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—

1 PLACE INTRO HERE. DON'T FORGET TO PUT HYPOTHESES

put ethics here

2 RELATED WORK

2.1 Correlation Perception

see Strain et al for a brief review of the history of correlation perception testing with scatterplots

2.2 Point Size

Hong et al paper will be useful here

2.3 Dot Pitch and Crowdsourced Experiments

3 METHODOLOGY

3.1 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`).

3.2 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 [?], and in accordance with previously published guidelines [?], 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

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

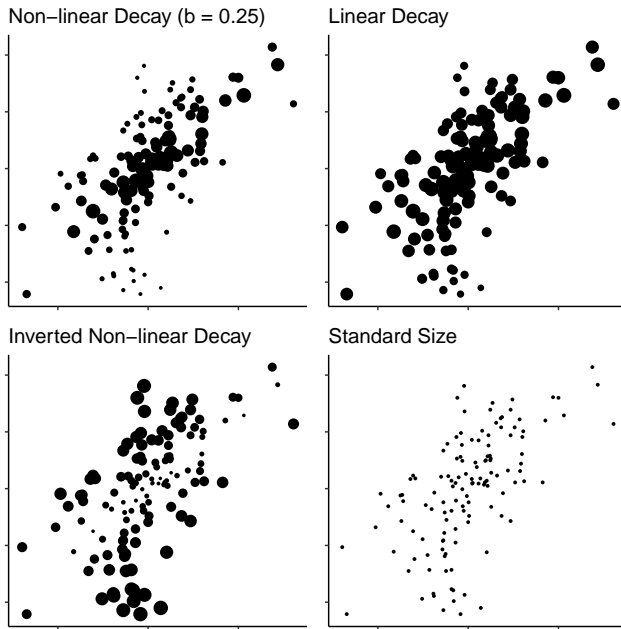average time taken to complete the experiment was 39 minutes (SD = 14 minutes).

3.3 Stimuli

The data used to generated the scatterplots in the current study was identical to that used previously [?]. 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 section 2.3 for a more detailed discussion of precise point sizes and dot pitch in crowd-sourced experiments.

As in our previous study [?], 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 [?]. Again, see section 2.3 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.

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

3.4 Dot Pitch and Crowdsourced Experiments

In our previous study [?], 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 [?]. 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 section 4 for analyses including dot pitch as a predictor.

3.5 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

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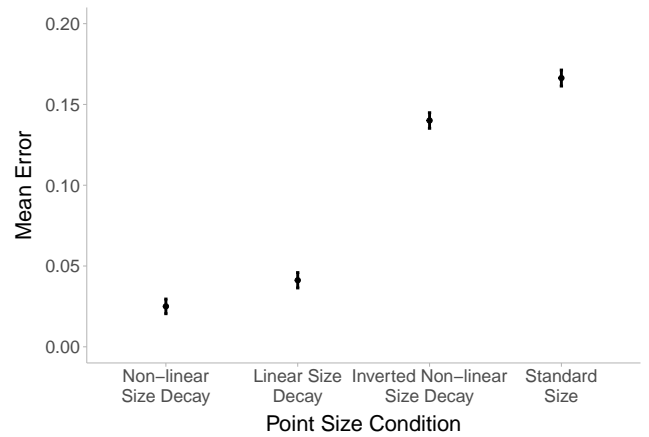

Figure 2: Mean error in correlation estimates across the four size manipulation conditions, with 95% confidence intervals shown.

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

to complete. Regardless, we consider these results to be indicative of a sufficient level of point visibility for the current experiment.

3.6 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`.

3.7 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, and their gender identity. Participants then completed the 5-item Subjective Graph Literacy (SGL) test [**?**], followed by the visual threshold testing described above. Participants then completed the screen scaling task described in section 2.3. Participants were given instructions, and then shown examples of *r* = 0.2, 0.5, 0.8, and 0.95, as in our previous work [**?**]. Section **??** 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. Interspersed were six attention check trials which asked participants to set the slider to 1 or 0 and ignore the scatterplot.

4 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 [**?**]) 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 are displayed in Table 1
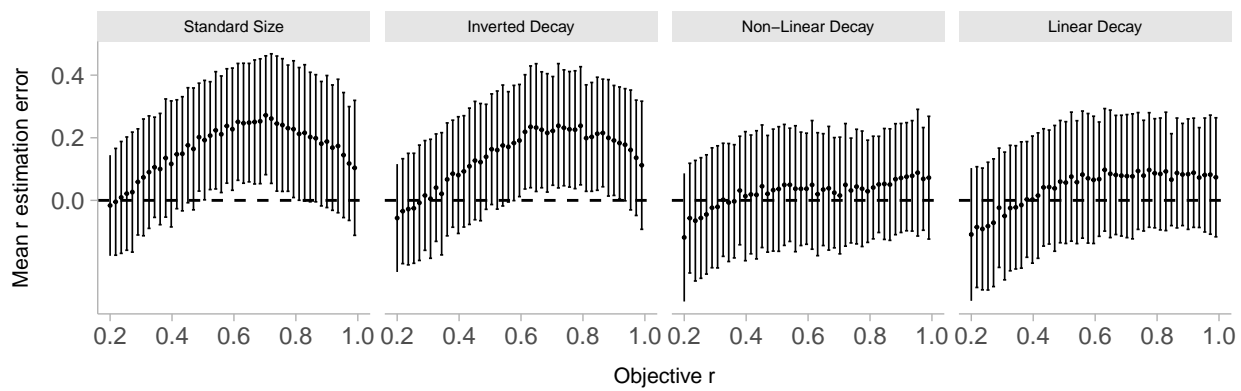
5 DISCUSSION

Figure 3: hello