Novel Effects of Size and Contrast Adjustments in Scatterplots

ANONYMOUS AUTHOR(S)

Changing the size and contrast of points on scatterplots can be used to systematically improve viewers' perceptions of correlation. Evidence points to these effects being similar with regards to the mechanics behind them, so one would expect that their combination would produce a simple additive effect on correlation estimation. We present a fully reproducible study in which we combine techniques for influencing correlation perception to show that in reality, effects of changing point size and contrast interact in a non-additive fashion. We show that there are few limits to the extent to which we can use visual features to change viewers' perceptions of data visualizations, and use our results to further explain the perceptual mechanisms at play when changing point size and contrast in scatterplots.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: correlation, scatterplot, perception, crowdsourced

ACM Reference Format:

Anonymous Author(s). 2018. Novel Effects of Size and Contrast Adjustments in Scatterplots. In . ACM, New York, NY, USA, 13 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Scatterplots are common bivariate representations of data. They have been extensively studied and are used for a variety of communicative tasks. They are most commonly used to represent linear correlation, or the degree of linear relatedness between two variables, but are also used to represent different groups (clustering), to aid in the detection of outliers, to characterize distributions, and to visualize non-linear correlations. Figure 1 contains examples of scatterplots optimised for different tasks. There is evidence that people generally interpret them in similar ways [16], and that they support the interpretation of correlation significantly better than other data visualizations [19]. Rapid interpretation by viewers [29] along with low levels of interindividual variance render scatterplots particularly suited for experimental work; they provide important insights into perception and visualization design while being simple to study.

While our interpretations of scatterplots are generally similar, the accuracy of those interpretations is generally low. Viewers systematically underestimate the correlation displayed in positively correlated scatterplots. This holds true for direct estimation tasks [5, 9, 10, 17, 18, 21, 33], and estimation via bisection tasks [30], and is particularly pronounced between $0.2 < r < 0.6$. The COVID-19 pandemic demonstrated that even outside of professional/office environments, lay populations are now expected to be able to use and accurately interpret data visualizations on a daily basis [3]. This expectation confers a responsibility on the part of data visualization designers to design in such a way that people with little statistical or graph training can be expected to correctly interpret data visualizations. Doing this requires us to understand human perception and perceptual phenomena, apply this understanding to data visualization design,

Manuscript submitted to ACM

and test those designs in rigorous empirical work. We therefore present a fully-reproducible, crowdsourced online experiment in which we systematically alter visual features in scatterplots to correct for a historic bias. We combine two techniques for correcting for correlation underestimation in scatterplots and show that the combined effect is stronger than one would expect were they linearly additive. Through this work we also present a framework for visualization design informed from the ground up by human perception.

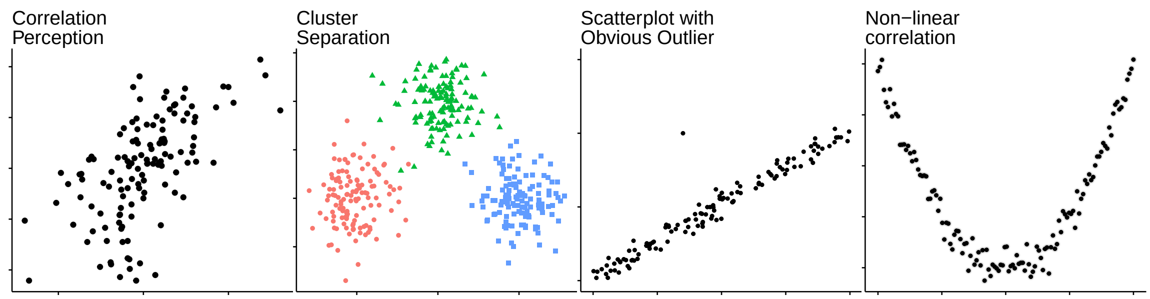

Fig. 1. Examples of scatterplots designed for different scatterplot-associated tasks. Both colour and point shape have been used to delineate different clusters in the cluster separation plot.

2 RELATED WORK

2.1 Testing Correlation Perception

The testing of linear correlation perception in scatterplots has a long and rich history, and has explored a wide variety of plot types, tasks, and modelling methods. Work has had participants make discriminative judgements between scatterplots with different correlations [11, 26], finding that performance on such a task became better as the objective r value increased. This performance on a discriminative judgement task can also be modelled by deep neural networks [39]. Extensive work throughout the 1970's to 1990's focused primarily on having participants produce a numerical estimate of correlation, finding evidence for a systematic underestimation for positive r values besides 0 and 1. This underestimation was especially pronounced for $0.2 < r < 0.6$ [5, 9, 10, 17, 18, 21, 33] and is demonstrated using an equation relating objective to subjective correlation [30] in Figure 2. More recent work has attempted to model participants' correlation estimation performance by using a combination of bisection task, in which participants were asked to adjust a plot until its correlation was halfway between two reference plots, and a staircase task designed to produce Just Noticeable Differences between scatterplots such that they are discriminable 75% of the time [31]. This work has also been extended to incorporate Bayesian data analysis [16]. The current experiment takes similar techniques from previous work [34**?**] and combines them to further push the envelope of how systematically adjusting visual features in scatterplots can radically alter people's perceptions of correlation. For this reason we use the same direct estimation paradigm to collect responses. This paradigm allows for a large number of judgements to be collected, and is simple enough that participants need little to no training.

2.2 Drivers of Correlation Perception

Evidence points towards correlation perception being driven by the shape of the underlying probability distribution represented by scatterplot points, however it should be noted that this is very much still an open question, and it may be the case that there are different contributory perceptual mechanisms operating at different levels based on task-specific differences such as viewing time and levels of graph-training. Increasing the x and y scales on a scatterplot such that the size of the point cloud decreases [9] is associated with an increase in a viewer's judgements of bivariate association, despite the objective r value remaining the same. It was suggested in this case that viewers may have been using the area of the point cloud to judge association. Later work found that the relationship between objective and perceived r values could be described by a function that included the mean of the geometric distances between the points and the regression line [22]. Investigations of the idea that people use visual features to judge correlation provide evidence that, among others features, the standard deviation of all perpendicular distances from scatterplot points to the regression line was predictive of performance on a correlation estimation task [38]. Equations for both discrimination and magnitude estimation of r in scatterplots include a quantity that is small when $r = 1$ and increases as r approaches 0 [30]. This quantity is indifferent to the type of visualization used, and is functionally similar to that found in work mentioned above [9, 22, 38]. Regarding scatterplots, this quantity represents the average distance between data points and the regression line, and can be thought of as representing the width of the underlying distribution. Findings from deep neural networks also support the idea that viewers are using an aspect of scatterplot shape to judge correlation, or some measure of what has been termed *dot entropy* [39], again considered a candidate visual proxy for correlation judgements [28, 30].

Recent work investigating the use of decay functions that change point size or contrast in scatterplots as a function of residual distance provide further evidence for both point density and salience/perceptual weighting being drivers of correlation perception. The use of an inverted contrast decay function [34], such that point contrast decreased as a point approached the regression line, resulted in significantly lower and less accurate correlation estimates compared to data-identical plots with the same overall shape, implying that the center of the scatterplot not being **filled** biased viewers' perceptions down. When point contrast or size are reduced as a function of distance from the regression line [34**?**], viewers rate correlation as significantly higher and are significantly more accurate, which supports a low-level data salience account. It is more difficult to comment on higher level perceptual mechanisms, and in any case this is not the intended contribution of the present work; we aim to test the impact of systematically altering visual features on correlation perception, and to provide empirically-derived tools for visualization designers to design better visualizations through the use of a simple, reproducible framework.

2.3 Transparency and Contrast

Changing the contrast of scatterplot points is standard practice to deal with issues of overplotting or clutter [4, 20]; scatterplots with very large numbers of points, especially with high degrees of overlap, suffer from low individual-point visibility caused by high point density. Lowering the contrast of all points addresses this, and makes data trends and distributions easier to see and interpret (see Figure 3).

In the present study we use the **ggplot2** package [37] to create our stimuli. This package uses an alpha parameter, or the level of linear interpolation [32] between foreground and background pixel values, to set the contrast of points. As demonstrated in Figure 4, an alpha value of 0 or 1 results in no interpolation and renders either the background or foreground pixel values respectively. Psychophysical definitions of perceived contrast are often based on what is being presented (i.e gratings) or are modelled to take into account phenomena of human vision (i.e visibility limits) [40]. Our crowdsourced methodology gives us no control the exact luminances of our stimuli, only over the *relative* differences in luminance between scatterplot points and backgrounds. For this reason we do not report absolute luminance nor make any attempt to adopt a formal definition of contrast; instead we simply report the alpha value used. Given that

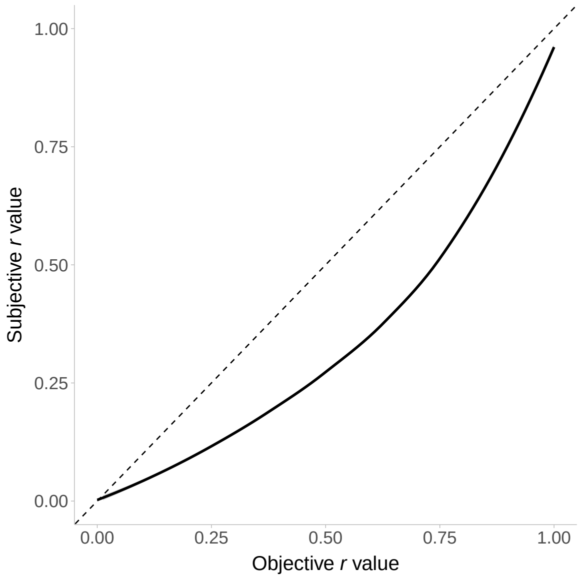

Fig. 2. Using a function relating objective to perceived r value [30] provides a visualization of the nature of correlation underestmiation reported in previous work. An identity line has been included to illustrate where viewers are most and least accurate.

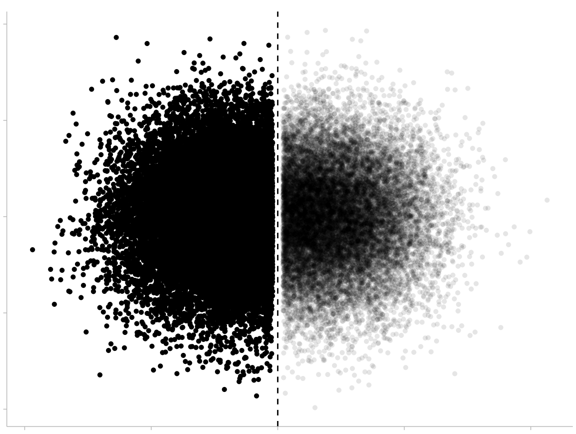

Fig. 3. Adjusting point contrast to address overplotting. Contrast between the points and the background is full (alpha = 1, L) or low (alpha = .1, R). The dataset used has 40,000 points.

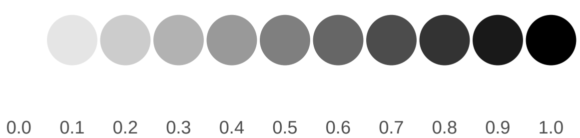

Fig. 4. Demonstrating the effects of different alpha values on point contrast.

our work aims to improve correlation perception without removing data from the scatterplot, we also incorporate point visibility testing (see Section 3.5). Informal point visibility piloting suggested that our smallest, lowest contrast points had very low visibility. We therefore implemented an alpha = 0.2 floor for these points, which we felt conferred a sufficient level of point visibility.

Previous work has found that lowering the contrast of *all* scatterplot points relative to the background can increase the level of underestimation error relative to full contrast, and that lowering point contrast *as a function of distance from the regression line* is able to bias correlation estimates upwards to partially correct for the underestimation bias [34]. Evidence suggests point salience/perceptual weighting and spatial uncertainty as drivers for this effect. Lower stimulus contrast is associated with lower salience [13], can bias judgements of mean point position [14] and increase error in positional judgements [36], and can result in greater uncertainty in speed perception [6]. Due to effects of both salience/perceptual weighting and spatial uncertainty operating in the same direction, previous work [34] has been unable to determine the extent to which each is responsible for the observed reduction in correlation estimation error.

2.4 Point Size

For discriminability reasons, scatterplots visualizing large datasets tend also to have smaller points. Bubble charts are a subclass of scatterplot which uses point size to describe a third variable, but what little experimental work there is on the impact of point size on correlation perception is inconclusive. Some work has found bias and variability in correlation perception to be invariant to changes in point size [27, 29], while elsewhere a strong effect of changing point size as a function of distance from the regression line has been reported [**?**]. Evidence points towards a salience-dominant mechanism in the latter case, albeit with a small effect of spatial uncertainty. There is evidence that stimulus size is associated with lower levels of spatial certainty [2] but higher levels of salience [13], and so the opposing directionality of predicted effects of salience/perceptual weighting and spatial uncertainty on correlation estimation has allowed researchers to provide evidence for the mechanistic dominance of salience when point size is used.

2.5 Hypotheses

Based on previously established effects of adjusting point size and point contrast using identical decay functions in isolation, we presently make the following hypotheses about the combination of these functions. We hypothesize that; (H1) an increased reduction in correlation estimation error will be observed when standard orientation decay functions are used; (H2) the use of congruent inverted orientation decay functions will produce the least accurate estimates of correlation; and (H3) that owing to the greater strength of the size channel observed in previous work [**?**], there will be a significant difference in correlation estimates between the two incongruent orientation conditions.

3 METHODOLOGY

3.1 Crowdsourcing

Much prior work into correlation perception in scatterplots has taken place in-person, most often with graduate students with experience in statistics. While this work is valuable, especially to perception audiences, it can struggle to provide data that is resilient to different lay viewers and viewing contexts. In addition, the ease and low-cost afforded to us by online, crowdsourced experimental work is unmatched. Given our intended HCI and design audience, we therefore choose to crowdsource all participants. We acknowledge however, that the technique has been affected by low quality data and skewed demographics in the past [7, 8, 24]. In light of these issues we follow published guidelines

[24] to ensure the collection of high quality data. Namely, we use the Prolific.co platform [1] with stringent pre-screen restrictions; participants were required to have completed at least 100 studies using Prolific, and were required to have a prolific score of 100, representing a 99% approval rate. This is more strict than the 95% suggested in previous work [24], but has served the authors well in prior studies.

3.2 Open Research

This study was conducted according to the principles of open and reproducible research. All data and analysis code are available at (repository link removed for anon). This repository contains instructions for building a docker image to fully reproduce the computational environment used. This allows for full replications of stimuli, analysis, and the paper itself. Ethical approval was granted by (removed for anon). Hypotheses and analysis plans were pre-registered with the OSF (links removed for anon) and there were no deviations from them.

3.3 Stimuli

The data used to generate the scatterplots were identical to that used in previous work [34**?**]. 45 scatterplot datasets were generated corresponding to 45 r values uniformly distributed between 0.2 and 0.99, as there is evidence that very little correlation is perceived below $r = 0.2$ [5, 9, 33]. Using so many values for r allows us to paint a broader picture of people's perceptions than work using fewer values. Scatterplot points were generated based on bivariate normal distributions with standard deviations of 1 in each direction. Each scatterplot had a 1:1 aspect ratio, was generated as a 1000*1000 pixel .png image, and was scaled up or down according to a participant's monitor such that they always occupied the same proportion of the screen. We used equation 1 to map residuals to size and contrast values. When adjusting point size, we further transform values using a scaling factor of 4 and a constant of 0.2 t ensure that the minimum point size in the present study is consistent with that of previous work [34**?**].

$$point_{size/contrast} = 1 - b^{residual} \tag{1}$$

0.25 was chosen as the value of b. This is both due to its previous usage in studies that the present work builds upon, and its production of a curve approximating the inverse around the identity line of the underestimation curve reported in previous work [30, 34**?**]. We acknowledge that there may be other, more suitable values of b, however extensive testing of these is outside the scope of the present work. We used 2x2 combinations of this equation applied to point size and contrast in standard and inverted orientation forms. Examples of the stimuli used can be seen in Figure 5.

3.4 Modelling

We use linear mixed effects models to model the relationships between the combination of size and contrast decay conditions and participants' errors in correlation estimates. Models such as these allow us to compare differences in our IV across the full range of participant responses, as opposed to relying purely on aggregate data, as in ANOVA. These models also afford us the ability to include random effects for participants and items. As per our pre-registrations we preferred maximal models, including random intercepts and slopes for participants and items. The structures of these models was identified using the **buildmer** package in R (version xx, [35]). This package takes a maximal random effects structure and then identifies the most complex model that converges, dropping terms that fail to explain a significant amount of variance.

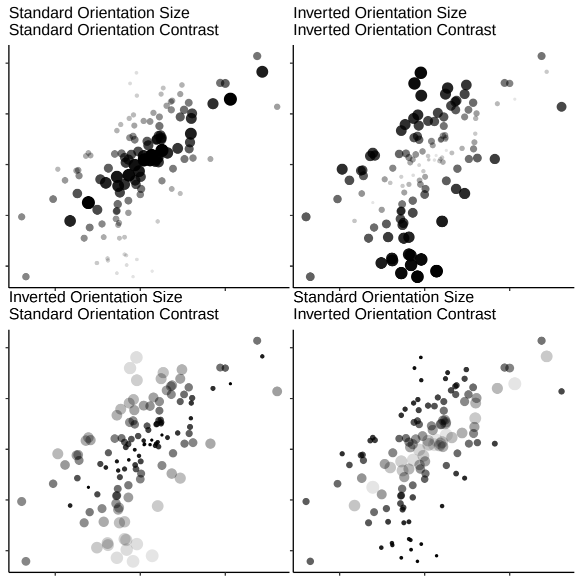

Fig. 5. Examples of the experimental stimuli used.

3.5 Point Visibility Testing

Discussions about the size and contrast of particular scatterplot points are inherently difficult in the context of online, crowdsourced experiments; controlling the devices participants use to participate in these kind of experiments, beyond insisting on laptop or desktop computers, is impossible. While this may result in a lack of consistency in scatterplot point sizes, luminances, or contrast ratios between participants, it also provides results that are more resilient to different viewing contexts than traditional lab-based experimental work. In addition to measures implemented to ensure high quality participant data (see Section 3.1), it is also key that we do not inadvertently remove data from scatterplots by including points whose size or contrast renders them invisible. We therefore included point visibility testing to ensure this. Participants viewed six scatterplots that were made up of a certain number of points. These points were of the same size and contrast as the smallest and lowest contrast points used in the experimental items. Participants were asked to enter in a textbox how many points were present. Participants scored an average of 74.89% (SD = 32.25). Despite our use of the contrast floor detailed in Section 2.3, it is clear that some of our small, low contrast points were not reliably visible, most likely due to low contrast between the point and background, as previous work [**?**] found point visibility largely invariant to size. We suggest this is due to differences in monitors between participants. In reality this contrast floor would need to be calibrated on a per-monitor basis. Figure 6 shows distributions of participants' performances on the visibility tests. We also include performance on the point visibility test as a fixed effect in Section 4.1.

3.6 Dot Pitch

We employed a method for obtaining the dot pitch of participants' monitors [23]. Combining this with monitor resolution information allows us to calculate the physical on-screen size of scatterplot points. Participants were asked to hold a standard size credit/ debit/ID card (ISO/IEC 7810 ID-1) up to their screen and resize an on-screen card until the

Fig. 6. Histogram of point visibility testing performance.

two matched. We assumed a widescreen 16:9 aspect ratio and calculated dot pitch based on these measurements. Mean dot pitch was 0.34 (*SD* = 0.05). We include analyses with dot pitch as a fixed effect in Section 4.1.

3.7 Procedure

Both experiments were built using PsychoPy [25] and hosted on Pavlovia.org. Participants were only permitted to complete the experiment on a desktop or laptop computer. Each participant was first shown the participant information sheet and provided consent through key presses in response to consent statements. They were asked to provide their age in a free text box, followed by their gender identity. Participants completed the 5-item Subjective Graph Literacy test [12], followed by the visual threshold task described in Section 3.5 and the screen scale task described in Section 3.6. Participants were given instructions, and were then shown examples of scatterplots with correlations of r = 0.2, 0.5, 0.8, and 0.95, as piloting of a previous experiment indicated some of the lay population may be unfamiliar with the visual character of scatterplots. Section 4 contains further discussion of the potential training effects of this. Two practice trials were given before the experiment began. Participants worked through a randomly presented series of 180 experimental trials and were asked to use a slider to estimate correlation to 2 decimal places. Visual masks preceded each scatterplot. Interspersed were 6 attention check trials which explicitly asked participants to ignore the scatterplot and set the slider to 0 or 1.

3.8 Participants

150 participants were recruited using the Prolific.co platform. Normal to corrected-to-normal vision and English fluency were required for participation. In addition, participants who had completed any of our previous studies into correlation estimation in scatterplots (references removed for anon) 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. Data from the remaining 150 participants were included in the full analysis (50.7% male, 48.7% female, and 0.7% non-binary). Participants mean age was 30.6 (*SD* = 8.6). Participants' mean graph literacy score was 22.5 (*SD* = 3.5). The average time taken to complete the experiment was 37 minutes (*SD* = 12.3).

3.9 Design

We used a fully repeated-measures 2*2 factorial design. Each participant saw each combination of size and contrast decay condition plots for a total of 180 experimental items. Participants viewed these experimental items, along with

Table 1. Significances of fixed effects and the interaction between them. Semi-partial R^2 for each fixed effect and the interaction term is also displayed.

	Estimate	Standard Error	df	t-value	p	R^2
(Intercept)	0.08	0.013	103.32	6.27	<0.001	
Size Decay	-0.14	0.005	148.39	-25.77	<0.001	0.104
Contrast Decay	0.12	0.002	26327.21	63.71	<0.001	0.087
Size Decay x Contrast Decay	0.15	0.004	26327.13	38.47	<0.001	0.034

6 attention check items, in a fully randomized order. All experimental code, materials, and instructions are hosted at (link removed for anon).

4 RESULTS

Our first two hypotheses were fully supported in this experiment. The combination of standard orientation size and contrast decay functions produced the most accurate estimates of correlation, although this also resulted in a large correlation overestimation for many values of r (see Figure 8). Our second hypothesis was also supported; the combination of inverted size and inverted contrast decay conditions produced the least accurate estimates of correlation. We found no support for our third hypothesis; there was no significant difference in correlation estimates for standard orientation size/inverted orientation contrast decay plots and inverted orientation size/standard orientation contrast decay plots (Z = -2.26, p = .11), however we did find a significant interaction effect that provides evidence that the combination of size and contrast decay functions is not additive in nature.

All analyses were conducted using R (version 4.3.1). Deviation coding was used for each of the experimental factors. We used the **buildmer** and **lme4** packages to build a linear mixed effects model where the difference between objective and rated r value was predicted by the size and contrast decay conditions used. A likelihood ratio test revealed that the model including point size and contrast decay conditions as fixed effects explained significantly more variance than the null ($\chi^2(3) = 5{,}286.81$, $p < .001$). There were significant fixed effects of size decay and contrast decay conditions, as well as a significant interaction between the two. The experimental model has random intercepts for items and participants, and a random slope for the size decay factor with regards to participants. Due both to our use of a linear mixed model with an interaction term, and our lack of comparative baseline condition (i.e no size or contrast function used), we do not report a measure of effect size. Instead we report the amounts of variance explained by each fixed effect term and the interaction term as semi-partial R^2 [**?**]. These values were calculated using the **r2glmm** package [15] can be see in Table 1 along with all model statistics.

4.1 Additional Analyses

We find no effects of graph literacy ($\chi^2(1) = 3.50$, $p = .061$) or performance on the visual threshold task ($\chi^2(1) = 1.29$, $p = .257$), or dot pitch ($\chi^2(1) = 1.52$, $p = .218$) on participants' errors in correlation estimation.

5 DISCUSSION

Our findings here provide further confirmatory evidence of what has been found previously with regards to the effects of point size and contrast manipulations on correlation estimation in scatterplots. Namely, that while both manipulations have a significant effect, the effect of changing point sizes is stronger, and that while we can influence correlation

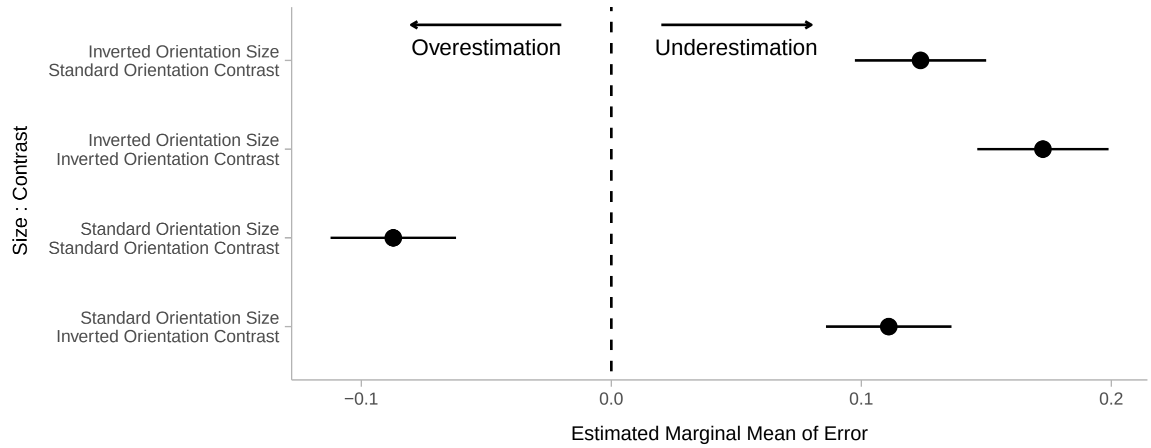

Fig. 7. Estimated marginal means for each combination of size and contrast decay conditions, including asymptotic lower and upper confidence limits calcualted by emmeans

Table 2. Pairwise comparisons. SO = Standard Orientation, IO = Inverted Orientation. The interaction is driven by the non-additive nature of combining point size and contrast decay functions, and the only nonsignificant contrast is found when incongruent decay functions are tested.

Contrast	Z.ratio	\textit{p}
SO Size x IO Contrast <-> IO Size x IO Contrast	-10.95	<0.001
SO Size x IO Contrast <-> SO Size x SO Contrast	72.29	<0.001
SO Size x IO Contrast <-> IO Size x SO Contrast	-2.26	0.108
IO Size x IO Contrast <-> SO Size x SO Contrast	46.13	<0.001
IO Size x IO Contrast <-> IO Size x SO Contrast	17.84	<0.001
SO Size x SO Contrast <-> IO Size x SO Contrast	-37.44	<0.001

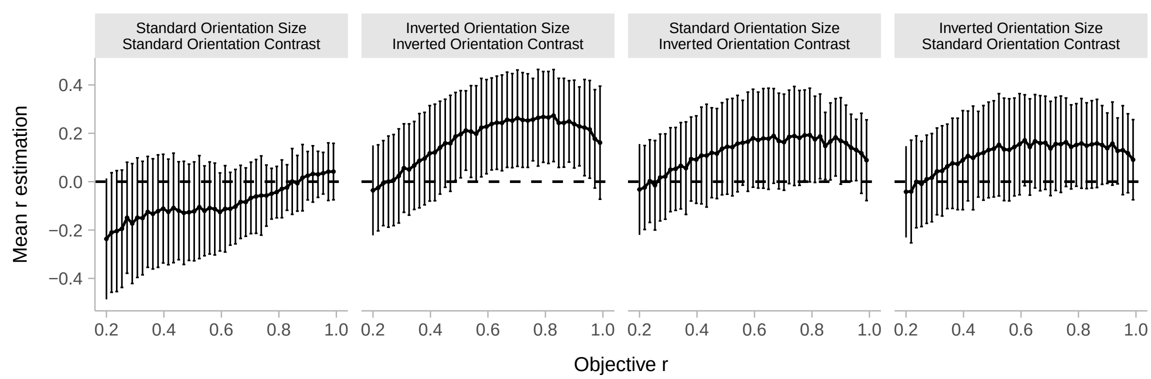

Fig. 8. Plots showing how participants' correlation estimation errors change as a function of the *r* value for each combination of size and contrast decay factors.

estimates in either direction, standard orientation manipulations are more powerful than inverted ones [34**?**]. As one would expect, we also see an effect of orientation congruency on the extent to which a manipulation can bias correlation estimates.

The lack of support for our third hypothesis, that there would be a difference in correlation estimates between incongruent conditions, was surprising given the greater strength of the size channel relative to contrast demonstrated in previous work [34**?**]. Despite the lack of support for this hypothesis, we did find that the size decay channel explained more variance (.104) in our model than contrast decay (.087) was able to. If the combination of size and contrast decay functions was strictly additive, we would expect a significant difference between incongruent conditions owing to the greater power of the size channel. Their non-additive nature, however, has resulted in each channel attenuating the other.

5.1 Combining Manipulations

Figure 8 and Figure 7 show how, on average, the combination of standard orientation size and contrast decay conditions has resulted in an overestimation of r for the majority of values. While this does not directly solve the underestimation problem. It does demonstrate that with regards to using point size and contrast to bias viewer's estimates of correlation in scatterplots, there would appear to be few limitations. The issue here is not one of ability to change people's perceptions, but of *tuning* the use of these visual factors to be able to change people's perceptions in systematic ways. We explore what further work would need to be done to do this in Section 5.6.

5.2 Contributions of Size and Contrast Decay

Incorporating data from previous work [34**?**] that used similar decay functions and experimental paradigms allows us to compare estimation curves for size decay and contrast decay in isolation and in combination (Figure 9). We can then derive new curves that describe the effect that each manipulation and the combination of manipulations has on people's estimates of correlation (Figure 10). Throughout the course of this section it should be noted that the analyses include data from several separate experiments. We argue that their methodological similarities render comparison appropriate, but we acknowledge the potential for overstated conclusions.

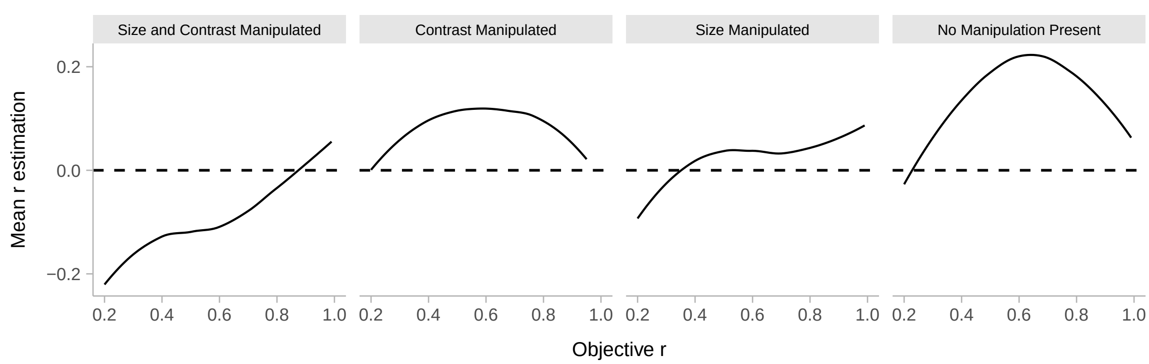

Fig. 9. From left to right, plotting r estimation error against the objective r value for the standard orientation condition in the present study, for standard orientation size and contrast manipulations in previous work, and for normal scatterplots averaged over identical conditions in previous work. Error bars have been left off this plot to make interpretation more simple.

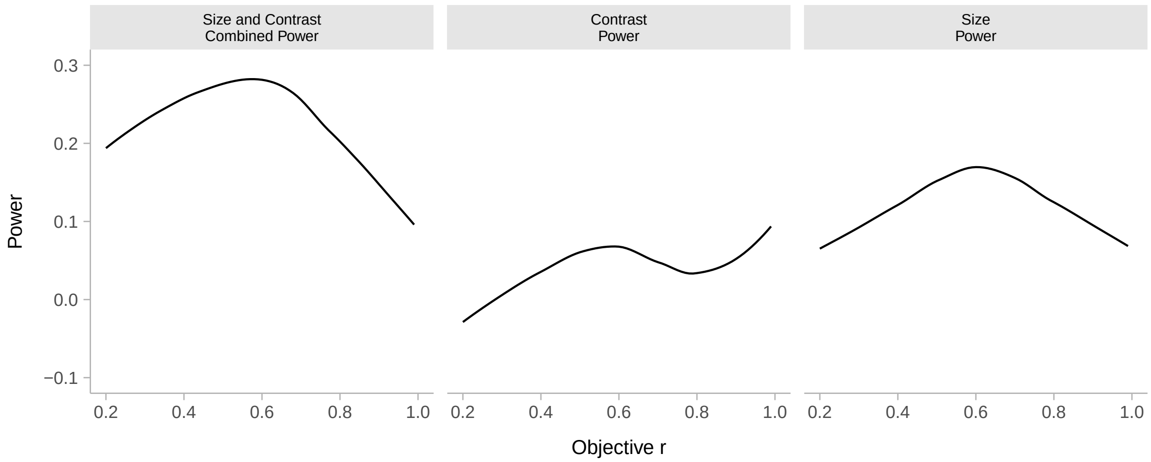

Fig. 10. additive_raw_pl = observed values for present study. standard_curve = no manipulation averaged across all experiments

5.3 Mechanisms

5.4 Further Tuning

5.5 Limitations

5.6 Future Work

REFERENCES

[1] 2023. Prolific. https://www.prolific.co

[2] David Alais and David Burr. 2004. The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current biology: CB* 14, 3 (Feb. 2004), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

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

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

[5] Philip Bobko and Ronald Karren. 1979. The Perception of Pearson Product Moment Correlations from Bivariate Scatterplots. *Personnel Psychology* 32, 2 (1979), 313–325. https://doi.org/10.1111/j.1744-6570.1979.tb02137.x

[6] Rebecca A. Champion and Paul A. Warren. 2017. Contrast Effects on Speed Perception for Linear and Radial Motion. *Vision Research* 140 (Nov. 2017), 66–72. https://doi.org/10.1016/j.visres.2017.07.013

[7] Nick Charalambides. 2021. We Recently Went Viral on TikTok - Here's What We Learned.

[8] Michael Chmielewski and Sarah C. Kucker. 2020. An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results. *Social Psychological and Personality Science* 11, 4 (May 2020), 464–473. https://doi.org/10.1177/1948550619875149

[9] W. S. Cleveland, P. Diaconis, and R. Mcgill. 1982. Variables on Scatterplots Look More Highly Correlated When the Scales Are Increased. *Science* 216, 4550 (June 1982), 1138–1141. https://doi.org/10.1126/science.216.4550.1138

[10] Charles E. Collyer, Kerrie A. Stanley, and Caroline Bowater. 1990. Psychology of the Scientist: LXIII. Perceiving Scattergrams: Is Visual Line Fitting Related to Estimation of the Correlation Coefficient? *Perceptual and Motor Skills* 71, 2 (Oct. 1990), 371–378E. https://doi.org/10.2466/pms.1990.71.2.371

[11] Michael E. Doherty, Richard B. Anderson, Andrea M. Angott, and Dale S. Klopfer. 2007. The Perception of Scatterplots. *Perception & Psychophysics* 69, 7 (Oct. 2007), 1261–1272. https://doi.org/10.3758/BF03193961

[12] Rocio Garcia-Retamero, Edward T. Cokely, Saima Ghazal, and Alexander Joeris. 2016. Measuring Graph Literacy without a Test: A Brief Subjective Assessment. *Medical Decision Making* 36, 7 (2016), 854–867. https://doi.org/10.1177/0272989X16655334

[13] Christopher Healey and James Enns. 2011. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics* 18, 7 (July 2011), 1170–1188. https://doi.org/10.1109/TVCG.2011.127

[14] Matt-Heun Hong, Jessica K. Witt, and Danielle Albers Szafir. 2021. The Weighted Average Illusion: Biases in Perceived Mean Position in Scatterplots. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. https://doi.org/10.1109/TVCG.2021.3114783

[15] Byron Jaeger. 2017. *r2glmm: Computes R Squared for Mixed (Multilevel) Models.* https://github.com/bcjaeger/r2glmm R package version 0.1.2.

[16] Matthew Kay and Jeffrey Heer. 2015. Beyond Weber's Law: A Second Look at Ranking Visualizations of Correlation. *IEEE transactions on visualization and computer graphics* 22 (Sept. 2015). https://doi.org/10.1109/TVCG.2015.2467671

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

[18] Thomas W. Lauer and Gerald V. Post. 1989. Density in Scatterplots and the Estimation of Correlation. *Behaviour & Information Technology* 8, 3 (June 1989), 235–244. https://doi.org/10.1080/01449298908914554

[19] Jing Li, Jean-Bernard Martens, and Jarke J van Wijk. 2010. Judging Correlation from Scatterplots and Parallel Coordinate Plots. *Information Visualization* 9, 1 (Jan. 2010), 13–30. https://doi.org/10.1057/ivs.2008.13

[20] Justin Matejka, Fraser Anderson, and George Fitzmaurice. 2015. Dynamic Opacity Optimization for Scatter Plots. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 2707–2710. https://doi.org/10.1145/2702123.2702585

[21] Joachim Meyer and David Shinar. 1992. Estimating Correlations from Scatterplots. *Human Factors* 34, 3 (June 1992), 335–349. https://doi.org/10.1177/001872089203400307

[22] Joachim Meyer, Meirav Taieb, and Ittai Flascher. 1997. Correlation Estimates as Perceptual Judgments. *Journal of Experimental Psychology: Applied* 3, 1 (1997), 3–20. https://doi.org/10.1037/1076-898X.3.1.3

[23] Wakefield L Morys-Carter. 2023. ScreenScale. https://doi.org/10.17605/OSF.IO/8FHQK

[24] Eyal Peer, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. 2021. Data Quality of Platforms and Panels for Online Behavioral Research. *Behavior Research Methods* (Sept. 2021). https://doi.org/10.3758/s13428-021-01694-3

[25] Jonathan Peirce, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in Behavior Made Easy. *Behavior Research Methods* 51, 1 (Feb. 2019), 195–203. https://doi.org/10.3758/s13428-018-01193-y

[26] Irwin Pollack. 1960. Identification of Visual Correlational Scatterplots. *Journal of Experimental Psychology* 59 (1960), 351–360. https://doi.org/10.1037/h0042245

[27] Ronald Rensink. 2012. Invariance of Correlation Perception. In *Journal of Vision*, Vol. 12. 433–433. https://doi.org/10.1167/12.9.433

[28] Ronald Rensink. 2022. Visual Features as Carriers of Abstract Quantitative Information. *Journal of Experimental Psychology General* 151 (Jan. 2022). https://doi.org/10.1037/xge0001165

[29] Ronald A. Rensink. 2014. On the Prospects for a Science of Visualization. In *Handbook of Human Centric Visualization*, Weidong Huang (Ed.). Springer New York, New York, NY, 147–175. https://doi.org/10.1007/978-1-4614-7485-2_6

[30] Ronald A. Rensink. 2017. The Nature of Correlation Perception in Scatterplots. *Psychonomic Bulletin & Review* 24, 3 (2017), 776–797. https://doi.org/10.3758/s13423-016-1174-7

[31] Ronald A. Rensink and Gideon Baldridge. 2010. The Perception of Correlation in Scatterplots. *Computer Graphics Forum* 29, 3 (Aug. 2010), 1203–1210. https://doi.org/10.1111/j.1467-8659.2009.01694.x

[32] Maureen Stone. 2008. Alpha, Contrast and the Perception of Visual Metadata. (2008), 5.

[33] Robert F. Strahan and Chris J. Hansen. 1978. Underestimating Correlation from Scatterplots. *Applied Psychological Measurement* 2, 4 (Oct. 1978), 543–550. https://doi.org/10.1177/014662167800200409

[34] Gabriel Strain, Andrew J. Stewart, Paul Warren, and Caroline Jay. 2023. The Effects of Contrast on Correlation Perception in Scatterplots. *International Journal of Human-Computer Studies* 176 (Aug. 2023), 103040. https://doi.org/10.1016/j.ijhcs.2023.103040

[35] Cesko C. Voeten. 2023. *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression.* https://CRAN.R-project.org/package=buildmer R package version 2.9.

[36] C. Wehrhahn and G. Westheimer. 1990. How Vernier Acuity Depends on Contrast. *Experimental Brain Research* 80, 3 (May 1990). https://doi.org/10.1007/BF00228001

[37] Hadley Wickham. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.

[38] F. Yang, L. T. Harrison, R. A. Rensink, S. L. Franconeri, and R. Chang. 2019. Correlation Judgment and Visualization Features: A Comparative Study. *IEEE Transactions on Visualization and Computer Graphics* 25, 3 (March 2019), 1474–1488. https://doi.org/10.1109/TVCG.2018.2810918

[39] Fumeng Yang, Yuxin Ma, Lane Harrison, James Tompkin, and David H. Laidlaw. 2023. How Can Deep Neural Networks Aid Visualization Perception Research? Three Studies on Correlation Judgments in Scatterplots. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–17. https://doi.org/10.1145/3544548.3581111

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