Perceptually-Optimal Visualization Choice via Just Noticeable Differences

Lane Harrison

1 Introduction

Much of visualization design remains more art than science [1]. Years of visualization research has led to systems [4] and guidelines [7] that aid the designer in choosing visual representations based on general data characteristics such as dimensionality and data-type. Yet the designer still faces multiple ambiguous choices for creating visualizations that accurately communicate data.

We believe that experimental methods from vision science can be leveraged to disambiguate visualization choice at a perceptual level. Vision science has developed experiment methodologies that allow us to determine how much change in a visual scene (e.g. correlation coefficient changes in scatterplots) is required for a typical person to be able to perceive the change. Changes exactly large enough to be perceived by a typical person 75% of the time are referred to as *just-noticeable difference (jnd)*.

Through a series of experiments, we leverage jnd to build perceptual models that describe how different visualization designs (such as scatterplots and parallel coordinates) represent data commonly perceived data characteristics (such as correlation-coefficient). If the perceptual models for scatterplots and parallel coordinates differ, our results will demonstrate that a *perceptually-optimal* visualization can be chosen for a given dataset.

2 One-sentence description

Leveraging just-noticeable differences to create models of perceived quantities commonly communicated via visualizations, we demonstrate that a perceptually-optimal visualization can be chosen based on data characteristics.

3 Project Type

Experiment (crowdsourced)

4 Audience

> Who is the audience for this project? How does it meet their needs? What happens if their needs remain unmet?

If jnd allows us to make perceptually-optimal visualization choices, our results will impact anyone who wants to accurately communicate information via visual representations. Given a choice between two or more visual representations and a dataset to be shown, this technique will determine which visual representation is optimal for a certain type of perceptual judgment (e.g. correlation coefficient).

This extends the work of Mackinlay et al.'s Show Me work [4]. In their system, appropriate visualizations are chosen based on general data characteristics (e.g. dimensionality). In our system, however, we will further refine the list of visualizations by calculating commonly perceived values and choosing the optimal visual representation.

5 Approach

> What is your approach and why do you think its cool and will be successful?

A few years ago Ron Rensink published "The Perception of Correlation in Scatterplots" [5], where he described how jnd can be used to build a model of how well we can distinguish between correlation values. I believe this technique can be extended to disambiguate between visualization design choices. If different visualizations produce different perceptual curves, we can rank them on perceptual accuracy for a given dataset.
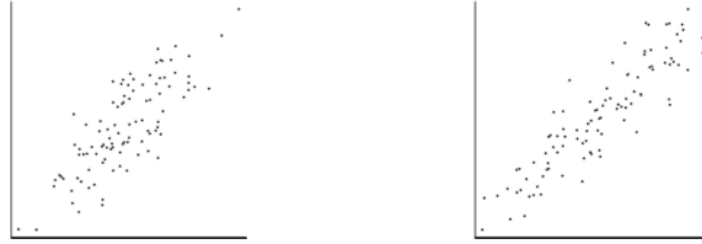
Figure 1: *Example of scatterplots used in the discrimination task. Observers were asked to choose which scatterplot is more highly correlated. In this example, base correlation is 0.8; jnd is from above.*

Figure 1: The technique Rensnik used to model how we perceive correlation in scatterplots.
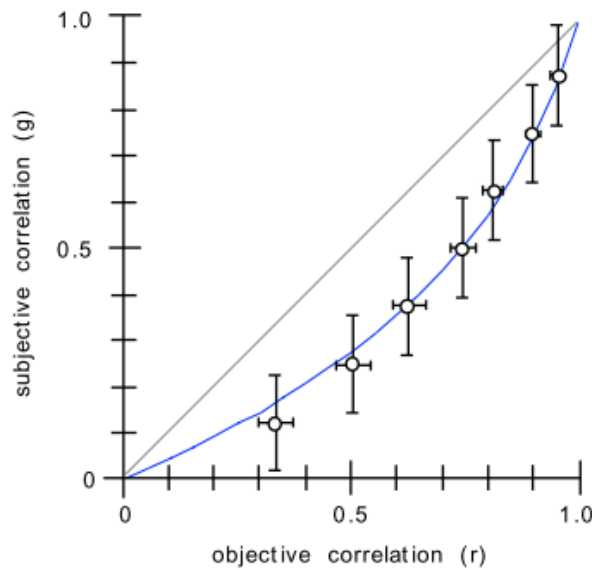

Figure 2: The resulting curve from Rensink's study, showing the perception of correlation (vs. actual correlation) in scatterplots.

The experiment approach Rensink used is illustrated in figure 1, and the resulting curve in figure 2.

6   Best-case Impact Statement

> In the best-case scenario, what would be the impact statement (conclusion statement) for this project?

In the best-case, we will show that the perception of correlation in scatterplots and parallel coordinates differ such that if we are given a dataset, we can compute a perceptually-optimal choice between them. In doing so, we further demonstrate that Rensink's original experiment is a generalizable technique for disambiguating visualization design choices at a perceptual level.

If successful, future experiments using the same technique may be able to answer key questions like how small a visualization can be while still allowing users to accurately perceive important metrics.

7   Major Milestones

- Replicate Rensink's results on Mechanical Turk.

- Determine a set of visualizations to compare: scatterplots, parallel coordinates, polar representations 2D data, and what else?

- Pilot to compare scatterplot vs. parellel coordinates correlation judgment.

- Determine a set of metrics to examine: correlation, outliers(?), data-size(?), width-height(?).

- Full experiment to compare all chosen visualizations across all chosen metrics.

8   Obstacles

8.1   Major obstacles

- Replicating Rensink's study on Turk may fail (due to variance in Turker judgments), or at least lead us to a significantly different curve. We would have to consider how our sample populations differ, and whether Turk is truly an appropriate platform for this type of perceptual study.

- In the pilot (where we compare scatterplots to parallel coordinates) we may find that the curves are too similar (effect size too small) for us to

make claims that adapting actually benefits end-users.

8.2 Minor obstacles

- The jnd experiment procedure Rensink uses is fairly complex, so it will require a non-trivial amount to setup and test.

- The original set of visualizations I wanted to compare (scatterplots of varying data size: regular, binned, alpha, jittered) ended up being an imperfect choice since each technique distorts the dataset, which may affect perception. The new set includes: scatterplots and parallel coordinates (horizontal/vertical variations), and polar representations ($\theta, radius$ and $radius, \theta$).

9 Resources Needed

What additional resources do you need to complete this project?

- Code to generate datasets for varying correlation-coefficients and data-sizes.

- Code to make scatterplots similar to the style of Rensink's.

- A thorough literature review of this space.

- Someone interested in running web-based experiments, especially with my existing experiment platform (which uses JavaScript, Node.js, d3, and R).

10 5 Related Publications

List 5 major publications that are most relevant to this project, and how they are related.

- Rensink applied JND to generate a model of how people perceive correlation in scatterplots [5].

- Mackinlay developed a system that chooses a visualization based on data characteristics [4].

- Wickham demonstrated a sort of visual t-test, where users had to distinguish between plots of the real data and null plots [6].

- Hoffman showed that cartesian representations are superior to polar representations by seeing how well participants could perceive different values between a lineup of charts [2].

- An earlier study from Jing compared how well people can perceive correlation in scatterplots versus parallel coordinates [3].

11 Define Success

When / How do you know if you have succeeded in this project?

If, after a pilot, we find that the perceptual curves for correlation in scatterplots are significantly different from those in parallel coordinates, we will have enough evidence to publish a paper that extends the Show Me work in a compelling way.

References

[1] W. S. Cleveland, P. Diaconis, and R. McGill. Variables on scatterplots look more highly correlated when the scales are increased. Technical report, DTIC Document, 1983.

[2] H. Hofmann, L. Follett, M. Majumder, and D. Cook. Graphical Tests for Power Comparison of Competing Designs. *IEEE Transactions on Visualization and Computer Graphics*, (12), 2012.

[3] J. Li, J.-B. Martens, and J. J. van Wijk. Judging Correlation from Scatterplots and Parallel Coordinate Plots. *IVI*, 2010.

[4] J. Mackinlay, P. Hanrahan, and C. Stolte. Show Me: Automatic Presentation for Visual Analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 2007.

[5] R. A. Rensink and G. Baldridge. The perception of correlation in scatterplots. *Computer Graphics Forum*, 2010.

[6] H. Wickham, D. Cook, H. Hofmann, and A. Buja. Graphical inference for infovis. *IEEE Transactions on Visualization and Computer Graphics*, 2010.

[7] J. Zacks and B. Tversky. Bars and lines: a study of graphic communication. *Memory & cognition*, 1999.