Effects of Alternative Scatterplot Designs on Belief

ANONYMOUS AUTHOR(S)

Viewers tend to underestimate correlation in positively correlated scatterplots. However, systematically changing the size and opacity of scatterplot points can bias estimates upwards, correcting this underestimation in a simple estimation paradigm. Here we examine whether application of these visualization techniques goes beyond a simple perceptual effect and could actually influence beliefs about information from trusted news sources. We present a fully-reproducible study in which we demonstrate that scatterplot manipulations that can correct for the correlation underestimation bias can also induce stronger levels of belief change compared to unadapted scatterplots presenting identical data. Consequently, we show that novel visualization techniques can be used to drive belief change, and suggest future directions for extending this work with regards to altering attitudes and behaviours.

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

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

ACM Reference Format:
Anonymous Author(s). 2025. Effects of Alternative Scatterplot Designs on Belief. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (CHI).* ACM, New York, NY, USA, 17 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Utilized for communication in a wide variety of contexts, scatterplots are simple representations of (usually) bivariate data. In 1983, they were estimated to account for between 70 and 80 percent of data visualizations in scientific publications [50], and while there is no doubt that the range of visualizations employed in and beyond science is now far broader, scatterplots remain an important tool for the visualization designer. Evidence that scatterplots are interpreted rapidly [40] facilitates the quick collection of large amounts of data, and their ubiquity [50] and low levels of interindividual variance [22] make them particularly suitable for studying perceptual and cognitive phenomena regarding data visualization.

While most commonly used to communicate the linear correlation, or level of relatedness, between a pair of variables, scatterplots can also be designed to facilitate the detection of outliers, to convey differences between clusters, or to display non-linear correlations [42]. The suitability of scatterplots for such a range of tasks, and the opportunity for designers to design with multiple tasks in mind, plays a large part in their popularity. Building on previous work, we elect to focus on the use of scatterplots for the communication of linear, bivariate, positive correlation. There is evidence that, while correlation perception in scatterplots is characterized by low levels of interindividual variance (especially when compared to other visualizations

Manuscript submitted to ACM

that communicate the same idea [19, 22]), our accuracy in interpretation is poor. Studies asking participants to numerically estimate correlation [6, 11, 13, 24, 25, 31, 44] or estimate it via a bisection task [41] find consistent levels of underestimation, particularly when $0.2 < r < 0.6$. If scatterplots were used solely for communication between those trained in statistics and data visualization, this would not be particularly problematic, however, this is not the case; lay people are expected to be able to use and interpret data visualizations on an almost daily basis. It is thus the duty of those who design such visualizations to design with the naive, inexperienced viewer in mind. Doing so requires us to understand *how* visualizations work, and to gain an appreciation for the hidden processes that allow pictorial representations to convey more than words and numbers ever could.

Recent work has sought to address the correlation underestimation bias in positively correlated scatterplots through the use of novel point encodings. Recently, Strain et al. [45–47] exploited the notion that viewers use the width of a probability distribution conveyed by the arrangement of scatterplot points as a proxy for their judgements of correlation to successfully correct for the underestimation bias. At the time of writing, this work has only provided evidence about perceptual effects using a simple direct estimation paradigm, and while successful, has not investigated whether, and to what extent, these techniques can influence cognition in the context of real-world data visualizations and the relatedness between variables.

Data visualization is a powerful tool. After all, if numerical data were sufficient for understanding, there would be no need to visualize data beyond aesthetic preference. Pattern recognition, attention, and familiarity are all aspects of human perception and cognition that can be exploited by visualization designers to facilitate more efficient, enjoyable, and effective communication [15]. This, however, is a double-edged sword; poor design, be it malevolent or misguided, can cause distrust, confusion, and misunderstanding amongst viewers. It is for these reasons that we choose to study belief change in scatterplots as a consequence of alternative designs. Scatterplots, like many other data visualizations, have been submitted as evidence in court cases [6], and play key roles in organizational decision-making, including in healthcare [38]. It is reasonable to assume that data visualizations are used to make decisions that result in positive or negative outcomes with regard to health and policy more generally, especially given findings that in certain contexts, they are more persuasive than textual information [34]. Studying the potential for new designs to alter beliefs about relatedness facilitates better visualization techniques, but also allows us to understand how these designs might be used by malevolent actors with a view to inoculating those who engage with them. To this end, we present a two-experiment study. First, we use crowdsourcing to select part of our experimental stimuli, then we test the propensity for previously established alternative scatterplot designs to alter beliefs about relatedness, taking into account the emotional content of the statement and the graph literacy and defensive confidence of participants.

2 RELATED WORK

In this section, we briefly discuss related work on correlation perception and estimation, the history and current state of the use of point size and opacity adjustments in scatterplots, including how these visual features have been used to correct for the underestimation bias, and perception and cognition in data visualization.

2.1 Correlation Perception

Correlation describes the level of relatedness between a number of variables. Here we focus exclusively on the most commonly used correlation metric - Pearson's r. This statistic takes values between -1 and 1 depending on the direction of the relationship being described. The perceptual and cognitive mechanisms that drive the interpretation of correlation from scatterplots are not well understood, however, some experimental results point towards the shape of the underlying probability distribution represented by the point cloud as a likely candidate. Scatterplots with lower area point clouds produce increased judgements of correlation [11], suggesting that it is the area of the point cloud that may influence perception. Work exploring the relationship between subjective and objective r values in scatterplots found that this relationship could be described by a power function that included the mean of the geometrical distances between scatterplot points and a regression line [32]. Other work includes some representation of the shape of a scatterplot's point cloud in equations describing magnitude estimation and correlation discrimination [32, 41], and work on visual features as proxies for correlation found that a similar quantity again is predictive of performance on correlation judgement tasks [54]. While we cannot say that this *is* the process of correlation perception in scatterplots, we can conclude that the shape of the point cloud is a good proxy for what is actually occurring during judgements of correlation. The goal of the work presented is to investigate whether well-established perceptual effects can have impacts on higher-level cognitive factors, in particular, beliefs.

2.2 Scatterplots: Opacity, Size, and Recent Developments

Changing the opacities and sizes of points in scatterplots are standard practices during the design process. Regarding opacity, this is often uniformly lowered to address overplotting issues that arise when visualizing very large datasets [29]. Similarly, scatterplots describing large datasets tend to have smaller points to maintain individual point discriminability. Point size has also been used to encode an additional third variable in what are known as *bubble charts*. Despite these techniques being established, there is relatively little experimental work on the effects of changing point opacities and sizes on correlation estimation. Some studies have found that correlation estimation is invariant to changes in point opacities and sizes [40, 41], while more recent work reports strong effects of the systematic adjustment of each visual feature [45–47]. The idea that it is the shape of the point cloud, and the probability distribution it represents that informs judgements of correlation has received support from recent work exploiting visual features to make correlation estimation more accurate. Strain et al. [45–47] changed the sizes and opacities of points in scatterplots as a function of their distance from the regression line and achieved success in biasing correlation estimates in positive and negative directions. When point opacities [46] or point sizes [45] were reduced with residual distance, participants were significantly more accurate on a correlation estimation task; employing both of these manipulations simultaneously [47] resulted in an overshoot of correction, biasing participants further. Figure 1 contains a summary of previously tested scatterplot manipulations and their effects on performance on a correlation estimation task. In those works, the opacities and sizes of scatterplot points are changed using equation 1:

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

Our study aims to investigate the potential for alternative scatterplot designs to have effects on cognition. For this reason, and to facilitate comparison to previous work, we utilize the same protocols here to produce the stimuli for our main study. This includes the number of points (n = 128), the value of b (0.25), and the size scaling factor and opacity floor. For our atypical scatterplot condition, we choose to use the previously established manipulation that has demonstrated the most dramatic change in participants' estimates of correlation: a combination of point opacity and size adjustments [47] (see the right pair of plots in Figure 1).

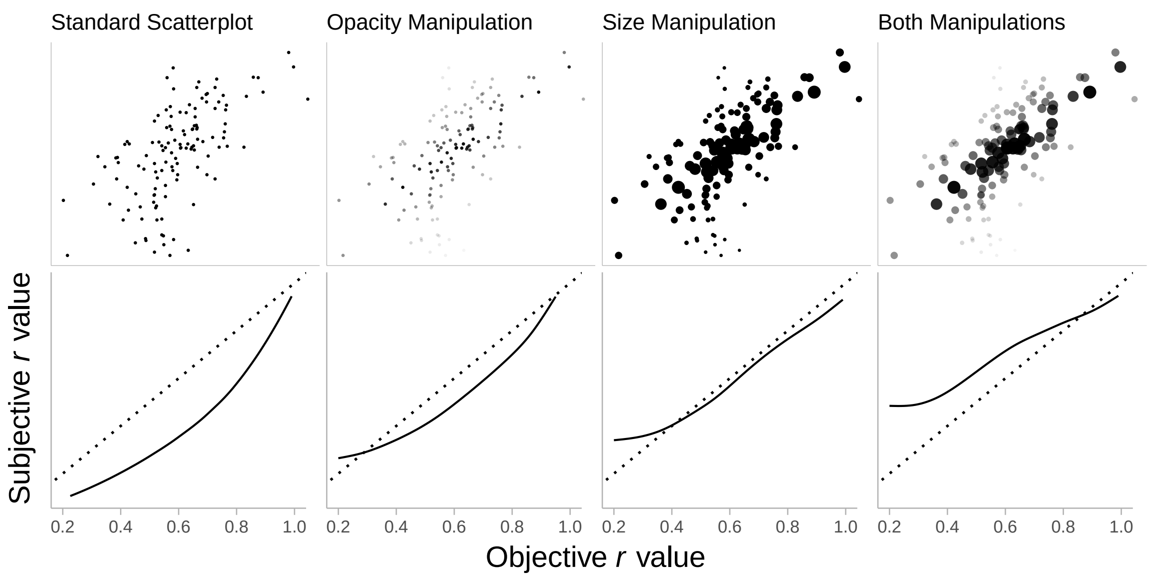

Fig. 1. Top row: Examples of scatterplot manipulations from previous work using an r value of 0.6. Bottom row: the corresponding correlation estimation behaviour across values of r between 0.2 and 0.99. The dashed diagonal line represents perfect estimation, while the solid line is what is observed when participants are asked to estimate correlation.

2.3 Perception & Cognition in Data Visualization

Interacting with data visualization is a complex process involving bottom-up and top-down mechanisms [15, 43, 53]. Previous work investigating alternative scatterplot designs has predominantly focused on perceptual factors and mechanisms; here we introduce the potential for top-down effects to bias participants. Data visualization does not take place without context, and so the investigation of top-down effects is critical for providing designers with the tools to design visualizations that work as intended in the field. We therefore present a two-experiment study investigating the propensity for recently established scatterplot visualization techniques to bias participants' beliefs about the levels of relatedness between variables.

3 GENERAL METHODS

In this section, we discuss our general research methods, including our implementations of open research practices and our approach to and justification for crowdsourcing.

3.1 Open Research

Both our pre- and main studies were conducted according to the principles of open and reproducible research [4]. We pre-registered hypotheses and analysis plans with the Open Science Framework (OSF) for the pre-study[1] and the main experiment[2]. All data and analysis code are included in a GitHub repository[3]. This repository contains instructions for building a Docker container [30] that reproduces the computational environment the paper was written in. This allows for full replication of stimuli, figures, analyses, and the paper itself. Ethical approval was granted by the (removed for anon).

3.2 Crowdsourcing

While much prior work into correlation perception in scatterplots has taken place in person, there is precedent for work that explores cognition and perception to take place online using crowdsourced participants [53]. Crowdsourcing affords us recruitment of samples from across our lay population of interest, and it is considerably quicker and less expensive than in-person testing. Previous work has reported issues of data quality and skewed demographics [8, 9, 35], so we follow published guidelines [35] to give us the best chance of collecting high-quality data. We use the Prolific.co platform [1] with strict pre-screening criteria; 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.

4 PRE-STUDY: INVESTIGATING BELIEFS ABOUT RELATEDNESS STATEMENTS

The goal of the present study is to investigate to what extent a novel scatterplot design can alter participants' beliefs about the level of relatedness between variables. There is evidence that belief change can be affected by prior beliefs and attitudes [28, 53], and that emotion, including the content of a visualization [18, 37] and the emotional state of a participant [49] can have perceptual and cognitive effects on participants. We were unable to find resources for correlative statements that included ratings for belief strength and statement emotionality, so elected to create our own. To control for these factors as much as possible, we ran our pre-study with the intent of finding a correlative statement that was matched on emotional content and level of belief strength. Instead of creating these statements ourselves, we chose to streamline the process by using the ChatGPT4 Large Language Model [2]. We used the following prompt:

> "Generate 100 statements that describe the correlation between two variables, such as: "X is associated with a higher level of Y" or "As X increases, Y increases". Try to match all the statements on emotionality."

The full list of these statements can be found in the supplementary materials. Two authors rated each statement on emotionality and strength of relatedness using Likert scales from 1 to 7. Both statement emotionality and strength of relatedness were anchored at points 1 and 7: *Very Negative* and *Very Positive* for the former, and *Not Related At All* and *Strongly Related* for the latter. All other points were unlabelled. We calculated a quadratic weighted Cohen's Kappa between the two raters using the **irr** package (version 0.84.1 [16]) to penalize larger magnitude disagreements more harshly. We found agreement above chance for both statement emotionality ($\kappa = 0.49$, $p < .001$) and strength of correlation ($\kappa = 0.51$, $p < .001$), indicating

[1] removed for anon
[2] removed for anon
[3] removed for anon

moderate levels of agreement in both cases [12, 14]. Following this, we selected strongly and weakly correlated statements with the highest level of absolute agreement, resulting in 14 strongly correlated statements and 11 weakly correlated statements that can be seen in the supplementary materials. We then tested these 25 statements with a representative UK sample to ascertain consensus on both statement emotionality and strength of relatedness. Doing so allows us to minimize the impact of these factors when we analyse the effects of atypical scatterplot design on the propensity for belief change in our main experiment. We hypothesized that:

- H1: there will be a significant difference in average ratings of emotionality between statements.
- H2: there will be a significant difference between average ratings of strength of relatedness between statements.

4.1 Method

4.1.1 Participants. 100 participants were recruited using the Prolific.co platform [1]. English fluency and UK residency were required for participation, as our main experiment relied on familiarity with data visualizations from a popular British news source. In addition to 25 experimental items, we included six attention check items asking participants to ignore the scatterplot and provide specific answers. No participants failed more than 2 out of 6 attention check items, and therefore data from all 100 were included in the full analysis (52% male and 48% female). Participants' mean age was 41.1 ($SD = 12.3$). The average time taken to complete the survey was 7.6 minutes ($SD = 2.9$ minutes).

4.1.2 Design. Each participant saw all survey items (see supplementary material), along with the six attention check items, in a fully randomized order. All experimental code, materials, and instructions are hosted on GitLab[4].

4.1.3 Procedure. The experiment was built using Psychopy [36] and hosted on Pavlovia.org. Participants were permitted to complete the experiment using a phone, tablet, desktop, or laptop computer. Participants were first shown the participant information sheet and were asked to provide 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 were told that they would be asked to read statements about the relatedness between a pair of variables, after which they would have to indicate their beliefs about statement emotionality and the strength of relatedness suggested using a pair of sliders. To familiarize themselves with the sliders, they were asked to complete a practice round in response to the statement: "As participation in online experiments increases, society becomes happier." The Likert scales used in the online experiment were identical to those described in Section 4.

4.2 Results

All analyses were conducted using R (version 4.4.1 [39]). We use the **irr** package to calculate Fleiss' Kappa to measure interrater agreement on statement emotionality and strength of relatedness for the 25 experimental items. This analysis revealed that participants agreed above chance on statement emotionality ($\kappa = 0.07$, $p < .001$) and strength of relatedness ($\kappa = 0.06$, $p < .001$).

[4]removed for anon

4.3 Selecting Statements for the Main Experiment

To control for the potential effects of statement emotionality in the main experiment, we first select statements that represent neutral emotional valence. Statements with average emotionality ratings between 3 and 5 are statements 2, 10, 22, 16, and 23. To ascertain which statements represent the greatest consensus, we add standard deviations in ratings for statement emotionality and strength of relatedness. Due to concerns about experimental power, and in line with evidence that propensity for belief change is highest when prior beliefs are not strongly held [28, 53], we elected at this point to test only the statement corresponding to weak beliefs about the strength of correlation between the variables in question. We therefore test statement number 22, “Higher consumption of spicy foods is associated with a lower risk of certain types of cancer”, however we modify the wording so that the variables (food consumption and cancer risk) are positively correlated, as while the manipulations we use in the atypical scatterplot condition can affect estimates of correlation in positively correlated scatterplots, no work regarding the effects of these manipulations in negatively correlated scatterplots has been completed.

4.4 Discussion

Fleiss’ Kappa values for interrater agreement on both statement emotionality and strength of correlation scales are low ($\kappa = 0.07$ and $\kappa = 0.06$ respectively), however do exceed that which would be expected by chance. We suggest this may be due to Fleiss’ Kappa not being designed with ordinal (Likert scales in this case) data in mind. In light of this we do not make decisions regarding which statement to use based on the values of Fleiss’ Kappa observed, but rather on the standard deviations of ratings across all raters. We also test statement emotionality and strength of correlation with participants in the main study and include these ratings as part of our analysis.

5 MAIN STUDY: POTENTIAL FOR BELIEF CHANGE USING ATYPICAL SCATTERPLOTS

We test the statement that exhibited the lowest average level of belief about correlation and the 2nd highest level of consensus. Modified for directionality, this statement is therefore: “Higher consumption of plain (non-spicy) foods is associated with a higher risk of certain types of cancer.” To give ourselves the best chance of detecting an effect of viewing atypical scatterplots, we elected to design our scatterplots based on a popular British news source and to falsely credit the data as being provided by the British National Health Service (NHS). Participants were informed that said news source had requested that their identity be obscured, and were debriefed that this was not the case and that the data were fictional, following completion of the experiment. We hypothesized that:

- H1: there will be a significant difference in ratings of strength of relatedness before and after participants viewed the experimental items.
- H2: this difference will be greatest when participants are exposed to scatterplots in the atypical condition.

5.0.1 Defensive Confidence. In line with evidence that those who are more confident in their ability to defend their own positions are more susceptible to having those positions changed [3], we measured participants’ defensive confidence using a 12-item scale. This scale is replicated from previous work in the supplemental material, and has been utilized more recently [28] to explore the potential for attitude change specifically with regard to correlations in scatterplots. Participants provide answers to the 12 scale items using a 5-point Likert scale anchored at points 1 (*not at all characteristic of me*) and 5 (*extremely characteristic of me*), with all other points being unlabelled. Analysis including participants' defensive confidence scores is included in Section 5.4.1.

5.1 Stimuli

Having selected a correlative statement describing a weak relationship and with a high level of consensus between participants in the pre-study, we used **ggplot2** (version 3.5.1 [52]) in R to create our stimuli; scripts for the creation of stimuli can be found in the repository associated with this project. As our statement was rated as describing a low level of relatedness, we utilize scatterplots that describe a strong relationship ($0.6 > r > 0.99$) to induce belief change. Our plots were created in line with guidance provided by previous research and detailed in Section 2.2. We used 45 values of r uniformly distributed between 0.6 and 0.99 to create 45 scatterplots for each condition. Examples of stimuli using an r value of 0.6 for both the typical and atypical scatterplot conditions can be seen in Figure 2.

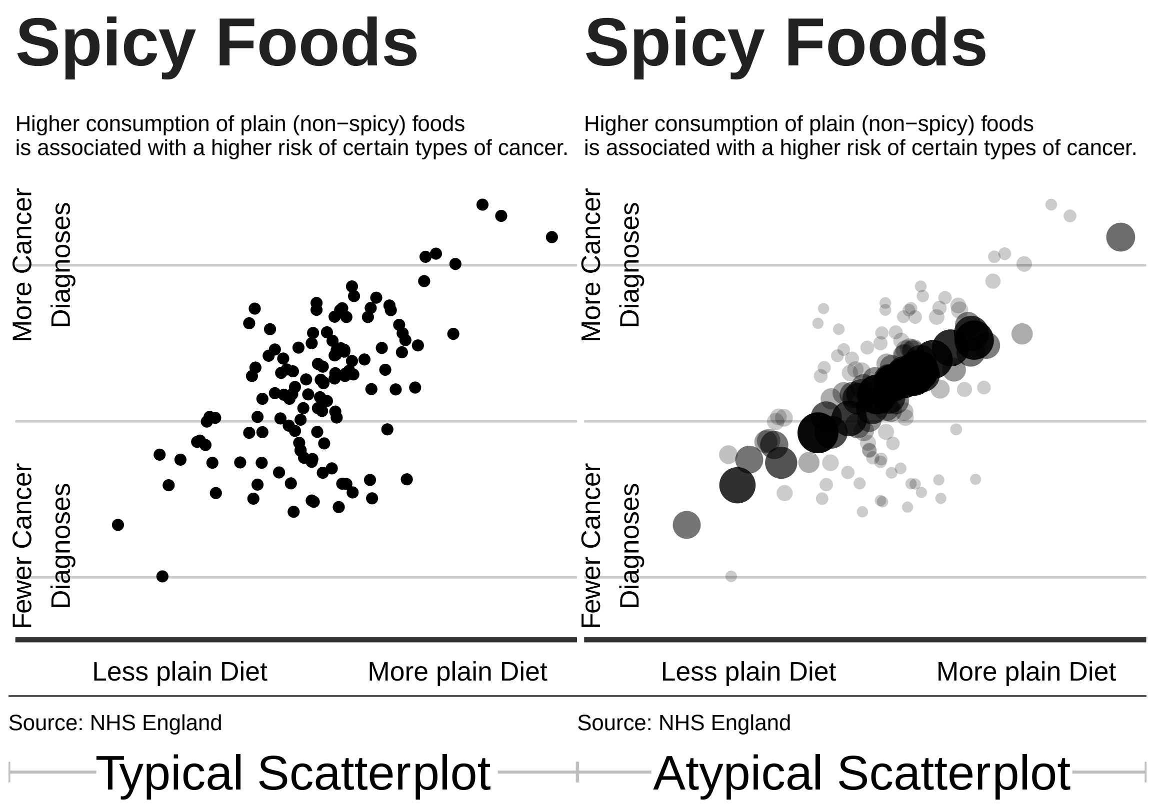

Fig. 2. Examples of the experimental stimuli used with an r value of 0.6.

5.2 Method

5.2.1 Participants. Participants were recruited using Prolific.co [1]. English fluency and UK residency were required for participation, as well as normal or corrected-to-normal vision, and having not participated in any of our previous studies regarding correlation perception in scatterplots [refs removed for anon]. Data were collected from 77 participants for each condition. 2 participants failed more than 2 out of 4 attention check questions for each condition, meaning their data were excluded per pre-registration stipulations. Data from the remaining 150 participants were included in the full analysis (48.7% male, 48.7% female, and 2.7% non-binary). Participants' mean age was 39.3 ($SD = 11.5$). Participants' mean graph literacy score was 21.3 ($SD = 4.3$) out of 30, their mean defensive confidence score was 43.0 ($SD = 6.8$) out of 60, and their mean rating of statement emotionality was 2.9 ($SD = 1.3$) on a 7-point Likert scale. On average, participants took 14.2 minutes to complete the experiment ($SD = 6.41$).

5.2.2 Design. We employed a between-participants design. Each participant was randomly assigned to either group A, in which case they viewed typical scatterplots, or group B, in which they viewed atypical scatterplots designed deliberately to elicit higher levels of belief change. Participants saw all 45 experimental items for their group, along with 4 attention check items, in a fully randomized order. Our dependent variable was the level of belief change induced by viewing the scatterplot visualizations, so participants were tested on how strongly related they believed the variables described by the correlative statement were **before** and **after** viewing the experimental items. All experimental code, materials, and instructions are hosted on GitLab as two separate experiments [5] [6].

5.2.3 Procedure. We used PsychoPy [36] to build our experiment and Pavlovia.org to host it. Participants were permitted to complete the experiment on a desktop or laptop computer. We elected to prevent participants from using a phone or tablet to complete the experiment as there is evidence that differences in the on-screen sizes of data visualizations can alter perceptions [11]. Participants were first shown the participant information sheet and asked to provide consent through key presses in response to consent statements. They were, again, asked to provide their age and gender identity. Participants then completed the 12-item Defensive Confidence scale described by Albarracín and Mitchell [3] and the 5-item Subjective Graph Literacy scale [17] [7]. To give legitimacy to our data visualizations with the hope of maximizing any potential belief change, participants were told that the graphs were taken from a well-known British news source, but that the identity of this source had been obscured. To promote engagement with the visualizations, participants were instructed to use a slider to estimate the correlation displayed in each scatterplot; no hypotheses were made based on these data and therefore we do not analyse them further. Following instructions, which included descriptions of scatterplots and Pearson's r, participants had a chance to practice using the slider, before being asked to indicate their belief about emotionality and the relatedness between variables described in our chosen statement. We captured these data using Likert scales identical to those described in Section 4. After completing the experimental trials, participants were tested again on their beliefs about relatedness and then debriefed that the data they were shown were fictional. Interspersed

[5] removed for anon
[6] removed for anon
[7] The inclusion of this scale was not specified in the pre-registration.

Table 1. Statistics for the significant main effect of rating time. Odds ratio and the equivalent Cohen's *d* value is also supplied.

	Estimate	Standard Error	Z-value	p	Odds Ratio	Cohen's d
Rating Time	3.77	0.049	76.62	<0.001	43.2	2.08

among the experimental items were 4 attention check trials which explicitly asked participants to set the slider to 0 or 1.

5.3 Results

All analyses were conducted using R (version 4.4.1 [39]). Likert scales capture whether one rating is higher or lower than another, however they do not quantify the difference between levels of rating. Metric modelling assuming equal levels of difference between ratings, such as linear regression, is therefore inappropriate [27]. In light of this, we use the **ordinal** package (version 2023.12-4.1 [10]) to build cumulative link mixed effects models to analyse Likert scale data [8]. We use the **buildmer** (version 2.11 [51]) package to automate the selection of the random effects structure; we provide a maximal model, which includes random intercepts for participants, and **buildmer** identifies the most complex model that successfully converges. We present odds ratios and equivalent Cohen's d effect sizes that were calculated using the **effectsize** package (version 0.8.9 [5]).

To test the first hypothesis, that ratings of strength of relatedness would be different before and after participants viewed experimental items, we build a model whereby the rating of strength of relatedness the participant made is predicted by whether it was made **before** or **after** viewing the experimental items. Our first hypothesis was supported; there was a significant difference in ratings of strength of relatedness made before and after participants viewed the experimental plots. A likelihood ratio test revealed that the model including time of rating as a predictor explained significantly more variance than the null ($\chi^2(1) = 8{,}046.95$, $p < .001$). This model has random intercepts for participants. Statistical testing providing support for this hypothesis is shown in Table 1. Figure 3 shows means and boxplots for ratings of strength of relatedness before and after viewing scatterplots in either the typical or atypical condition.

Our second hypothesis, that the difference between ratings of strength of relatedness before and after viewing experimental plots would be greater when participants were assigned to the atypical scatterplot condition, also received support. Deviation coding was used for each of the experimental factors of rating time (pre- or post-) and condition, which allows us to compare means of ratings to the grand mean. We built a cumulative link mixed effects model whereby the rating of strength of relatedness the participant made was predicted by the condition the participant was assigned to and the time they made the rating. A likelihood ratio test revealed that the model including condition and rating time as a predictor explained significantly more variance than the null ($F(3) = 8{,}151.94$, $p < .001$). This model had random intercepts for participants. We again found a main effect of rating time, found no main effect of condition, and found an interaction between rating time and condition. Test statistics, along with odds ratios and equivalent Cohen's

[8]Our pre-registration specified linear mixed effects models (metric). Conclusions are identical when using said models, and code for this analysis is included in the repository associated with this paper.

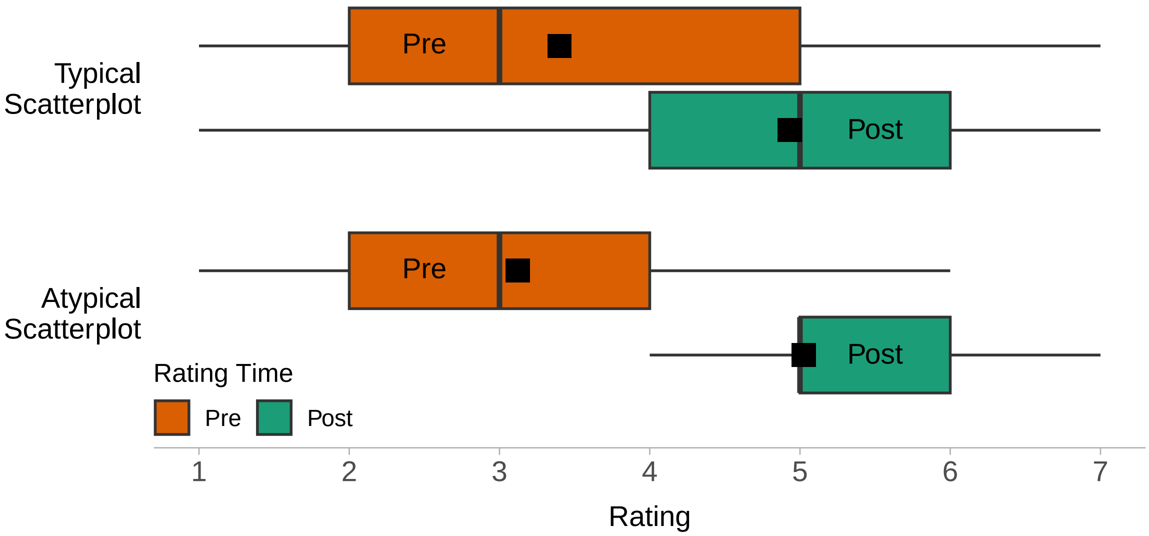

Fig. 3. Boxplots showing ranges, interquartile ranges, medians (vertical lines) and means for participants' ratings of strength of relatedness before and after viewing either typical or atypical scatterplots.

Table 2. Statistics for the significant main effect of condition on the difference between pre- and post- scatterplot viewing ratings for typical and atypical plots. Odds ratios and equivalent Cohen's *d* are also shown. NB: the odds ratio for the effect of condition is caluclated based on the absolute value of the estimate.

	Estimate	Standard Error	Z-value	*p*	Odds Ratio	Cohen's *d*
Rating Time	3.79	0.049	76.69	<0.001	44.10	2.09
Condition	-0.13	0.390	-0.33	0.745	1.14	0.07
Rating Time x Condition	0.72	0.071	10.22	<0.001	2.06	0.40

Table 3. Pairwise comparisons. The interaction is driven by there being greater differences in ratings of belief strength made before and after viewing plots in the atypical condition compared to the typical.

Contrast	Z ratio	*p*
Post-Viewing x Atypical | Pre-Viewing x Atypical	66.32	<0.001
Post-Viewing x Atypical | Post-Viewing x Typical	0.60	0.932
Post-Viewing x Atypical | Pre-Viewing x Typical	9.32	<0.001
Pre-Viewing x Atypical | Post-Viewing x Typical	-9.96	<0.001
Pre-Viewing x Atypical | Pre-Viewing x Typical	-1.25	0.596
Post-Viewing x Typical | Pre-Viewing x Typical	58.12	<0.001

d can be seen in Table 2. To further explore the interaction, we used the **emmeans** package (version 1.10.4 [26]) to calculate pairwise comparisons between levels of the condition and rating time factors, which can be seen in Table 3. The interaction we found is driven by participants' beliefs changing more between pre-viewing and post-viewing times for atypical as opposed to typical plots.

5.3.1 Additional Analyses. We also found effects of participants' scores on the defensive confidence test ($F(4) = 69.73$, $p < .001$), participants' scores on the graph literacy test ($F(4) = 42.66$, $p < .001$), and of how emotional participants rated the chosen correlative statement before beginning the block of trials ($F(4) = 43.51$, $p < .001$). We discuss the interactions between the main effect and graph literacy, defensive confidence, and statement emotionality in Section 5.4.1.

5.4 Discussion

Both of our hypotheses were supported in this experiment. Participants reliably updated their beliefs after viewing scatterplots, and the difference between pre- and post-viewing beliefs was greatest for the participants who viewed scatterplots in the atypical condition. These results suggest that the effects described in previous work can go beyond perception and have impacts on higher-level cognitions, specifically, participants' strength of beliefs about relatedness. These findings are encouraging for data visualization designers who wish to design scatterplots such that correlation perception more closely matches the underlying statistics, however further work is required before developing guidelines for the use of alternative scatterplot designs with regards to producing more persuasive visualizations.

Fig. 4. Illustrating how differences in beliefs about strength of relatedness change as a function of participants' scores on the graph literacy test (Left), their scores on the defensive confidence test (Middle), and their ratings of statement emotionality (Right). Localled smoothed cruves are shown separately for typical and atypical scatterplot viewing conditions. Lower ratings of Difference in Belief (y axis) correspond to lower levels of belief change between pre- and post- scatterplot viewing times.

5.4.1 Graph Literacy, Defensive Confidence, and Statement Emotionality. Mean differences in pre- and post-plot-viewing ratings of strength of relatedness by Subjective Graph Literacy score can be seen in Figure 4. Generally, participants who scored higher on a graph literacy test experienced smaller changes in their ratings of strength of relatedness. While the effect size associated with this interaction is small (~ 0.01), it is in line with previous work suggesting that those with higher levels of graph or visualization literacy show better performance in inference tasks related to visualizations [7], are more capable of describing effects that visualizations aim to communicate [43], and can preferentially attend to relevant features of visualizations to a greater degree [33], than those with lower levels of graph literacy. In the present study, we provide evidence that those with greater levels of graph literacy are *less susceptible* to having their beliefs changed by visualizations. The use of the plot manipulation largely removes this effect, suggesting that there is less systematic reliance on graph literacy when participants are faced with an unfamiliar data visualization.

We observe an opposing pattern of results when examining the effects of defensive confidence on participants' propensity for belief change. Generally, participants who scored more highly on the defensive confidence test experienced greater levels of belief change. This is in line with evidence that those who are more confident in their ability to defend their own beliefs are more liable to having those beliefs changed in light of evidence [3]. This effect has previously been explained as being due to those with a greater degree of confidence in their own ability to defend their ideas engaging with information with lower levels of attention to the fact it opposes their beliefs. The present study provides additional evidence in favour of this phenomenon. While the general pattern of results is expected based on previous work, the interaction present between defensive confidence and scatterplot condition is novel (see Figure 4). It would appear that despite following the normal pattern of results for low to moderate levels of defensive confidence, those participants who viewed the typical scatterplots experienced a drop in belief change as defensive confidence increased past ~ 36/60. We suggest that the unfamiliar nature of the atypical scatterplots was protective against an unexpected, standard behaviour whereby very high levels of defensive confidence decrease susceptibility to belief change.

The effect of statement emotionality on belief change is also illustrated in Figure 4. There is a broad research space regarding emotionality and data visualization [23], and it is clear from previous work that emotion affects perception, cognition, and behaviour [18, 37, 49] regarding to data visualization. Harrison et al. [18] found that participants who were positively primed performed better on a low-level visual judgement task. Comparison of this work to the current is difficult, as *success* on our task is hard to define.

Further experimental work is required to provide more comprehensive explanations for the interactive effects of graph literacy, defensive confidence, and statement emotionality in the current experimental paradigm.

6 GENERAL DISCUSSION

The most parsimonious explanation for the results we observe in the present study is as follows; things that *look* more related will be *judged* as being more related, and are therefore *more* able to change beliefs about the levels of relatedness between variables. Given the frequent real-world usage of scatterplots, and the role of data visualizations in decision-making, it is particularly important to empirically test whether perceptual effects may be extended to influence beliefs. Doing so is a necessary step in broadening the data visualization design space and bringing novel designs closer towards use cases with the potential for real-world consequences while maintaining a strong foundation of experimental evidence. Having controlled as far as possible for factors such as the emotional content of the graph, the consensus on how related the variables in question were, and the general design (bar the points themselves) of the scatterplot, we can conclude with strong evidence that atypical scatterplot design was responsible for increasing the level of belief change amongst participants.

Previous work has provided support for the idea that it is the shape of the point cloud, more specifically, the width of the probability distribution it represents, that drives correlation perception in scatterplots. If this mechanism were valid, our results would be expected. These results are broadly consequential. For data visualization designers, they provide strong evidence that utilizing alternative scatterplot designs described here and in previous work can affect beliefs about levels of relatedness without requiring the removal of data. For researchers, these results pave the way for work in a number of directions, which we discuss in Section 7.

7 FUTURE WORK

Because alternative scatterplot designs have not been tested before with regard to belief change, we elected to design our study with the intention of capturing effects, should they exist. This meant that our design was simple; we did not investigate multiple correlative statements, the propensity for strongly held beliefs to be changed, nor the effect of topics with strong or polarized emotional components. We chose to use a simple, blunt measure of belief about relatedness, and did not investigate variations within the alternative scatterplot design space, such as using different values of equation 1, or using opacity or size manipulations in isolation. Each of these components deserves study, and each is ripe for future work to investigate the contributions of each factor to the effects we have seen here.

In Section 5.4.1, we describe the effects that graph literacy, defensive confidence, and participants' ratings of the emotionality of the correlative statement have on the propensity for belief change. Future work may wish to investigate these factors, along with others that affect perceptions of correlation, such as educational background or spatial abilities [48]. Xiong et al. [53] describe how correlation estimation may differ according to the context the data are presented in; this could be extended to instead investigate statements with differing emotional contents and how alternative scatterplot designs might interact with emotional valence. Similarly, selecting matched participant groups with low or high graph literacy or defensive confidence would facilitate understanding of how we might use alternative designs to cater for people with different levels of experience, or who differ in terms of their faith in their own ideas and abilities.

Previous works investigating beliefs with regard to correlation estimation have made distinctions between beliefs and attitudes [28, 53]. We elected not to do so due to our utilization of alternative designs. Markant et al. [28] found that while beliefs about correlations changed in participants as a result of interaction with scatterplots, attitudes did not. Future work may wish to investigate whether this finding would persist with scatterplots utilizing the alternative designs described here. Finally, while changing perceptions, beliefs, and attitudes are promising early steps, changing people's behaviours would be the real test of the power of alternative visualization techniques; while this may be difficult to study, future work should investigate whether what we have found here may be used to induce behaviour change.

8 LIMITATIONS

Our commitment to finding an effect, should one exist, is also our biggest limitation. The exploratory nature of the work results in us being unable to comment specifically on how different forms of size and opacity manipulation in scatterplots may change beliefs in different ways, although addressing this using the framework we present here would be simple to accomplish. To date, there has been no qualitative work performed on alternative scatterplot designs such as those we utilize here; it may be that any perceptual or cognitive benefits are outweighed by distrust or unfamiliarity with novel designs. We chose to use a simple, blunt, 7-point Likert scale. While we argue that this is not particularly problematic given that we intended only to find an effect (and succeeded in that), future work may wish to use techniques that provide further scope for analysis, such as the graphical elicitation method developed by Karduni et al. [20, 21].

9 CONCLUSION

We investigated the effects of alternative scatterplot designs that vary the opacities and sizes of points as a function of their distance from the regression line on the propensity for belief change following visualization viewing. We presented these designs, and corresponding typical scatterplots, as if they were part of a news item and used a real-world variable pair that had been selected by our population of interest as being representative of a weakly-held, emotionally neutral correlation. We found that participants who viewed scatterplots employing alternative designs experienced greater levels of belief change than those who viewed typical plots. In addition, we found small interactive effects of a number of participant characteristics. Our results suggest that visualization techniques that have previously been employed to improve perception amongst participants are deserving of study with regards to their potential to change beliefs.

REFERENCES

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

[2] 2024. ChatGPT. https://chat.openai.com/

[3] Dolores Albarracín and Amy L. Mitchell. 2004. The Role of Defensive Confidence in Preference for Proattitudinal Information: How Believing That One Is Strong Can Sometimes Be a Defensive Weakness. *Personality & social psychology bulletin* 30, 12 (Dec. 2004), 1565–1584. https://doi.org/10.1177/0146167204271180

[4] 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

[5] Mattan S. Ben-Shachar, Daniel Lüdecke, and Dominique Makowski. 2020. effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software* 5, 56 (2020), 2815. https://doi.org/10.21105/joss.02815

[6] 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

[7] Matt Canham and Mary Hegarty. 2010. Effects of Knowledge and Display Design on Comprehension of Complex Graphics. *Learning and Instruction* 20, 2 (April 2010), 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

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

[9] 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

[10] Rune H. B. Christensen. 2023. *ordinal—Regression Models for Ordinal Data*. https://CRAN.R-project.org/package=ordinal R package version 2023.12-4.1.

[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] Jacob Cohen. 1968. Weighted Kappa: Nominal Scale Agreement Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin* 70, 4 (1968), 213–220. https://doi.org/10.1037/h0026256

[13] 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

[14] Joseph L. Fleiss, Jacob Cohen, and B. S. Everitt. 1969. Large Sample Standard Errors of Kappa and Weighted Kappa. *Psychological Bulletin* 72, 5 (Nov. 1969), 323–327. https://doi.org/10.1037/h0028106

[15] Steven L. Franconeri, Lace M. Padilla, Priti Shah, Jeffrey M. Zacks, and Jessica Hullman. 2021. The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest* 22, 3 (Dec. 2021), 110–161. https://doi.org/10.1177/15291006211051956

[16] Matthias Gamer, Jim Lemon, and Ian Fellows Puspendra Singh <puspendra.pusp22@gmail.com>. 2019. *irr: Various Coefficients of Interrater Reliability and Agreement*. https://CRAN.R-project.org/package=irr R package version 0.84.1.

[17] 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

[18] Lane Harrison, Drew Skau, Steven Franconeri, Aidong Lu, and Remco Chang. 2013. Influencing Visual Judgment through Affective Priming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 2949–2958. https://doi.org/10.1145/2470654.2481410

[19] L. Harrison, F. Yang, S. Franconeri, and R. Chang. 2014. Ranking Visualizations of Correlation Using Weber's Law. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 1943–1952. https://doi.org/10.1109/TVCG.2014.2346979

[20] Alireza Karduni, Douglas Markant, Ryan Wesslen, and Wenwen Dou. 2021. A Bayesian Cognition Approach for Belief Updating of Correlation Judgement through Uncertainty Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (Feb. 2021), 978–988. https://doi.org/10.1109/TVCG.2020.3029412

[21] Alireza Karduni, Ryan Wesslen, Douglas Markant, and Wenwen Dou. 2023. Images, Emotions, and Credibility: Effect of Emotional Facial Expressions on Perceptions of News Content Bias and Source Credibility in Social Media. *Proceedings of the International AAAI Conference on Web and Social Media* 17 (June 2023), 470–481. https://doi.org/10.1609/icwsm.v17i1.22161

[22] 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

[23] Xingyu Lan, Yanqiu Wu, and Nan Cao. 2023. Affective Visualization Design: Leveraging the Emotional Impact of Data. arXiv:2308.02831 [cs]

[24] 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

[25] 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

[26] Russell V. Lenth. 2024. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://CRAN.R-project.org/package=emmeans R package version 1.10.4.

[27] Torrin M. Liddell and John K. Kruschke. 2018. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology* 79 (Nov. 2018), 328–348. https://doi.org/10.1016/j.jesp.2018.08.009

[28] Doug Markant, Milad Rogha, Alireza Karduni, Ryan Wesslen, and Wenwen Dou. 2023. When Do Data Visualizations Persuade? The Impact of Prior Attitudes on Learning about Correlations from Scatterplot Visualizations. arXiv:2302.03776 [cs]

[29] 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

[30] Dirk Merkel. 2014. Docker: lightweight Linux containers for consistent development and deployment. *Linux J.* 2014, 239, Article 2 (mar 2014).

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

[32] 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

[33] Yasmina Okan, Mirta Galesic, and Rocio Garcia-Retamero. 2016. How People with Low and High Graph Literacy Process Health Graphs: Evidence from Eye-tracking: Graph Literacy and Health Graph Processing. *Journal of Behavioral Decision Making* 29, 2-3 (April 2016), 271–294. https://doi.org/10.1002/bdm.1891

[34] Anshul Vikram Pandey, Anjali Manivannan, Oded Nov, Margaret Satterthwaite, and Enrico Bertini. 2014. The Persuasive Power of Data Visualization. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 2211–2220. https://doi.org/10.1109/TVCG.2014.2346419

[35] 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

[36] 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

[37] Elizabeth A. Phelps, Sam Ling, and Marisa Carrasco. 2006. Emotion Facilitates Perception and Potentiates the Perceptual Benefits of Attention. *Psychological science* 17, 4 (April 2006), 292–299. https://doi.org/10.1111/j.1467-9280.2006.01701.x

[38] Eleftheria Polychronidou, Ilias Kalamaras, Konstantinos Votis, and Dimitrios Tzovaras. 2019. Health Vision: An interactive web based platform for healthcare data analysis and visualisation. In *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* (Siena, Italy). IEEE Press, 1–8. https://doi.org/10.1109/CIBCB.2019.8791462

[39] R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[40] 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

[41] 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

[42] Alper Sarikaya and Michael Gleicher. 2018. Scatterplots: Tasks, Data, and Designs. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (Jan. 2018), 402–412. https://doi.org/10.1109/TVCG.2017.2744184

[43] Priti Shah and Eric G. Freedman. 2011. Bar and Line Graph Comprehension: An Interaction of Top-Down and Bottom-Up Processes. *Topics in Cognitive Science* 3, 3 (2011), 560–578. https://doi.org/10.1111/j.1756-8765.2009.01066.x

[44] 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

[45] Gabriel Strain, Andrew J. Stewart, Paul Warren, and Caroline Jay. 2023. Adjusting Point Size to Facilitate More Accurate Correlation Perception in Scatterplots. In *2023 IEEE Vis X Vision*. IEEE, Melbourne, Australia, 1–5. https://doi.org/10.1109/VisXVision60716.2023.00006

[46] 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

[47] Gabriel Strain, Andrew J. Stewart, Paul A. Warren, and Caroline Jay. 2024. Effects of Point Size and Opacity Adjustments in Scatterplots. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3613904.3642127

[48] Sara Tandon, Alfie Abdul-Rahman, and Rita Borgo. 2024. Effects of Spatial Abilities and Domain on Estimation of Pearson's Correlation Coefficient.

[49] John C. Thoresen, Rebecca Francelet, Arzu Coltekin, Kai-Florian Richter, Sara I. Fabrikant, and Carmen Sandi. 2016. Not All Anxious Individuals Get Lost: Trait Anxiety and Mental Rotation Ability Interact to Explain Performance in Map-Based Route Learning in Men. *Neurobiology of Learning and Memory* 132 (July 2016), 1–8. https://doi.org/10.1016/j.nlm.2016.04.008

[50] Edward Tufte and P. Graves-Morris. 1983. The Visual Display of Quantitative Information. *Information Design Journal* 4, 3 (1983), 235–236. https://doi.org/10.1075/idj.4.3.12cos

[51] Cesko C. Voeten. 2023. *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. https://CRAN.R-project.org/package=buildmer R package version 2.11.

[52] Hadley Wickham. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

[53] Cindy Xiong, Chase Stokes, Yea-Seul Kim, and Steven Franconeri. 2022. Seeing What You Believe or Believing What You See? Belief Biases Correlation Estimation. https://doi.org/10.48550/arXiv.2208.04436 arXiv:2208.04436 [cs]

[54] 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