Point Size and Correlation Perception in Scatterplots

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

The University of Manchester

ABSTRACT

Place abstract here.

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

Scatterplots, utilised in scientific communication for a variety of tasks, are some of the most widely used and widely studied data visualizations. Viewers mostly interpret them in similar ways [9], and they are simple enough to be easily studied while providing important insights into visualization design, human-computer interaction, and human perception. In a previous study [15], we showed that a novel point contrast manipulation, in which the contrast of a certain scatterplot point was reduced as the size of the point's residual increased, could be used to partially correct for a systematic underestimation bias present in the literature [**?**, **?**, 3, 4, 10, 11, 14]. We suggested that this was due to the narrowing of the perceived probability distribution of a particular plot brought on by the lower contrast (and therefore higher uncertainty) in those outer areas. We tested linear, non-linear, and non-linear inverted functions relating point contrast to residual size, finding that the non-linear function produced the most accurate estimates of correlation, and that the non-linear inverted produced the least accurate.

Of course, point contrast is not the only available visual feature that might be used to influence viewer's perceptions of the width of a probability distribution. Instead, we might use the size of a point. There is evidence in the literature that larger points are more salient [7], can bias judgements of point position [8], and can result in faster reaction times [6] to peripherally presented stimuli. In addition, smaller stimuli are associated with greater levels of spatial uncertainty [1], and if this is driving the reduction in error we saw in our previous work [15], we would expect a similar effect when point size is used instead of point contrast. We therefore hypothesize that correlation estimation will be most accurate when viewers are presented with the non-linear size decay condition, and will be least accurate when presented with the non-linear inverted size decay condition.

We present a single experiment study in which we demonstrate that the use of a non-linear size decay function relating to the residuals of points on scatterplots can be employed to correct for a systematic underestimation of correlation by viewers in scatterplots. We find no evidence for an effect of graph literacy (put something about training here). The effect we observe here is much stronger, both with regards to effect size and in terms of the observed reduction in error, than the effect we observed in our previous study [15]. We suggest that this function can be used to facilitate more accurate correlation perception in scatterplots, and provide exciting future avenues for the continuation and refinement of these techniques.

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

Ethical approval was granted by the University of Manchester's Computer Science Departmental Panel (Ref: 2022-14660-24397)

0.1 Correlation Perception

Scatterplots have been widely studied, especially as mediums for the communication of correlation (see [15] for a detailed review of this history). Much previous work has found evidence for a pronounced underestimation in judgements of correlation in positively correlated plots, especially between $0.2 < r < 0.6$. The nature of this investigation has varied, ranging from direct estimation techniques, to discriminative judgement, bisection, and staircase tasks. As in our previous work, we choose to use the direct estimation paradigm owing to its simplicity, and its suitability to online experimentation. The use of this paradigm renders the judgements we collect comparative by nature, although such work does allow us to inform design guidelines as well as perception.

0.2 Point Size

Hong et al paper will be useful here

0.3 Dot Pitch and Crowdsourced Experiments

1 Methodology

1.1 Open Research Statement

The experiment was conducted according to the principles of open and reproducible research. All data and analysis code are available at `https://github.com/gjpstrain/size_contrast_and_scatterplots`. This repository contains instructions for building a docker image to fully reproduce the computational environment used, allowing for full replications of stimulus generation, analyses, and the paper itself. The experiment was pre-registered with the OSF (`https://osf.io/k4gd8`).

1.2 Participants

150 participants were recruited using the Prolific.co platform. Normal to corrected-to-normal vision and English fluency were required for participation. As in [15], and in accordance with previously published guidelines [13], participants were required to have completed at least 100 studies on Prolific, and were required to have a Prolific score of at least 100, indicating acceptance on at least 100/101 previously completed studies. Participants who took part in any of our previous studies 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 check questions, and, as per pre-registration stipulations, were rejected from the study. Data from 150 participants was included in the analysis (48.00% male, 50.00% female, and 2.00% non-binary). Mean age of participants was 29.56 (SD = 8.54). 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).

1.3 Stimuli

The data used to generated the scatterplots in the current study was identical to that used previously [15]. Scatterplots were generated based on 45 uniformly distributed r values between 0.2 and 0.99.

Scatterplot points were generated based on bivariate normal distributions with standard deviations of 1 in each direction. Each scatterplot had a 1:1 aspect ratio, was generated as a 1200 x 1200 pixel .png image, and was scaled up or down according to the participant's monitor. See Sect. 0.3 for a more detailed discussion of precise point sizes and dot pitch in crowd-sourced experiments.

As in our previous study [15], we used equation 1 to map residuals to point sizes. We used a scaling factor of 4 and a constant of 0.2 to achieve a minimum point size of 12/13 pixels, which is consistent with the point size on a 1920 x 1080 monitor for both experiments in [15]. Again, see Sect. 0.3 for a discussion of dot pitch. Scripts detailing scatterplot and mask generation can be found in the item preparation folder in the repository linked below.

$$point - size = 1 - b^R \tag{1}$$

1.4 Dot Pitch and Crowdsourced Experiments

In our previous study [15], we had no way of obtaining dot pitch or participant to monitor distance due to the online, crowdsourced nature of the experiments. Since then we have adopted a method for obtaining the height of a participant's monitor in inches [12]. Combining this with the monitor resolution fetched from Psychopy 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. While dot pitch is not necessary information for the present study, as we are only interested in the relative differences in correlation estimates between point size conditions, the fact that we can collect it at all is indicative of the gap being narrowed with regards to psychophysical testing between in-person and online experiments. Given that the latter are cheaper, easier, and immeasurably quicker, anything that can be done be done to narrow this gap in capability is a boon. See Sect. 2 for analyses including dot pitch as a predictor.

1.5 Visual Threshold Testing

It is key that our manipulation does not functionally removing data from the scatterplot, thus, in order to test that all our points were visible across a range of viewing contexts and on a range of apparatus, we included visual threshold testing prior to the experimental items in the study. Participants were shown six scatterplots with a number of points, and were asked to enter in a textbox 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 visual threshold 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 did not answer at all for this particular questions, which is more suggestive of either a mis-click or an initial misunderstanding of the task they were asked to complete. Regardless, we consider these results to be indicative of a sufficient level of point visibility for the current experiment.

1.6 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 randomised order. There were four scatterplots for each of the 45 r values corresponding to the four levels of the size 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`.

1.7 Procedure

Each participants was shown the participants information sheet (PIS) and provided consent through key presses in response to consent statements. They were asked to provide their age in a free text box,

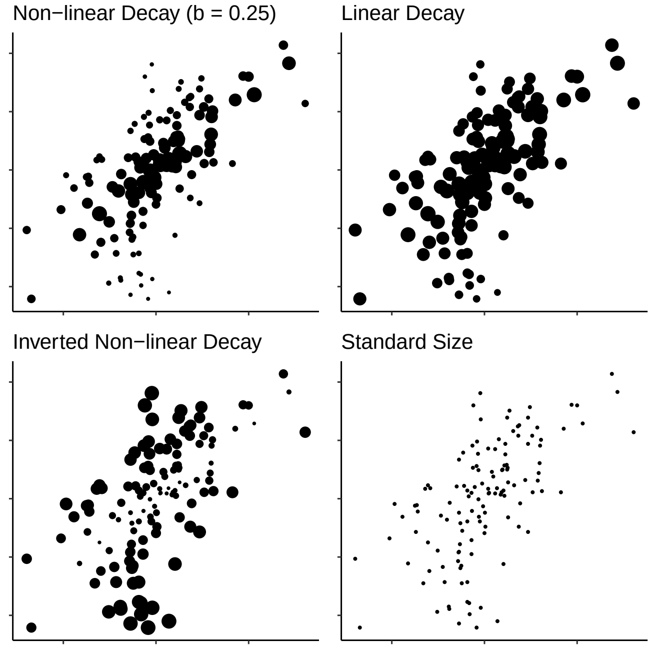

Figure 1: Four levels of the point size condition, demonstrated with an r value of 0.6

and their gender identity. Participants then completed the 5-item Subjective Graph Literacy (SGL) test [5], followed by the visual threshold testing described above. Participants then completed the screen scaling task described in Sect. 0.3. Participants were given instructions, and then shown examples of r = 0.2, 0.5, 0.8, and 0.95, as in our previous work [15]. **??** includes a discussion of the potential effects of this training. Two practice trials were given before the experiment began. Participants worked through a series of 180 trials in which they were asked to use a slider to estimate the correlation shown in the scatterplot. Visual masks preceded each plot. Interspersed were six attention check trials which asked participants to set the slider to 1 or 0 and ignore the scatterplot.

2 Results

All analyses were conducted using R (version 4.2.3 [**?**]). Models were built using the **buildmer** (version 2.8 [**?**]) and **lme4** (version 1.1-32 [2]) packages, with size manipulation being set as the predictor for participants' errors in correlation estimates

Mean errors in correlation estimates for the four size manipulation conditions can be seen in figure 2. A likelihood ratio test revealed that the model including size manipulation as a predictor explained significantly more variance than a null model ($\chi^2(3) = 3{,}508.84$, $p < .001$). This model has random intercepts for items and participants. The effect here is driven by participants' errors being lower for scatterplots with the non-linear size decay manipulation than for all other conditions, for error being lower for scatterplots with linear size decay than for plots with inverted non-linear decay or standard size, and for errors being higher for scatterplots with standard size than for plots with inverted non-linear decay.

Testing for contrasts between the four levels of the size manipulation condition were performed with the **emmeans** package (version 1.8.5 [**?**]), and of correlation estimates are shown in Figure 2. The **EMAtools** package was used to calculate effects sizes in Cohen's d, the results of which can be seen in Table 2.

In addition, we find no significant difference between the experimental model and another including graph literacy as a fixed effect ($\chi^2(1)$ 0.16, $p = .690$). These results suggest the effect we found

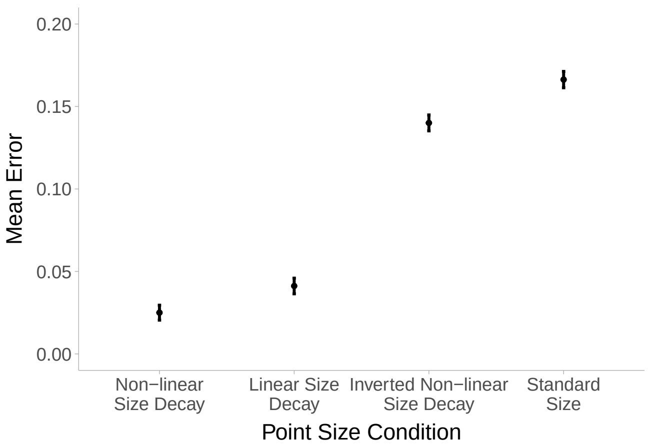

Figure 2: Mean error in correlation estimates across the four size manipulation conditions, with 95% confidence intervals shown.

Table 1: contrasts table

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

Table 2: The effects sizes of our point size manipulations when compared to the standard size condition.

Comparison	Cohen's d
Non-linear Size Decay	-0.61
Linear Size Decay	-0.54
Inverted Non-linear Size Decay	-0.11

was not driven by difference in graph literacy.

Figure 3 shows how participants' mean errors in correlation estimates change with the objective Pearson's r value, plotted separately for each size decay condition. Note the close-to-zero errors present in the non-linear size decay condition.

Acknowledgments

lah di dah

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*, 67:1–48, Oct. 2015. doi: 10.18637/jss.v067.i01

[3] 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

[4] W. S. Cleveland and R. McGill. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984. Publisher: [American Statistical Association, Taylor & Francis, Ltd.]. doi: 10.2307/2288400

[5] 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

[6] 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

[7] 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. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2011.127

[8] 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

[9] 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

[10] 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

[11] 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

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

[13] 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

[14] 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

[15] 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

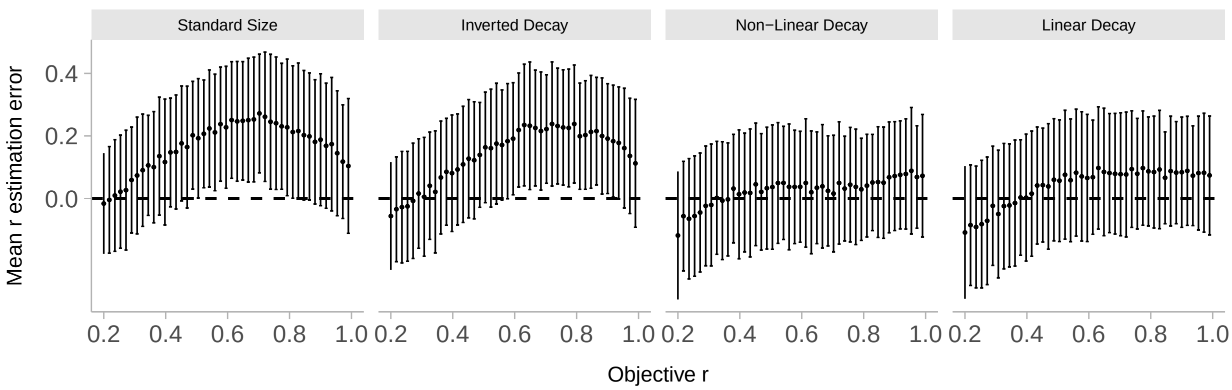

Figure 3: hello