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, primarily the correlation between two variables. Despite their ubiquity, people typically do not perceive correlations between variables accurately, tending to underestimate the strength of 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

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 brings with it obligations, not only to design and communicate with honesty, but also to 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 phenomena (see BBC (2022) for examples of the types of visualizations many saw daily). As data visualizations 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 correlation estimation tasks. 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 correct for a systematic underestimation by viewers of correlations in scatterplots that has a long-standing basis in the literature. Through this we also present a framework for the effective, inexpensive, and high-powered study of data visualizations.

*Corresponding author

Email addresses: Gabriel.Strain@manchester.ac.uk (Gabriel Strain), Andrew.J.Stewart@manchester.ac.uk (Andrew J. Stewart), Paul.Warren@manchester.ac.uk (Paul Warren), Caroline.Jay@manchester.ac.uk (Caroline Jay)

1.1. *Why Study Scatterplots?*

Scatterplots, estimated in 1984 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.

2. Related Work

2.1. *Testing Correlation Perception*

The primary concept communicated by scatterplots is correlation, (i.e the strength of the relationship between two variables). 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 seen. 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) allows 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. Mathematical modelling is useful with regards to discovering perceptual processes and laws for correlation perception. Given that this paper seeks to provide design guidelines, we do not model the relationship between r and perceived r. We are instead interested in comparative, naturalistic judgements of correlation, and therefore 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, which we will discuss below. 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 judged association increased as the size of the point cloud decreased, 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 has also had the effect of narrowing the width of the distribution displayed, as the length of the minor axis has decreased.

Furthermore, 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 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

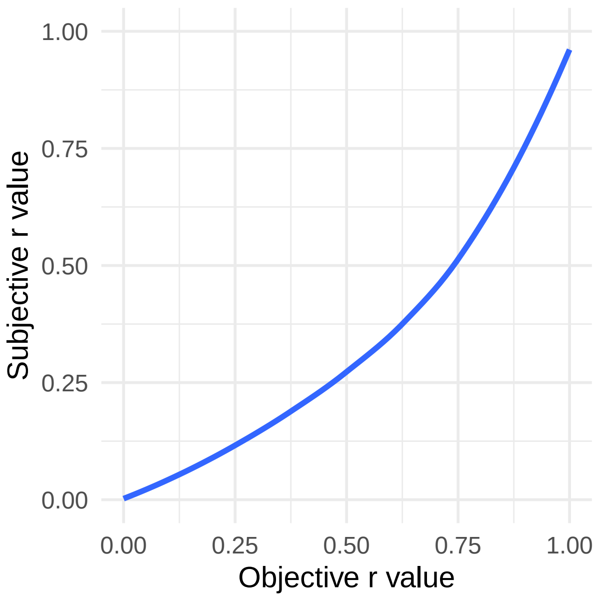

Figure 1: Using a function that relates objective to subjective r supplied in (Rensink, 2017) allows us to visualize the nature of the underestimation curve found in correlation perception studies. The curve represents the underestimation of correlation.

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 in Rensink (2017), that is small when $r = 1$, and increases as r approaches 0. The usefulness 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. Contrast

Adjusting the contrast in scatterplots has been used extensively to solve problems of overplotting (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. Despite the popularity of this technique, 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 low sample sizes (n = 12), and using only bisection/JND methodologies.

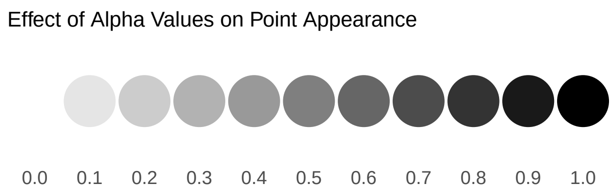

Figure 2: Higher alpha values result in greater contrast between the foreground (scatterplot point) and background. When alpha = 0, the foreground is ignored and the background is rendered.

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), this implies 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.

Intuitively, the best way to correct for an underestimation in correlation would be to simply remove outer data points until correlation perception aligned with what the actual correlation value was. This is arguably dishonest, and crucially our methodology allows us to begin to correct for the underestimation bias **without** removing data points, thereby preserving the integrity of the scatterplot for numerical retrieval or cluster separation tasks, among others.

3. The Current Experiments

Before we can begin attempting to use contrast manipulations to address the underestimation bias, we must first test what the effect of contrast is on correlation perception in our paradigm more generally. In both of the experiments described below, we use a 1-factor, 4-level design to test the effects of different contrast levels on participants' estimates of correlation.

4. 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 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 colour 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; for a single target against a uniform background (Weber contrast); models that take into account visibility limits (CIELAB lightness); and contrast in periodic patterns such as sinusoidal gratings (Michelson's contrast), to name but a few. 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, 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.

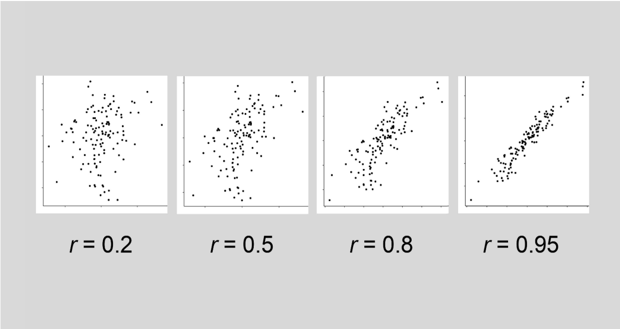

Figure 3: Participants viewed this for at least 8 seconds before being allowed to continue onto the practice trials.

5. Overview of Experiments

Experiments 1 and 2 share multiple aspects of their procedures. To save space these will be described here, while aspects that differ will be described during more specific discussion of the experiments.

Both experiments were built using PsychoPy (Peirce et al., 2019), and hosted on pavlovia.org. Participants were only permitted to complete the experiments on a desktop computer or laptop. As with the luminances of scatterplots and their points, we had no control over the distance participants were from their monitors. Again, as we are measuring relative differences in both of our experiments, we do not consider this to be an issue.

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 that can be seen in the experimental repository (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. 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 1/0"[1] (in red, attention check trial, n = 6). 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.

5.1. Plot Generation

Scatterplots were randomly generated from bivariate normal distributions. All presented were 1200 * 1200 pixels in size, and included no titles, tick labels, or axis labels. Each plot contained 128 data points. 45 r values were generated on a uniform distribution between 0.2 and 0.95. 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.

[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.

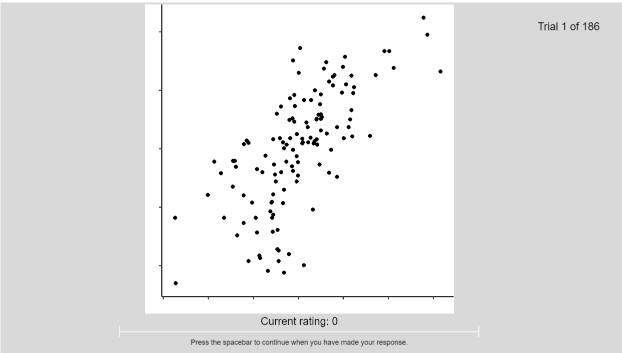

Figure 4: An example of an experimental trial.

6. 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 contains instructions for building a Docker container that fully reproduces the computational environment this paper was written in. 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. Consistent with our pre-registrations, when producing models for our analyses, we aimed for the most complex random effects structures. The structure of these models was identified using the **buildmer** package in R (Voeten, 2022). This package takes the most complex random effects model and subsequently removes random effects terms that do not substantially contribute to explaining variance in correlation ratings. 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.

7. Experiment 1: Changing the Contrast of All Points

7.1. Introduction

Our first experiment varied the contrast of every point on a scatterplot in a uniform manner. For each of the 45 r values there were 4 versions of each plot corresponding to the 4 levels of point contrast, examples of which can be seen in Figure 5. We hypothesized that there would be a wider spread of correlation estimates for plots with lower contrast compared to plots with higher contrast.

The experiment is hosted at https://gitlab.pavlovia.org/Strain/exp_uniform_adjustments. This repository contains all the experimental code, materials, and instructions needed to run the experiment in full.

7.2. 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 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.

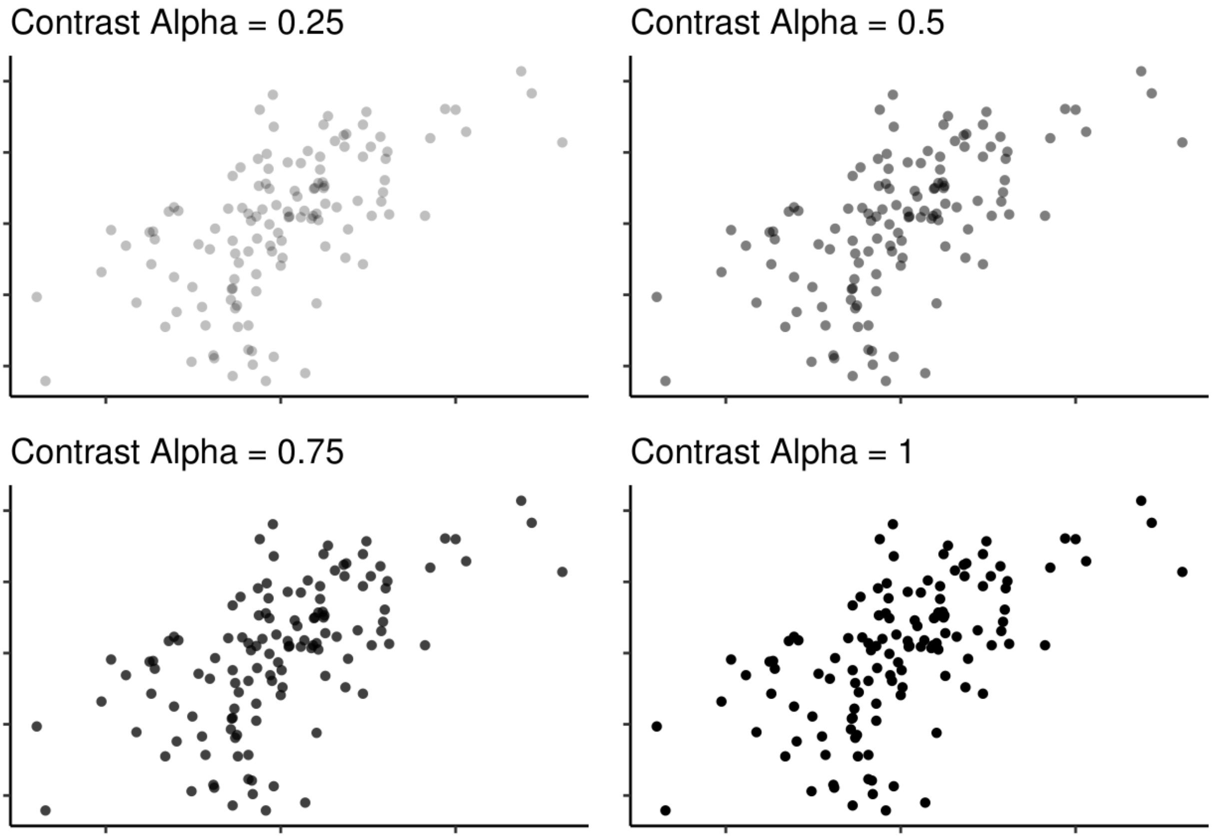

Figure 5: The four levels of the contrast condition in experiment 1, demonstrated with an r value of 0.6.

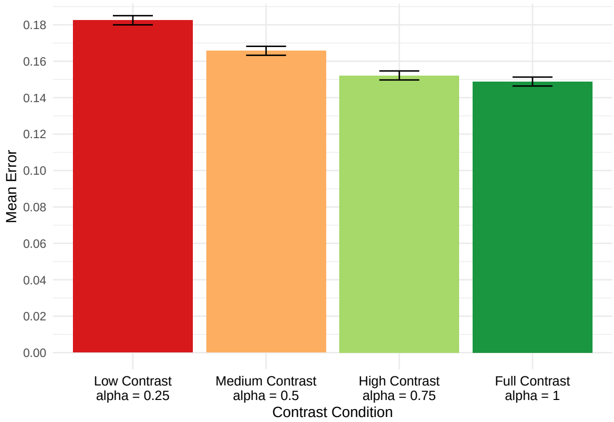

Figure 6: Mean error in correlation estimation across the four contrast conditions in E1.

Table 1: This table shows the contrasts between different levels of the contrast factor in E1. A = Low contrast (alpha = 0.25), Medium contrast (alpha = 0.5), C = High contrast (alpha = 0.75), D = Full contrast (alpha = 1).

Contrast	Z.ratio	p.value
A - D	13.48	<0.001
A - C	12.12	<0.001
A - B	6.72	<0.001
D - C	-1.35	0.528
D - B	-6.76	<0.001
C - B	-5.40	<0.001

7.3. Results

All analyses were conducted using R (version 4.2.1, (R Core Team, 2022)).

Models were built using the **buildmer** (version 2.7 (Voeten, 2022)) and **lme4** (version 1.1-30 (Bates et al., 2015)) packages. **buildmer** takes the most complex model, in terms of random effects intercepts for participants and items, and corresponding slopes for fixed effects terms (Barr et al., 2013) as an input. It then identifies the most complex model that successfully converges, and drops 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.

Figure 6 shows the mean error in correlation estimation for the 4 contrast conditions. A likelihood ratio test reveals 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 testing for contrasts between the 4 levels of the contrast condition was performed with the **emmeans** package (version 1.8.1-1, (Lenth, 2022)) and are shown in table 1. Means and standard deviations of correlation judgements are shown in table 2.

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 one, but includes graph literacy as a fixed effect.

Table 2: This table shows means and standard errors for the contrast conditions in E1. A = Low contrast (alpha = 0.25), Medium contrast (alpha = 0.5), C = High contrast (alpha = 0.75), D = Full contrast (alpha = 1).

Contrast	Mean	SD
A	0.183	0.206
D	0.149	0.201
C	0.152	0.202
B	0.166	0.201

We found no significant differences between the original model and the one including graph literacy as a fixed effect ($\chi^2(1) = 0.00$, $p = .995$). These results suggest that the effect we found was not driven by differences in graph literacy between participants.

7.4. 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 table 2, there was little difference in standard deviations between the 4 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 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) describes what is, to our knowledge, the only other direct testing of correlation perception and point contrast, and reports invariance in accuracy and precision with regards to contrast. Our results contradict these previously reported findings, which we argue is due to crucial differences in methodology and experimental power. Rensink used bisection and JND tasks to fit equations for accuracy and precision. Our methodology simply asked participants to estimate correlation, producing comparative judgements more suited to informing scatterplot design than the absolute relationship between correlation and perceived correlation illuminated upon by other methodologies. The major strength of this study is the large sample size (n = 150). Compared with the relatively small sample size in Rensink (2014) (n = 12), our study's heightened experimental power represents strong evidence that total contrast in a scatterplot can affect people's perceptions of correlation.

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.

8. Experiment 2: Spatially-Dependent Contrast Adjustments

8.1. Introduction

Our second experiment varied the contrast of scatterplot points using spatially dependent functions. For each of the 45 r values there were four versions of each plot, which can be seen in Figure 7. Three used functions relating point residuals to point contrast, which we henceforth refer to as decay parameters, with the remaining condition being the contrast alpha = 1 uniform contrast condition from the first experiment. 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

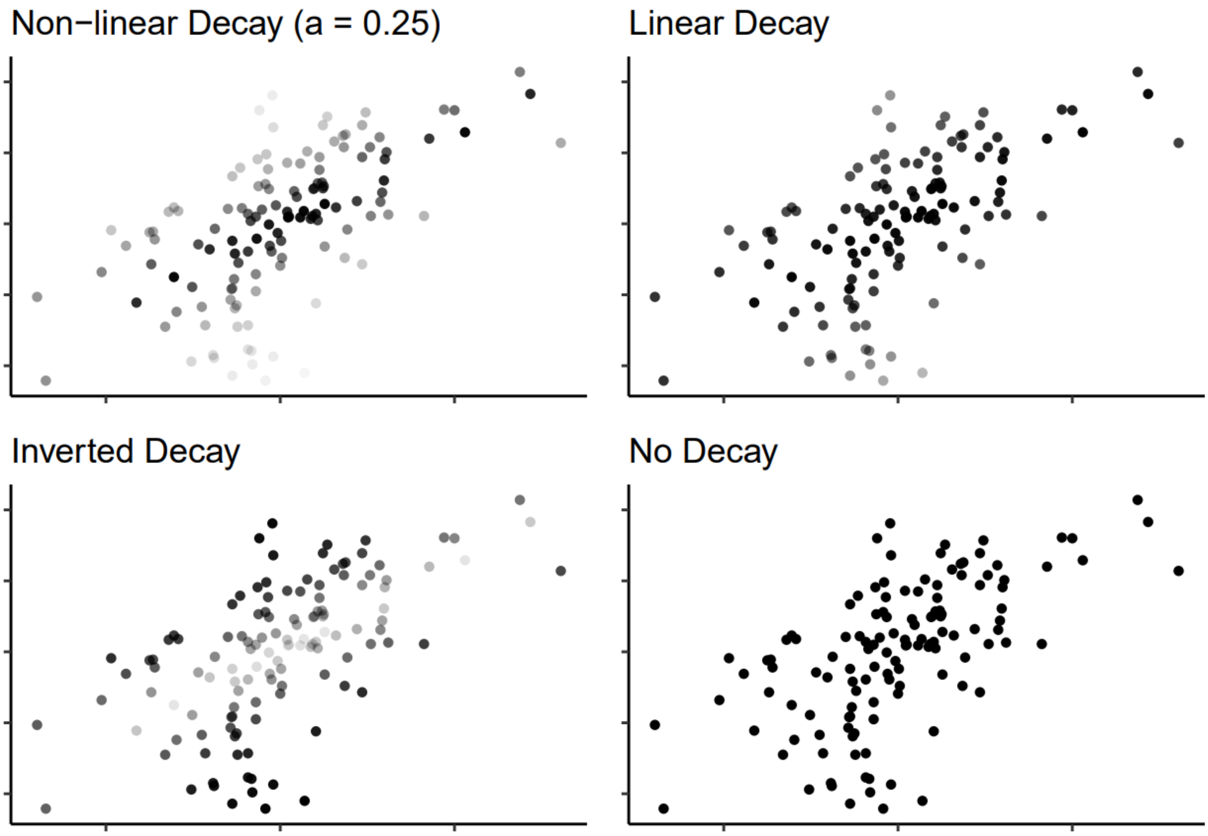

Figure 7: The 4 levels of the contrast condition in experiment 2, demonstrated with an r value of 0.6.

residual distance, would result in higher mean errors than all other conditions. We used the following equation 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, which represented an ideal balance between point legibility and contrast decay severity.

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

Figure 8 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.

8.2. 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 *(preprint)* 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

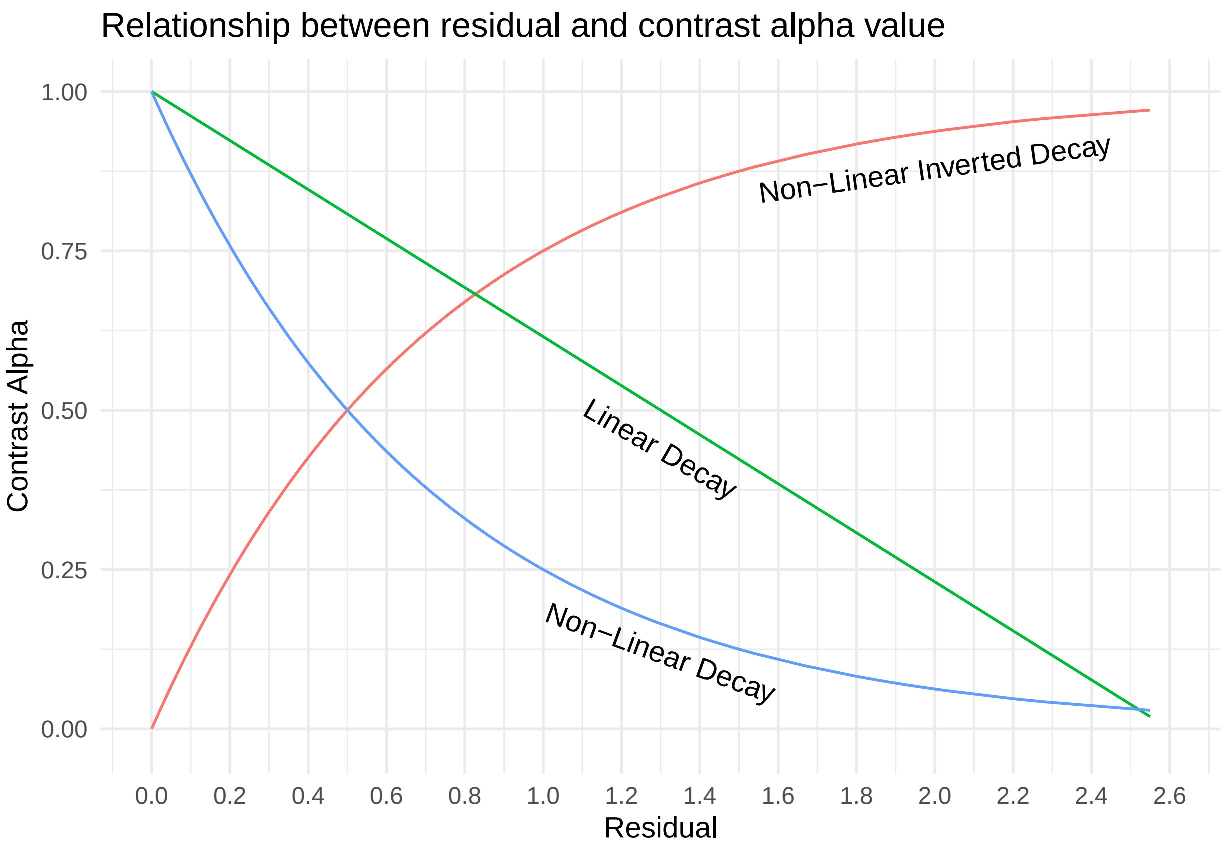

Figure 8: Here we use an r value of 0.2 to demonstrate the relationship between the size of a point's residual and the contrast alpha value that translates to.

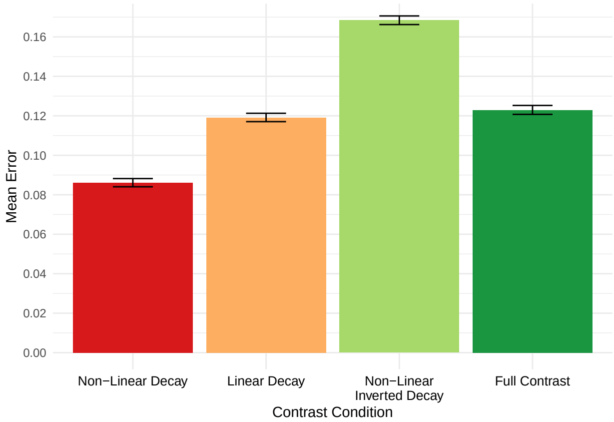

Figure 9: Mean error in correlation estimation across the four contrast conditions in E2.

Table 3: This table shows the contrasts between different levels of the contrast factor. A = Non-linear contrast decay (a = 0.25). B = Linear decay, C = Non-linear inverted decay, D = Full contrast.

Contrast	Z.ratio	p.value
D - C	-18.88	<0.001
D - B	1.55	0.405
D - A	15.30	<0.001
C - B	20.43	<0.001
C - A	34.17	<0.001
B - A	13.74	<0.001

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.

8.3. Results

All analyses were conducted using R (version 4.2.1, (R Core Team, 2022)).

As in E1, models were built using the **buildmer** (version 2.7, (Voeten, 2022)) and **lme4** (version 1.1-30 (Bates et al., 2015)) packages.

Figure 9 shows mean correlation estimation errors for the 4 contrast conditions. A likelihood ratio test reveals 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.

Statistical testing for significant contrasts between the 4 levels of the contrast conditions was performed with the **emmeans** package (version 1.8.1-1, (Lenth, 2022)) and are shown in table 3. Means and standard deviations of correlation estimates are shown in table 4.

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 one, but includes graph literacy as a fixed effect.

Table 4: This table shows means and standard errors for the contrast conditions in E2. A = Non-linear contrast decay (a = 0.25). B = Linear decay, C = Non-linear inverted decay, D = Full contrast.

Contrast	Mean	SD
D	0.123	0.185
C	0.168	0.179
B	0.119	0.173
A	0.086	0.170

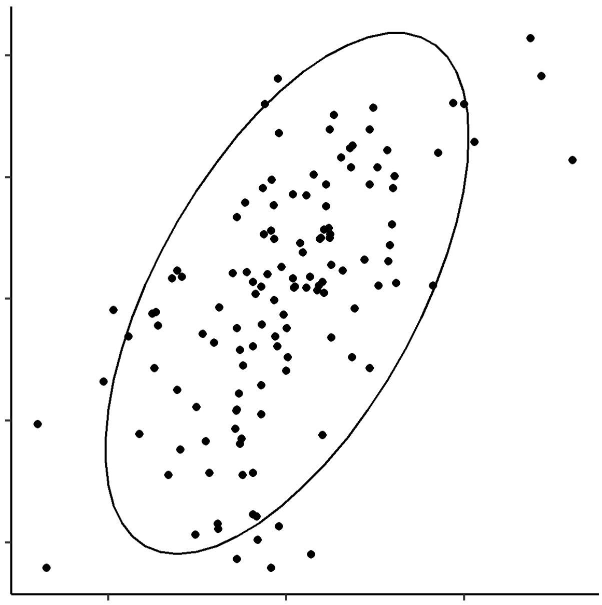

Figure 10: Here we have drawn a 95% prediction ellipse over a scatterplot with an r value of 0.6.

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.

8.4. 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. Filtering out lower r values, those with naturally higher 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 finding that the use of the non-linear inverted decay parameter, in which contrast was increased with distance from the regression line, 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

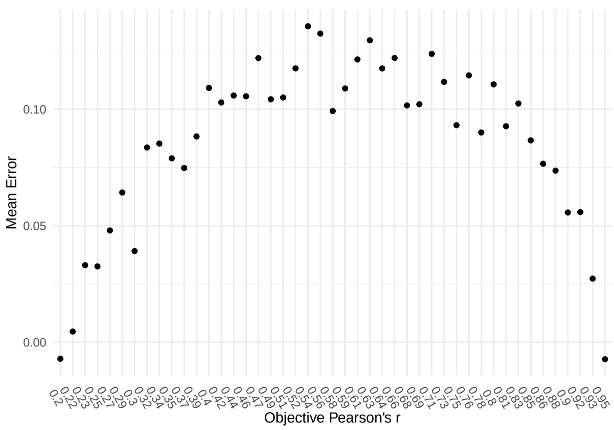

Figure 11: Plotting objective r values against mean error in correlation estimates shows that accuracy increases with r.

predict new observations assuming a bivariate normal distribution (see Figure 10 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 significantly 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).

9. General Discussion

In this paper we detail a novel visualizations technique that significantly improves viewer's performance in correlation estimation tasks. 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 visualizations 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, 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 11 and 12 plot the objective r value against the mean correlation estimate and standard deviation of estimates for the non-linear contrast decay condition in experiment 2, and illustrate, as in many studies cited, the markedly lower levels of precision and accuracy when correlations are not nearer 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. We also confirm the systematic underestimation of correlation, and suggest a strategy to correct for it. Ultimately the primary contribution of this work is a set of recommendations for visualizations designers:

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.

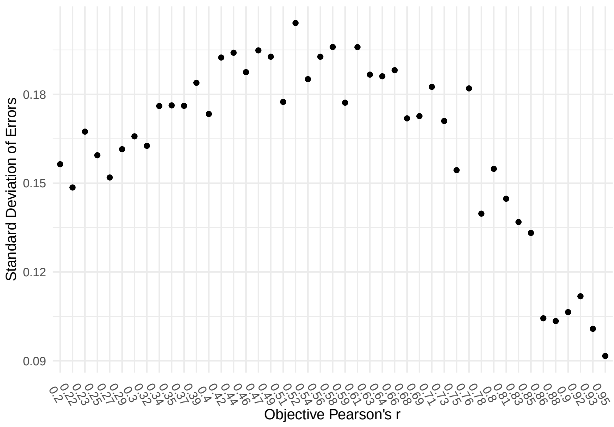

Figure 12: Plotting objective r values against the standard deviation of errors in correlation estimates shows that precision increases with r.

2. The use of a non-linear contrast decay parameter, in which contrast falls as a function of residual size, can be used to correct for the underestimation seen in correlation estimation in scatterplots.

Scatterplots, being as widely used as they are, are designed often with a number of communicative concepts in mind. When one of these concepts is illustrating to people the degree of association between two variables, we would argue that designers should utilise the technique we have described here to give visualizations viewers the best chance of interpreting the correlation displayed as accurately as possible.

10. 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 tables 2 and 4, and as more specifically illustrated in Figure 12, 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 in isolation.

11. Future Work

Although we have presented a novel visualizations technique that has resulted in participants' estimates of correlation in scatterplots being more accurate, we have not completely solved the problem. In our experiments, the underestimation bias was reduced, but it was still present. It may be the case that the value of b we have used in equation 1 is not optimised. Conducting further research using a variety of b values for equation 1 would be trivial following the framework we have established in this paper.

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. Given, however, previous work (Sher et al., 2017) which 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 also. 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. Again, testing this would be trivial using our framework.

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.

Dale J. Barr, Roger Levy, Christoph Scheepers, and Harry J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):10.1016/j.jml.2012.11.001, April 2013. ISSN 0749-596X. doi: 10.1016/j.jml.2012.11.001. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3881361/`.

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`.

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.

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.

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.

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*, 2022. URL `https://CRAN.R-project.org/package=emmeans`. R package version 1.7.3.

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.

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 Rensink. Invariance of Correlation Perception. volume 12, pages 433–433, May 2012. doi: 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, 2022. URL `https://CRAN.R-project.org/package=buildmer`.

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.