The Absolute Judgement of Statistical Graphics: A Proposal

Naomi Martinez

In his paper *The Future of Data Analysis*, Tukey (1962) praised statistical graphs for their ability to provide more information than any other device; Cleveland and McGill (1984) reverberated this point by suggesting that graphical techniques be used to analyse data. Despite the clear communication that statistical graphs are imperative to research, there is a lack of communication between the graph itself and the audience. In this proposal, I will first review the research literature on how the trained and untrained eye alike interpret statistical graphs and will propose a study to attempt what statistical graphs do best: further the discovery of knowledge.

Statistical Graphs

Published in 1637, it was Descartes' *La Géométrie* that provided the foundation of statistical graphs in its introduction of the idea of a Cartesian coordinate system (Lewandowsky & Spence, 1989). William Playfair later created some of the most used graphs today: the histogram and line graph in 1786 and the pie chart in 1801 (Wainer, 2005). It was not until 1832 that Herschel created the next major graph, the scatterplot (Friendly & Denis, 2005). Although more have been created since, this proposal will focus on these foundations of modern statistical graphics.

Statistical graphs had not become commonplace in the scientific community until the 19th century. Even then, progress in their use was slow given that the creation of graphs was time consuming (Lewandowsky & Spence, 1989) and the visuals were difficult to understand, even to the scientifically trained mind (Hoff & Geddes, 1962). Nevertheless, these visual creations helped immensely in the furthering of scientific knowledge (Larkin & Simon, 1987; Tukey, 1975). It was only after graphing the absolute magnitude of stars as a function of their spectral class did the modern theories of stellar evolution come about.

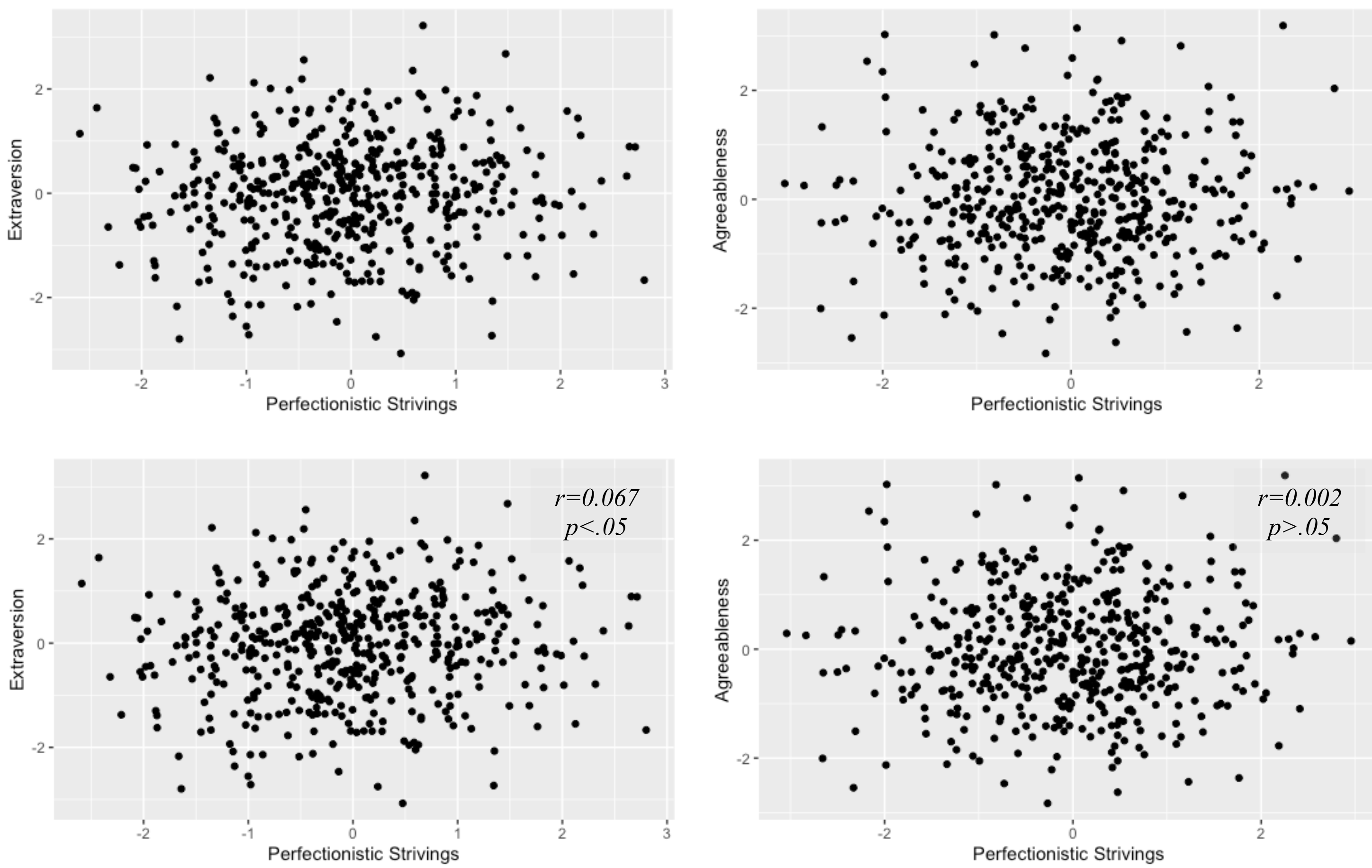
Statistical graphs have also been useful for the communication of knowledge (Tukey, 1962; Gelman & Stern, 2006) with a strength being the ability to visualize differences and similarities (e.g., does A have a higher proportion compared to B?) or the amount of anything that is of interest (i.e., an effect size; see Cumming & Fidler, 2009). Scatterplots, for example, have been successfully used as evidence in court cases on equal employment opportunities (Bobko and Karren, 1979).

Despite the recent movement in Psychology from null hypothesis statistical testing (NHST) toward effect sizes (Cumming, 2014; Flora & Pek, 2017; Wilkinson and the Task Force on Statistical Inference [TFSI], 1999), psychological research tends to contain less graphs than the natural sciences (Best et al., 2001; Smith et al., 2002) and primarily relies on graphs that lack context on the distribution(s) (Lane & Sandor, 2009). One recommendation for more information is to include inferential statistics (e.g., marking differences as significant or non-significant; or, simply including a label for the *p*-values within the graph itself). As Lane and Sandor (2009) put it, "The failure of graphs to portray inferential information successfully is a serious problem because inferential statistics are important for interpreting a graph." (p. 246)

On the contrary, including inferential statistics may only encourage the audience to confuse statistical significance with meaningfulness or importance. A relationship between two variables may seem meaningless unless its *p*-value is significant; likewise, a relationship may seem meaningful unless its *p*-value is non-significant. The scatterplots in Figure 1, for example, show two separate associations: perfectionistic strivings versus extraversion and perfectionistic strivings versus agreeableness. The top two scatterplots are context-free whereas the bottom two introduce inferential context, as per Lane and Sandor's (2009) guidelines. At face value, it may seem that there is no meaningful association between perfectionistic strivings and neither extraversion nor agreeableness. Introducing inferential context, however, may convince otherwise: there is a significant association between perfectionistic strivings and extraversion whereas the association between perfectionistic strivings and agreeableness is non-significant (Stricker et al. (2019) was used as a reference for the Big Five correlations).

Figure 1

Scatterplots for the association of extraversion and agreeableness to perfectionistic strivings. Note that the data are simulated.



Rather than us interpreting the visual, the inclusion of inferential statistics may be interpreting the graph for us. An alternative proposed by Cumming and Finch (2005) is to simply use confidence intervals to infer by eye, which does not solve the problem of bias introduced by inferential statistics as significance can easily be concluded from the confidence interval (e.g., $M_2 - M_1 = 0.5$ with a 95% CI [0.2, 0.7] is significant because it does not include the null effect of zero). Furthermore, the graph proposed by Cumming and Finch (2005) introduced a second y-axis for the difference between (in)dependent groups which may confuse audiences.

Visuals lacking inferential context live up to the ideal posed by Tukey (1977) that graphs "should force us to notice what we never expected to see." Considering only the summary and inferential statistics shown in Figure 1, we would have expected to see a noticeable difference between a significant and non-significant association. On the contrary, little to no differences are visible. The discrepancy between the quantity of the phenomena of interest and its visual representation makes for an important research topic that could have serious implications for how we conceptualize effect sizes (e.g., what is a meaningful difference or association, really?) In the next section, we review the literature on estimating magnitude from context-free statistical graphs that provide information on sample distributions (e.g., boxplots, histograms, and scatterplots).

Context-free Magnitude Estimation

Scatterplots

There have been various investigations on discriminative and absolute judgements of associations (Anderson et al., 2007). Discriminative judgements require the participants to judge associations on a greater than or less than basis. For example, Pollack (1960) was the first to investigate how scatterplots were perceived on an oscilloscope by having participants compare a scatterplot to a reference scatterplot on the grounds of graph size, sign of the correlation, and so on. Absolute judgments, on the other hand, require participants to provide a point estimate on the correlations presented in the shown scatterplots. The first investigation to study the direct estimation of a correlation from a scatterplot was by Strahan and Hansen (1978).

Strahan and Hansen (1978) presented booklets of 13 randomized positive bivariate normal scatterplots (200 points each) to Psychology graduate students and faculty. Half of these booklets had scatterplots that were equally spaced in r (r = .01, .092, .173, etc) and half had scatterplots that were equally spaced in r^2 (r^2 = .1, .301, .416, etc). Participants were told to write down for each scatterplot what they perceived to be the best estimate of the actual r up to two decimal places. The authors found that participants underestimated the degree of correlation, especially when the observed scatterplots were equally spaced in r^2. Furthermore, discrepancies between the actual and estimated r were more pronounced for scatterplots between the extreme values of 0 and 1. However, the hypothesis that participants might perceive r as r^2 was not supported.

Unlike Strahan and Hansen (1978), Wainer and Thissen (1979) found that graduate students in an introduction to statistics course did not underestimate the positive correlation coefficients presented in nine scatterplots. However, the authors had provided the participants prototypes of four bivariate normal scatterplots (r = 0, .25, .5, .75) prior to estimation.

Cleveland, Diaconis and McGill (1982) also found that participants with prior statistical skills underestimated the linear associations shown in 19 scatterplots (200 points each) within a printed booklet. However, these scatterplots did not present values of r but of $w(r)$, defined as $1 - \sqrt{1 - r^2}$ of which the authors hypothesized would be closer to participants' judgement than r. Participants had to judge ten levels of these associations on a scale from 0 (no association) to 1 (perfect association). Using a 10% trimmed mean, it was found that although $w(r)$ was a better fit for describing participants' perceived associations than r, it was not adequate. Similarly, although not the purpose of the research conducted by Lane, Anderson and Kellam (1985), undergraduate psychology participants tended to underestimate correlation coefficients presented in scatterplots (9 points) on a scale of 0 (no relatedness) to 100 (perfect relatedness).

Lauer and Post (1989) also analyzed the perception of bivariate normal associations presented in scatterplots by recruiting participants in technical and managerial positions with prior statistics knowledge. Four levels of correlation coefficients were shown via scatterplots (r = .25, .5, .75, .95) and the variance and number of points displayed were manipulated. Because these scatterplots were shown in an AT&T 6300 and Compaq portable computer, participants had to move the cursor along a number line printed below each scatterplot. As with most of the previous studies, it was found that the correlation coefficient was consistently underestimated; in fact, these estimations were lower than r^2.

Another study that found an underestimation of actual correlation coefficient was that by Collyer, Stanley and Bowater (1990) wherein undergraduates (prior statistical knowledge was not a requirement) were asked to estimate the actual correlation coefficient from 16 scatterplots presented in an Apple IIe computer, which ranged from an r of 0 to 1. The scatterplots varied such that there were 4 levels of correlation and 4 levels of standard deviation. Each scatterplot had 19 data points, and a visual line which participants could manipulate using game controllers until they judged it to be a good fit to the data. It was found that participants tended to underestimate the actual correlation coefficient 94% of the time and underestimated the coefficient of determination 70% of the time.

Meyer and Shinar (1992) also replicated the finding that people consistently underestimate r and r^2 by conducting two experiments. In the first experiment, university faculty involved in quantitative research and undergraduates that had finished an introductory to statistics course were given a booklet containing scatterplots that varies on levels of correlation, point cloud dispersion and regression line. Specifically, there were three levels of r^2: .4, .65, .9. In the second experiment, participants were senior engineering and management students that had previous statistical experience, and senior high school students with no experience. Similar to Cleveland et al. (1982), six values of $w(r)$ were presented in scatterplots: .3, .4, .5, .6, .7, .8. In both experiments, participants had to estimate the true correlation coefficient for each scatterplot on a scale of 0 to 100, with ticks for every 10 units. It was found the statistically savvy participants' estimations tended to be closer to the correct coefficient; overall, however, all participants consistently underestimated.

Bobko and Karren (1979) also investigated this topic by mailing a random sample of researchers (n=89) from Division 5 of the American Psychological Association eight scatterplots (n=100, equal variance) to estimate the magnitude and sign of: -.64, -.35, -.19, 0, .19, .35, .50, and .64. Participants underestimated all graphs, and the difference between the estimated r and true r were greatest in the mid-range values of $.2 \leq |r| \leq .6$. Furthermore, absolute values of $|r| \leq .2$ were perceived as virtually having no relationship. It is also interesting to note that the estimated rs are closer to their respective value of r^2 than r itself.

Histograms and Boxplots

There are no studies on the absolute judgment of the difference between two distributions as shown on a statistical graph. However, there is some research on discriminative judgements such that participants were required to judge two distributions on a greater than or less than basis. Lem et al. (2014) presented students that had completed introductory statistics 40 histogram pairs. The participants' task was to determine which of the two histograms for each pair displayed the greater mean. In addition to the mean, each pair varied in bar height (i.e., one histogram had greater frequencies compared to the other). Students tended to incorrectly use bar height to deduce the mean in a histogram (i.e., histograms with greater frequencies were perceived to have bigger means). Although experts (i.e., researchers in statistics) were more accurate, they also took more time in answering when presented with histograms varying in bar height.

Lem et al. (2013) presented five paper-and-pencil stimuli to first-year students that had likewise completed an introductory to statistics course. The first item compared two histograms, two boxplots or two descriptive statistics summaries. Students were asked whether one distribution contained more participants (i.e., a greater sample) than the other (note that the correct answer was 'indeterminable' when comparing between two box plots). The second item consisted of either one histogram, boxplot or descriptive summary and students were asked whether the mean was the same in two of the shown figures. Items 3 and 4 inquired on whether a certain quantity was observed (e.g., whether the minimum was three) and the last item was on whether the distributions exceeded a certain limit. Similar to the findings in Lem et al. (2014), students mistakenly used bar height in histograms to assess the distributions' means. In regard to boxplots, students tended to ignore the whiskers and falsely attributed the area of the box itself to be equal to frequencies (i.e., larger box area was perceived as frequencies or proportions).

Other research primarily focused on the misinterpretation of univariate distributions (i.e., a single boxplot or a single histogram) have found similar misinterpretations. For example, delMas et al. (2005) found that students had difficulty reading the data from histograms and preferred bar charts; Meletieu and Lee (2002) reported that some students use bar height as indicators of variability; and Bakker et al. (2004) highlighted that the area of boxplots was again misinterpreted for frequencies.

An explanation for the misinterpretations of histograms and boxplots could be found in Tversky's graphic design principles (1977). The directionality principle states that we primarily give focus to the vertical rather than the horizontal dimensions of a graph because we are oriented vertically, and gravity defines the vertical axis of our perceived world. This may explain the perceived importance of bar height in histograms despite the task being to focus on the distribution along the horizontal axis. (Furthermore, the boxplots in the previous research were oriented horizontally).

The spatial metaphors principle states that we naturally give meaning to areas such that areas in a graph are associated with properties of the variable (e.g., frequencies or proportions). This may explain why the area within a traditional boxplot is typically misinterpreted to represent the number of observations in a given distribution.

Proposed study

Considering that there is no research on the absolute judgements of the difference between means of two distributions, the proposed study seeks to fill that gap by surveying via a Shiny app trained participants on the magnitude of the mean differences that they perceive via histograms and boxplots. Furthermore, we will take Tversky's (1977) directionality and spatial metaphors principles into account by incorporating histograms and boxplots displayed vertically and horizontally and vary the boxplot area (see Figures 2-5).

Although we've shown that there is substantial research on absolute judgements of associations, we will nevertheless include scatterplots to replicate previous findings that actual correlations are consistently underestimated and take advantage of the benefits associated with Shiny. For example, unlike paper-and-pencil tests and other software, Shiny will allow participants to manipulate the visuals. Furthermore, no two graphs will be the same as they will be generated in a reactive environment.

It is important to note that only one graph will be shown to the participant at a time. Our samples will be simulated in R from normal distributions with a mean of zero and a standard deviation of one. Considering the latter, we can treat the mean differences as standardized (i.e., Cohen's *d*). Our research questions therefore are:

(1) What are the estimated Cohen's *d* values perceived from the overlapping histograms, back-to-back histograms, and boxplots?
(2) What are the estimated correlation values perceived from the scatterplots?
(3) Are there discrepancies between the estimated values perceived from an overlapping histogram versus a back-to-back histogram?
(4) Are there discrepancies between the estimated values perceived from a vertical boxplot versus a horizontal boxplot?
(5) Does the discrepancy between estimated and actual values differ by the type of graph (e.g., histogram versus boxplot)?
(6) At what point does a visual association and mean difference become meaningful?

Measures

The histograms will display two distributions separated by a true *d* value ranging from 1 to 2 by .05 increments (41 overlapping histograms and 41 back-to-back histograms) with a sample size of N = 500. Likewise, the boxplots will display two distributions separated by a true *d* value ranging from 1 to 2 by .05 increments (41 vertical and 41 horizontal) with a sample size of N = 500.

Using the *runif* function in R, the first histogram/boxplot in each pair will have a random value that could range from 0 to 2 assigned to its mean. The mean of the second histogram/ boxplot in each pair will be the sum of the mean of the first plot and its assigned *d.* For example, if the mean of the first histogram/boxplot was 1.6 then the mean of the second would be 1.6 + 1 to display a *d* or mean difference of one. This will be done to prevent the participants from easily calculating the true *d* value in each graph. Lastly, each histogram/boxplot will be shown with a slider bounded by an upper limit of 3 and a lower limit of 0 with increments of .05.

The scatterplots will display ρ values ranging from -1 to 1 by .05 increments (41 total scatterplots), with a sample size of N = 500. These figures will be created using the *rmvnorm* function in R. In each instance, the two variables will be randomly sampled from a multivariate normal distribution with a mean of zero. The sliders associated with each scatterplot will be bounded by an upper limit of 1 and a lower limit of -1 and increments of .05.

Figure 2

An example of a back-to-back histogram that will be shown to participants (*d* = 2)

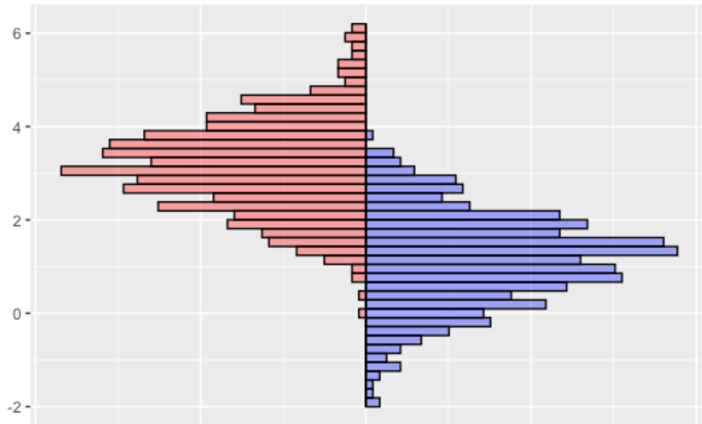


Figure 3

An example of an overlapping histogram that will be shown to participants (d = 2)

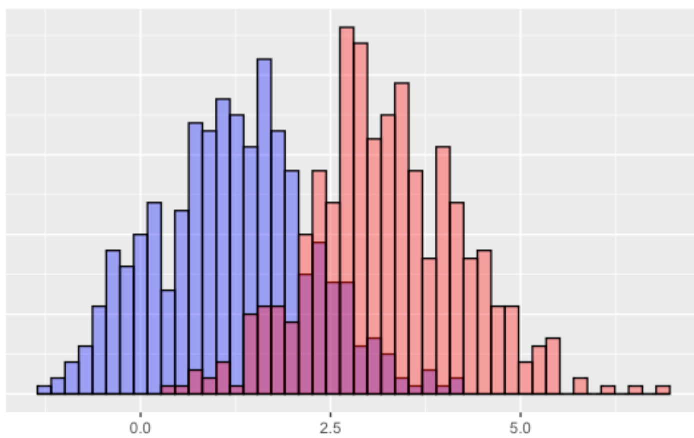


Figure 4

An example of a vertical boxplot that will be shown to participants (d = 2). X marks the sample mean.

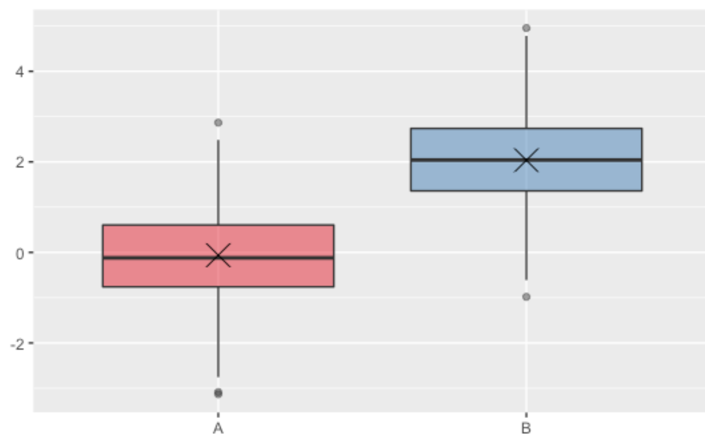
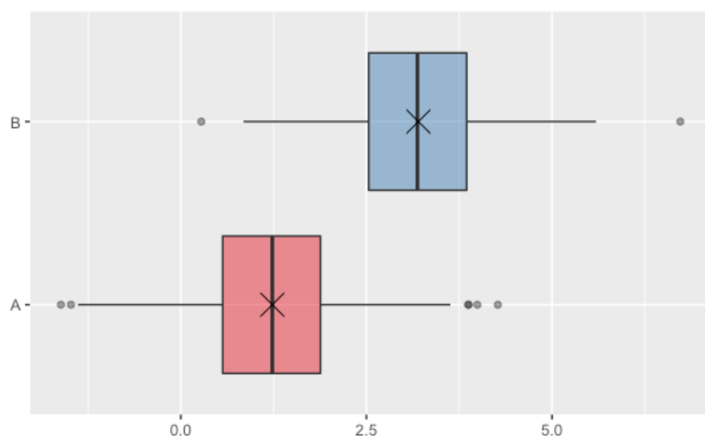


Figure 5

An example of a horizontal boxplot that will be shown to participants (d = 2). X marks the sample mean.



Method

Procedure

Participants will be directed to the application hosted by the Shiny Server (http://shinyapps.io). After providing consent for participation in the study, participants will be asked to move a practise slider from its default value of zero to a value of three to provide evidence of general understanding.

Training

To account for potential misunderstandings of the statistical and graphical concepts involved, we will include statistical training sessions prior to each module. That is, prior to the modules on histograms and boxplots, participants will be taught what a histogram is and how it should be interpreted (e.g., the height of the bars pertain to frequencies). Furthermore, the anatomy of the traditional box plot will be broken down to its parts (e.g., the median and the IQR).

Three examples will be provided on what a fairly large (*d*= 2), considerable (*d* = 1) and trivial (*d* = 0) difference in means looks like visualized via boxplots, overlapping histograms and back-to-back histograms. The 'Next' button will only be enabled when participants' estimates are within an acceptable range. That is, participants will only be able to continue the study if their estimated *d* pertaining to true *d* = 0 is $0 \leq | d | \leq .5$; the estimated *d* for true *d* = 1 is $.5 \leq | d | \leq 1.5$; and, the estimated *d* for true *d* = 2 is $1.5 \leq | d | \leq 2.5$.

Likewise, we will incorporate a short introduction to scatterplots and showcase three examples to teach participants what a fairly strong positive (ρ = .7), fairly strong negative (ρ = -.7) and no association (ρ = 0) between two variables looks like. Participants will only be able to continue the study if their estimated *r* pertaining to ρ = 0 is $-.5 \leq | r | \leq .5$; the estimated *r* for ρ = -.7 is $-.9 \leq | r | \leq -.3$; and, the estimated *r* for ρ = .7 is $.3 \leq | r | \leq .9$.

Modules

There will be three modules—one for histograms, another for boxplots and a third for scatterplots (in that order). Participants will first conclude the training session on histograms prior to starting the first module. The histogram module, specifically, will consist of 41 overlapping histograms with *d* values ranging from 0 to 2 by .05 increments. Following this batch, 41 back-to-back histograms with *d* values ranging from 0 to 2 by .05 increments will be shown. Participants will be asked to estimate the true *d* displayed in each plot.

The training session on boxplots will commence after the first module. Afterwards, participants will be shown 41 vertical boxplots with *d* values ranging from 0 to 2 by .05 increments followed by 41 horizontal boxplots with *d* values ranging from 0 to 2 by .05 increments. As with the first module, participants will be asked to estimate the true *d* displayed in each plot.

Participants will need to complete the training on scatterplots prior to the last module. Within this third module, participants will be shown 41 scatterplots displaying ρ values ranging from -1 to 1 by .05 increments and they will be asked to estimate the correlation (*r*) displayed in each scatterplot.

Lastly, participants will be asked to manipulate histograms, boxplots and scatterplots end the end of the last module to indicate at what point they think an association or mean difference would become meaningful.

Conclusion

Answers to the six questions posed could have various implications for future research. For example, if we more accurately estimate effects when a graph is oriented vertically versus horizontally, should we introduce statistical graphs to novice students in a vertical fashion? If perceived meaningful estimations congregate around a certain value, what does that mean for the guidelines provided by Cohen? These are only some questions that could come about from this proposal. Although we have been using these statistical graphs in our own research for decades, very little is known about our understanding of them. In fact, rather than the graph itself, perhaps it is our understanding of how we make sense of these visuals that will finally force us to notice what we never expected to see.

References

Bakker, & Hoffmann, M. H. G. (2005). Diagrammatic Reasoning as the Basis for Developing Concepts: A Semiotic Analysis of Students' Learning about Statistical Distribution. Educational Studies in Mathematics, 60(3), 333–358. https://doi.org/10.1007/s10649-005-5536-8

Best, Smith, L. D., & Stubbs, D. A. (2001). Graph use in psychology and other sciences. Behavioural Processes, 54(1), 155–165. https://doi.org/10.1016/S0376-6357(01)00156-5

Bobko, & Karren, R. (1979). The Perception of Pearson Product Moment Correlations From Bivariate Scatterplots. Personnel Psychology, 32(2), 313–325. https://doi.org/10.1111/j.1744-6570.1979.tb02137.x

Cleveland, Diaconis, P., & McGill, R. (1982). Variables on Scatterplots Look More Highly Correlated When the Scales are Increased. Science (American Association for the Advancement of Science), 216(4550), 1138–1141. https://doi.org/10.1126/science.216.4550.1138

Cleveland, & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Journal of the American Statistical Association, 79(387), 531–554. https://doi.org/10.1080/01621459.1984.10478080

Collyer, Stanley, K. A., & Bowater, C. (1990). Psychology of the Scientist: LXIII. Perceiving Scattergrams: Is Visual Line Fitting Related to Estimation of the Correlation Coefficient? Perceptual and Motor Skills, 71(2), 371–378E. https://doi.org/10.2466/pms.1990.71.2.371

Cumming, & Fidler, F. (2009). Confidence Intervals: Better Answers to Better Questions. Zeitschrift Für Psychologie Mit Zeitschrift Für Angewandte Psychologie, 217(1), 15–26. https://doi.org/10.1027/0044-3409.217.1.15

Cumming, & Finch, S. (2005). Inference by Eye: Confidence Intervals and How to Read Pictures of Data. The American Psychologist, 60(2), 170–180. https://doi.org/10.1037/0003-066X.60.2.170

Delmas, R.C., Garfield, J., & Ooms, A. (2005). Using assessment items to study students' difficulty reading and interpreting graphical representations of distributions.Retrieved from https://www.causeweb.org/cause/archive/artist/articles/SRTL4_ARTIST.pdf

Doherty, Anderson, R. B., Angott, A. M., & Klopfer, D. S. (2007). The perception of scatterplots. Perception & Psychophysics, 69(7), 1261–1272. https://doi.org/10.3758/BF03193961

Gelman, & Stern, H. (2006). The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant. The American Statistician, 60(4), 328–331. https://doi.org/10.1198/000313006X152649

Friendly, & Denis, D. (2005). The early origins and development of the scatterplot. Journal of the History of the Behavioral Sciences, 41(2), 103–130. https://doi.org/10.1002/jhbs.20078

Hebbel E. Hoff, & L. A. Geddes. (1962). The Beginnings of Graphic Recording. Isis, 53(3), 287–324. https://doi.org/10.1086/349594

Lane, & Sándor, A. (2009). Designing Better Graphs by Including Distributional Information and Integrating Words, Numbers, and Images. Psychological Methods, 14(3), 239–257. https://doi.org/10.1037/a0016620

Larkin, J.H. and Simon, H.A. (1987), Why a Diagram is (Sometimes) Worth Ten Thousand Words. Cognitive Science, 11: 65-100. https://doi.org/10.1111/j.1551-6708.1987.tb00863.x

Lauer, & Post, G. V. (1989). Density in scatterplots and the estimation of correlation. Behaviour & Information Technology, 8(3), 235–244. https://doi.org/10.1080/01449298908914554

Lem, Onghena, P., Verschaffel, L., & Van Dooren, W. (2013). On the misinterpretation of histograms and box plots. Educational Psychology (Dorchester-on-Thames), 33(2), 155–174. https://doi.org/10.1080/01443410.2012.674006

Lem, Onghena, P., Verschaffel, L., & Van Dooren, W. (2014). Interpreting histograms. As easy as it seems? European Journal of Psychology of Education, 29(4), 557–575. https://doi.org/10.1007/s10212-014-0213-x

Lewandowsky, & Spence, I. (1989). The Perception of Statistical Graphs. Sociological Methods & Research, 18(2-3), 200–242. https://doi.org/10.1177/0049124189018002002

Meletiou-Mavrotheris, Maria. (2002). Student understanding of histograms: a stumbling stone to the development of intuitions about variation. Retrieved from https://www.researchgate.net/publication/253978752_STUDENT_UNDERSTANDING_OF_HISTOGRAMS_A_STUMBLING_STONE_TO_THE_DEVELOPMENT_OF_INTUITIONS_ABOUT_VARIATION

Meyer, & Shinar, D. (1992). Estimating Correlations from Scatterplots. Human Factors, 34(3), 335–349. https://doi.org/10.1177/001872089203400307

Pek, & Flora, D. B. (2018). Reporting Effect Sizes in Original Psychological Research: A Discussion and Tutorial. Psychological Methods, 23(2), 208–225. https://doi.org/10.1037/met0000126

Pollack. (1960). Identification of visual correlational scatterplots. Journal of Experimental Psychology, 59(6), 351–360. https://doi.org/10.1037/h0042245

Strahan, & Hansen, C. J. (1978). Underestimating Correlation from Scatterplots. Applied Psychological Measurement, 2(4), 543–550. https://doi.org/10.1177/014662167800200409

Smith, Best, L. A., Stubbs, D. A., Archibald, A. B., & Roberson-Nay, R. (2002). Constructing Knowledge: The Role of Graphs and Tables in Hard and Soft Psychology. The American Psychologist, 57(10), 749–761. https://doi.org/10.1037/0003-066X.57.10.749

Stricker, Buecker, S., Schneider, M., Preckel, F., & Kandler, C. (2019). Multidimensional Perfectionism and the Big Five Personality Traits: A Meta-analysis. European Journal of Personality, 33(2), 176–196. https://doi.org/10.1002/per.2186

Stukas, & Cumming, G. (2014). Interpreting effect sizes: Toward a quantitative cumulative social psychology: Interpreting effect sizes. European Journal of Social Psychology, 44(7), 711–722. https://doi.org/10.1002/ejsp.2019

Tukey. (1962). The Future of Data Analysis. The Annals of Mathematical Statistics, 33(1), 1–67. https://doi.org/10.1214/aoms/1177704711

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975* (Vol. 2, pp. 523-531).

Tukey. (1977). Exploratory data analysis. Addison-Wesley Pub. Co.

Tversky, B. (1997). Cognitive principles of graphic displays.Retrieved from https://aaai.org/Library/Symposia/Fall/fs97-03.php

Wainer, & Thissen, D. (1979). On the Robustness of a Class of Naive Estimators. Applied Psychological Measurement, 3(4), 543–551. https://doi.org/10.1177/014662167900300411

Wainer. (2005). Graphic discovery: a trout in the milk and other visual adventures. Princeton University Press.

Wilkinson. (1999). Statistical Methods in Psychology Journals: Guidelines and Explanations. The American Psychologist, 54(8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594