Title
(In)Accuracy of Human-Generated Correlations in A Scatterplot Drawing Task

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Cui, Lucy
Wang, Ching-Yi
Wang, Yiwei
et al.

Publication Date
2024

(In)Accuracy of Human-Generated Correlations in A Scatterplot Drawing Task

Lucy Cui[1] (cuil@rpi.edu)
Ching-Yi Wang[2] (ian0504@ucla.edu)
Yiwei Wang[2] (yiweiwang2001@outlook.com)
Peike Li[2] (peike239@ucla.edu)
Medha Kini[3] (medhakini@ucla.edu)
Zili Liu[2] (zili@psych.ucla.edu)
[1]Department of Cognitive Science, Rensselaer Polytechnic Institute, Troy, NY 12180
[2]Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095
[3]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA 90095

Abstract

Previous research on perception of correlation of scatterplots used scatterplots as stimuli and asked participants to estimate or compare correlations of those scatterplots. This literature has shown a tendency for people to underestimate correlation in some correlation ranges. We flipped the task: instead of estimating correlation from visual stimuli, participants drew a scatterplot based on a given correlation: 0, 0.25, 0.5, 0.75 and 1 using 20 dots. Participants drew greater correlations for r = 0.25 and r = 0.5 (0.59 and 0.71 respectively), which is analogous to underestimating correlation in previous viewing tasks. Drawn correlations for r = 0, 0.75 and 1 were more accurate. The number of statistics courses taken did not improve correlation drawing accuracy in a strong or meaningful way. We discuss possible interpretations of these results and future directions.

Keywords: scatterplots, drawing, correlation, perception, bias

Introduction

In a world of big data, graph literacy is becoming more important than ever. Scatterplots, specifically, are very important for teaching statistical concepts (e.g., correlation, regression), testing assumptions for data analyses, and communicating data and trends in industry, academia and beyond. When learning, teaching and communicating in the space of graphs and data visualizations, it is important to consider people's misconceptions about statistics, misperceptions from graphs and more importantly, the visual features involved and the visual system (see Cui & Liu for a review). For example, with scatterplots, many perceptual systems may be involved, such as mean perception and variation perception of x-coordinates and y-coordinates, or centroid location such as with dot clouds. For these reasons, many researchers are interested in people's ability to estimate/perceive correlation from scatterplots.

The perception of correlation in scatterplots is typically studied through presenting viewers with scatterplots and asking them to judge the correlation displayed. Typical methods of assessing people's perception of correlations from scatterplots include having participants make discriminative judgments or absolute judgments (see Doherty, Anderson, Angott & Klopfer (2007) for review). The researchers manipulate different features, beyond correlation, of the scatterplots in these methods.

People's ability to accurately perceive correlation from scatterplots is very finicky. Many features influence or bias people's estimation of correlation from scatterplots. Slope, for example, can bias people's estimates to be higher if the slope is steeper (Bobko & Karen, 1979, Lane, Anderson & Kellam, 1985). The number of data points included in a scatterplot can also influence estimates (Lane, Anderson & Kellam, 1985). Variance in different forms: error, x-values (Lane, Anderson & Kellam, 1985), and y-values (Lauer & Post, 1989) is also important. Other influential features include density of dot clouds (Boynton, 2000), outlier, heteroscedasticity, and restriction of range (Bobko & Karen, 1979, Lauer & Post, 1989). Size, color, saturation of dots could also bias correlation estimations (e.g., Hong, Witt & Szafir, 2021; Tseng, Quadri, Wang & Szafir, 2023). Finally, people tend to attend to few visual features when discriminating correlation (Yang, Harrison, Rensink & Franceroni, 2018).

Beyond these features, it has also been observed consistently that people tend to underestimate correlations (Meyer & Shinar, 1992, Strahan & Hansen, 1978) as opposed to overestimating (Meyer, Taieb, & Flascher, 1997). This is the case regardless of whether the viewer has statistics expertise (Meyer & Shinar, 1997). Underestimation of correlation is most pronounced in the range between r = 0.2 and r = 0.5 (Cleveland, Diaconis & McGill, 1982) and r < 0.2 show greater difficulty for perceiving a correlation (Bobko & Karen, 1979, Cleveland et al., 1982). Perceiving correlation accurately with so many features that can influence or bias correlation estimation is very difficult. It takes thousands of trials of perceptual training in order to overcome these perceptual biases and to correct for the tendency to underestimate for r < 0.5 (Cui, Massey, & Kellman, 2018).

In all these studies on correlation estimation, all of the data points in the scatterplot were available at once in the visual stimulus that is a presented scatterplot. Thus, the viewer has many options for estimating a correlation, either holistically - incorporating all information at once, or sequentially, by sampling or resampling the data points or shifting one's attention. If we ask people to draw a scatterplot from scratch, they would need to produce one data point at a time, gradually adding information and gradually having more and more information that they can incorporate into their drawing. These two processes look qualitatively different but

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

viewing and drawing scatterplots could produce different results in terms of correlation estimation.

The present study investigates whether people can draw scatterplots with a requested correlation accurately and whether the generated scatterplots have similar themes to the literature on viewed scatterplots. We anticipate a few possible outcomes: people generate scatterplots with correlations (a) that are consistent (i.e., draw greater correlations than requested) with the patterns of underestimation seen with viewed scatterplots, (b) that reflect some of the underestimation seen with viewed scatterplots but at a lesser degree, or (c) that have different patterns of over- or under- estimation than viewed scatterplots. If there is less tendency to underestimate with drawn scatterplots, it suggests that it is the viewing process of scatterplots and its potential perceptual biases from scatterplot characteristics that produces the underestimation. In other words, because the drawer has control over the scatterplot characteristics (e.g., outliers, density of clouds, etc.), they would draw a prototypical drawing of the correlation free of these biasing features and thus being closer to the actual correlation requested. If people draw greater correlations than requested and this deviation pattern is consistent with previous literature with visual viewing and estimating of correlation, then it suggests (a) that the drawing process is not that different from viewing in correlation estimation (i.e., they are viewing their drawn scatterplot and biasing their drawing throughout the drawing process), and/or (b) that people may have a poor idea of what low correlations look like, making them harder to imagine, recreate and perceive, such that they appear in drawings and perceptions of scatterplot stimuli. Both could be in play and would need to be further disentangled.

Given the novel task, we decided to select a limited number of correlation coefficients for participants to draw, equally spread out over the range 0 – 1. As correlation judgment precision differs between positively correlated and negatively correlated data (Harrison, Yang, Franceroni, & Chang, 2014), we only used the positive range of correlations. Participants drew five scatterplots, one for each correlation: 0, 0.25, 0.5, 0.75 and 1, by drawing 20 data points on blank graph. They reviewed the concepts of scatterplots and correlation before they started their drawings. Afterwards, we calculated the drawn correlations from the x, y coordinates of the datapoints that they drew.

Methods

Participants

Participants were 120 undergraduates (89 Female, 29 Male, 2 prefer not to say, M_{age}= 20.61, SD_{age}= 2.68) from the University of California, Los Angeles. They participated for partial course credit. Most students have taken some statistics class(es) (M = 1.75 classes, SD = 0.75). Only four participants have not taken a statistics class before.

Materials

We used an Amazon Fire HD 10 tablet and a compatible stylus for our in-person experiment. We used Nearpod to administer the experiment and collect the drawings and responses. We provided the participants a review of the necessary statistical concepts for drawing scatterplots: scatterplot and correlation. The "scatterplot screen" provided a definition of scatterplot along with two example scatterplots. We did not provide any correlation coefficients for these examples. On the "correlation screen" we provided a definition of correlation, the range of correlations, and a pictorial example of a negative and of a positive correlation. For the drawing task, the participants were provided with a blank graph, with only an x-axis and a y-axis with 0 through 1, in 0.2 increments, labelled. There were no gridlines provided. Participants were asked to draw 20 data points for their scatterplots. We chose 20 data points because the number is high enough to calculate a correlation but not so high to cause fatigue in the participant or them to lose count.

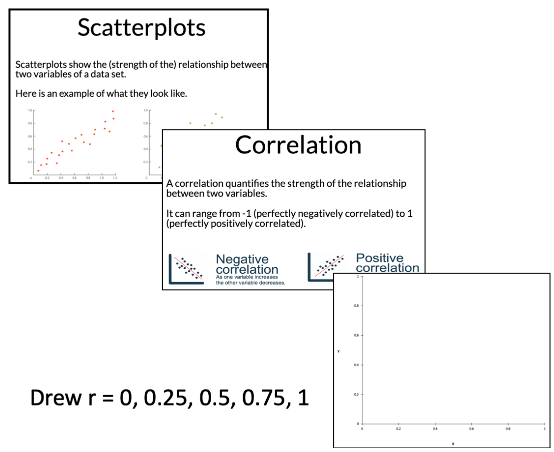

Figure 1: Concept review screens and drawing task screen.

Procedure

Participants completed the experiment on a digital tablet with a stylus. First, they had an opportunity to play around with the drawing interface, such as drawing dots and erasing them. Next, the participants reviewed the concept of scatterplots and correlation. We provided some examples of what scatterplots look like and also what negative and positive correlations look like. Participants were able to review the "scatterplots screen" and the "correlation screen" on their own pace. See Figure 1 for screenshots of the screens the participants viewed before and during their task. For the task, the participants were asked to draw a scatterplot using 20 dots in order to create a specified correlation coefficient. The participants started with r = 1 and then r = 0 in order to set a baseline. The next three correlation coefficients: 0.25, 0.5 and 0.75 that they drew were counterbalanced in order. After they were done with their scatterplot drawings, the research assistant asked the participant to explain how they drew the scatterplots for the different correlations. The research assistant than took notes in a participant log, summarizing their descriptions. After this interview portion, the participants filled out their demographics information, such as age, gender and number of statistics classes taken before.

Image Processing

Image processing was constrained to the bar-of-interest for the task (i.e., the rightmost purple bar in Figure 1). We used the FindContours function in the OpenCV Computer Vision library for object (dot) detection. The original image was inverted into binary coding (black dots on white background). Once the contour (i.e., dot) is detected, we used the center of the contour as the coordinate for the point. The coordinates were scaled based on the x- and y-axis labels. Then these x- and y-coordinates were used to calculate the correlation coefficients.

Qualitative Coding

Immediately after the experiment, the research assistant recorded whether the participant committed a slope-correlation conflation error. Mistaking correlation for slope (drawing steeper slopes for greater correlations) for all correlation coefficients was coded as 1, mistaking correlation for slope for r = 0.25, 0.5 and 0.75 but drawing r = 0 and r = 1 correctly was coded as 0.5. All other participants were coded as 0. We will refer to this as Slope Mistake.

Two independent coders qualitatively coded the interviewer's (research assistant) notes from the participants' explanations of their scatterplot drawings using the grounded theory approach with no a priori categories. In other words, the primary coder went through the responses, noted trends as they read through, and created code categories based on the existing trends. The secondary coder coded a randomly sampled 30 responses to get our interrater reliability scores, displayed in Table 4. The following is the resulting codebook used for our qualitative data.

Table 1. Resulting codebook. All binary-coded.

Code Name	Description
Wrong r	Participant had wrong interpretation of the correlation coefficient R (catch-all code, more specific codes below)
Sign	Participants incorrectly drew positive correlations with a different trend direction, such as plotting the dots as a negative correlation
Slope Positive	Participants mistook correlation as slope for r = 0.25, 0.5, 0.75 and 1 and drew steeper slopes as correlation increased
Memory	Participants applied their prior knowledge on correlation and scatterplot drawing
x/y-axis	Participant mistook correlation as the value of x-axis or y-axis
Slope Zero	Participant mistook correlation as slope for r = 0 or drew incorrectly for r = 0
Confusion	Participants expressed confusion on the instructions
Correct	Participant interpreted the correlation and slope correctly, such as correlation increases means more closer dots

Results

We had 111 participants with valid drawings (calculatable correlation) for all correlations requested (0, 0.25, 0.5, 0.75, 1). Only data from these participants were used for our quantitative analyses. All 120 participants were used for our qualitative analyses. Some of the qualitative codes were used to subdivide our quantitative data for analyses.

Quantitative Data

There were no order effects on drawn correlations, $p > .10$, so we excluded it as a variable for further analysis.
We conducted a repeated-measures (requested correlation: 0, 0.25, 0.5, 0.75, 1) ANOVA on drawn correlation. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $W = .43$, $\chi^2(9) = 90.24$, $p < .001$. Since *epsilon* < 0.75, we used the Greenhouse-Geisser correction. There was a main effect of requested correlation, $F(2.97,327) = 171.23$, $p < .001$, $\eta_p^2 = .61$. The relationship between requested correlation and drawn correlation was significant for a linear trend, $F(1, 110) = 372.21$, $p < .001$, quadratic trend, $F(1, 110) = 76.14$, $p < .001$, cubic trend, $F(1, 110) = 39.12$, $p < .001$ and order 4 trend, $F(1, 110) = 8.58$, $p = .004$. The trend is consistent with that of Lauer & Post (1989), see Figure 2.

Post hoc pairwise comparisons of requested correlation levels with Bonferroni correction revealed significant differences, $p < .001$, between each pair of requested correlations except for between r = 0.75 and r = 1, $M_{diff} = .078$, $SE = 0.038$, $p = .42$.

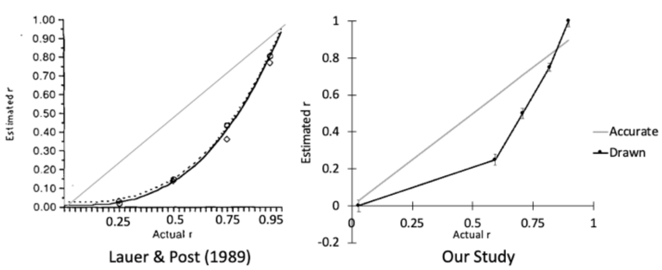

Figure 2: Our results (right; estimated r = requested r, actual r = drawn r) compared to Lauer & Post (1989; left)

The average correlation drawn by participants for each requested correlations can be found in Table 2. For a sense of what the drawn scatterplots looked like, we have created a heatmap showcasing where dots were typically drawn throughout a coordinate plane, analogous to the one that the participants received during the experiment. You can read Figure 3 as where most participants placed their datapoints to create a scatterplot for each of the requested correlations. A brighter spot means more participants drew dots in that area.

Table 2. Drawn correlations based on the requested correlations (top row)

r = _	0	0.25	0.5	0.75	r 1
Mean	0.029	0.593	0.706	0.818	0.896
SE	0.030	0.029	0.026	0.022	0.030

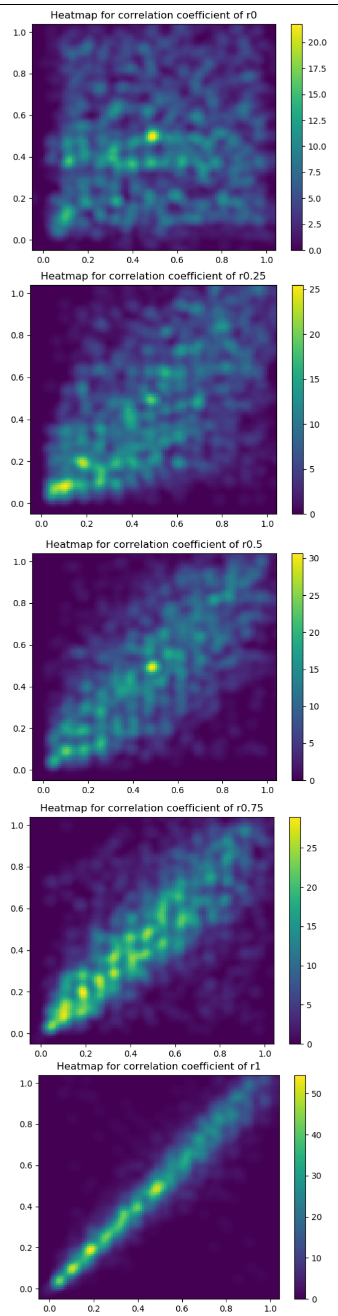

Figure 3: Heatmaps showing where most participants drew dots for each of the requested correlations (top to bottom: r = 0, 0.25, 0.5, 0.75, 1).

The above heatmaps show, from top to bottom, increasing requested correlations. As expected, the dispersion of drawn dots decreased as the requested correlation increases. We also see, on average, good compliance to the slope = 1 instruction for r = 0.25 – 1 as the slopes for the last four heatmaps have similar slopes. We can also see how much of the coordinate plane participants used for each requested correlation. For example, for r = 0.75 and r = 1 (last two heatmaps), participants seem to concentrate their dots in the first half of possible x- and y-values (bottom-left quadrant).

We can also see a faint horizontal banner-like pattern for r = 0 (first heatmap), suggesting a strategy of drawing a line with slope = 0 to create r = 0 for many participants. The midpoint of coordinate plane was also popular for r = 0, 0.5 and 1, perhaps to give a point of reference for their other drawn dots.

Statistics Expertise. We also investigate the relationship between drawn correlations and statistics expertise (number of statistics classes taken). Since there were only 4 participants who have not taken a statistics class, we excluded them from this analysis. We conducted a repeated-measures (requested correlation) ANOVA on drawn correlations with number of statistics classes taken, which ranged 1- 3. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $W = .42$, $\chi^2(9) = 77.66$, $p < .001$. Since *epsilon* < 0.75, we used the Greenhouse-Geisser correction. We found a significant main effect of requested correlation, $F(2.90, 266.54) = 117.93$, $p < .001$, $\eta_p^2 = .56$. There was a significant interaction between requested correlation and number of statistics classes taken on drawn correlation, $F(5.79, 266.54) = 2.90$, $p = .01$, $\eta_p^2 = .06$. There was no main effect of number of statistics classes taken (between-subjects), $F(2, 92) = 1.15$, $p = .32$, $\eta_p^2 = .02$.

The relationship between requested correlation and drawn correlation was significant for a linear trend, $F(1, 92) = 254.63$, $p < .001$, quadratic trend, $F(1, 92) = 49.03$, $p < .001$, cubic trend, $F(1, 92) = 18.56$, $p < .001$ but not for order 4 trend, $F(1, 92) = 2.17$, $p = .14$.The interaction between requested correlation and number of statistics classes taken was quadratic, $F(2, 92) = 3.84$, $p = .025$, with no other significant trends, $F(2, 92) < 2.90$, p's > .06. See Figure 4 for correlation drawing performance based on number of statistics classes taken.

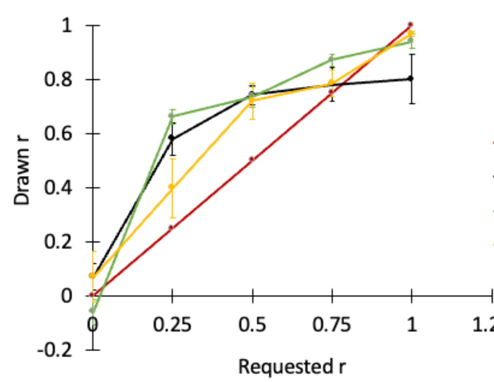

Figure 4: Drawn correlations separated based on number of statistics classes taken. Red line refers to perfectly drawing the correlations.

The apparent differences that can be seen in Figure 4 are better performance at lower correlations (r = 0 and r = 0.25) and better performance at higher correlations (r = 0.75 and r = 1) for participants who have completed more statistics classes (i.e., 2-3 classes). Participants who only completed 1 statistics class drew lower correlations than r = 1 and their drawn correlations varied.

There was a relationship between drawn correlations for r = 0.25 and the number of statistics classes taken, $F(2, 100) = 4.88$, $p = .01$, with only one significant comparison: 1 statistics class vs. 3 statistics classes (Bonferroni corrected), $M_{diff} = .25$, $SE = 0.081$, $p = .007$. There was no reliable relationship between drawn correlations for r = 1 and statistics classes, $F(2, 99) = 2.27$, $p = .11$.

Correct Drawings Only. For this analysis, we only included the participants who did not commit any listed errors in the Codebook in Table 1 and got Correct = 1 in the coding process. Only 70 participants qualified for this description.

We conducted a repeated-measures ANOVA on drawn correlation. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $W = .59$, $\chi^2(9) = 35.64$, $p < .001$. Since *epsilon* > 0.75, we used the Huynh-Feldt correction. There was a main effect of requested correlation, $F(3.43, 236.41) = 227.95$, $p < .001$, $\eta_p^2 = .77$. The relationship between requested correlation and drawn correlation was significant for a linear trend, $F(1, 69) = 751.21$, $p < .001$, quadratic trend, $F(1, 69) = 96.35$, $p < .001$, cubic trend, $F(1, 69) = 77.94$, $p < .001$ but not for the order 4 trend, $F(1, 69) = 2.66$, $p = .11$.

Post hoc pairwise comparisons of requested correlation levels with Bonferroni correction revealed significant differences, $p < .001$, between each pair of correlations and between r = 0.75 and r = 1, $M_{diff} = .128$, $SE = 0.037$, $p = .011$.

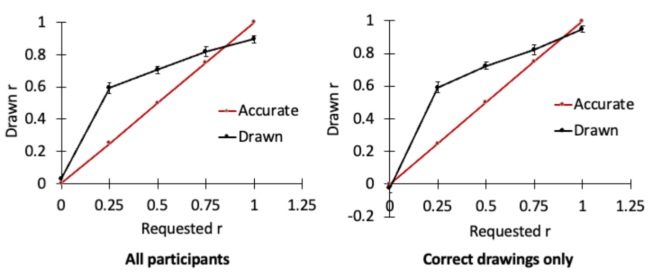

Figure 5: Same data as Figure 2, depicted differently (left). Data from participants who did not commit any correlation-understanding errors (right).

The main differences that can be seen in Figure 5 are better performance at r = 1 for participants who did not commit a correlation-understanding error, such as conflating slope for correlation, drawing the opposite direction/sign correlation, etc. This difference is significant, $t(115) = 2.22$, $p = .028$. Otherwise, the trend looks very similar, which is consistent with earlier findings that statistics experts are not less likely to underestimate (Meyer & Shinar, 1992).

Qualitative Data

We qualitatively coded drawing before and during the experiment (Slope Mistake) and coded the drawings as a group (looking at each participant's five scatterplot drawings as a set) after the experiment (see resulting Codebook in Table 1).

As a recap, Slope Mistake coded for whether participants conflated slope for correlation: 0 for not at all, 0.5 for only r = 0.25 – 0.75 (drawing r = 0 and 1 correctly), and 1 for all requested correlations. Table 3 shows how often participants made each mistake. Majority of participants (66.1%) did not make a slope mistake, based on research assistant observation during the experiment. About 20% of participants drew r = 0 and 1 correctly but became confused with correlations between 0.25 and 0.75 and drew based on slope instead.

Table 3: Slope Mistake frequency table

Code	Frequency	Percentage
No slope mistake (0)	80	66.1%
r = 0 and 1 drawn correctly, all other correlations mistaken as slope (0.5)	24	19.8%
For all correlation coefficients, correlations mistake as slope (1)	13	10.7%

After looking at the drawn scatterplots as a set for each participant, we have identified the most common mistakes as Slope Zero (36.4%, drawing r= 0 incorrectly), Slope Positive (26.4%, steeper slopes as correlations increased), and X/Y-axis (13.2%; mistaking x-axis and y-axis). These codes reveal the common misconceptions and confusions students have about scatterplots and correlations. Surprisingly, very few participants included in their explanations their use of their memory of correlations in order to recreate a correlation (Memory, 3.3%), which suggests most participants created a best guess of what a correlation would look like. See Table 4 for exact frequency and percentages for all codes from Table 1.

Table 4: Frequency table using Codebook in Table 1, displaying only frequency (n) and percentage (%) for code = 1. Reliability is interrater reliability, defined as percentage agreement between two independent coders.

Code	n	%	Reliability
Wrong r	35	36.1%	0.83
Sign	2	1.7%	0.97
Slope Positive	32	26.4%	0.87
Memory	4	3.3%	0.97
X/Y-axis	16	13.2%	0.97
Slope Zero	44	36.4%	0.90
Correct	74	62.2%	0.80
Confuse	2	1.7%	0.97

Since number of statistics classes seems to influence drawn correlation, we were curious whether participants who took more statistics classes were less likely to commit errors. We conducted a Chi-square Test between Correct drawing (0 or 1) and number of statistics classes taken (0 – 3) and found no relationship between the two variables, $\chi^2(3) = 1.91$, $p = .56$. There was also no relationship between statistics classes and Wrong r drawing, $\chi^2(3) = 1.60$, $p = .66$, Slope Positive mistake, $\chi^2(3) = 4.49$, p = .21, and Slope Zero mistake, $\chi^2(3) = 2.87$, $p = .41$.

Discussion

The purpose of the present study was to see how accurately people can draw scatterplots with a requested correlation and whether the generated scatterplots have similar themes to the literature on viewed scatterplots. We hypothesized that either drawn correlations would have similar themes as correlation estimation of viewed scatterplots or have different patterns. If we observe similar themes, this suggest that the generative process has similar components to the viewing and estimating process. If we observe greater drawn correlations for r = 0.25 and r = 0.5 much like the underestimation of r = 0.25 – 0.50 of viewed scatterplots, then this suggests that people's sense of r = 0.25 – 0.50 could be flawed and/or they viewed and adjusted their scatterplots throughout the drawing process. Finally, if we do not see pronounced discrepancy between drawn correlation and requested correlation (i.e., drawn > requested), then it suggests that the drawing process lends itself less opportunities for perceptual biases.

In each of the ways we have looked at the data (all the data, separating into number of statistics classes, and including only the participants who did not commit a correlation-understanding error), we found correlation estimation for r = 0.25 and 0.50 consistent with those found in previous literature (Lauer & Post, 1989). On average, all participants drew r = 0 accurately but r = 1 was drawn most accurately in participants who have taken 3 statistics classes or have not committed a correlation-understanding error. Participants who have taken 3 statistics classes had less underestimation for r = 0.25 and 0.5. It is important to note that the average drawn correlation for r = 1 is only 0.896, which could reflect difficulty drawing a perfectly straight line.

From the qualitative data, we see that most participants (66.1%) did not mistake correlation for slope. The common misunderstandings that participants had about correlation and scatterplots is not knowing how to draw r = 0, drawing steeper slopes as correlation increases, and mistaking which axis to draw on. These are very concerning errors, but what is more concerning is that once we removed these participants and only focused on the other participants, we still found a similar pattern of correlation estimations.

In addition, the number of statistics classes did not relate to correctness in drawing or any of the common mistakes committed, suggesting that there is something else distinguishing between those who drew correctly and those who drew with misconceptions. Perhaps some of the more statistics-educated may have been stumped by the drawing task because this is a low ecological validity task – one would never be asked to draw a scatterplot to match a requested correlation in real life. However, it is still concerning that one would commit errors that suggest misunderstanding different components (e.g., axes, slope, correlation) in a scatterplot.

Our results and its interpretation are limited to the quality of the notes that the research assistants took from interviewing the participants as well as the self-awareness and reporting accuracy from the participants. In addition, we do not have information such as the timestamps for each datapoint that the participants drew and how many times they erased their drawn dots. This information would help us understand their drawing process (e.g., whether it was dynamic) and whether there were readjustments. We can only analyze the submitted drawings and not each stage of the drawing process. To address these concerns, future studies could record and transcribe the interview between the participant and the research assistant and record the timestamps of the drawn dots in order to separate more intentional drawers from those less intentional and capture hesitation or confusion.

It is possible that the participants were perceiving as they were drawing and planning their next drawn datapoint based on what they have seen already. Unfortunately, we do not have information on their intentionality of drawing and their decision process point-by-point. Nonetheless, the participants still submitted a drawing that is consistent to what they would have viewed and estimated as a correlation.

To address these concerns, future studies could try to blindfold the participant and ask them to draw a scatterplot with a requested correlation. This would force participants to imagine a correlation and pre-plan all their dots before they draw it on the tablet. In addition, this would eliminate the opportunity to view previously drawn dots and be biased by them and to erase drawn dots to readjust their drawing. Future studies could also have participants draw more correlations between r = 0.25 and r = 0.5 to see whether underestimation (through drawings) in this range is consistent with the underestimation observed when participants estimate correlation from viewing scatterplots.

While our drawing task may seem low in ecological validity, the generative process of drawing scatterplots with certain correlations can reveal insights to the cognitive and perceptual processes involved in understanding and interpreting scatterplots. There is some evidence that correlation estimation is not purely perceptual. Correlation estimation can be influenced by cognitive biases (e.g., overestimating the correlation of two variables in a scatterplot that should have a strong correlation; Xiong, Stokes, Kim & Franceroni, 2022). Future work on the generative process could disentangle the cognitive and perceptual components. Finally, there is some value in exploring other generative tasks of scatterplots. As some have pointed out problems with correlation estimation tasks (Surber, 1986) and that trend judgment of scatterplots is a better task (Ciccione & Dehaene, 2021, Ciccione et al., 2023), trend generation can also be explored.

References

Bobko, P. & Karen, R. (1979). The perception of pearson product moment correlations from bivariate scatter plots," *Personnel Psychology,* 32(2), 313-325.

Boynton, D. (2000). The psychophysics of informal covariation assessment: Perceiving relatedness against a background of dispersion. *Journal of Experimental Psychology: Human Perception and Performance, 26*(3), 867-876.

Ciccione, L., & Dehaene, S. (2021). Can humans perform mental regression on a graph? Accuracy and bias in the perception of scatterplots. *Cognitive Psychology, 128*, 101406. https://doi.org/10.1016/j.cogpsych.2021.101406

Ciccione, L., Sablé-Meyer, M., Boissin, E., Josserand, M., Potier-Watkins, C., Caparos, S., & Dehaene, S. (2023). Trend judgment as a perceptual building block of graphicacy and mathematics, across age, education, and culture. *Scientific Reports, 13*(1), 10266. https://doi.org/10.1038/s41598-023-37172-3

Cleveland, W., Diaconis, P. & McGill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. *Science*, *216*(4550), 1138-1141.

Cui, L. & Liu, Z. (2021). Synergy between research on ensemble perception, data visualization, and statistics education: A tutorial review. *Attention, Perception & Psychophysics, 83*, 1290-1311.

Cui, L., Massey, C. M., & Kellman, P. J. (2018). Perceptual learning in correlation estimation: The role of learning category organization. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (p. 262 – 267). Madison, WI: Cognitive Science Society.

Doherty, M., Anderson, R., Angott, A. & Klopfer, D. (2007) The perception of scatterplots. *Perception & Psychophysics, 69*(7), 1261-1272.

Harrison, L., Yang, F., Franconeri, S., & Chang, R. (2014). Ranking visualizations of correlation using weber's law. *IEEE transactions on visualization and computer graphics, 20*(12), 1943-1952.

Hong, M. H., Witt, J. K., & Szafir, D. A. (2021). The weighted average illusion: Biases in perceived mean position in scatterplots. *IEEE Transactions on Visualization and Computer Graphics, 28*(1), 987-997.

Lane, D., Anderson, C., & Kellam, K. (1985). Judging the relatedness of variables: The psychophysics of covariation detection. *Journal of Experimental Psychology: Human Perception & Psychophysics, 11*, 640-649.

Lauer, T. & Post, G. (1989). Density in scatterplots and the estimation of correlation. *Behavior & Information Technology, 8*(3), 235-244.

Meyer, J. & Shinar, D. (1992). Estimation correlations from scatter-plots. *Human Factors, 34*, 335-349.

Meyer, J., Taieb, M., & Flascher, I. (1997). Correlation estimates as perceptual judgments. *Journal of Experimental Psychology: Applied, 3*, 3-20. Strahan, R. & Hansen, C. (1978). Underestimating correlation from scatter plots. *Applied Psychological Measurement, 2*, 543-550.

Strahan, R. & Hansen, C. (1978). Underestimating correlation from scatter plots. *Applied Psychological Measurement, 2*, 543-550.

Surber, C. (1986). Model Testing Is Not Simple: Comments on Lane, Anderson, and Kellam. Journal of Experimental Psychology*: Human Perception and Performance, 12*(1), 108–109.

Tseng, C., Quadri, G. J., Wang, Z., & Szafir, D. A. (2023, April). Measuring categorical perception in color-coded scatterplots. *Proceedings of the 2023 CHI conference on Human Factors in Computing Systems* (pp. 1-14).

Rensink, R. (2017). The nature of correlation perception in scatterplots. *Psychonomic Bulletin & Review, 24*(3), 776-797.

Xiong, C., Stokes, C., Kim, Y.-S., & Franconeri, S. (2022). Seeing What You Believe or Believing What You See? Belief Biases Correlation Estimation. *IEEE Transactions on Visualization and Computer Graphics*, Article arXiv:2208.04436. http://arxiv.org/abs/2208.04436

Yang, F., Harrison, L. T., Rensink, R. A., Franconeri, S. L., & Chang, R. (2018). Correlation judgment and visualization features: A comparative study. *IEEE transactions on visualization and computer graphics, 25*(3), 1474-1488.