
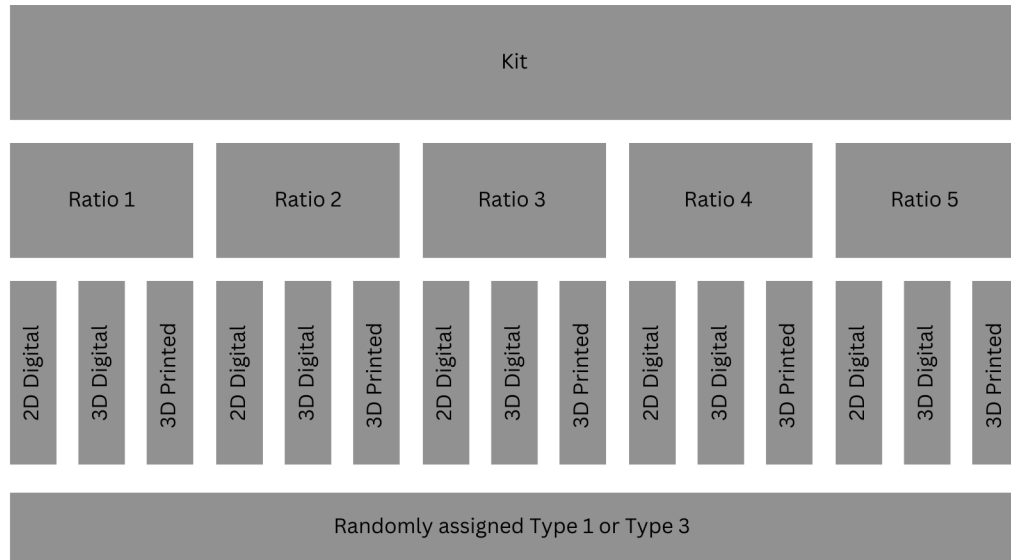
Figure 2: A graphical representation of the study design. Only five of the seven ratios were used in each kit where the ratios are randomly selected. A total of 21 kits were created to include all combinations of the five ratios.
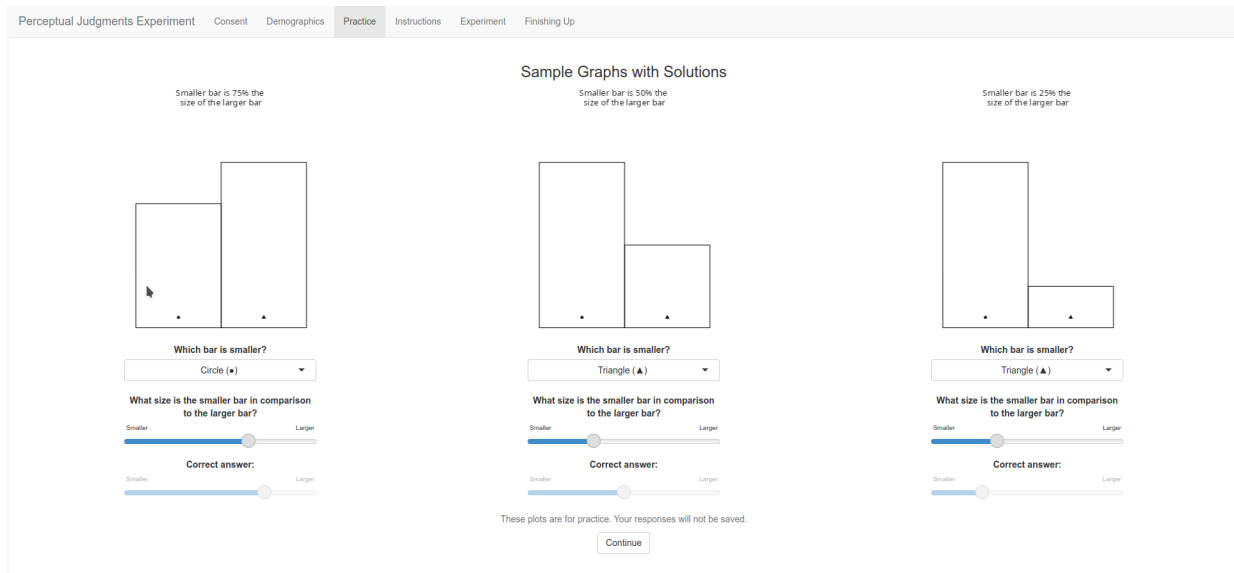

Figure 3: Screenshot of Shiny application practice screen. Three 2D bar charts with different ratios were provided, along with sliders indicating the correct proportion. Participants could practice with the sliders and preview the questions that would be asked as part of the task.
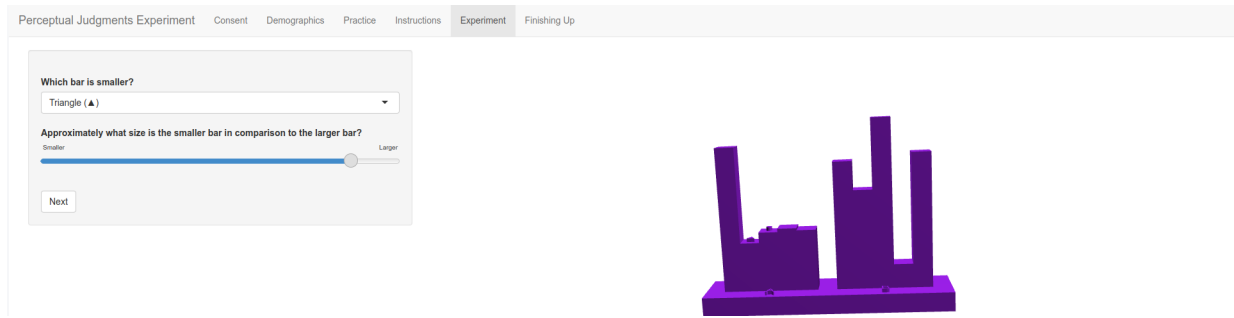
Figure 4: Screenshot of the applet collecting data for a 3D rendered chart task. Participants were asked to select which bar (circle or triangle) was smaller, and then to estimate the ratio of the smaller bar to the larger bar.

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 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 4.

3   Results

Cleveland and McGill used

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

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.

Our data was analyzed with the same log error that was used by Cleveland and McGill. We used a linear mixed model to account for variation attributed to the participants and to quantify effect sizes.

This is presented as the following model.

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

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 ith subject
- R_j is the effect of the jth ratio
- $G(R)_{(k)j}$ is the effect of the kth graph type nested in the jth ratio
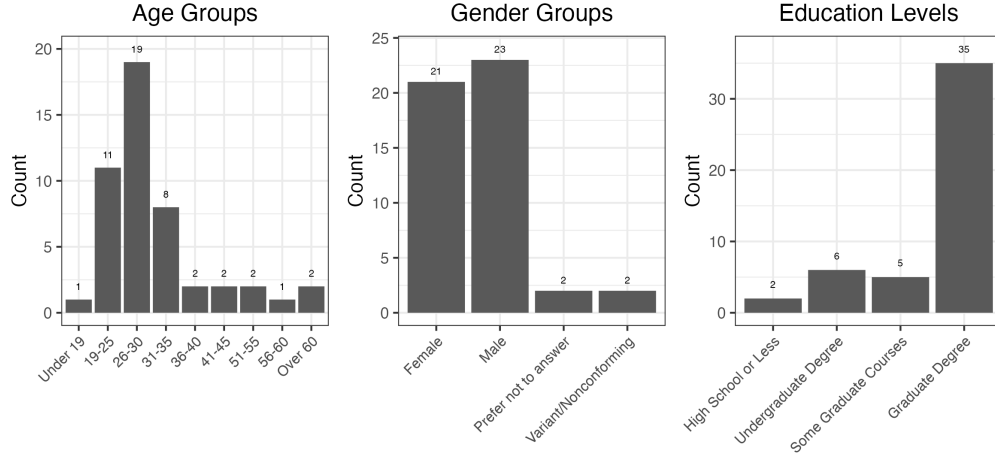- T_l is the effect of the lth comparison type

Figure 5: Demographic breakdown of participants in the study. All subjects were recruited from faculty and students in the statistics department at University of Nebraska-Lincoln.

- $\epsilon_{ijklm} \sim N(0, \sigma_\epsilon^2)$ is the random error

  A total of 49 subjects participated in the study.

  All responses that incorrectly identified the smaller bar were removed. No differences were detected for the true ratio of bars (p-value = .689), whether the bars were adjacent or separated (p-value = .220), or for the plot nested within the true ratio (p-value = .837).

  Figure 6 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.

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 is that this study is underpowered - the effect size is small, and our 49 participants were insufficient; the original study included 51 participants, which is slightly larger. 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 and provide the option for online participants to remove the 3D printed graphs from the graphs in their kit. Stimuli, code, and data for this experiment are provided at https://github.com/TWiedRW/2023-JDS-3dcharts along with instructions for replicating the analysis.
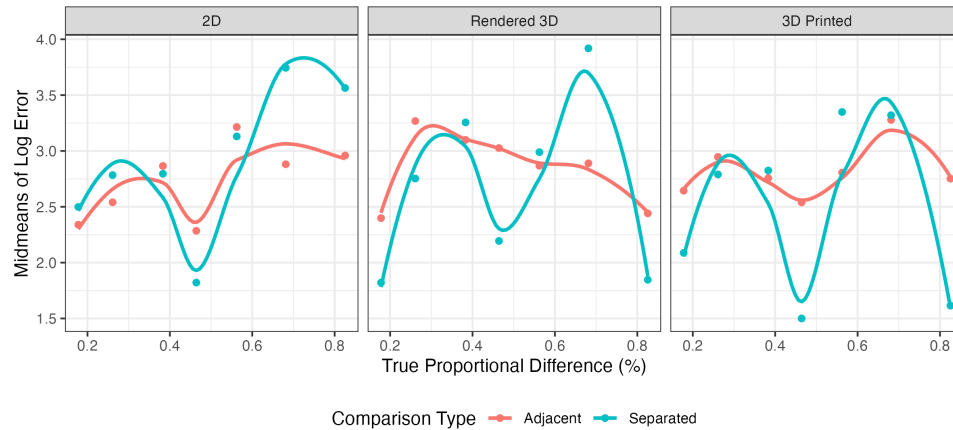
Figure 6: Midmeans of log absolute errors for the true ratio of bars. Each overlaying line are computed with loess.

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.

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

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.

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.

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

1170–1190.

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