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 mechanisms behind them, so one may 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 is a great deal of scope for using visual features to change viewers' perceptions of data visualizations. Additionally, we use our results to further interrogate the perceptual mechanisms at play when changing point size and contrast in scatterplots.

CCS Concepts: • **Human-centered computing** → *Visualization design and evaluation methods*; **Empirical studies in HCI**; *HCI design and evaluation methods*; **Empirical studies in visualization**.

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

ACM Reference Format:

Anonymous Author(s). 2024. Novel Effects of Size and Contrast Adjustments in Scatterplots. In . ACM, New York, NY, USA, 18 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. While most commonly used to represent linear correlation, or the degree of linear relatedness between two variables, they can also be 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 optimized for different tasks. There is evidence that people generally interpret scatterplots in similar ways [18], and that they support the interpretation of correlation significantly better than other data visualizations [22]. Rapid interpretation by viewers [33], 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 [33].

While our interpretations of scatterplots are generally similar, the accuracy of those interpretations is generally poor. In particular, viewers systematically underestimate the correlation displayed in positively correlated scatterplots. This holds true for direct estimation tasks [7, 11, 12, 19, 20, 24, 38], and estimation via bisection tasks [34], and is particularly pronounced for values of r between 0.2 and 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 [5]. This expectation

Manuscript submitted to ACM

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 expect to be able to correctly interpret data visualizations. Doing this requires us to understand human perception and perceptual phenomena, apply this understanding to data visualization design, and test those designs in rigorous empirical work. Here, we present a fully-reproducible, crowdsourced online experiment in which we systematically alter visual features in scatterplots to correct for a well-known bias. We combine two techniques for correcting for correlation underestimation in scatterplots and show that the effects are stronger than what might be expected from simple additive combination. Building on this, we present a framework for visualization design informed from the ground up by human perception.

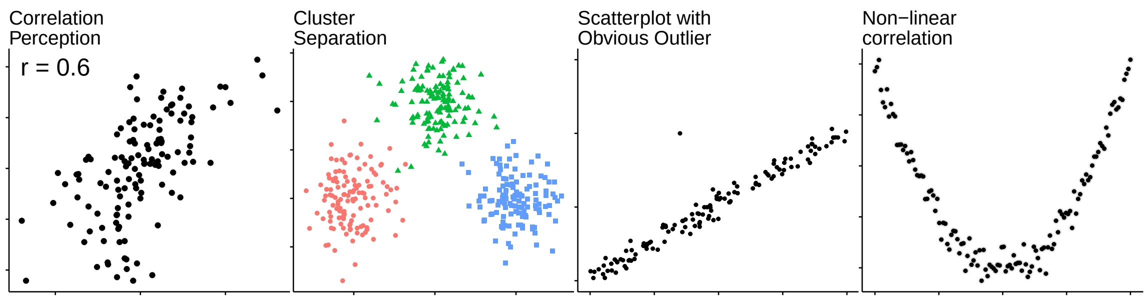

Fig. 1. Examples of scatterplots designed for different scatterplot-associated tasks. Both color and point shape have been used to delineate different clusters in the cluster separation plot.

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 modeling methods. Work has involved participants engaged in making discrimination judgements between scatterplots with different correlations [13, 30], finding that performance on such tasks became better as the objective r value increased. This performance on a discrimination judgement task can also be modeled by deep neural networks [48]. Extensive work throughout the 1970s to 1990s focused primarily on having participants produce a numerical estimate of correlation, and found evidence for a systematic *underestimation* for positive r values besides 0 and 1. This underestimation was especially pronounced for $0.2 < r < 0.6$ [7, 11, 12, 19, 20, 24, 38] and is demonstrated using an equation relating objective to subjective correlation [34] in Figure 2. More recent work has attempted to model participants' correlation estimation performance by using a combination of a bisection task, in which participants are asked to adjust a plot until its correlation is halfway between that of two reference plots, and a staircase method task designed to produce Just Noticeable Differences between scatterplots such that they are discriminable 75% of the time [35]. This work has also been extended to incorporate Bayesian data analysis [18], which is in agreement with other literature identifying scatterplots as particularly suited for the communication of correlation [22]. The current experiment adapts techniques from previous work [39, 40] 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, especially with regards to what low level perceptual mechanisms may be at play. It is possible that there are different contributory perceptual mechanisms operating at different levels based on task-specific differences such as viewing time or levels of graph-training. Increasing the x and y scales on a scatterplot such that the size of the point cloud decreases [11] 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 [25]. Investigation of the idea that people use visual features to judge correlation provides evidence that, among three other equally predictive features, the standard deviation of all perpendicular distances from scatterplot points to the regression line is predictive of performance on a correlation estimation task [47]. 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 [34]. This quantity is indifferent to the type of visualization used, and is functionally similar to that found in work mentioned above [11, 25, 47]. 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 probability distribution. Findings from a convolutional neural network that learnt visual features related to correlation perception 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* [48], again considered a candidate visual proxy for correlation judgements [32, 34].

Recent work investigating the use of decay functions that change point size or contrast in scatterplots as a function of residual distance provide evidence for both point density and salience/perceptual weighting being drivers of correlation perception. The use of an inverted contrast decay function [40], such that the contrast of a point decreased the closer it was to the regression line, resulted in significantly lower and less accurate correlation estimates compared to uniformly full-contrast plots with the same overall shape. This finding suggests that the lower contrast in the center of the scatterplot biased viewers' perceptions down. When point contrast or size are reduced as a function of distance from the regression line [39, 40], viewers rate correlation as significantly higher and are significantly more accurate. This supports a low-level data salience account. Our aims are 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 informed by 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 [6, 23]; scatterplots with very large numbers of points, especially those with high degrees of overlap, suffer from low individual-point visibility caused by high point density. Lowering the contrast of all points addresses

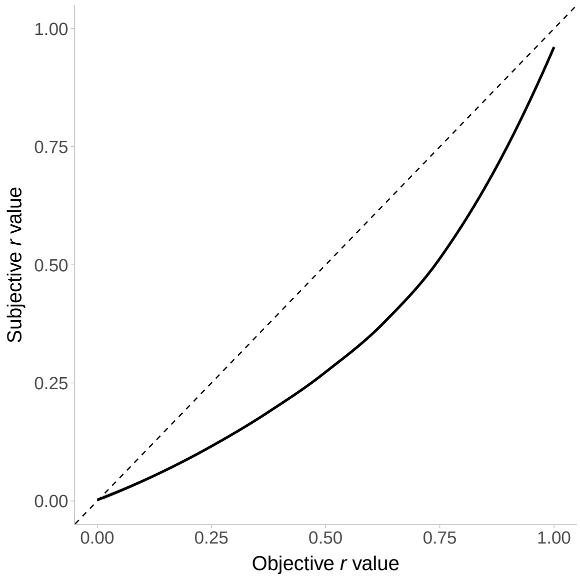

Fig. 2. Using a function relating objective to perceived r value [34] provides a visualization of the nature of correlation underestimation reported in previous work. An identity line has been included to illustrate where viewers are most and least accurate.

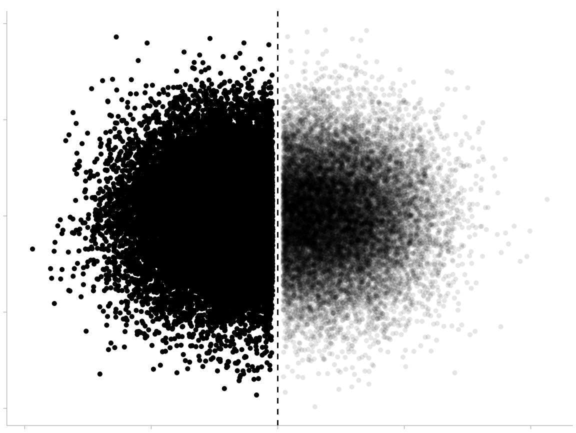

Fig. 3. Adjusting point contrast to address overplotting. Contrast between the points and the background is full (alpha = 1, L) or low (alpha = .1, R). The dataset used has 40,000 points.

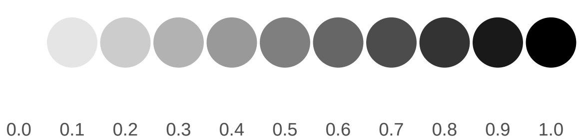

Fig. 4. Demonstrating the effects of different alpha values on point contrast.

this, and makes data trends and distributions easier to see and interpret (see Figure 3). In the present study we use the **ggplot2** package [46] to create our stimuli. This package uses an alpha parameter, or the level of linear interpolation [37] 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 rendering of either the background or foreground pixel values respectively. Psychophysical definitions of perceived contrast are often based on what is being presented (e.g., gratings) or are modeled to take into account aspects of human vision (e.g., visibility limits) [49]. Our crowdsourced methodology gives us no control over 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 report the alpha value used. Given that 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 judged as conferring 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 observed [40]. Evidence suggests point salience/perceptual weighting or spatial uncertainty as drivers for this effect. Lower stimulus contrast is associated with lower salience [15], can bias judgements of mean point position [16], increase error in positional judgements [45], and can result in greater uncertainty in speed perception [8]. Due to mechanistic accounts of both salience/perceptual weighting and spatial uncertainty predicting results in the same direction, previous work [40] 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 use 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 [31, 33], while elsewhere a strong effect of changing point size as a function of distance from the regression line has been reported [39]. Evidence points towards a salience-dominant mechanism in the latter case, albeit with a small effect of spatial uncertainty. There is evidence that larger stimulus size is associated with lower levels of spatial certainty [2], but higher levels of salience [15]. The predicted effects of these drivers on correlation perception work in the opposite direction to one another, which has allowed researchers to provide evidence for the mechanistic dominance of salience when point size is used. When an inverted size decay function is used such that smaller points are located nearer the regression line, correlation estimation is significantly more accurate than when all points are the same size [39]. In this case it was suggested that the higher spatial uncertainty brought on by larger exterior points caused a perceptual downweighting during correlation estimation, which is in line with work suggesting our perceptual systems make robust use of visuo-spatial information [39, 43, 44]. This effect was small, meaning we do not take it into account when making our hypotheses.

2.5 Hypotheses

We present a single experiment based on established effects of adjusting point size and point contrast. In our study, we combine previously independently tested point size and point contrast decay functions in both standard orientation (contrast/size is reduced with residual magnitude) and inverted orientation (contrast/size is increased with residual magnitude). Throughout we refer to *congruent* and *incongruent* conditions with respect to the combination of orientations of size and contrast decay functions. We hypothesize that: (H1) an increased reduction in correlation estimation error will be observed when congruent 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 [39], 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 viewing contexts and the wide range of levels of statistical and graph experience present in lay populations. 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 [9, 10, 28]. In light of these issues we follow published guidelines [28] 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 [28], but has served the authors well in prior work.

3.2 Open Research

This study was conducted according to the principles of open and reproducible research [3]. All data and analysis code are available at (repository link removed for anon). This repository contains instructions for building a Docker container 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

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 $r = 0.2$ [7, 11, 38]. Using so many values for r allows us to paint a broader picture of people's perception 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 x 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 to ensure that the minimum point size in the present study is both visible and consistent with that of previous work [39, 40].

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

0.25 was chosen as the value of b. This value was used in previous studies that the present work builds upon, and it produces a curve approximating the inverse around the identity line of the underestimation curve reported in previous work [34, 39, 40]. 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 this equation in standard and inverted orientation forms applied to point size and contrast in a factorial 2 x 2 design. Examples of the stimuli used can be seen in Figure 5.

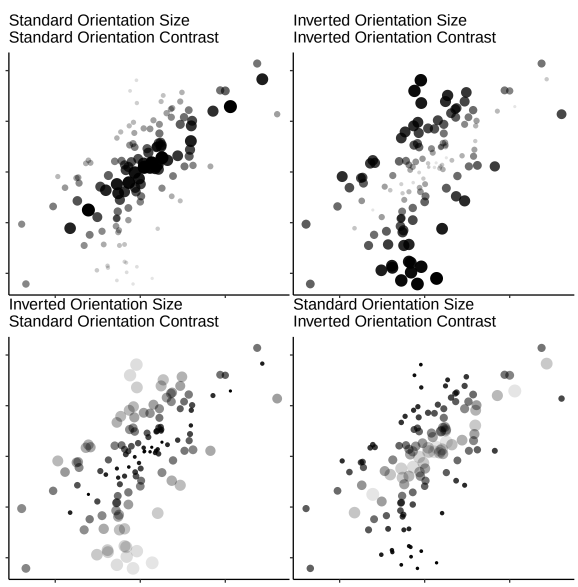

Fig. 5. Examples of the experimental stimuli used. Congruent conditions are at the top, while incongruent conditions are below.

3.4 Modeling

We use linear mixed effects models to model the relationships between the combination of size and contrast decay functions 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 were identified using the **buildmer** package (version 2.10.1, [42]) in R. This package takes a maximal random effects structure and then identifies the most complex model that converges by dropping terms that fail to explain a significant amount of variance.

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 kinds 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 check this. Participants viewed six scatterplots that were made up of a certain number of points between 2 and 7. 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 [39] found point visibility largely invariant to size. We suggest this is due to differences in monitor specifications between participants. In reality minimum visible size and contrast would need to be calibrated on a per-monitor basis. 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 [26]. We define dot pitch as the distance in millimetres between the centers of the pixels on a participants' monitor. 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 their sizes matched. We assumed a widescreen 16:9 aspect ratio and calculated dot pitch based on these measurements. Mean dot pitch was 0.34mm ($SD = 0.05$), corresponding to a physical on-screen size of 4.39mm on a 1920 x 1080 pixel monitor for the smallest points displayed. We include analyses with dot pitch as a fixed effect in Section 4.1.

3.7 Procedure

Both experiments were built using PsychoPy [29] 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 [14], followed by the point visibility 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 4.1 contains further analysis 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. Participants were asked to complete each trial as quickly and accurately as possible, although there were no time limits on individual trials. Visual masks were displayed for 1 second in between each scatterplot presentation. 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 x 2 factorial design. Each participant saw each combination of size and contrast decay function plots for a total of 180 experimental items. Participants viewed these experimental items, along with 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 over-correction and consequent overestimation for many values of r (see Figure 7). Our second hypothesis was also supported; the combination of inverted size and inverted contrast decay functions produced the least accurate estimates of correlation. We found no support for our third hypothesis; there was no significant difference in correlation estimates between standard orientation size/inverted orientation

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 in lieu of effect sizes.

	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

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** (version 1.1-34 [4]) 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 functions 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, 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 a 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 [27]. These values were calculated using the **r2glmm** package (version 0.1.2 [17]) and can be see in Table 1 along with all model statistics. The **emmeans** package (version 1.8.8 [21]) was used to calculate pairwise comparisons between levels of the size and contrast decay factor, and can be seen in Table 2.

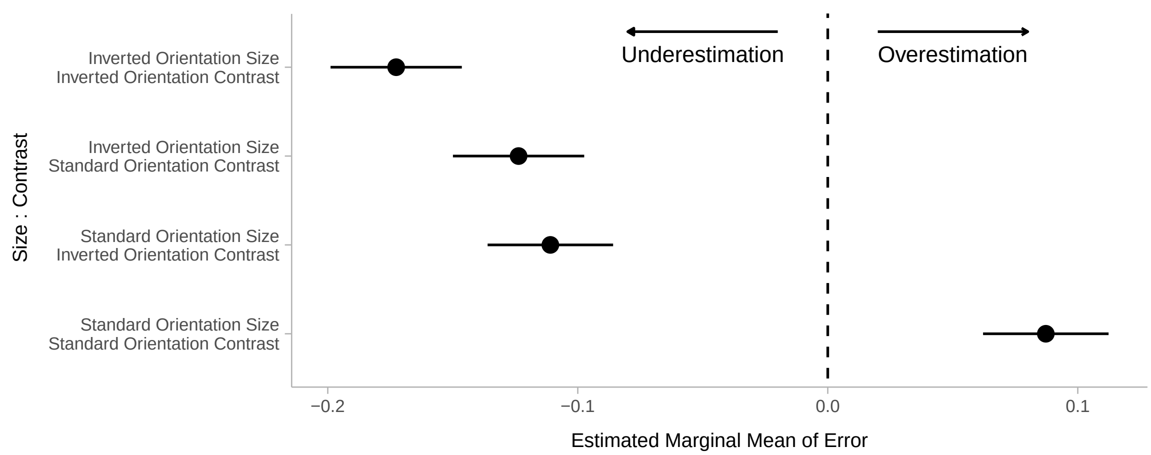

Fig. 6. Estimated marginal means for each combination of size and contrast decay conditions, including asymptotic lower and upper confidence intervals.

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 compared.

Contrast	Z.ratio	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

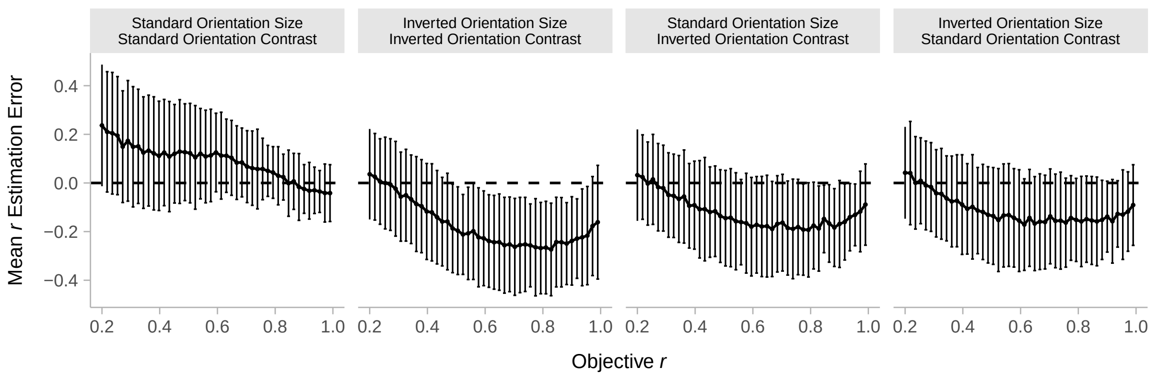

Fig. 7. Plots showing how participants' correlation estimation errors change as a function of the r value for each combination of size and contrast decay factors. Overestimation occurs above the dashed line.

4.1 Additional Analyses

We find no effects of graph literacy ($\chi^2(1) = 3.50$, $p = .061$), performance on the point visibility task ($\chi^2(1) = 1.29$, $p = .257$), or dot pitch ($\chi^2(1) = 1.52$, $p = .218$) on participants' errors in correlation estimation. We did find a significant effect of training ($\chi^2(1) = 23.78$, $p < .001$), with participants rating correlation .01 lower during the second half. This may imply that having more recently viewed the four training plots described in Section 3.7 increased participants' estimates of correlation. While this effect is statistically significant, further discussion of the effects of training on correlation perception in scatterplots is outside the scope of the present work.

5 DISCUSSION

The finding of a significant interaction between point size and contrast decay provides evidence that these functions interact in non-additive ways. In addition, we provide further confirmatory evidence of what has been found previously. Namely, that while manipulations of both point size and contrast have significant effects, the effect of changing point size is stronger, and that while we can influence correlation estimates in either direction, standard orientation manipulations are more powerful than inverted ones [39, 40]. 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 [39, 40], although this may be a facet of the non-additive interaction between size and contrast manipulations we found. 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.

Taking into account the present work, which manipulates point size and contrast together, and previous work manipulating the same visual features in isolation, we provide recommendations to visualization designers and researchers:

- The size decay function in isolation should be used when r is between approximately 0.3 and 0.75. In this range it produces the most accurate r estimates in participants.
- Outside of this range, the contrast decay function should be used in isolation, as here it produces the most accurate r estimates.
- There exists a combination of size and contrast decay functions that produces accurate correlation estimates while maintaining the increased r estimation precision that we would expect to see with high r values. Finding this will require extensive future testing.

5.1 Combining Manipulations

Figure 6 and Figure 7 show how, on average, the combination of standard orientation size and contrast decay functions 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. If we can over-correct correlation estimates, then we certainly have the ability to correct appropriately. The issue here is not one of our *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. Combining inverted size and contrast decay functions also had the predicted effect in this case, producing the lowest and least accurate estimates of correlation. Combining inverted manipulations did not, however, significantly change the shape of the estimation curve (see Figure 7). In addition to interacting non-additively, the effects we observe operate differently depending on the direction of the change induced in perception. This finding can also explain the lack of support found for our hypothesis that there would be a significant difference in r estimation error between the two incongruent conditions. Despite the size channel being more powerful with regards to influencing correlation estimates, the fact that this power depends on the direction the function is set causes incongruent functions to act against each other in ways we would not expect. Indeed, the incongruent condition that used a standard orientation size decay function did exhibit lower mean error than the one using inverted orientation size decay (see Figure 7), however in each case the contrast decay appears to have blunted the power of the size decay function to the extent that the difference in errors is not statistically significant.

5.2 Estimation Precision

Much previous work is consistent with regards to the finding that r estimation precision increases with the objective r value [13, 31, 33–35]. More recent work using size or contrast decay functions similar to the ones used here [39, 40] has found that in some cases, precision in r estimation is *constant* across the range of r values investigated. For example, the use of a size decay function, whether using standard/inverted orientation non-linear functions or a linear decay function, results in **no** change in r estimation precision [39]. When contrast is used in the same ways, only an inverted decay function **does not** exhibit the conventional increase in precision with r. In the present work, precision in r estimation increased whenever a standard orientation contrast decay function was used. We suggest this is part of the moderating effect of the point contrast decay function on the size decay function; the visual character of scatterplots with high r values that use the size decay function eliminates the usual increase in precision we would expect, however the introduction of the contrast decay function normalizes this to the point where precision is restored.

5.3 Contributions of Size and Contrast Decay

Incorporating data from previous work [39, 40] 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 8 shows correlation estimation error curves in the present experiment, in two previous studies that used decay functions applied solely to size or contrast, and with no manipulations present. This final plot is averaged over conditions from previous work.

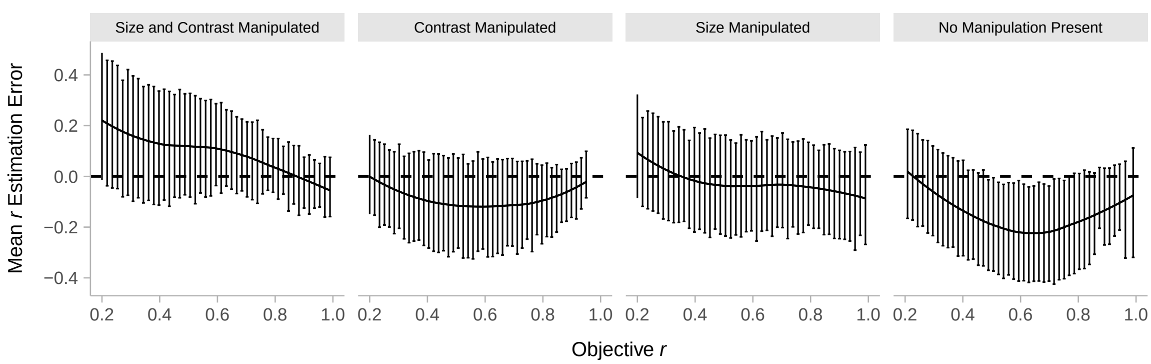

Fig. 8. Plotting r estimation error against the objective r value for contrast and size decay in isolation from previous work, and for their combination in the present study. The dashed line represents 0 error in correlation estimation, and standard deviations are shown as error bars. Note that these curves have been smoothed. Overestimation occurs above the dashed line. The plot line for the plot without manipulation was averaged from previous work.

Using contrast decay alone does not significantly change the shape of the estimation error curve, whereas using size decay does. When size and contrast decay functions are combined, however, the shape appears similar to that of size alone. This is in line with previous work establishing size as a more potent channel for the manipulation of correlation estimates [39]. It would appear that the addition of the contrast curve moderates the effect of the size curve as a function of the objective r value itself, without affecting the general shape of the curve. In the following, we briefly discuss the effects of each manipulation in isolation and the manipulations together, before making a case for the inclusion of both when tuning scatterplots for correlation estimation due to the complementary benefits each confers. 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.

Using a contrast decay function in isolation has a small effect on correlation estimation. It does little to change the shape of the underestimation curve (see Figure 2), but slightly biases r estimates up to partially correct for the underestimation observed with standard scatterplots [40]. Importantly, it also preserves the increase in correlation estimation precision that we would expect to find during correlation estimation tasks. Using the size decay function in isolation has a more dramatic effect. The shape of the estimation curve is altered quite radically, and there is no increase in precision with the objective r value [39]. Size decay over-corrects at lower values of r, leading to an overestimation effect, while at high values the curve begins to change direction, leading to a more severe underestimation. In the middle range of r values, however, the size decay function in isolation performs well. One option for tuning correlation estimation using these functions would therefore be to use the size decay function in isolation for mid-range values of r (0.3 to 0.75), and to use the contrast decay function in isolation outside of this range. Used together however, we can exploit the power of the size decay function whilst maintaining the expected increase in precision with r that the contrast decay function confers. It is clear that the simple combination we have used in the present study does not represent an ideal tuning, as participants overestimated r for the majority of values, but this confirms that there is the scope to bias r estimates significantly using the functions supplied here. Further work would be required to obtain precise measures of the power of each decay function and their power together. Doing this would allow us to tune each function according to both objective r value and the tuning of the other function to produce accurate correlation estimates. We can derive new curves that describe the effect that each manipulation and the combination of manipulations has on correlation estimates (Figure 9) by comparing them to estimates without any manipulation present. We term this 'power'. The dotted line on each plot shows the power we would need to correct for the standard underestimation curve (see Figure 2). As we can see, size alone provides the closest to the requirement, and combining size and contrast decay functions results in gross overestimation.

5.4 Mechanisms

Previous work has made the case for contrast and size decay acting primarily through salience/perceptual weighting, with the caveat that spatial certainty also plays a small part in the mechanism behind size decay [39, 40]. Our results are supportive of this notion, with our highest and lowest estimates being observed in the congruent standard and congruent inverted conditions respectively. These findings also support dot density [48] and feature-based attentional bias accounts [16, 41]. As all of these mechanisms would be expected to operate in the same direction, making conclusions about the relative contributions of each is difficult. The body of evidence generally points to a high-level probability distribution account [32, 34]. On a lower level, numerous candidate mechanisms exist, which are mostly expected to act in the same direction. Previous work concluded that spatial certainty [39] may play a small role with regards to the effects of size decay on correlation perception; our results neither confirm nor refute this, but instead provide further evidence for salience/perceptual weighting/dot density changing participants' perceptions of the width of a probability distribution to affect correlation estimates.

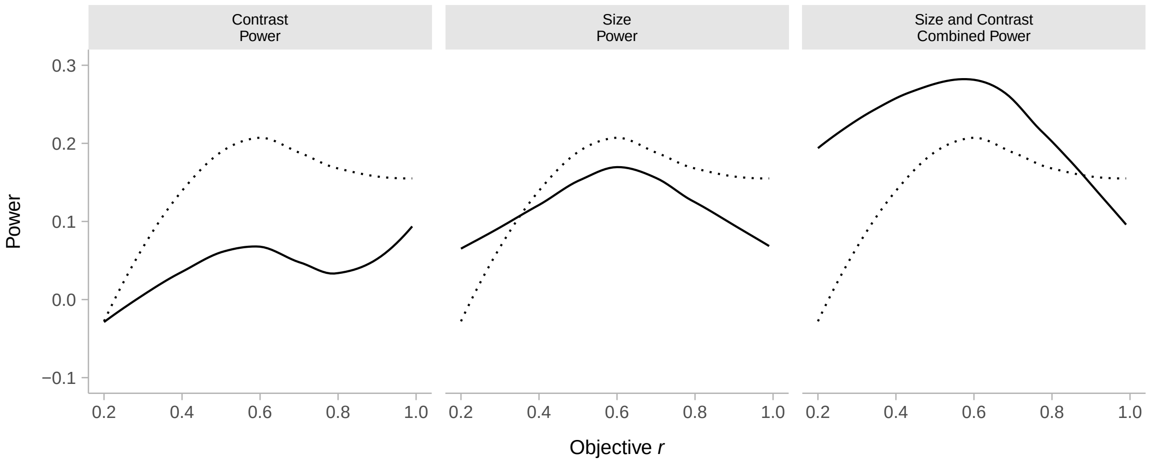

Fig. 9. Power is the difference between what is observed when a decay function/combination of decay functions is used and what is observed when no manipulation is used. The dashed line represents the power that would be required to correct for the observed underestimation of correlation in scatterplots.

5.5 Limitations

Firstly, our participants' performance on the point visibility task was poor, with an average score of only 74.89%. It is clear from these results that for many participants, the smallest and lowest contrast points we used were simply not visible, although it would seem that this low visibility had no significant effect on correlation estimates. Regardless, for many of our participants it will have appeared that we were removing data, which goes against our intended aims. Solving this would require a by-participant calibration of point size and point contrast, which is beyond the scope of our current methodology, as it would require stimuli to be fully re-generated for each participant. We aim to implement this in the future, although it would require a different platform, such as a server-based instance of Rstudio being run in the background to re-generate stimuli per a calibration task completed by the participant. We cannot say precisely what proportions of the observed effect in the standard orientation congruent condition were due to size or contrast decay. We can conclude that these effects are not linearly additive, but must suggest further work to define precisely each of their contributions.

5.6 Future Work

There is evidence that viewers overestimate correlation in negatively correlated scatterplots [36]. Future work may make use of the functions and findings we have provided to begin solving this problem. We found evidence that the influence of size and contrast decay functions changes according to the direction they are operating in, meaning experimental work with negatively correlated scatterplots would be required, and results may differ significantly from findings related to the underestimation of correlation in positively correlated scatterplots. For size and contrast decay in the present work we used equation 1. Given our finding that the combination is non-additive, there are a multitude of adjustable parameters for each decay function that require rigorous testing in order to produce concrete values of the contributions of each. The value of b is one such parameter. We used the same value of b (0.25) as in previous work [39, 40]. Changing this value can increase or decrease the severity of the effect in question. Additionally, we used a constant and a scaling factor with the size decay manipulation to ensure our points were visible. These values could also be changed. Aside from changing aspects of equation 1, there are other equations that could be used, including ones that take into account objective r value to change the values used to set point size or contrast. The present work opens the door for this future work, as it provides the necessary additional data to previous experiments using only size [39] or contrast [40] decay functions. Further testing of these manipulations in isolation and combination using different decay function parameters will allow researchers to build a more complete picture of how these visual features impact correlation estimation, and how we can exploit them to correct for well-known biases.

Through this work we also provide an example of an experimental framework that we argue should be employed to test a wider range of data visualizations, statistical summaries, and task types. Our framework is fully open source, and can be easily adapted for other charts and modalities. Doing this furthers the cause of empirically-informed data visualization design.

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] Paul Ayris, Isabel Bernal, Valentino Cavalli, Bertil Dorch, Jeannette Frey, Martin Hallik, Kistiina Hormia-Poutanen, Ignasi Labastida i Juan, John MacColl, Agnès Ponsati Obiols, Simone Sacchi, Frank Scholze, Birgit Schmidt, Anja Smit, Adam Sofronijevic, Jadranka Stojanovski, Martin Svoboda, Giannis Tsakonas, Matthijs van Otegem, Astrid Verheusen, Andris Vilks, Wilhelm Widmark, and Wolfram Horstmann. 2018. LIBER Open Science Roadmap. (July 2018). https://doi.org/10.20350/digitalCSIC/15061

[4] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

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

[6] 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

[7] 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

[8] 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

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

[10] 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

[11] 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

[12] 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

[13] 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

[14] 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

[15] 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
[16] Matt-Heun Hong, Jessica K. Witt, and Danielle Albers Szafir. 2022. The Weighted Average Illusion: Biases in Perceived Mean Position in Scatterplots. *IEEE Transactions on Visualization and Computer Graphics* 28, 01 (2022), 987–997. https://doi.org/10.1109/TVCG.2021.3114783
[17] Byron Jaeger. 2017. *r2glmm: Computes R Squared for Mixed (Multilevel) Models.* https://github.com/bcjaeger/r2glmm R package version 0.1.2.
[18] 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, 1 (Aug. 2015), 469–478. https://doi.org/10.1109/TVCG.2015.2467671
[19] 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
[20] 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
[21] Russell V. Lenth. 2023. *emmeans: Estimated Marginal Means, aka Least-Squares Means.* https://CRAN.R-project.org/package=emmeans R package version 1.8.8.
[22] 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
[23] 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
[24] Joachim Meyer and David Shinar. 1992. Estimating Correlations from Scatterplots. *Human Factors* 34, 3 (June 1992), 335–349. https://doi.org/10.1177/001872089203400307
[25] 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
[26] Wakefield L Morys-Carter. 2023. ScreenScale. https://doi.org/10.17605/OSF.IO/8FHQK
[27] Shinichi Nakagawa and Holger Schielzeth. 2013. A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-Effects Models. *Methods in Ecology and Evolution* 4, 2 (2013), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x
[28] 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* 54, 4 (Sept. 2021), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3
[29] 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
[30] Irwin Pollack. 1960. Identification of Visual Correlational Scatterplots. *Journal of Experimental Psychology* 59 (1960), 351–360. https://doi.org/10.1037/h0042245
[31] Ronald Rensink. 2012. Invariance of Correlation Perception. In *Journal of Vision*, Vol. 12. Vision Sciences Society, 433–433. https://doi.org/10.1167/12.9.433
[32] Ronald Rensink. 2022. Visual Features as Carriers of Abstract Quantitative Information. *Journal of Experimental Psychology General* 151, 8 (Jan. 2022), 1793–1820. https://doi.org/10.1037/xge0001165
[33] 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
[34] 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
[35] 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
[36] Varshita Sher, Karen G. Bemis, Ilaria Liccardi, and Min Chen. 2017. An Empirical Study on the Reliability of Perceiving Correlation Indices Using Scatterplots. *Computer Graphics Forum* 36, 3 (2017), 61–72. https://doi.org/10.1111/cgf.13168
[37] Maureen Stone and Lyn Bartram. 2008. Alpha, Contrast and the Perception of Visual Metadata. *Color and Imaging Conference* 16, 355-355 (2008), 5.
[38] 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
[39] Gabriel Strain, Andrew Stewart, Paul Warren, and Caroline Jay. 2023. Adjusting Point Size to Facilitate More Accurate Correlation Perception in Scatterplots. (9 2023). https://doi.org/10.6084/m9.figshare.24117987.v1

[40] 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

[41] Peng Sun, Charles Chubb, Charles E. Wright, and George Sperling. 2016. The Centroid Paradigm: Quantifying Feature-Based Attention in Terms of Attention Filters. *Attention, Perception & Psychophysics* 78, 2 (Feb. 2016), 474–515. https://doi.org/10.3758/s13414-015-0978-2

[42] Cesko C. Voeten. 2023. *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. https://CRAN.R-project.org/package=buildmer R package version 2.10.1.

[43] Paul A. Warren, Laurence T. Maloney, and Michael S. Landy. 2002. Interpolating Sampled Contours in 3-D: Analyses of Variability and Bias. *Vision Research* 42, 21 (Sept. 2002), 2431–2446. https://doi.org/10.1016/S0042-6989(02)00266-3

[44] Paul A. Warren, Laurence T. Maloney, and Michael S. Landy. 2004. Interpolating Sampled Contours in 3D: Perturbation Analyses. *Vision Research* 44, 8 (April 2004), 815–832. https://doi.org/10.1016/j.visres.2003.11.007

[45] C. Wehrhahn and G. Westheimer. 1990. How Vernier Acuity Depends on Contrast. *Experimental Brain Research* 80, 3 (May 1990), 618–620. https://doi.org/10.1007/BF00228001

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

[47] 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

[48] 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

[49] 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