The Effects of Contrast on Correlation Perception in Scatterplots

Gabriel Strain[a,*], Andrew J. Stewart[a], Paul Warren[b], Caroline Jay[a]

[a]*Department of Computer Science, Faculty of Science and Engineering, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom*
[b]*Division of Psychology, Communication and Human Neuroscience, School of Health Sciences, Faculty of Biology, Medicine, and Health, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom*

Abstract

Scatterplots are common data visualizations that can be used to communicate a range of ideas, the most intensively studied being the correlation between two variables. Despite their ubiquity, people typically do not perceive correlations between variables accurately from scatterplots, tending to underestimate the strength of the relationship displayed. Here we describe a two-experiment study in which we adjust the visual contrast of scatterplot points, and demonstrate a systematic approach to altering the bias. We find evidence that lowering the total visual contrast in a plot leads to increased bias in correlation estimates and show that decreasing the salience of points as a function of their distance from the regression line, by lowering their contrast, can facilitate more accurate correlation perception. We discuss the implications of these findings for visualization design, and provide a framework for online, reproducible, and large-sample-size (N = 150 per experiment) testing of the design parameters of data visualizations.

Keywords: Scatterplot, Correlation Perception, Crowdsourced, Data visualization

This submission is approximately 7700 words.

1. Introduction

In one form or another, data visualizations have been used for thousands of years to aid analysis, to supplement narrative prose, and to communicate ideas (Azzam et al., 2013). Where once they were the preserve of those working directly with data, it is now expected that most professionals, and indeed many members of the public, are comfortable and familiar with an array of different data visualizations. The widespread adoption of data visualizations is positive for science as effective data visualizations can aid communication, but it also confers obligations, not only to design and communicate with honesty, but to also study how people understand and work with data visualizations.

In the last two centuries, the use of data visualizations has become increasingly common (Friendly and Denis, 2005; Azzam et al., 2013). The speed of the adoption of visualizations has meant rigorous scientific study of how they are comprehended by a viewer has often failed to keep pace. For many people, the COVID-19 pandemic has made data visualizations an everyday phenomenon (see BBC, 2022, for examples of the types of visualizations many saw daily). As data visualization designers, we have a duty to design in such a way that viewers with little to no formal statistical or data training can understand the message that visualizations are trying to convey.

In this paper we present a novel visualization technique that significantly increases the accuracy of people's performance on a correlation estimation task. In our first experiment, we show that manipulating the contrast of scatterplot points can bias participants' estimates of correlation. In our second we leverage

---

this effect to partially correct for a systematic underestimation by viewers of correlations in scatterplots that has a long-standing basis in the literature. Through this work we also present a framework for the effective, inexpensive, and high-powered study of data visualizations.

1.1. Why Study Scatterplots?

Scatterplots, estimated in 1983 to account for between 70 and 80 percent of data visualizations in scientific publications (Tufte and Graves-Morris, 1983), are simple representations of bivariate data that people generally interpret in the same way (Kay and Heer, 2015). Rensink (2014) identifies them as similar to a fruit fly for geneticists; simple enough to be easily studied, but complex enough to provide interesting insights. These features make them ideal candidates for controlled, empirical study, in addition to providing us with insights into visualization design and perception. Designers of data visualizations have a duty to design in such a way that viewers with limited statistical, data, or graph training can be expected to correctly interpret the messages that visualizations are trying to convey. It is for this reason that we have chosen our population; lay participants with varying levels of statistical training and graph literacy, as these are the people that we should be designing for.

2. Related Work

2.1. Testing Correlation Perception

The most intensively studied aspect of scatterplots is correlation perception, (i.e. the strength of the relationship between two variables), although they are used for a wide variety of other tasks, including cluster separation, outlier detection, and trend identification (Behrisch et al., 2018). Throughout this paper we refer to an r value, or the Pearson product moment correlation coefficient. Pearson's r takes a value between 0 and $|1|$, and is positive or negative depending on the direction of the relationship between the two variables in question.

Scatterplots have been extensively studied in a variety of experimental paradigms. Very early work (Pollack, 1960) asked participants to make discriminative judgements between scatterplots with different correlations, and found that people were more easily able to discriminate as the magnitude of the r value increased. Subsequent work focused on asking participants to provide a numerical estimate of the r value, with studies finding evidence for a systematic underestimation for positive r values apart from 0 and 1. In several studies this effect was particularly pronounced for $0.2 < r < 0.6$ (Strahan and Hansen, 1978; Bobko and Karren, 1979; Cleveland et al., 1982; Lane et al., 1985; Lauer and Post, 1989; Collyer et al., 1990; Meyer and Shinar, 1992), see Figure 1 for an approximation of the underestimation observed. Micallef et al. (2017) tested an automated system for varying scatterplot aspect ratio, point size, and point contrast, finding that a 1:1 aspect ratio produced the best performance on a correlation estimation task, and reporting no effect of adjusting point size or contrast, although this was carried out uniformly to deal with overplotting issues. In addition to studies employing discriminative judgement or direct estimation tasks, several more recent investigations have employed a combination of bisection tasks, in which participants are asked to adjust a test plot so that its correlation is halfway between two reference plots, and a staircase discriminative judgement task that allows researchers to find the just-noticeable-difference (JND) between scatterplots such that their correlations are distinguishable 75% of the time. This novel approach (Rensink and Baldridge, 2010) allowed researchers to obtain measurements for participants' precision and accuracy in correlation estimation, and to begin to fit mathematical models that describe the relationship between objective and perceived correlation. Given that this paper seeks to provide design guidelines, and is interested in comparative, naturalistic judgements of correlation, we have elected to use a direct estimation paradigm.

2.2. What Drives Correlation Perception?

Several key pieces of evidence point to correlation perception being driven by the shape of the probability distribution relayed by the points. A study investigating the effect of increasing the x and y scales on scatterplots (thereby decreasing the size of the point cloud, Cleveland et al., 1982) found that a viewer's judgement of the association between the two variables increased as the size of the point cloud decreased,
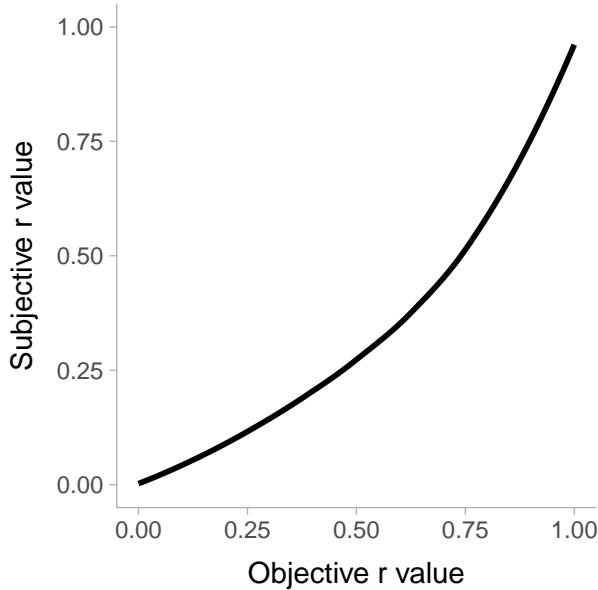
Figure 1: Using a function that relates objective to subjective r supplied in Rensink (2017) allows us to visualize the nature of the underestimation curve found in correlation perception studies. The curve represents the underestimation of correlation.

despite the r value remaining the same between conditions. The authors suggest this may be due to participants using the area of the point cloud, or the ratio of the minor and major axes of it, to judge association. Decreasing the size of the point cloud here also had the effect of narrowing the width of the distribution displayed, as the length of the minor axis decreased.

Another study asked participants to provide estimates of correlation in scatterplots (Meyer et al., 1997). It found that the relationship between objective and perceived r values could be accurately described by a function that included the mean of the geometrical distances between the points and the regression line. This is intuitive, as scatterplots with higher correlation will generally have lower average distances between their points and the regression line.

A more recent study investigating the hypothesis that people use visual features to judge correlation (Yang et al., 2019) found evidence that several visual features were predictive of correlation estimation performance. Among these was the standard deviation of all perpendicular distances from the points to the regression line, a quantity similar to that in Meyer et al. (1997), which on an individual level was more predictive of participants' estimates of correlation than the objective r value itself.

Bringing together work that has sought to model the relationship between objective and perceived r values, Rensink (2017) notes that equations for both discrimination and magnitude estimation include a parameter, termed u, that is small when $r = 1$, and increases as r approaches 0. The utility of this parameter in modelling correlation perception is indifferent to the type of data visualization used, which implies that the width of the probability distribution, summarised by the aforementioned parameter, is key to how people estimate correlation. Within the context of scatterplots however, this parameter can also be expressed as the average distance between the points and the regression line (the X parameter in Meyer et al., 1997).

None of the above is proof that people are using *only* the mean or standard deviation of geometrical distances between the points and the regression line to estimate correlation. However, taken with findings that correlation is perceived rapidly by viewers (Rensink, 2014), what we have discussed thus far suggests that this parameter is at least a good proxy for what people are really attending to, insofar as changing it has the ability to influence how people estimate correlation. From this evidence, a good candidate for influencing people's perceptions of correlation is changing the perceived width of this probability distribution by changing the perceived distance between points and the regression line.

2.3. Transparency

Adjusting the contrast in scatterplots has been used extensively to solve problems of overplotting or clutter (Matejka et al., 2015; Bertini and Santucci, 2004), in which scatterplots with very large numbers of data points suffer from visibility issues caused by excessive point density. Lowering the contrast of all points makes the underlying distributions and trends much easier to discern for the user. Increasing the transparency leads to a reduction of the contrast of isolated points with the background, and for regions with overlapping points color intensities are summed. Our stimuli had only 128 small points, meaning the majority of points were clearly visible at all times, and we do not consider the small degree of overlap problematic. For this reason, and due to the fundamentality of contrast as a feature of human visual perception (Ginsburg, 2003) we elected to use the term contrast to describe our point manipulations. The approach we describe may not be useful with much larger datasets where clutter becomes problematic.

2.4. Contrast

Despite the popularity of adjusting contrast to address overplotting issues, little investigation has taken place into the effects of reducing point contrast on people's perceptions of correlation; what has been found is that correlation perception seems to be invariant to changes in point contrast (Rensink, 2012), although this work took place with small sample sizes (n = 12), and using only bisection/JND methodologies.

Changing the contrast of a visual stimulus effectively reduces the strength of that signal. A likely consequence of this is increased uncertainty in aspects of that stimulus, for example the locations of points in scatterplots. Consequently, one might anticipate that increased noise could lead to altered perception of correlation and/or more noise in correlation estimates due to effects on the perceived position of points within the cloud. While there is indeed evidence (Wehrhahn and Westheimer, 1990) that the perception of stimulus position becomes exponentially worse as contrast is reduced (as measured by vernier acuity tasks), this is only true for a narrow range of low contrasts just above the detection threshold. For higher contrasts vernier acuity appears largely robust to such changes. Nonetheless, there is clear evidence that other perceptual estimates become more uncertain with reduced contrast, for example speed perception (Champion and Warren, 2017). With this in mind, and the relatively small sample size used in Rensink (2012), we suggest that the effects of contrast on perceived correlation warrant further investigation.

A recent study (Hong et al., 2021) used contrast and size to encode a third variable in trivariate scatterplots. The authors then asked participants to use a mouse to click on the average position of all the points displayed. They found that participants' estimates of average point position were biased towards larger or darker points, which the authors term the *weighted average illusion.* Together with evidence that darker and larger points are more salient (Healey and Enns, 2012), the implication is that we can use contrast to reduce the salience of the points representing the widest parts of the probability distribution; if this is successful, and participants perceive a narrower distribution, we would expect this to be able to correct for a viewer's underestimation bias.

One way to correct for an underestimation in correlation would be to simply remove outer data points until correlation perception is aligned with the actual correlation value. However, this would necessitate hiding data and thus changing the information presented to the viewer. An alternative approach is to manipulate the contrast of only some of the points; it would seem most sensible to do so for the points that are more extreme relative to the underlying regression line.

In the present study we address the issues raised above in two online experiments with large sample sizes. In the first we consider the effects of point contrast over the entire scatterplot on correlation estimates. In the second experiment we examine how changing contrast as a function of distance to the regression line affects perceived correlation. To pre-empt our results we find clear effects of both manipulations.

3. General Methods

3.1. Formalising Contrast

We use the **ggplot2** (Wickham, 2016) package for plot creation in both experiments, which uses an alpha parameter to set contrast. Alpha here refers to the linear interpolation (Stone, 2008) between foreground and
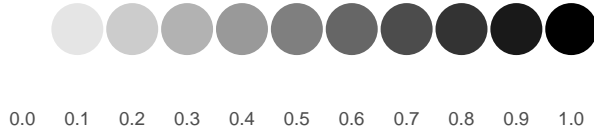
Figure 2: The relationship between alpha values and rendered point contrast. Higher alpha values result in greater contrast between the foreground (scatterplot point) and background. When alpha = 0, the foreground is ignored and the background is rendered.

background pixel values; alpha values of 0 (full transparency) and 1 (full opacity) result in no interpolation and rendering of either the background or foreground pixel values respectively. Alpha values between 0 and 1 correspond to different ratios of interpolation between foreground and background pixel values.

There are numerous psychophysical definitions of perceived contrast (Zuffi et al., 2007) based on what is being presented, for example models that take into account visibility limits (CIELAB lightness), and contrast in periodic patterns such as sinusoidal gratings (Michelson's contrast). The common thread running through these definitions is the use of the ratio between target and background luminances. Our experiment was fully online, with participants completing it on their personal laptop or desktop computers. This meant we had no control over the exact luminances of our stimuli, only over the *relative* luminance between targets (scatterplot points) and backgrounds. Given that we are interested in *relative* differences in correlation perception averaged over a series of 180 single-plot presentation trials, we do not consider this a shortcoming. In light of this however, it would be inappropriate to report absolute luminance values. Instead, we simply report the alpha value, which is representative of the luminance ratio. Figure 2 illustrates the contrasts created by alpha values between 0 and 1. For clarity, we henceforth refer to the alpha value as "contrast alpha" throughout.

3.2. Overview of Experiments

Experiments 1 and 2 share multiple aspects of their procedures. Both experiments were built using PsychoPy (Peirce et al., 2019), and hosted on pavlovia.org. Both use 1-factor, 4-level designs. Participants were only permitted to complete the experiments on a desktop computer or laptop. As with the luminances of scatterplots and their points, our crowd-sourced approach renders the measurement of participant-to-monitor distance and the recording of exact apparatus impossible. We do not consider this a shortcoming however, as it allows for findings that are robust to various displays and viewing contexts and allows us to make conclusions that are of particular relevance to the HCI audience.

Ethical approval for both experiments was granted by the University of Manchester's Computer Science Department Panel (Ref: 2022-14660-24397). Each participant was shown the participant information sheet (PIS) and provided consent through key presses in response to consent statements. They were then asked to provide their age in a free text box, and their gender identity. Following this they were asked to complete the 5-item Subjective Graph Literacy (SGL) test (Garcia-Retamero et al., 2016). Participants were then shown 7 instructional slides, the text for which can be seen in the experimental repository (https://gitlab.pavlovia.org/Strain/exp_uniform_adjustments/blob/master/instructions.csv). Ad-hoc piloting with a graduate student in humanities suggested people might be unfamiliar with what different correlations looked like in scatterplots. They were therefore shown examples of $r = 0.2$, 0.5, 0.8, and 0.95, which can be see in Figure 3. See Section 6.1 for a discussion of the potential effects of this. Participants were then given two practice trials before the experiment began.

Each trial was preceded by text that either told the participant "Please look at the following plot and use the slider to estimate the correlation" (in black, experimental trial), or "Please IGNORE the correlation displayed and set the slider to 1"[1] (n = 3) or "Please IGNORE the correlation displayed and set the slider to

---

[1]The word "ignore" was in lowercase in the first experiment, but was changed to uppercase in the second to increase the salience of attention check items.
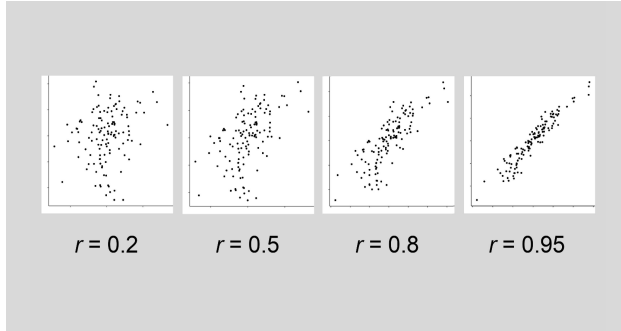
Figure 3: Participants viewed this for at least 8 seconds before being allowed to continue onto the practice trials.
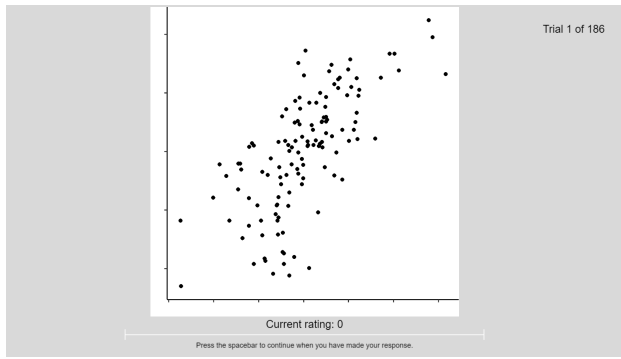

Figure 4: An example of an experimental trial.

0" (n = 3) (in red, attention check trial). Each plot was preceded by a visual mask displayed for 2.5 seconds. Figure 4 shows an example of an experimental trial. There was no time limit per trial, but participants were instructed to make their judgements as accurately and quickly as possible.

Both experiments described here use a fully repeated measures, within-participants design. Participants saw all 180 plots, corresponding to ~ 27,000 individual judgements per experiment. Presentation order was randomised.

3.3. Plot Generation

Scatterplots were randomly generated from bivariate normal distributions, being generated from data with standard deviations of 1 in each direction. All plots presented had a 1:1 aspect ratio, were generated as 1200x1200 pixel .png images, and were then scaled up or down according to the participant's monitor, always occupying the same proportion of monitor space. Plots included no titles, tick labels, or axis labels. Points were 10 pixels in diameter on a 1280x720 resolution monitor, and 13 pixels in diameter on a 1920x1080 resolution monitor. Due to our crowdsourced approach we were unable to obtain monitor resolution or physical screen size measurements for participants, however we are working on a method to accomplish this in future work. Each plot contained 128 data points. 45 r values were generated on a uniform distribution between 0.2 and 0.95. The same dataset was used for plot generation in both experiments. We chose these boundaries due to evidence that very little correlation is perceived below 0.2 (Bobko and Karren, 1979; Cleveland et al., 1982; Strahan and Hansen, 1978). Scripts detailing item and mask creation for each experiment can be found in the item_preparation folder.

3.4. Open Research Statement

Both experiments were conducted according to the principles of open and reproducible research. All data and analysis code are available at https://github.com/gjpstrain/contrast_and_scatterplots. This repository
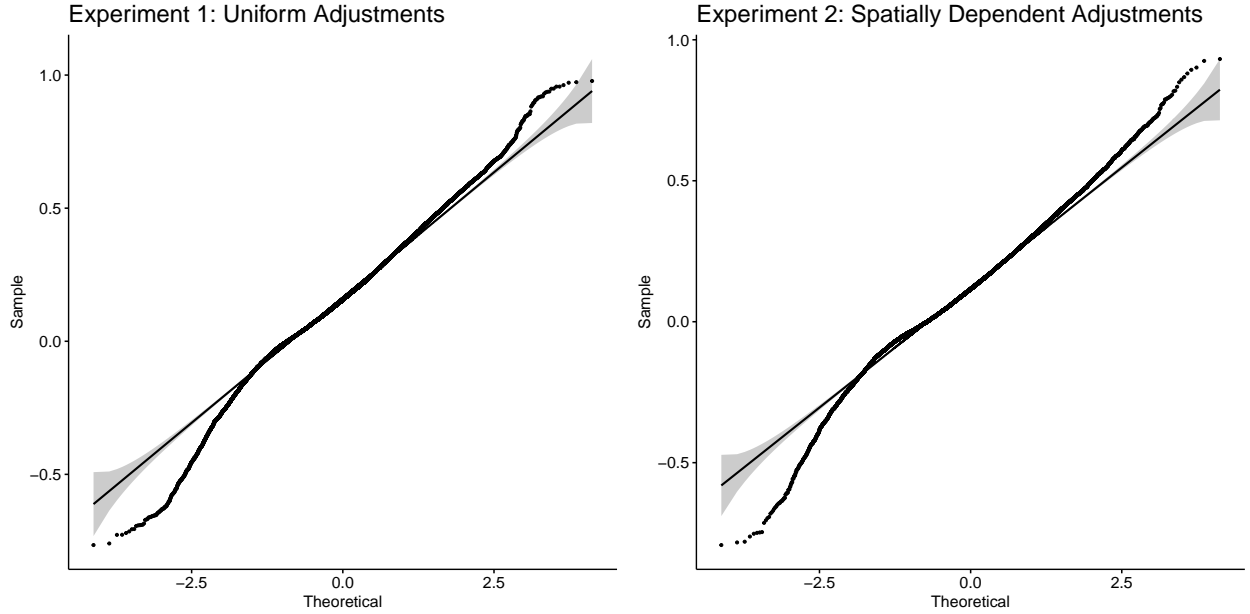
6

Figure 5: qqplots showing the distribution of errors in experiments 1 and 2.

contains instructions for building a Docker container that fully reproduces the computational environment this paper was written in, allowing for a full replications of stimulus generation, analyses, and the paper itself. Both experiments and their related hypotheses and analysis plans were pre-registered with the OSF (https://osf.io/v23e9/), and there were no deviations from them.

3.5. Modelling

We used linear mixed-effects models to model the relationships between our independent variables (point contrast in experiment 1 and the type of contrast decay function in experiment 2) and our dependent variable (participants' estimates of correlation). Linear mixed-effects models allow us to compare differences between levels of our independent variable across the full range of participant responses on the dependent variable, as opposed to simply analysing the differences between means that would be afforded to us with ANOVA. Linear mixed-effects models also allow us to include random effects for both participants and experimental items in our modelling. Consistent with our pre-registrations, we aimed for the most complex random effects structures when producing models. The structure of these models was identified using the **buildmer** package in R (Voeten, 2023). This package takes the most complex random effects model, in terms of intercepts for participants and items and corresponding slopes for fixed effects terms as an input. It then identifies the most complex model that successfully converges, dropping terms that fail to explain a significant amount of variance as assessed with likelihood ratio tests. This provides a simple and reproducible methodology for the construction of linear mixed-effects models. This approach does mean that the final model used is not always the most complex one possible, but rather is the most complex that substantially explains variance and converges.

3.6. Normality

Figure 5 shows qqplots testing the normality of participants' errors in correlation estimation for both experiments. While error is not entirely normally distributed, the size of our sample renders this a non-issue with regards to statistical testing (Ghasemi and Zahediasl, 2012), so no transform was used.

4. Experiment 1: Changing the Contrast of All Points

4.1. Introduction

Our first experiment varied the contrast of every point on a scatterplot in a uniform manner. Given the effects of contrast on perception described above (Champion and Warren, 2017; Wehrhahn and Westheimer, 1990), we hypothesized that there would be a more variable spread of correlation estimates for plots with lower contrast compared to plots with higher contrast, potentially due to the greater spatial uncertainty induced by lower contrast.

4.2. Method

4.2.1. Participants

150 participants were recruited using the Prolific.co platform. Normal to corrected-to-normal vision and English fluency were required for participation. In accordance with guidelines published in Peer et al. (2021), participants were required to have previously completed at least 100 studies on Prolific, and were required to have a Prolific score of at least 100, indicating acceptance on at least 100/101 studies previously completed. In addition, participants who had completed an earlier, similar study that was run by the authors of the current paper were prevented from participating.

Data were collected from 158 participants. 8 failed more than 2 out of 6 attention check questions, and, as per pre-registration stipulations, were rejected from the study. The remaining 150 participants' data were included in the full analysis (51.01% male, 47.65% female, and 1.34% non-binary). Mean age of participants was 28.29 ($SD = 8.59$). Mean graph literacy score was 21.76 ($SD = 4.47$) out of 30. The average time taken to complete the experiment was 33 minutes (SD = 10 minutes).

4.2.2. Design

For each of the 45 r values there were four versions of each plot corresponding to the four levels of point contrast, examples of which can be seen in Figure 6. The experiment is hosted at https://gitlab.pavlovia.org/Strain/exp_uniform_ This repository contains all the experimental code, materials, and instructions needed to run the experiment in full.

4.3. Results

All analyses were conducted using R (version 4.2.2, R Core Team, 2022). Models were built using the **buildmer** (version 2.8, Voeten, 2023) and **lme4** (version 1.1-31, Bates et al., 2015) packages, with contrast condition being set as the predictor for participants' error in correlation estimates.

Figure 7 shows the mean error in correlation estimation for the four contrast conditions. A likelihood ratio test revealed that the model including contrast as a fixed effect explained significantly more variance than a model not including contrast as a fixed effect ($\chi^2(3) = 224.25$, $p < .001$). This model has random intercepts for items and participants. This effect was driven by: low contrast scatterplots (contrast alpha = 0.25) being rated on average as having lower correlation than medium contrast (contrast alpha = 0.5), high contrast (contrast alpha = 0.75), and full contrast (contrast alpha = 1) plots; and medium contrast plots being rated on average as having lower correlation than high and full contrast plots. There was no significant difference in correlation estimates between high and full contrast plots.

Statistical tests for contrasts between the four levels of the contrast condition were performed with the **emmeans** package (version 1.8.4-1, Lenth, 2023) and are shown in Table 1. Means and 95% confidence intervals of correlation judgements are shown in Figure 7. The **EMAtools** package (version 0.1.4, Kleiman, 2021) was used to calculate effect sizes in Cohen's d. For the difference in correlation ratings between the lowest contrast (contrast alpha = 0.25) and highest contrast plots (contrast alpha = 1), an effect size of d = -0.17 was obtained. While this effect is small, it is not statistically insignificant, and given the lack of reported effect on correlation perception of global point contrast adjustments (Rensink, 2012) is an unsurprising result.

We also generate an additional model to test whether the results we found could be explained by differences in graph literacy. This model is identical to the experimental model, but includes graph literacy as a fixed effect. We found no significant differences between the original model and the one including graph literacy as
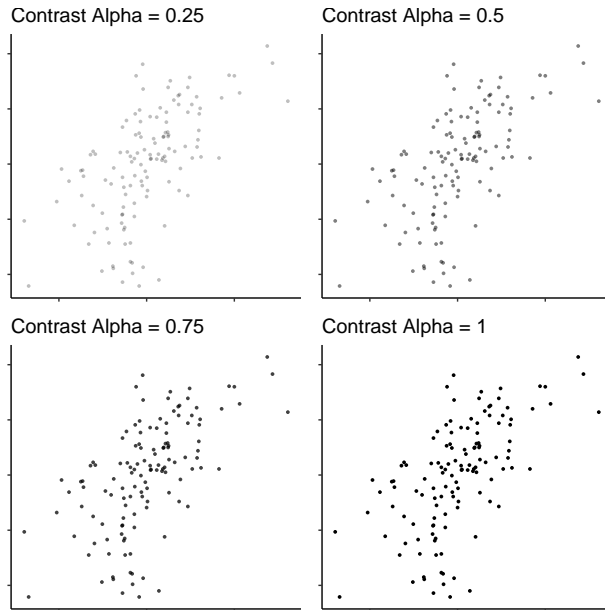
Figure 6: The four levels of the contrast condition in experiment 1, demonstrated with an r value of 0.6. Some points overlap, resulting in increased contrast in those areas, although this effect is negligible
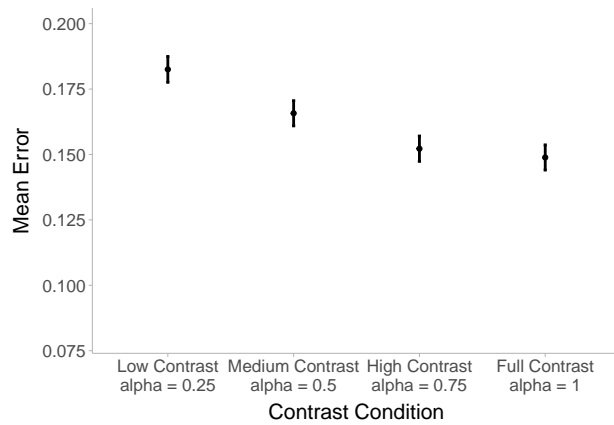

Figure 7: Mean error in correlation estimation across the four contrast conditions in E1, with 95% confidence intervals shown.

Table 1: Table showing the contrasts between different levels of the contrast factor in E1.

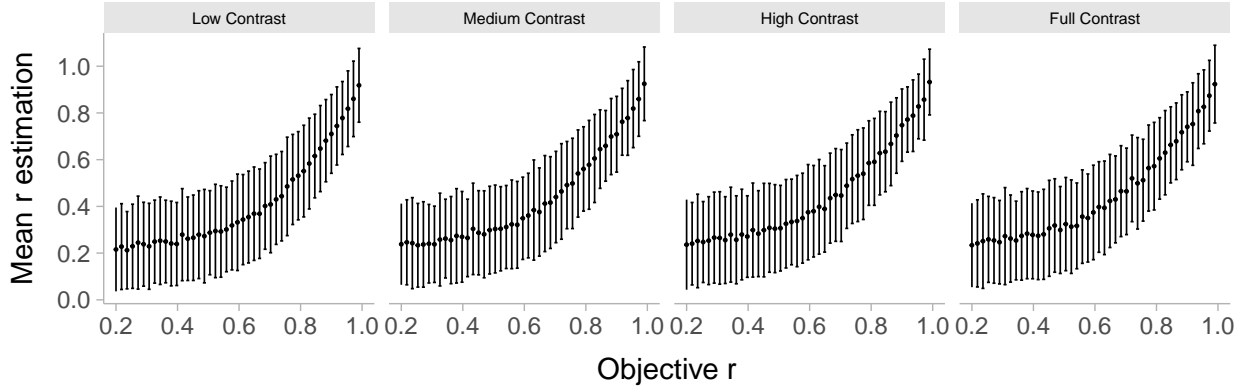| Contrast	Z.ratio	p.value
Low Contrast : Full Contrast	13.48	<0.001
Low Contrast : High Contrast	12.12	<0.001
Low Contrast: Medium Contrast	6.72	<0.001
Full Contrast : High Contrast	-1.35	0.528
Full Contrast : Medium Contrast	-6.76	<0.001
High Contrast : Medium Contrast	-5.40	<0.001

Figure 8: The mean of participants' estimates of r compared with objective Pearson's r in experiment 1, plotted by contrast condition. Error bars show standard deviations of mean estimates.
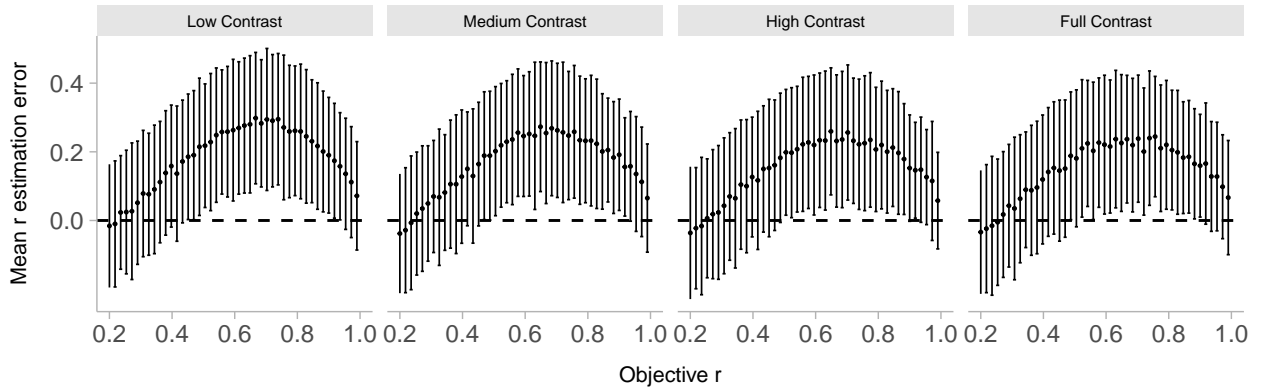

Figure 9: Plots showing how participants' mean correlation estimation errors change with Pearson's r in experiment 1. Participants were on average most accurate at estimating the r value when presented with the full contrast scatterplots (condition D). Error bars show standard deviations of estimates.
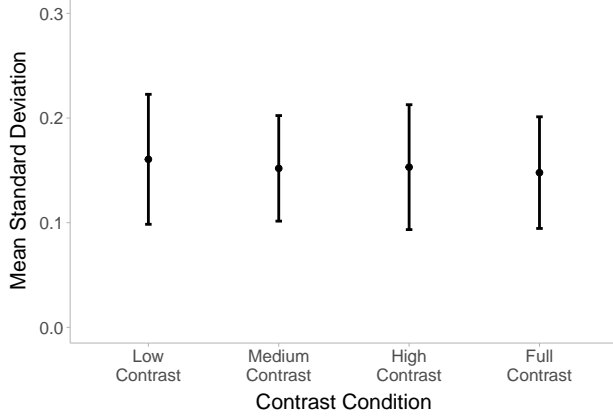
Figure 10: The mean of standard deviations of participants in experiment 1, grouped by contrast condition. Error bars show spreads of standard deviations.

a fixed effect ($\chi^2(1) < .001$, $p = .995$). These results suggest that the effect we found was not driven by differences in graph literacy between participants.

Figure 8 shows how participants' mean estimates of correlation change with the objective Pearson's r value, plotted separately for each contrast condition. We observe underestimation curves similar to those reported in previous literature (see Section 2.1).

4.4. Discussion

Our hypothesis was not supported in this experiment. We hypothesized that plots with lower contrast would have a wider spread of correlation estimates than plots with higher contrast. As shown in Figure 10, there was little difference in mean standard deviations between the four contrast conditions. Participants' errors in correlation estimation were significantly higher when the contrast of all points was lower compared to when it was higher. This was true up until contrast alpha was set to 0.75, implying a threshold around contrast alpha = 0.75 past which there is little variation in the perception of contrast, at least as far as it is associated with correlation estimation. This lack of significant difference in correlation estimation between our two highest contrast values fits with the logarithmic nature of contrast/brightness perception (Varshney and Sun, 2013; Fechner, 1948); despite there being equal linear distance between the contrast values we used, the perceptual distance between them was clearly non-linear.

As mentioned previously, Rensink (2012) and Rensink (2014) describe what is, to our knowledge, the only other direct testing of correlation perception and point contrast, and reports no difference in either bias or variability in correlation perception with regards to contrast manipulations. Our results, in comparison to these previously reported findings, do find an effect, which can be quantitatively explained by differences in experimental power, although we argue that methodological differences may have played a role also. Given the small effect size (d = -0.17), the small sample in Rensink (2014) (n = 12) would have made finding the effect very unlikely, having a power of only .08. The current study, having a power of 0.54 and large sample of n = 150, has found the effect. While differences in power can explain this discrepancy, we also argue that our methodology is more representative of the normal usage of scatterplots, and is therefore more suited to informing scatterplot design than investigating the mathematical relationship between correlation and perceived correlation. While our effect size is small, and may be of little practical relevance with regards to correcting for the correlation underestimation effect in scatterplots, it demonstrates that differences in the contrasts of points *can* affect correlation estimates in scatterplots. The fact that this difference is present suggests that it can be used to optimise perception, which we start attempting to do in experiment 2. We suggest that correlation perception functions similarly to speed perception (Champion and Warren, 2017) with regards to changes in contrast; the greater spatial uncertainty brought on by reduced contrast, while

not eliciting the greater spread in correlation estimates that we hypothesized, might be responsible for the results observed via an increase in the perceived width of the probability distribution in the plot.

From our results it is clear that a scatterplot optimized for correlation perception should have maximum contrast between the foreground (points) and background. That we found significant differences in correlation estimation between data-identical scatterplots with different point contrasts however, suggests that we might be able to leverage this effect to further improve participants' estimates of correlation.

5. Experiment 2: Spatially-Dependent Contrast Adjustments

5.1. Introduction

In experiment 1 we found that contrast has a clear effect on the perception of correlation such that scatterplots with higher levels of point contrast are rated as being more correlated. Given this result, the important question arises of whether we might see additional changes in correlation perception as a function of the spatial arrangement of point contrast in scatterplots. With this question in mind, we hypothesized that participants' estimates of correlation would exhibit lower mean error with the decay parameter in which contrast falls with residual distance in a non-linear fashion (the non-linear decay parameter), and that the use of a non-linear inverted decay parameter, in which contrast increased with residual distance, would result in higher mean errors than all other conditions.

5.2. Methods

5.2.1. Participants

150 participants were recruited using the Prolific.co platform. Normal to corrected-to-normal vision and English fluency were required for participation. In accordance with guidelines published in (Peer et al., 2021), participants were required to have previously completed at least 100 studies on Prolific, and were required to have a Prolific score of at least 100, indicating acceptance on at least 100/101 previously completed studies. In addition, participants who had completed an earlier, similar study that was run by the authors of the current paper, or the first experiment described above were prevented from participating.

Data were collected from 157 participants. 7 failed more than 2 out of 6 attention check questions, and, as per pre-registration stipulations, were rejected from the study. The remaining 150 participants data were included in the full analysis (51.33% male, 46.00% female, and 2.67% non-binary). Mean age of participants was 27.05 ($SD = 7.37$). Mean graph literacy score was 21.71 ($SD = 4.06$) out of 30. The average time taken to complete the experiment was 33 minutes (SD = 13 minutes).

5.2.2. Design

For each of the 45 r values there were four versions of each plot, which can be seen in Figure 11. Three used functions relating point residuals to point contrast, which we refer to as decay parameters, with the remaining condition being the contrast alpha = 1 uniform contrast condition from the first experiment. We used the equation 1 to non-linearly map residuals to alpha values, where R is the residual of the point in question. 0.25 was chosen as the value of b. Given the shape of the underestimation curve reported in previous literature (see Figure 1), intuition suggested that we use a symmetrically opposing curve (see Figure 12) to relate point contrast with residuals. We tested a number of values of b, and felt that 0.25 rendered plots that maintained point legibility while also allowing a large enough range in contrast that, if an effect was present, one would be found.

$$alpha = 1 - b^R \tag{1}$$

Figure 12 illustrates the relationship between the size of a residual and the contrast produced. The experiment is hosted at https://gitlab.pavlovia.org/Strain/exp_spatially_dependent. This repository contains all the experimental code, materials, and instructions needed to run the experiment in full.
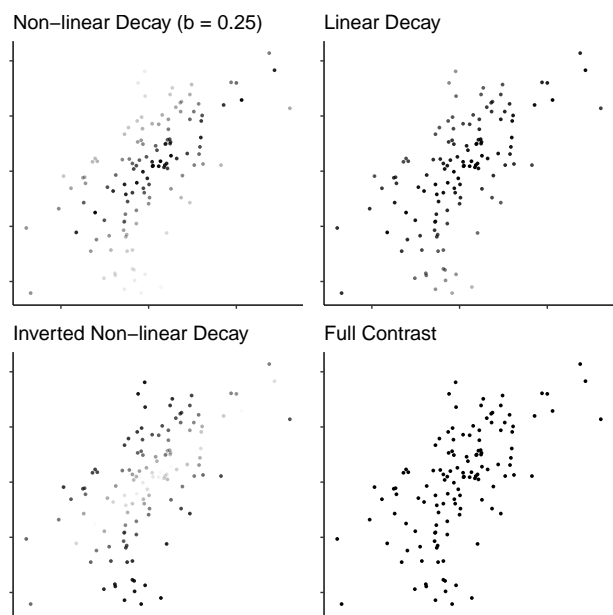
Figure 11: The four levels of the contrast condition in experiment 2, demonstrated with an r value of 0.6.
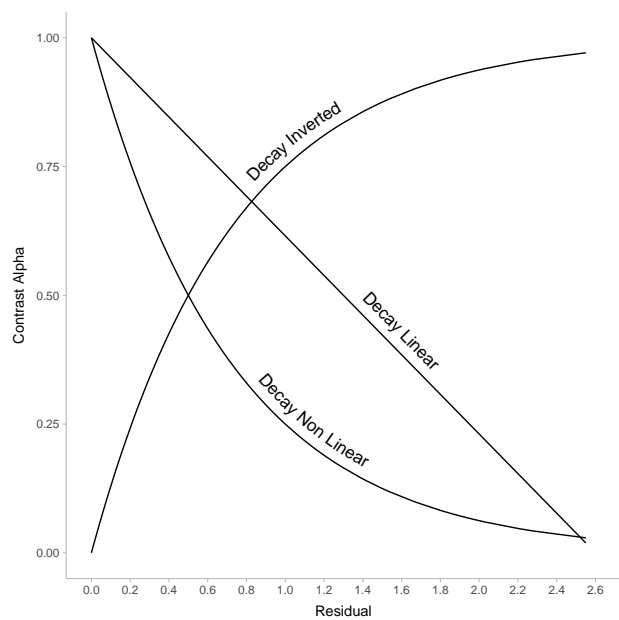

Figure 12: Here we use an r value of 0.2 to demonstrate the relationship between the size of a point's residual and the contrast alpha value that translates to.
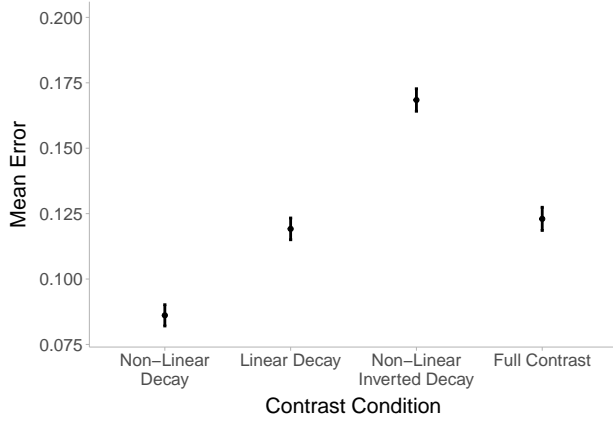
Figure 13: Mean error in correlation estimation across the four contrast conditions in E2, with 95% confidence intervals shown

Table 2: Table showing the contrasts between different levels of the contrast condition factor.

Contrast	Z.ratio	p.value
Full Contrast : Non-linear Inverted Decay	-18.88	<0.001
Full Contrast : Linear Decay	1.55	0.405
Full Contrast : Non-linear Contrast Decay	15.30	<0.001
Non-linear Inverted decay : Linear Decay	20.43	<0.001
Non-linear Inverted decay : Non-linear Contrast Decay	34.17	<0.001
Linear Decay : Non-linear Contrast Decay	13.74	<0.001

5.3. Results

All analyses were conducted using R (version 4.2.2, R Core Team, 2022) As in experiment 1, models were built using the **buildmer** (version 2.8, Voeten, 2023) and **lme4** (version 1.1-31, Bates et al., 2015) packages, with contrast decay function being set as the predictor for participants' error in correlation estimates.

Figure 13 shows the mean correlation estimation errors for the four contrast conditions. A likelihood ratio test revealed that the model including contrast condition as a fixed effect explained significantly more variance than a model not including contrast as a fixed effect ($\chi^2(3) = 1{,}157.62$, $p < .001$). This model has random intercepts for items and participants. This effect was driven by participants' correlation estimates being on average more accurate for the non-linear decay parameter than for the linear decay parameter, non-linear inverted decay parameter, and full contrast conditions; by estimates with linear decay being more accurate than estimates with non-linear inverted decay; and by full contrast estimates being more accurate than estimates with non-linear inverted decay. There was no significant difference in correlation estimates between linear decay and full contrast conditions. Similar to the lack of significant difference between high and full contrast conditions in experiment 1, we hypothesize that this is due to the difference between these two conditions becoming smaller as objective correlation approaches 1; it may be that the difference is not on the whole great enough to produce a significant effect.

Statistical tests for significant contrasts between the four levels of the contrast conditions were performed with the **emmeans** package (version 1.8.4-1, Lenth, 2023) and are shown in Table 2. Means and 95% confidence intervals of correlation estimates are shown in Figure 13. The **EMAtools** package (version 0.1.4, Kleiman, 2021) was used to calculate effect sizes in Cohen's d. For the difference in correlation ratings between full contrast and non-linear decay function plots, an effect size of d = 0.19 was obtained. Between full contrast and non-linear inverted decay conditions, an effect size of d = 0.23 was obtained. These effects sizes are small and medium, but again, not insignificant.
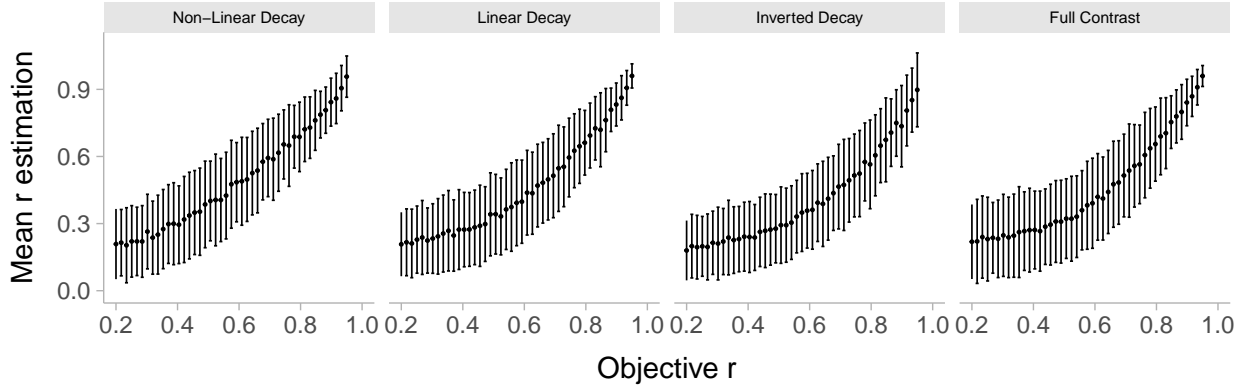
Figure 14: The mean of participants' estimations of r compared with objective Pearson's r, plotted by contrast condition
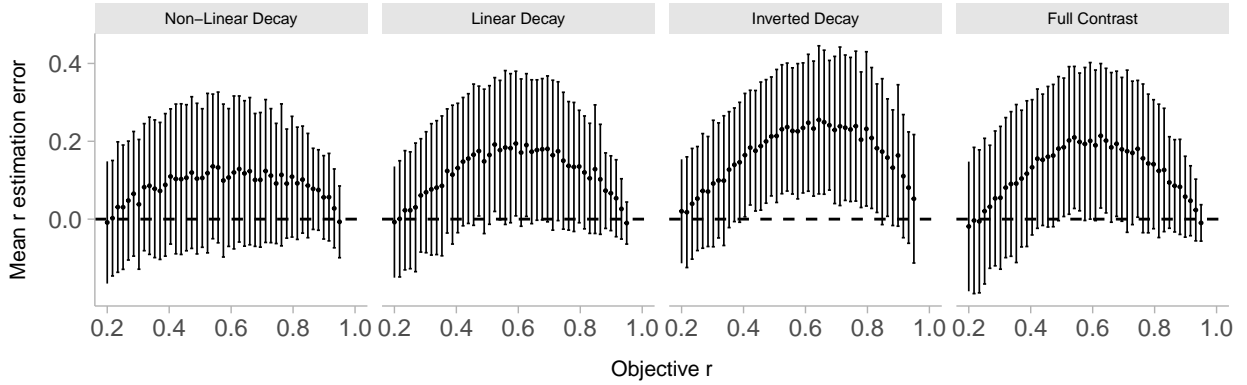

Figure 15: Plots showing how participants' mean correlation estimation errors change with Pearson's r in experiment 2. Participants were on average most accurate at estimating the r value when presented with the non-linear contrast decay scatterplots (condition A). Error bars show standard deviations of estimates.

Again, we also generated an additional model to test whether the results we found could be explained by differences in graph literacy. This model is identical to the experimental model, but includes graph literacy as a fixed effect. We found no significant differences between the original model and the one including graph literacy as a fixed effect ($\chi^2(1) = 0.24$, $p = .623$). These results suggest that the effect we found was not driven by differences in graph literacy between participants.

Figure 14 shows how participants' mean estimates of correlation change with the objective Pearson's r value, plotted separately for each of our contrast decay manipulation conditions. We again observe underestimation curves similar to those reported in previous literature (see Section 2.1 and Figure 1). Figure 16 shows how participants' mean estimation errors change with objective r value.

5.4. Discussion

Our hypotheses were fully supported in this experiment. Participants' errors in correlation estimation were lower when the non-linear decay parameter was used, and were highest when the non-linear inverted decay parameter was used. The only surprising result was the lack of significant difference in correlation estimates between the linear decay parameter and the full contrast condition. On closer inspection of the scatterplots included in the linear decay parameter condition however, it becomes clear why; the logarithmic nature of contrast perception (Varshney and Sun, 2013; Fechner, 1948) means that there is little perceptual distance between contrasts with high (> 0.75) contrast alpha values, which translates in our study to no perceived differences, on average, between plots with linear decay parameters and full contrast. This apparent threshold for contrast effects was also found in experiment 1. Filtering out lower r values, those with naturally higher
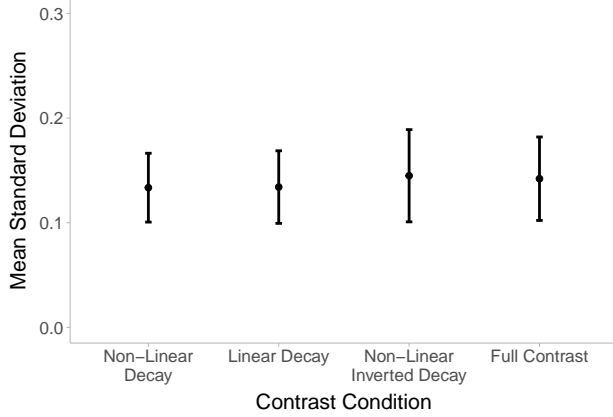
Figure 16: The mean of standard deviations of participants in experiment 2, grouped by contrast condition. Error bars show spreads of standard deviations

total residuals (arbitrarily $r < 0.6$), still produces no significant differences between correlation estimation errors for linear decay and full contrast conditions ($\chi^2(1) = 0.09$, $p = .769$). Our effects sizes were again small in this experiment, with the largest when compared to full contrast being for the inverted non-linear decay condition. This suggests that it is easier to induce greater bias in correlation estimates through a reduction in salience of the point cloud's center than it is to correct for the underestimation bias. Figure 16 shows how participant standard deviations vary with contrast condition.

Looking at the standard deviations of participants' r estimates in Figure 14, we can see that, as in experiment 1 (see Figure 8), aside from for the inverted non-linear decay condition, participants became more precise as r approached 1. This is in line with previous research. In the inverted non-linear decay condition, as r approaches 1 and correspondingly point residuals approach 0, the contrast of points also approaches 0. Just as the standard deviation of correlation judgements for the low contrast condition in experiment 1 was higher than for the other conditions (see Figure 10), having lower contrast points at the high r end of the non-linear inverted condition in experiment 2 resulted in fairly constant standard deviations across r values, as the usual reduction towards $r = 1$ was not seen.

Our finding that the use of the non-linear inverted decay parameter, in which contrast was increased with distance from the regression line, produces lower correlation estimation errors, adds perspective to suggestions (Yang et al., 2019) that, among other visual features, the area of a prediction ellipse (Yang et al., 2019; Cleveland et al., 1982), a region used to predict new observations assuming a bivariate normal distribution (see Figure 17 for an example) was a better predictor of people's performance on correlation judgement tasks than the objective r value was itself. In our non-linear inverted decay parameter condition, the area of this prediction ellipse did not change, yet people's estimates of correlation did. It would appear then that the apparent density of scatterplot points is also having an effect on people's perceptions of correlation, at least in our experimental paradigm. Previous research has found that more dense scatterplot displays are rated as having higher correlation, although this effect is weak (Lauer and Post, 1989; Rensink, 2014). To fully explore what is driving the effect seen in the non-linear inverted decay parameter condition, further work is needed on what exactly people attend to when completing correlation perception tasks. Eye-tracking studies would be well suited for this, but as of yet have only been used for simpler tasks such as the number of, or distance between points (Netzel et al., 2017).

6. General Discussion

In this paper we examine the effects of scatterplot point contrast on perceived correlation. We find that lower total contrast is associated with greater correlation underestimation error, that the use of the point contrast manipulation described above can partially correct for this error, and that inverting the
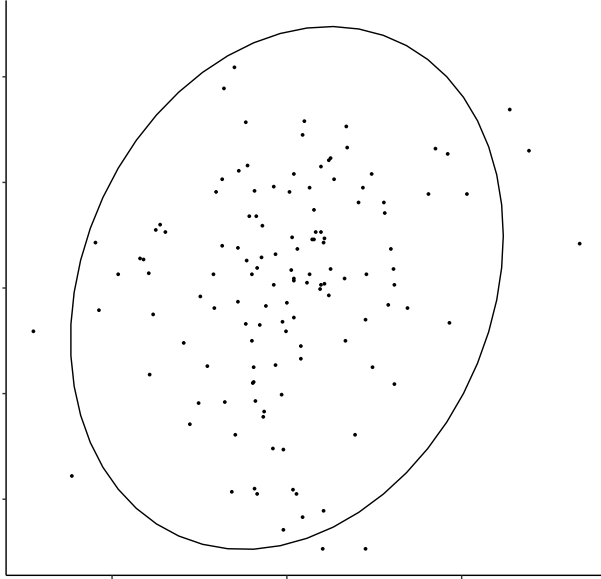
16

Figure 17: The above plot shows a 95% prediction ellipse over a scatterplot with an r value of 0.6.

aforementioned manipulation is associated with greater errors in correlation estimation. We suggest that these findings could be used to develop novel approaches to visualizing relationships between data while minimising the error in perceived correlation. The majority of the studies cited in this paper have used small samples of participants with experience in data science and statistics to draw their conclusions, often graduate students in visualization-heavy fields. We argue that this does not inform the design of commonly used data visualizations in a naturalistic way. In comparison, we have recruited from much more representative populations, including people from a variety of educational and national backgrounds, and have demonstrated that a simple framework can be used with these groups to gather high quality data and provide conclusions that can, by design, be thought of as far more naturalistic than studies that have taken place in labs with experienced participants.

In agreement with much previous research (Rensink and Baldridge, 2010; Rensink, 2012, 2014, 2017; Pollack, 1960), we found that participants were more accurate and more precise when the r value was higher. Figures 9 and 15 plot objective Pearson's r against the mean errors in correlation estimates for experiments 1 and 2 respectively. Standard deviations of estimates are shown as error bars. These plots illustrate, as in many studies cited, the lower levels of precision and accuracy for r values further from 0 or 1.

Our experiments contribute to a body of evidence that suggests participants are paying attention to the width of the probability distribution displayed in scatterplots (e.g Cleveland et al., 1982; Meyer et al., 1997; Yang et al., 2019; Rensink, 2017). We also confirm the systematic underestimation of correlation and suggest a strategy to correct for it. Through this work we do not attempt to redesign the scatterplot as a medium, but to provide a set of recommendations for visualization designers when designing scatterplots to support correlation perception:

1. Lowering the total contrast in a scatterplot can cause people to underestimate correlation compared to when contrast is maximal between the points and the plot background.
2. The use of a non-linear contrast decay parameter, in which contrast falls as a function of residual size, can be used to counteract the underestimation seen in correlation estimation in scatterplots.

Scatterplots, being as widely used as they are, are often designed with a number of communicative concepts in mind. When one of these concepts is illustrating to people the degree of association between

17

two variables, we would argue that designers should utilize the technique we have described here to give visualization viewers the best chance of interpreting the correlation displayed as accurately as possible.

6.1. Training

Both experiments described in the current work tested lay participants with varying levels of graph literacy. Because of this, participants first saw four plots depicting several correlations (see Figure 3) to familiarize them with the concept. To test if the patterns we observed in correlation estimation were as a result of this training, we built models including the half of the session (first or second) as a predictor. Comparing these to the original models not including session half as a predictor revealed a significant effect in experiment 1 ($\chi^2(1) = 7.13$, $p = 0.01$), but not experiment 2 ($\chi^2(1) = 2.20$, $p = 0.14$). In experiment 1, participants' errors in correlation estimation were on average higher in the second half of the experiment, suggesting that having recently viewed correctly labelled scatterplots helped participants make more accurate judgements of correlation. That we did not observe this effect in experiment 2 suggests that the influence of our manipulation (the contrast decay function) had a greater effect on correlation estimation than any effect of training. Future work might examine correlation estimation when no training is offered, or when training is incorrect. Given the quick and intuitive nature of correlation perception reported in the literature (Rensink, 2014), we would expect only a small effect of training, although the boundary conditions of such a manipulation are currently unknown.

6.2. Range Restriction

Pearson's r can only take a value between 0 and |1|. Because of this, we must necessarily consider whether or not the results we have seen could be at least partially attributed to range restriction. We hypothesize that the results seen previously in the literature, replicated here in Figures 9 and 15, in which participants' estimates of correlation were more accurate and less variable when r was closer to 0 and 1, may be partially due to the effects of range restriction. Given the variability of estimates, ratings of correlation may begin to truncate towards 0 or 1 as correlation gets towards 0 or 1. This may also be a result of scatterplots tending towards certain patterns when r is close to 1 (scatterplot tends towards a single line of points) and 0 (scatterplot tends towards an even random distribution). Range restriction cannot, however, explain the results we found, particularly in experiment 2, in which the use of the non-linear decay parameter significantly reduced participants' average errors in correlation estimation.

7. Limitations

The results in experiment 2 provide evidence that reducing the salience of points as they move further from the regression line can increase people's estimates of correlation, at least when plots like these are presented with other, conventional ones. Testing whether this phenomenon would exist with a plot in isolation would present a number of difficulties. As can be seen in Figures 7 and 13, and as more specifically illustrated in Figures 9 and 15, participants' estimates of correlation, especially between 0.2 and 0.7, suffer from high variance. Our high numbers of trials and participants ameliorate this to an extent, but this does by necessity mean we are unable to comment on judgements made on single plots.

We have made an important first step in the utilization of contrast changes to optimize perception of correlation in scatterplots, however there remains much to do. We were unable to quantitatively determine if 0.25 is indeed the most optimal value for b in equation 1, for example. It may well be the case that changing the value of b as a function of the objective Pearson's r value could produce more accurate correlation estimates in participants; our finding that participants were more accurate in correlation estimation when r was nearer 0 or 1 would suggest that the use of a decay parameter for these correlations is unnecessary.

The simplicity of the direct estimation task we have used does confer some limitations on the conclusions we can draw, although these limitations do not prevent that data we have gathered from being practically useful. We cannot comment on absolute perceptions of correlation in the same way that studies employing JND/bisection tasks can. Telling of this is the finding that participants' mean errors in judgements of correlation for the full contrast conditions were different in experiment 1 (0.149) vs experiment 2 (0.123).

We suggest that this is due to our direct estimation paradigms, in which, indirectly, participants are making comparative correlation judgements. While this is a limitation with regards to studying the perception of scatterplots from a psychophysical point of view, it still allows us to inform design research. We also found that standard deviations between these two conditions were not similar, which we suggest may be due to context; viewing full contrast plots in the context of plots exhibiting spatially-dependent contrast manipulations corresponded with more precise correlation estimates when compared to viewing them in the context of experiment 1.

8. Future Work

In this paper we worked strictly with positive r values, primarily because that is what has been investigated in the majority of the research that informed ours. However, given previous work (Sher et al., 2017) that found evidence for people overestimating negative correlation values, it would be reasonable to expect that the technique we have developed here could also be used to address this bias. We predict that the use of our non-linear inverted decay parameter will reduce the overestimation bias seen in estimates of correlation for negatively correlated scatterplots.

Our manipulations have used only the vertical distance between a particular point and the regression line to set contrast. Previous work, some of which has been inconclusive (Meyer et al., 1997), has generally suggested that the perpendicular distance between a point and the regression line may be a more accurate predictor of performance on correlation estimation tasks (Cleveland et al., 1982; Yang et al., 2019; Rensink, 2017). Future work investigating if a difference in correlation estimation accuracy might be found between contrast decay functions using vertical (residual) point-line distances or perpendicular ones could further hone the manipulation.

In addition to the range of scatterplot-associated tasks mentioned in Section 2.1, they may also be used to convey non-linear relationships between variables. Non-linearity renders Pearson's r unsuitable for the numerical representation of non-linear correlation, and a number of other measurements have been proposed (Laarne et al., 2021). The techniques we have used in the current paper could be extended for use with non-linear correlations in the future.

The final and perhaps most exciting avenue for future work is the possibility of combining the technique we have used here with other novel scatterplot techniques that have a basis in the literature. Liu et al. (2021) used triangular scatterplot points of different orientations to support the interactive fitting of a trend line. They found that the use of point orientations consistent with the regression line could reduce participants' errors in visual estimation. Combining this point orientation technique with the contrast decay function described here could be used to support both correlation and trend perception in scatterplots.

9. Acknowledgements

References

Tarek Azzam, Stephanie Evergreen, Amy A. Germuth, and Susan J. Kistler. Data Visualization and Evaluation. *New Directions for Evaluation*, 2013(139):7–32, 2013. ISSN 1534-875X. doi: 10.1002/ev.20065. URL http://onlinelibrary.wiley.com/doi/abs/10.1002/ev.20065. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ev.20065.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67:1–48, October 2015. ISSN 1548-7660. doi: 10.18637/jss.v067.i01. URL https://doi.org/10.18637/jss.v067.i01.

BBC. Covid map: Coronavirus cases, deaths, vaccinations by country. *BBC News*, July 2022. URL https://www.bbc.com/news/world-51235105.

M. Behrisch, M. Blumenschein, N. W. Kim, L. Shao, M. El-Assady, J. Fuchs, D. Seebacher, A. Diehl, U. Brandes, H. Pfister, T. Schreck, D. Weiskopf, and D. A. Keim. Quality Metrics for Information Visualization. *Computer Graphics Forum*, 37(3):625–662, 2018. ISSN 1467-8659. doi: https://doi.org/10.1111/cgf.13446. URL http://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13446. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13446.

Enrico Bertini and Giuseppe Santucci. Quality Metrics for 2D Scatterplot Graphics: Automatically Reducing Visual Clutter. In Andreas Butz, Antonio Krüger, and Patrick Olivier, editors, *Smart Graphics*, Lecture Notes in Computer Science, pages 77–89, Berlin, Heidelberg, 2004. Springer. ISBN 978-3-540-24678-7. doi: 10.1007/978-3-540-24678-7_8.

Philip Bobko and Ronald Karren. The Perception of Pearson Product Moment Correlations from Bivariate Scatterplots. *Personnel Psychology*, 32(2):313–325, 1979. ISSN 1744-6570. doi: 10.1111/j.1744-6570.1979.tb02137.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1744-6570.1979.tb02137.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1744-6570.1979.tb02137.x.

Rebecca A. Champion and Paul A. Warren. Contrast effects on speed perception for linear and radial motion. *Vision Research*, 140:66–72, November 2017. ISSN 1878-5646. doi: 10.1016/j.visres.2017.07.013.

W. S. Cleveland, P. Diaconis, and R. Mcgill. Variables on Scatterplots Look More Highly Correlated When the Scales Are Increased. *Science*, 216(4550):1138–1141, June 1982. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.216.4550.1138. URL https://www.sciencemag.org/lookup/doi/10.1126/science.216.4550.1138.

Charles E. Collyer, Kerrie A. Stanley, and Caroline Bowater. Psychology of the Scientist: LXIII. Perceiving Scattergrams: Is Visual Line Fitting Related to Estimation of the Correlation Coefficient? *Perceptual and Motor Skills*, 71(2):371–378E, October 1990. ISSN 0031-5125. doi: 10.2466/pms.1990.71.2.371. URL https://doi.org/10.2466/pms.1990.71.2.371. Publisher: SAGE Publications Inc.

Gustav Theodor Fechner. Elements of psychophysics, 1860. In *Readings in the history of psychology*, Century psychology series, pages 206–213. Appleton-Century-Crofts, East Norwalk, CT, US, 1948. doi: 10.1037/11304-026.

M. Friendly and Daniel J. Denis. The early origins and development of the scatterplot. *Journal of the history of the behavioral sciences*, 2005. doi: 10.1002/JHBS.20078.

Rocio Garcia-Retamero, Edward T. Cokely, Saima Ghazal, and Alexander Joeris. Measuring Graph Literacy without a Test: A Brief Subjective Assessment. *Medical Decision Making*, 36(7):854–867, 2016. ISSN 0272-989X. doi: 10.1177/0272989X16655334. URL https://doi.org/10.1177/0272989X16655334. Publisher: SAGE Publications Inc STM.

Asghar Ghasemi and Saleh Zahediasl. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism*, 10(2):486–489, 2012. ISSN 1726-913X. doi: 10.5812/ijem.3505. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3693611/.

Arthur P. Ginsburg. Contrast Sensitivity and Functional Vision. *International Ophthalmology Clinics*, 43(2):5, 2003. ISSN 0020-8167. URL https://journals.lww.com/internat-ophthalmology/fulltext/2003/43020/contrast_sensitivity_and_functional_vision.4.aspx.

Christopher Healey and James Enns. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, July 2012. ISSN 1941-0506. doi: 10.1109/TVCG.2011.127. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

Matt-Heun Hong, Jessica K. Witt, and Danielle Albers Szafir. The Weighted Average Illusion: Biases in Perceived Mean Position in Scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2021. ISSN 1077-2626, 1941-0506, 2160-9306. doi: 10.1109/TVCG.2021.3114783. URL https://ieeexplore.ieee.org/document/9556578/.

Matthew Kay and Jeffrey Heer. Beyond Weber's Law: A Second Look at Ranking Visualizations of Correlation. *IEEE transactions on visualization and computer graphics*, 22, September 2015. doi: 10.1109/TVCG.2015.2467671.

Evan Kleiman. *EMAtools: Data Management Tools for Real-Time Monitoring/Ecological Momentary Assessment Data*, 2021. URL https://CRAN.R-project.org/package=EMAtools. R package version 0.1.4.

Petri Laarne, Martha A. Zaidan, and Tuomo Nieminen. ennemi: Non-linear correlation detection with mutual information. *SoftwareX*, 14:100686, June 2021. ISSN 2352-7110. doi: 10.1016/j.softx.2021.100686. URL https://www.sciencedirect.com/science/article/pii/S2352711021000315.

David Lane, Craig Anderson, and Kathryn Kellam. Judging the Relatedness of Variables. The Psychophysics of Covariation Detection. *Journal of Experimental Psychology: Human Perception and Performance*, 11:640–649, October 1985. doi: 10.1037/0096-1523.11.5.640.

Thomas W. Lauer and Gerald V. Post. Density in scatterplots and the estimation of correlation. *Behaviour & Information Technology*, 8(3):235–244, June 1989. ISSN 0144-929X. doi: 10.1080/01449298908914554. URL https://www.tandfonline.com/doi/citedby/10.1080/01449298908914554. Publisher: Taylor & Francis.

Russell V. Lenth. *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2023. URL https://CRAN.R-project.org/package=emmeans. R package version 1.8.4-1.

Tingting Liu, Xiaotong Li, Chen Bao, Michael Correll, Changehe Tu, Oliver Deussen, and Yunhai Wang. Data-Driven Mark Orientation for Trend Estimation in Scatterplots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–16, Yokohama Japan, May 2021. ACM. ISBN 978-1-4503-8096-6. doi: 10.1145/3411764.3445751. URL https://dl.acm.org/doi/10.1145/3411764.3445751.

Justin Matejka, Fraser Anderson, and George Fitzmaurice. Dynamic Opacity Optimization for Scatter Plots. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2707–2710, Seoul Republic of Korea, April 2015. ACM. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702585. URL https://dl.acm.org/doi/10.1145/2702123.2702585.

Joachim Meyer and David Shinar. Estimating Correlations from Scatterplots. *Human Factors*, 34(3):335–349, June 1992. ISSN 0018-7208. doi: 10.1177/001872089203400307. URL https://doi.org/10.1177/001872089203400307. Publisher: SAGE Publications Inc.

Joachim Meyer, Meirav Taieb, and Ittai Flascher. Correlation estimates as perceptual judgments. *Journal of Experimental Psychology: Applied*, 3(1):3–20, 1997. ISSN 1939-2192. doi: 10.1037/1076-898X.3.1.3. Place: US Publisher: American Psychological Association.

Luana Micallef, Gregorio Palmas, Antti Oulasvirta, and Tino Weinkauf. Towards Perceptual Optimization of the Visual Design of Scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1588–1599, June 2017. ISSN 1941-0506.

doi: 10.1109/TVCG.2017.2674978. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

Rudolf Netzel, Jenny Vuong, Ulrich Engelke, Seán O'Donoghue, Daniel Weiskopf, and Julian Heinrich. Comparative eye-tracking evaluation of scatterplots and parallel coordinates. *Visual Informatics*, 1(2):118–131, June 2017. ISSN 2468-502X. doi: 10.1016/j.visinf.2017.11.001. URL `https://www.sciencedirect.com/science/article/pii/S2468502X17300293`.

Eyal Peer, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, September 2021. ISSN 1554-3528. doi: 10.3758/s13428-021-01694-3. URL `https://doi.org/10.3758/s13428-021-01694-3`.

Jonathan Peirce, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203, February 2019. ISSN 1554-3528. doi: 10.3758/s13428-018-01193-y.

Irwin Pollack. Identification of visual correlational scatterplots. *Journal of Experimental Psychology*, 59:351–360, 1960. ISSN 0022-1015. doi: 10.1037/h0042245. Place: US Publisher: American Psychological Association.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL `https://www.R-project.org/`.

Ronald A. Rensink. Invariance of Correlation Perception. *Journal of Vision*, 12(9):433–433, August 2012. ISSN 1534-7362. doi: 10.1167/12.9.433. URL `https://doi.org/10.1167/12.9.433`.

Ronald A. Rensink. On the Prospects for a Science of Visualization. In Weidong Huang, editor, *Handbook of Human Centric Visualization*, pages 147–175. Springer New York, New York, NY, 2014. ISBN 978-1-4614-7484-5 978-1-4614-7485-2. doi: 10.1007/978-1-4614-7485-2_6. URL `http://link.springer.com/10.1007/978-1-4614-7485-2_6`.

Ronald A. Rensink. The nature of correlation perception in scatterplots. *Psychonomic Bulletin & Review*, 24(3):776–797, 2017. ISSN 1069-9384. doi: 10.3758/s13423-016-1174-7. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5486871/`.

Ronald A. Rensink and Gideon Baldridge. The Perception of Correlation in Scatterplots. *Computer Graphics Forum*, 29(3):1203–1210, August 2010. ISSN 01677055. doi: 10.1111/j.1467-8659.2009.01694.x. URL `http://doi.wiley.com/10.1111/j.1467-8659.2009.01694.x`.

Varshita Sher, Karen G. Bemis, Ilaria Liccardi, and Min Chen. An Empirical Study on the Reliability of Perceiving Correlation Indices using Scatterplots. *Computer Graphics Forum*, 36(3):61–72, 2017. ISSN 1467-8659. doi: https://doi.org/10.1111/cgf.13168. URL `http://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13168`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13168.

Maureen Stone. Alpha, contrast and the perception of visual metadata. page 5, 2008.

Robert F. Strahan and Chris J. Hansen. Underestimating Correlation from Scatterplots. *Applied Psychological Measurement*, 2(4):543–550, October 1978. ISSN 0146-6216. doi: 10.1177/014662167800200409. URL `https://doi.org/10.1177/014662167800200409`. Publisher: SAGE Publications Inc.

Edward Tufte and P. Graves-Morris. The Visual Display of Quantitative Information. *Information Design Journal*, 4(3):235–236, 1983. ISSN 0142-5471, 1569-979X. doi: 10.1075/idj.4.3.12cos. URL `http://www.jbe-platform.com/content/journals/10.1075/idj.4.3.12cos`.

Lav R. Varshney and John Z. Sun. Why do we perceive logarithmically? *Significance*, 10(1):28–31, 2013. ISSN 1740-9713. doi: 10.1111/j.1740-9713.2013.00636.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1740-9713.2013.00636.x`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1740-9713.2013.00636.x.

Cesko C. Voeten. *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*, 2023. URL `https://CRAN.R-project.org/package=buildmer`. R package version 2.8.

C. Wehrhahn and G. Westheimer. How vernier acuity depends on contrast. *Experimental Brain Research*, 80(3), May 1990. ISSN 0014-4819, 1432-1106. doi: 10.1007/BF00228001. URL `http://link.springer.com/10.1007/BF00228001`.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL `https://ggplot2.tidyverse.org`.

F. Yang, L. T. Harrison, R. A. Rensink, S. L. Franconeri, and R. Chang. Correlation Judgment and Visualization Features: A Comparative Study. *IEEE Transactions on Visualization and Computer Graphics*, 25(3):1474–1488, March 2019. ISSN 1941-0506. doi: 10.1109/TVCG.2018.2810918. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

Silvia Zuffi, Carla Brambilla, Giordano Beretta, and Paolo Scala. Human Computer Interaction: Legibility and Contrast. In *14th International Conference on Image Analysis and Processing (ICIAP 2007)*, pages 241–246, September 2007. doi: 10.1109/ICIAP.2007.4362786.