Adjusting Point Size to Facilitate More Accurate Correlation Perception in Scatterplots

Gabriel Strain* Andrew J. Stewart† Paul Warren‡ Caroline Jay§

The University of Manchester

ABSTRACT

Viewers consistently underestimate correlation in positively correlated scatterplots. We use a novel data point size manipulation to correct for this bias. In a high-powered and fully reproducible study, we demonstrate that decreasing the size of a point on a scatterplot as a function of its distance from the regression line is able to correct for a systematic perceptual bias long present in the literature. We recommend the implementation of our technique when designing scatterplots that aim to communicate positive correlations.

Index Terms: Human-centered computing—Visualization—Empirical studies in visualization—; Human-centered computing—Human computer interaction (HCI)—Empirical studies in HCI—

1 INTRODUCTION

Scatterplots, utilized in scientific communication for a variety of tasks, are some of the most widely used and studied data visualizations. Viewers interpret them in similar ways [11], and they are simple to study while providing important insights into visualization design, human-computer interaction, and perception. Previously [27] we showed that a novel point contrast manipulation, in which the contrast of a certain scatterplot point was reduced as the size of that point's residual increased, could be used to partially correct for a systematic correlation underestimation bias present in the literature [4–6, 13, 14, 17, 26]. We suggested that this was due to a narrowing of the width of the perceived probability distribution of the data in a plot relative to the regression line, presumably driven by a reduction in the salience of, or weight given to, the lower contrast points in those outer areas. We tested linear, non-linear, and inverted non-linear decay functions relating point contrast to residual magnitude. As anticipated, we found that the non-linear function produced the most accurate estimates of correlation, and that the non-linear inverted produced the least accurate.

In this paper we utilize the same functions applied to point size to demonstrate that a non-linear decay function can be employed to correct for a systematic underestimation of correlation. We find no evidence for effects of graph literacy or training. The effect we observe here is stronger, both with regards to effect size and in terms of the observed reduction in error, than that observed in our previous study [27]. We suggest that this approach can be used to facilitate more accurate correlation perception in scatterplots while providing exciting future avenues for the continuation and refinement of these techniques.

1.1 Scatterplots and Correlation

Scatterplots have been widely studied, especially as mediums for the communication of correlation. Our previous work contains a review of the history of this literature [27]. Previous studies have found evidence for a pronounced underestimation in judgements of correlation in positively correlated scatterplots, especially between $0.2 < r < 0.6$. This is true for both direct estimation [6, 17] and estimation via bisection tasks [25]. Additionally, methods from psychophysics have been employed to explore how we discriminate between different correlations [24, 25]. Presently, we use the direct estimation paradigm owing to its simplicity and its suitability to online experimentation. The judgements we collect are therefore comparative by nature, as we analyse the differences in correlation estimation performance between data-identical scatterplots with different point size manipulations (see Figure 1). Such work allows us to inform design guidelines and affords us insights into perception. It is our duty as visualization designers to ensure that the messages we are trying to communicate are being interpreted accurately. To achieve this, we must understand human perception, apply that understanding to design, and test those designs in rigorous empirical studies.

1.2 Point Size

Contrast adjustments have been used extensively to solve issues of overplotting and clutter in scatterplots [3, 16]. Practicality dictates that scatterplots visualizing large datasets inherently require their points to be smaller to prevent obfuscation of the data. Point size changes (e.g bubble charts) have been used to represent trivariate data. There has, however, been little testing of the impact of point size on correlation perception. Studies have found invariance in the bias and variability of correlation perception with regards to changing point sizes, but these have been low-powered [23, 24]. Our motivation is to extend our previous work [27] to further improve correlation perception in scatterplots, while providing visualization designers with tools for more effective design. From the wider literature there is evidence that larger points can bias judgements of mean point position more strongly than point contrast [10] and can result in faster reaction times to peripherally presented stimuli [8]. There is some evidence that larger stimuli are associated with lower levels of spatial certainty [1] but higher levels of salience [9]. The current experiment should therefore allow us to distinguish between these candidate drivers of the effects observed in a way that was not possible when manipulating contrast, as in that case effects of salience or spatial certainty would operate in the same direction.

1.3 Hypotheses

We hypothesized that correlation estimates would be most accurate when viewers are presented with the non-linear size decay condition, and would be least accurate when presented with the non-linear inverted size decay condition.

2 METHODOLOGY

The experiment was conducted according to the principles of open and reproducible research. Data and analysis code are available at `https://github.com/gjpstrain/size_and_scatterplots`. This repository contains instructions for building a docker image to reproduce the computational environment used, allowing for full replications of stimulus generation, analyses, and

*Gabriel.Strain@manchester.ac.uk
†Andrew.J.Stewart@manchester.ac.uk
‡Paul.Warren@manchester.ac.uk
§Caroline.Jay@manchester.ac.uk

the paper itself. Ethical approval was granted by the University of Manchester's Computer Science Departmental Panel (Ref: 2022-14660-24397). Hypotheses and analysis plans were pre-registered with the OSF (`https://osf.io/k4gd8`).

2.1 Stimuli

45 uniformly distributed r values were generated between 0.2 and 0.99. For each of these values, a scatterplot consisting of a series of 128 data points was generated based on a normal distribution. The random seed used was the same as that used in our previous work [27], meaning the datasets were identical. This range of r values was chosen as there is evidence that little correlation is perceived below $r = 0.2$ [4, 5, 26]. A total of 45 r values were chosen so that we could build a more detailed picture of people's perceptions of correlations than previous work using fewer values, in addition to affording us a more objective method of comparing people's judgements beyond semantic labels such as 'weak' or 'strong' correlation as response measures. We use only positive r values, building on previous work that has done the same. Each scatterplot had a 1:1 aspect ratio, was generated as a 1000 x 1000 pixel .png image, and was scaled up or down according to the participant's monitor. See Sect. 2.2 for a more detailed discussion of precise point sizes and dot pitch in crowd-sourced experiments.

We used equation 1 to map residuals to point sizes in the two non-linear decay conditions. 0.25 was chosen as the value for b due to both its prior use with contrast decay [27] and its production of a curve approximating the inverse arodun the identity line of the underestimation curve reported in previous literature [25]. We acknowledge that more suitable values of b may exist, but extensive testing of this is outside the scope of the current study. We adjusted residual values using a scaling factor of 4 and a constant of 0.2 to achieve a minimum on-screen point size of 12 pixels, which is consistent with the point size on a 1920 x 1080 pixel monitor in our previous study [27]. In our fourth condition, which we refer to as *standard size*, point size was uniformly set to be consistent with that in our previous work. Scripts detailing scatterplot and mask generation are in the item preparation folder in the repository linked above.

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

2.2 Dot Pitch and Crowdsourced Experiments

Previously [27], we had no way of obtaining dot pitch or participant-to-monitor distance due to the online, crowdsourced nature of the experiments. We have since adopted a method for obtaining the height of a participant's monitor in inches [19]; participants are asked to hold up a standard size credit/debit/ID card up to the monitor, and then to resize a corresponding image until it matches the physical size of the card. These cards have a universal standard size (ISO/IEC 7810 ID-1), which when combined with the monitor resolution obtained from PsychoPy [21] and assuming a widescreen 16:9 aspect ratio, allows us to infer dot pitch and therefore the physical size of the points in our experiment. Mean dot pitch was 0.33mm ($SD = 0.06$), corresponding to a physical size on the screen of 4.32mm for the smallest points displayed. See Sect. 3 for analyses including dot pitch as a predictor.

2.3 Point Visibility Testing

It is key that our manipulation does not remove data from the scatterplot, thus, we include point visibility testing prior to the experimental items in the study. Participants were shown six scatterplots and were asked to enter in a text box how many points were being displayed. The points were the same size as the smallest points used in the experimental materials. 5% of participants were correct on 5 out of 6 point visibility questions, while 95% were correct on 6 out of 6. It should be noted that those participants scoring 5/6 did not answer incorrectly, rather they did not answer at all for a particular question, which is suggestive of a mis-click or an initial misunderstanding of the task. Regardless, we consider these results to be indicative of a sufficient level of point visibility.

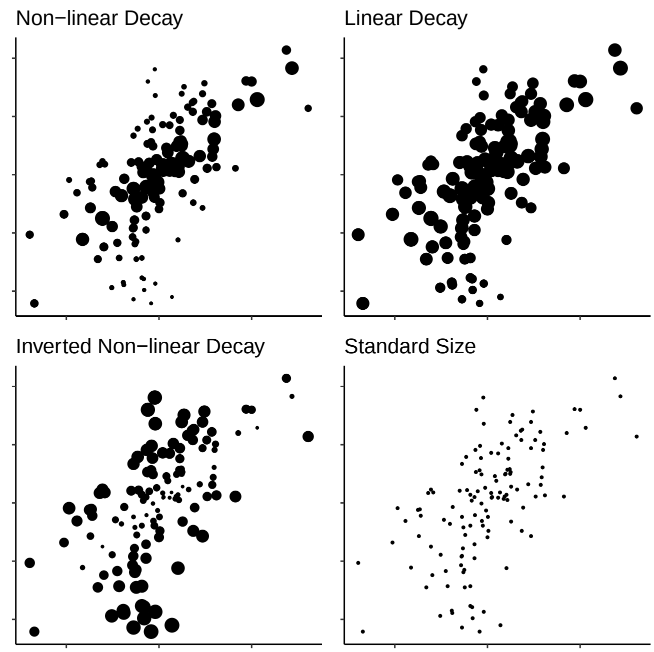

Figure 1: Four levels of the point size decay condition, demonstrated with an r value of 0.6

2.4 Design

The experiment used a fully repeated measures, within-participants design, with each participant seeing and responding to each of the 180 scatterplots in a randomized order. There were four scatterplots for each of the 45 r values corresponding to the four levels of the size decay condition, examples of which can be see in Figure 1. Everything needed to run the experiment, including code, materials, instructions, and scripts, is hosted at `https://gitlab.pavlovia.org/Strain/exp_size_only`.

2.5 Procedure

Each participant was shown the participant information sheet and provided consent through key presses in response to consent statements. They were asked to provide their age in a free text box, and their gender identity. Participants then completed the 5-item Subjective Graph Literacy test [7], followed by the point visibility test described above and the screen scaling task. Participants were given instructions, and then shown examples of $r = 0.2$, 0.5, 0.8, and 0.95. Sect. 4.3 includes a discussion of the potential training effects of viewing these examples. Two practice trials were given before the experiment began. Participants worked through a series of 180 trials and were asked to use a slider to estimate the correlation shown in the scatterplot to 2 decimal places. Visual masks preceded each plot. Interspersed were six attention check trials which explicitly asked participants to set the slider to 1 or 0 and ignore the scatterplot.

2.6 Participants

150 participants were recruited using the Prolific.co platform. Normal to corrected-to-normal vision and English fluency were required for participation. In accordance with published guidelines [20], participants were required to have completed at least 100 studies on Prolific, and were required to have a Prolific score of at least

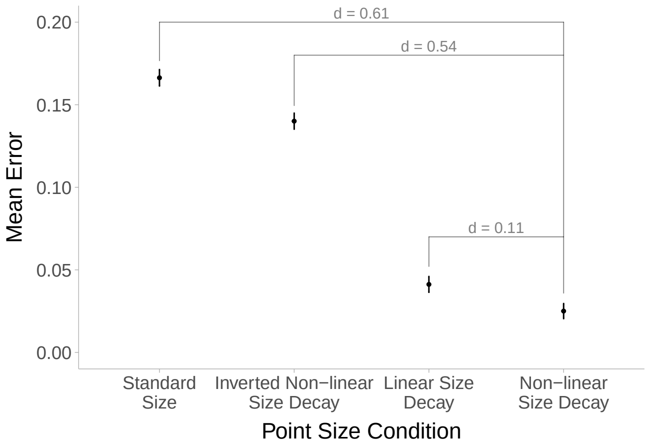

Figure 2: Mean error in correlation estimates across the four conditions, with 95% confidence intervals. Effect sizes between standard size and other conditions in Cohen's d are also displayed.

Table 1: Contrasts between the four levels of the size decay condition.

Contrast	Z.ratio	p.value
Standard Size : Inverted Non-linear Decay	9.35	<0.001
Standard Size : Linear Decay	44.40	<0.001
Standard Size : Non-Linear Decay	50.14	<0.001
Inverted Non-linear Decay : Linear Decay	35.06	<0.001
Inverted Non-linear Decay : Non-Linear Decay	40.79	<0.001
Linear Decay : Non-Linear Decay	5.73	<0.001

100, indicating acceptance on at least 100/101 previously completed studies. Participants who took part in any of our previous studies on correlation perception in scatterplots were prevented from participating, and participants were only permitted to complete the experiment on a desktop or laptop computer.

Data were collected from 164 participants. 14 failed more than 2 out of 6 attention checks, and, as per pre-registration stipulations, were rejected from the study. Data from 150 participants was included in the analysis (48% male, 50% female, and 2% non-binary). Mean age of participants was 29.6 (*SD* = 8.5). Mean graph literacy score was 21.77 (*SD* = 4.29) out of 30. The average time taken to complete the experiment was 39 minutes (*SD* = 14 minutes).

3 Results

All analyses were conducted using R (version 4.3.1 [22]). Linear mixed effects models were built using the **buildmer** (version 2.9 [29]) and **lme4** (version 1.1-34 [2]) packages, with size decay condition being set as the predictor for participants' errors in correlation estimates. The experimental model has random intercepts for items and participants.

Mean errors in correlation estimates and 95% confidence intervals for the four size decay conditions can be seen in Figure 2. A likelihood ratio test revealed that the model including size decay condition as a predictor explained significantly more variance than a null model ($\chi^2(3) = 3{,}508.84$, $p < .001$). The effect here is driven by significant differences in correlation estimation error between all levels of size decay condition. Figure 3 shows how participants' mean errors in correlation estimates change with the objective r value, plotted separately for each size decay condition. Overall condition means are also shown in the plot titles. Note the close-to-zero average errors present in the non-linear size decay condition.

Testing for contrasts between was performed with the **emmeans** package (version 1.8.8 [15]), and are shown in Table 1. The **EMAtools** (version 0.1.4 [12]) package was used to calculate effect sizes in Cohen's d, the results of which can be seen in Figure 2. The largest effect size we found was 0.61 when comparing the non-linear size decay and standard size decay conditions, compared to 0.19 for the equivalent comparison in our previous work [27].

We find no significant difference between the experimental model and one including graph literacy as a fixed effect ($\chi^2(1) = 0.16$, $p = .690$), suggesting the effect we found was not driven by differences in graph literacy.

We employed a method for obtaining a measurement of dot pitch from each participant. While participants performed well on a point visibility task, there may be some other facet of using a larger or smaller monitor with a higher or lower resolution that could have affected the estimates participants gave. To check this, we built a model including dot pitch as a fixed effect. Comparing this to the experimental model revealed a significant effect of dot pitch ($\chi^2(1) = 4.65$, $p = .031$). There was no interaction between size decay condition and dot pitch, with a decrease in dot pitch of 0.1 resulting in a decrease in estimated correlation of .03. Given that dot pitch range was only 0.13 to 0.63, we do not consider this effect substantial enough to warrant further discussion.

4 Discussion

As can be seen in Figure 3, participants' errors in correlation estimates were significantly lower in the the non-linear size decay condition (see Figure 1) compared to all other conditions, providing support for our first hypothesis. We found no support for our second hypothesis, that participants' estimates would be least accurate in the inverted non-linear size decay condition. Errors in this condition were significantly higher than for the other two size decay conditions, but were significantly lower than the error with the standard size condition.

4.1 Increased Correlation Estimation Accuracy

The mean signed error in correlation estimation for the non-linear size decay condition used in the present study was .025, while the equivalent condition in the second experiment of our previous study, which used the same equations applied to contrast, resulted in a mean error of .086 [27]. This provides evidence that point size is a stronger encoding channel for the manipulation of perceived correlation in scatterplots than point contrast. If these effects are driven by the lower salience of points in the outer regions of the plots, that we have found a larger effect of point size manipulations is congruent with research showing clear influences of stimulus size on object salience and perceptual weighting [8–10]. Our results therefore provide support for point salience/weight being the key driver the effects observed. Lower point salience in the outer regions of the plots then reduces the perceived width of the distribution of data points about the regression line [18, 25], leading in this case to more more accurate estimates. Simultaneously we acknowledge that other candidate mechanisms exist; similar results would be expected if a feature-based attentional bias was at play [10, 28]. It may be that both attention and perceived probability distribution width are responsible for the effects we see, however with our current methodology we are unable to comment as to the extent of each.

The lack of support for our second hypothesis, while surprising, suggests that point salience or perceived distribution width do not form the whole story. There is evidence that larger stimuli exhibit higher spatial uncertainty [1], and it is possible that this uncertainty causes the perceptual system to downweight the contribution of these points during correlation estimation. This is consistent with previous [30, 31] work suggesting the brain may make robust statistical use of visuo-spatial information. These mechanisms act to downweight the influence of more unreliable information (in this case the higher spatial uncertainty of larger exterior points) on subsequent perceptual estimates. Despite the size channel being more

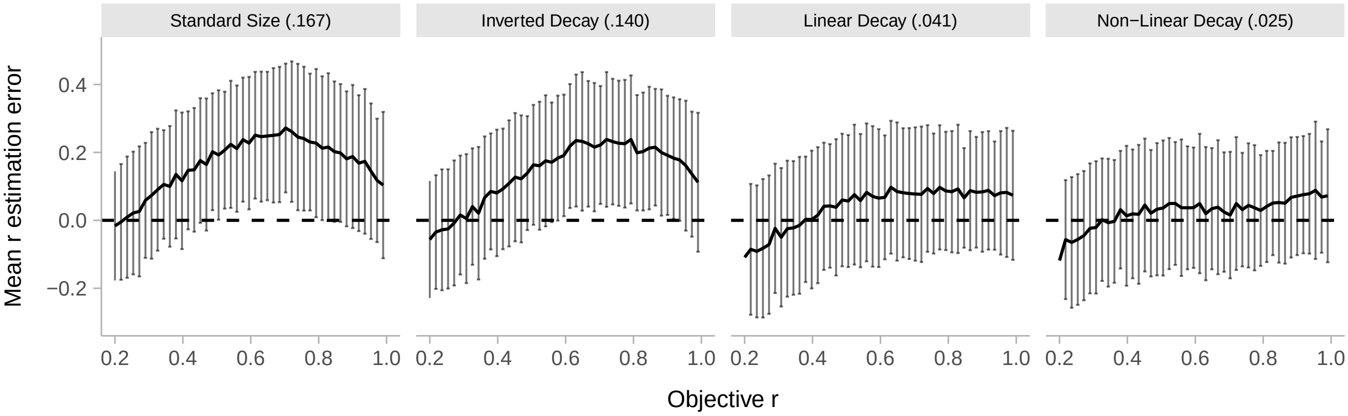

Figure 3: Participants' mean errors in *r* estimates plotted against the objective *r* value for each size decay condition. Error bars represent standard deviations. The dotted horizontal line represents accurate estimation. Means of errors by condition are also displayed next to plot titles.

powerful with regards to correcting for correlation underestimation, it appears unable to produce the opposing effect. In our previous work we suggested that inverted contrast manipulations could be used to correct for the *overestimation* of the correlation of negatively correlated scatterplots; in light of our findings we would not suggest the use of the size channel for this given the results here.

4.2 Precision in Correlation Estimation is Constant

Unlike our previous work, in which the standard deviations of errors generally became smaller as the objective *r* value increased, distributions of standard deviations of correlation estimates here remained mostly constant. This is unexpected, as previous work, including our own, finds precision in *r* estimation to increase as the objective *r* value increases. Given that we found this in our work manipulating point contrast [27], and its robustness in the literature, this result is surprising. We suggest that this is due to the nature of the stimuli. At high values of *r* there is a large amount of overlap between points in the non-linear, non-linear inverted, and linear size decay conditions. It may be that this is producing greater uncertainty and causing an absence of the increased precision we would expect. While the visual character of the scatterplots in the aforementioned conditions can account for the absence of higher precision at higher *r* values, this cannot be said for the standard size condition. Aside from the inverted non-linear decay condition in our previous work [27], the finding that precision increased with *r* was robust. Its absence here is curious given that the standard size decay condition here is identical to the full contrast conditions in our previous work. Relying on relative judgements means the interplay between scatterplots with different visual features must be accounted for within a particular experiment. The stimuli as *r* approaches 1 in the current study exhibit greater levels of visual variance than the stimuli in our previous work [27], which may explain the lack of increased precision here. Further testing is required for a more concrete explanation.

Ultimately we aim to provide tools for the design of visualizations more suited for the tasks they are intended to support. When that task is the perception of positive correlation, we would recommend the use of the non-linear size decay condition described here. We acknowledge that for other scatterplot tasks, such as cluster separation or numerosity perception, or other chart types, the use of the size manipulation may in fact be a hindrance, and in scenarios where the intended usage of a scatterplot includes tasks such as these, we would not recommend it. Instead we may recommend the use of a contrast decay condition [27], which also corrects for the underestimation bias (albeit to a lesser degree), while not impeding other tasks.

4.3 Training

Before the experiment, participants viewed plots for a minimum of eight seconds with examples of $r = 0.2$, 0.5, 0.8, and 0.95. This was to account for any unfamiliarity with scatterplots present in the samples that we recruited; this risk is inherent in recruiting from lay populations, but we would argue is acceptable given it leads to more generalizable and broadly applicable findings. Comparing a model including session half as a predictor with the original model revealed no significant effect ($\chi^2(1) = 1.06$, $p = 0.30$), suggesting that having more recently viewed the example plots did not have an effect on participants' performance.

4.4 Limitations

Despite confirming a method of obtaining dot pitch, we still have no method of obtaining head-to-monitor distances. This, along with the comparative judgements we collect, prevents us from drawing concrete psychophysical conclusions. Instead, our paradigm allows for findings that are rigorous to different viewing contexts and are of particular importance for the HCI and visualization design audiences. It may be that a high level perceptual phenomenon is responsible for the effects we have seen here; investigating this is beyond the scope of the current study and does not negate our findings. We acknowledge that there is the potential for misinterpretation of the scatterplots we present in the current study, especially given their similarity in form, but not purpose to bubble charts.

4.5 Future Work

At present, we have confirmed the potential for both point contrast and size manipulations to influence participants' perceptions of correlation in scatterplots, each to varying degrees. It is also clear that these manipulations are not necessary, and may be making perception worse, at certain values of *r*. We will therefore investigate the effect of manipulating *both* point size and contrast on correlation estimation, and will introduce a parameter to control the strength of this family of manipulations according to the objective *r* value itself. More qualitative work is needed to both address any misinterpretation of our adjusted scatterplots and to provide further insights into strategies participants may use. We have also demonstrated that our experimental framework [27] is transferable to other visual features of plot design, and in the future could be applied to test other chart types or statistical summaries.

Acknowledgments

This work was supported by funding from the University of Manchester Department of Computer Science and Division of Psychology,

Communication and Human Neuroscience.

REFERENCES

[1] D. Alais and D. Burr. The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3):257–262, Feb. 2004. doi: 10.1016/j.cub.2004.01.029

[2] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01

[3] E. Bertini and G. Santucci. Quality Metrics for 2D Scatterplot Graphics: Automatically Reducing Visual Clutter. In A. Butz, A. Krüger, and P. Olivier, eds., *Smart Graphics*, Lecture Notes in Computer Science, pp. 77–89. Springer, Berlin, Heidelberg, 2004. doi: 10.1007/978-3-540-24678-7_8

[4] P. Bobko and R. Karren. The Perception of Pearson Product Moment Correlations from Bivariate Scatterplots. *Personnel Psychology*, 32(2):313–325, 1979. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1744-6570.1979.tb02137.x. doi: 10.1111/j.1744-6570.1979.tb02137.x

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

[6] C. E. Collyer, K. A. Stanley, and C. Bowater. Psychology of the Scientist: LXIII. Perceiving Scattergrams: Is Visual Line Fitting Related to Estimation of the Correlation Coefficient? *Perceptual and Motor Skills*, 71(2):371–378E, Oct. 1990. Publisher: SAGE Publications Inc. doi: 10.2466/pms.1990.71.2.371

[7] R. Garcia-Retamero, E. T. Cokely, S. Ghazal, and A. Joeris. Measuring Graph Literacy without a Test: A Brief Subjective Assessment. *Medical Decision Making*, 36(7):854–867, 2016. Publisher: SAGE Publications Inc STM. doi: 10.1177/0272989X16655334

[8] G. R. Grice, L. Canham, and J. M. Boroughs. Forest before trees? It depends where you look. *Perception & Psychophysics*, 33(2):121–128, Mar. 1983. doi: 10.3758/BF03202829

[9] C. Healey and J. Enns. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, July 2012. doi: 10.1109/TVCG.2011.127

[10] M.-H. Hong, J. K. Witt, and D. A. Szafir. The Weighted Average Illusion: Biases in Perceived Mean Position in Scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2021. doi: 10.1109/TVCG.2021.3114783

[11] M. Kay and J. Heer. Beyond Weber's Law: A Second Look at Ranking Visualizations of Correlation. *IEEE transactions on visualization and computer graphics*, 22, Sept. 2015. doi: 10.1109/TVCG.2015.2467671

[12] E. Kleiman. *EMAtools: Data Management Tools for Real-Time Monitoring/Ecological Momentary Assessment Data*, 2021. R package version 0.1.4.

[13] D. Lane, C. Anderson, and K. Kellam. Judging the Relatedness of Variables. The Psychophysics of Covariation Detection. *Journal of Experimental Psychology: Human Perception and Performance*, 11:640–649, Oct. 1985. doi: 10.1037/0096-1523.11.5.640

[14] T. W. Lauer and G. V. Post. Density in scatterplots and the estimation of correlation. *Behaviour & Information Technology*, 8(3):235–244, June 1989. Publisher: Taylor & Francis. doi: 10.1080/01449298908914554

[15] R. V. Lenth. *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2023. R package version 1.8.8.

[16] J. Matejka, F. Anderson, and G. Fitzmaurice. Dynamic Opacity Optimization for Scatter Plots. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 2707–2710. ACM, Seoul Republic of Korea, Apr. 2015. doi: 10.1145/2702123.2702585

[17] J. Meyer and D. Shinar. Estimating Correlations from Scatterplots. *Human Factors*, 34(3):335–349, June 1992. Publisher: SAGE Publications Inc. doi: 10.1177/001872089203400307

[18] J. Meyer, M. Taieb, and I. Flascher. Correlation estimates as perceptual judgments. *Journal of Experimental Psychology: Applied*, 3(1):3–20, 1997. Place: US Publisher: American Psychological Association. doi: 10.1037/1076-898X.3.1.3

[19] W. L. Morys-Carter. ScreenScale, May 2021. https://doi.org/10.17605/OSF.IO/8FHQK.

[20] E. Peer, D. Rothschild, A. Gordon, Z. Evernden, and E. Damer. Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, Sept. 2021. doi: 10.3758/s13428-021-01694-3

[21] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203, Feb. 2019. doi: 10.3758/s13428-018-01193-y

[22] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023.

[23] R. Rensink. Invariance of Correlation Perception. vol. 12, pp. 433–433, May 2012. doi: 10.1167/12.9.433

[24] R. A. Rensink. On the Prospects for a Science of Visualization. In W. Huang, ed., *Handbook of Human Centric Visualization*, pp. 147–175. Springer New York, New York, NY, 2014. doi: 10.1007/978-1-4614-7485-2_6

[25] R. A. Rensink. The nature of correlation perception in scatterplots. *Psychonomic Bulletin & Review*, 24(3):776–797, 2017. doi: 10.3758/s13423-016-1174-7

[26] R. F. Strahan and C. J. Hansen. Underestimating Correlation from Scatterplots. *Applied Psychological Measurement*, 2(4):543–550, Oct. 1978. Publisher: SAGE Publications Inc. doi: 10.1177/014662167800200409

[27] G. Strain, A. J. Stewart, P. Warren, and C. Jay. The Effects of Contrast on Correlation Perception in Scatterplots. *International Journal of Human-Computer Studies*, 176:103040, Aug. 2023. doi: 10.1016/j.ijhcs.2023.103040

[28] P. Sun, C. Chubb, C. E. Wright, and G. Sperling. The centroid paradigm: Quantifying feature-based attention in terms of attention filters. *Attention, Perception & Psychophysics*, 78(2):474–515, Feb. 2016. doi: 10.3758/s13414-015-0978-2

[29] C. C. Voeten. *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*, 2023. R package version 2.9.

[30] P. A. Warren, L. T. Maloney, and M. S. Landy. Interpolating sampled contours in 3-D: analyses of variability and bias. *Vision Research*, 42(21):2431–2446, Sept. 2002. doi: 10.1016/S0042-6989(02)00266-3

[31] P. A. Warren, L. T. Maloney, and M. S. Landy. Interpolating sampled contours in 3D: perturbation analyses. *Vision Research*, 44(8):815–832, Apr. 2004. doi: 10.1016/j.visres.2003.11.007