

# Graphical Data Analysis

Paul Collett

*Shimonoseki City University*

## Reference Data

Collett, P. (2021). Graphical data analysis. In P. Clements, R. Derrah, & P. Ferguson (Eds.), *Communities of teachers & learners*. JALT. <https://doi.org/10.37546/JALTPCP2020-01>

Limitations in statistical significance testing are an issue of ongoing debate in many academic disciplines. Moving away from a reliance on their use in quantitative research is seen as an important step towards improving the quality of such research. One area that can help here is graphical data analysis, both as an exploratory and explanatory tool. This paper presents an overview of graphical techniques for quantitative data analysis. After outlining the rationale for the use of graphical data analysis, consideration of the appropriate types of graphs to use is provided. A number of useful graphs, created using the R statistical package, are introduced, along with a link to the full code to reproduce the examples. Suggestions are presented for how graphical techniques can help with both the exploration and confirmation stages in research.

統計的有意差検定の限界は、多くの学術分野で継続的に議論されている問題である。量的研究における統計的有意差検定への使用依存から脱却することは、研究の質を向上させるための重要なステップであると考えられる。ここで役立つのが、探索的・説明的なツールとしてのグラフデータ分析である。本稿は、定量的なデータ分析のためのグラフ技術の概要を説明する。グラフィカルなデータ分析を使用する理由を説明した後、使用するグラフの適切な種類について検討する。R統計パッケージを用いて作成された便利なグラフの数々を紹介し、例題を再現するための完全なコードへのリンクも掲載する。グラフィカルな技術が、研究における探索と確認の両方の段階でどのように役立つかを提案する。

Graphical data analysis, or the visualization of data, can play an important part in quantitative research. Following the data collection stage, one of the first steps in a research project should be to explore data to check if it fits the assumptions of tests such as a t-test or ANOVA (e.g., that the data is normally distributed), and to identify possible patterns or trends. Plotting data in different ways can be informative here, revealing

issues which might be obscured if the research focuses only on statistical outcomes. In guidelines into research practices for the behavioral sciences, Wilkinson, and Task Force on Statistical Inference (1999, p. 597) commented that “Graphics broadcast; statistics narrowcast...As soon as you have collected your data, before you compute any statistics, look at your data...if you assess hypotheses without examining your data, you risk publishing nonsense.” This holds true for research within any domain.

## Why Graphical Analysis?

Recent years have seen a call for alternatives to the reporting of  $p$ -values in quantitative research within many fields of research. This continues a decades-long discussion on the deficiencies of the standard hypothesis testing method (null hypothesis statistical testing, NHST) with its reliance on finding statistical significance. One of the main issues here relates to a misunderstanding or misinterpretation of exactly what getting a statistically significant result means: “A  $p$ -value is the probability of the observed data (or of more extreme data points), given that the null hypothesis,  $H_0$  is true” (Gigerenzer et al., 2004, p. 393). It does not mean you have disproved the null hypothesis, or that you have proved your experimental hypothesis. There are other issues that raise questions regarding the value of NHST. Cohen (1994) is one particularly authoritative source, outlining the general problems with the logic behind this methodology. These include a problem with the deductive syllogistic reasoning related to rejecting the null hypothesis, an issue with the probability behind acceptance/rejection of the null hypothesis, and the likelihood that the null hypothesis is always false. He posits alternative approaches, including the use of exploratory and graphic data analysis to better understand the data under study.

Haig (2018), in arguing for a reformulation of the hypothesis-testing approach, discusses other problems, in particular the hybrid nature of NHST, whereby the standard NHST theory and methodology presented in statistical texts or taught in methodology

courses is a mix of ideas from Fisher's original model and the later Neyman-Pearson approach, which misrepresents both. Cumming (2012) provides a comprehensive introduction to the "New Statistics" of estimation using effect sizes and confidence intervals (CIs), which should be used to augment, if not replace, *p*-values. CIs, which give an interval estimation of the sample means, are particularly suited for graphical presentation. While graphical data analysis approaches are not an answer to all the problems here, they can help researchers better understand data, moving beyond the limitations of reliance on a single point estimate like a *p*-value for drawing conclusions about the difference or otherwise between groups under study. One part of good statistical practice should include "a variety of numerical and graphical summaries of data" (Wasserstein & Lazar, 2016, p. 132).

Tukey (1980) outlines the need for an exploratory approach to data analysis which should guide theory generation, experimental design, data collection, and analysis. His point is "that to implement the very confirmatory paradigm properly we need to do a lot of exploratory work" (p. 23), here of course graphical analysis can play an important role. Haig (2013) meanwhile points out that relying too heavily on statistical inference "has had the effect of discouraging a concerted examination of datasets in terms of their quality and structure" (p. 9).

Research reports within the L2 field are not immune to these problems. The following draws on discussion from a number of different research fields, and the conclusions are broadly applicable across the spectrum of language learning and applied linguistics research. Al-Hoorie and Vitta (2019) elucidate trends in statistical reporting across a range of journals in the second language learning field. They found a number of recurrent issues across the journals they investigated. Such problems as not reporting assumption checks, or poor inferential conclusions from datasets are areas in which exploring the data graphically could help (Norris et al., 2015). Larson-Hall (2017) advocates for the use of data accountable graphics (graphics showing all the relevant details of the dataset, i.e., plotting each response from a sample) and data rich graphics (those that illustrate the data distribution and provide a large amount of information) to augment statistical results as "...such graphics will firmly establish the credibility of the statistical arguments" (p. 244).

One factor holding back the production of effective graphics has been limitations in the ability to produce suitable plots for dissemination. Larson-Hall and Herrington (2010) call for the use of modern developments in applied statistics to improve

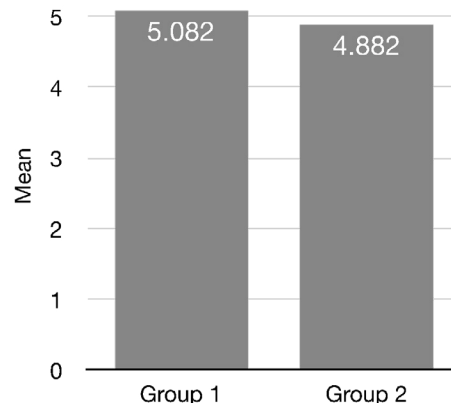
data analysis in SLA. This general idea was further developed by Mizumoto and Plonsky (2016) who recommend R, a computer programming and statistical software environment as a "lingua franca" for research into applied linguistics. R, in tandem with the RStudio integrated development environment (RStudio Team, 2021), is especially appropriate for the creation of graphs and charts thanks to numerous contributed packages, especially the ggplot2 package for graphics (Wickham, 2016). Echoing these views, this paper builds from the suggestions put forward by Larson-Hall and Herrington (2010), Mizumoto and Plonsky (2016), and Larson-Hall (2017), looking at some newer developments to augment their recommendations.

### What Kinds of Graphs?

The majority of graphics used in reporting second or foreign language learning research results are bar graphs (barplots) and line graphs (Hudson, 2015; Larson-Hall, 2017). One limitation with these kinds of graphs is that they obscure the trends in the data and may even misrepresent findings. An example follows, working with randomly-generated simulated datasets (the R code to generate the data, and carry out the *t*-test as well as produce the graphs can be found online at <https://rpubs.com/pcjapan/696485>). For illustrative purposes, we will assume the data represents results of a comprehension test with a total score of 10 given to two separate groups, group 1 ( $n = 100$ ) and group 2 ( $n = 100$ ). We wish to test the null hypothesis of no difference in the population mean ( $\mu$ ) between groups:  $H_0: \mu_1 = \mu_2$ .

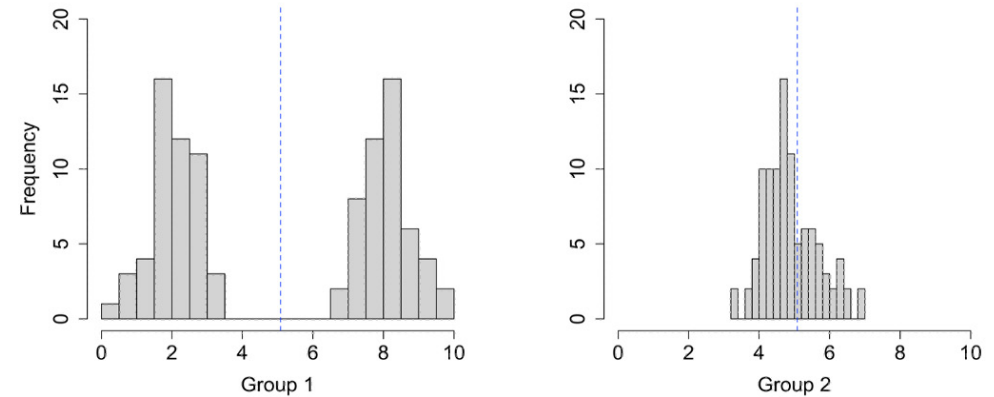
Figure 1 shows a bar graph for the means of the two groups. The graph suggests there is virtually zero difference between mean scores for the two groups. The *t*-test of the data supports this:  $t(110.68) = 0.625$ ,  $p = .533$ ,  $CI_{95\%} [-.432, .831]$ . As the *p* value is greater than .05, standard practice would then be to report that a non-statistically significant result for the difference between the two groups was found. Based on the graph and the result of the *t*-test, it would seem appropriate to say that the scores for the two groups are similar, and to draw conclusions based on this.

**Figure 1**  
*Means for Simulated Dataset 1*



However, if we look more closely at the data, some quite different possibilities emerge. Rather than focusing on the means, looking at how the data is distributed may be more helpful; this can be done with a histogram. As the histograms in Figure 2 illustrate, the data for the two groups display different distributions, with group 1 showing a bimodal distribution, and group 2 an approximately normal distribution. The two groups are not similar, contrary to the above-reported bar graph and *t*-test results. While this is an admittedly extreme scenario, it should illustrate how an appropriate graphical representation of the data can aid in understanding what exactly your data represents.

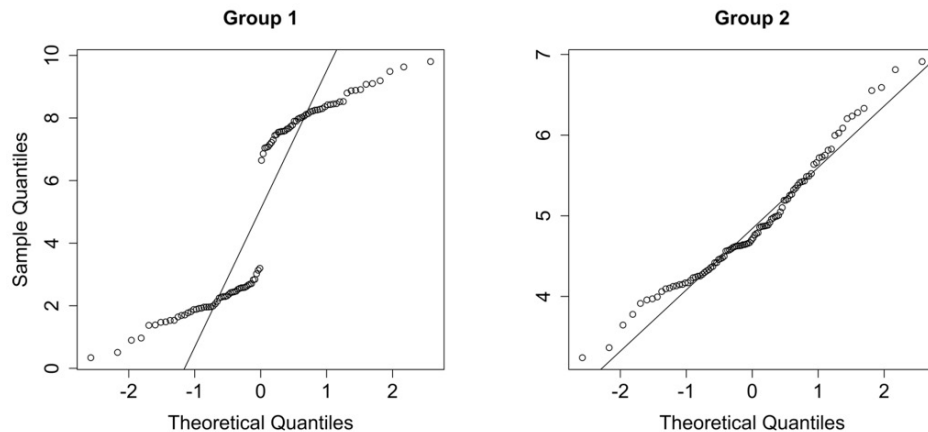
**Figure 2**  
*Distribution of Simulated Dataset 1*



*Note:* the blue dashed line shows the mean for each group.

An alternative to using a histogram is the Q-Q plot which shows the data ranked and plotted against a hypothetical normal distribution. The plots for our data are seen in Figure 3, with group 1 on the left and group 2 on the right. As with the histograms in Figure 2, it is easy to spot the problems with the distribution. The solid lines show the hypothetical normal distribution, the points show our data, a normally-distributed dataset will follow closely the hypothetical distribution.

**Figure 3**  
*Q-Q Plots for Simulated Dataset 1*



Both of these kinds of plots are useful for understanding the overall shape of your dataset, and are appropriate for the exploratory stage of research. They can be augmented with other graphs, to create a richer description of the data for explanatory purposes. In particular, Larson-Hall and Herrington (2010) suggest using boxplots and scatterplots with smoother (loess) lines. This advice is still applicable, and further developments have made an even wider range of graphical analysis techniques relatively easy to produce. We will turn to these in the following section.

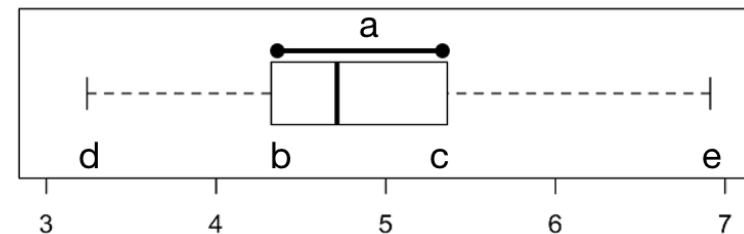
### Boxplots and Their Variants

Boxplots (also known as box-and-whisker plots) provide a useful way to make sense of the patterns of data in two (or more) samples, in cases where differences between the data is of interest. For example, they can be used when carrying out a t-test or ANOVA (or a non-parametric equivalent) to analyse the difference between means for a number of groups.

Figure 4 illustrates the components of a boxplot, plotted from group 2 of the hypothetical dataset. The solid black line in the box is the median. The length of the box (a) shows the distribution of the middle 50% of the data (the interquartile range). The

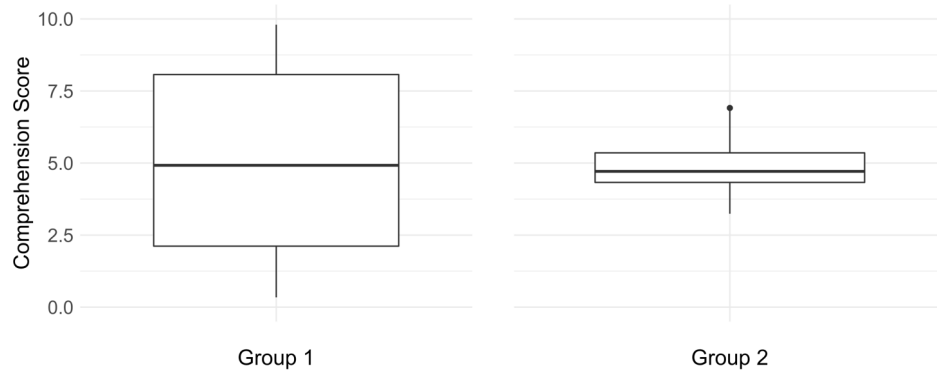
ends of the box (b, c) give the lower and upper quartile. The vertical bars (d, e) at the end of the dashed lines—the whiskers—represent the limit of the range of the lower and upper 25% of the data. So, from this particular example we can see that the lower 25% of the data is in the range from 3.2 to 4.3; the central 50% of the data (interquartile range) lies between 4.3 to 5.3; and the top 25% is between 5.3 and 6.9. 75% of the data is below the upper quartile (c) - i.e., 75% of the data is less than 5.3, and 25% below the lower quartile (b), as noted above in our case, below 4.3. Any outliers would be plotted as points above and below the limits of the top and bottom 25%. The length of the whiskers points to any skew in the distribution. Skew is when there is an uneven distribution of data about the mean with more data grouped on one side than the other, or a non-normal distribution, which can have unwanted effects on statistical outcomes. In this example the distribution has a longer right tail or positive skew; i.e., more of the data lies in the upper range of the distribution.

**Figure 4**  
*Boxplot Structure*



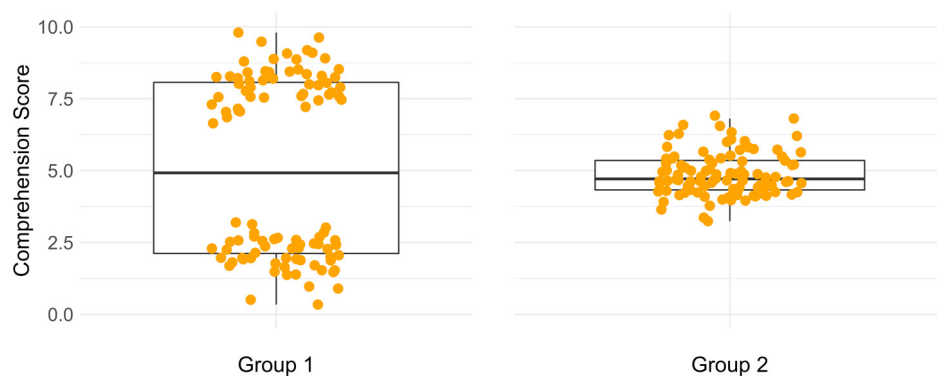
As Larson-Hall and Herrington (2010) point out, boxplots show more data than a bar graph; they suggest that “boxplots should always be preferred over barplots unless the data are strictly frequency data” (pp. 369-370). However, at least for the somewhat extreme example of our simulated datasets, boxplots could also lead to incorrect assumptions. Figure 5 shows boxplots for groups 1 and 2. Comparing these to Figure 1, it is obvious the boxplot supplies a lot more information. The interquartile range and the whiskers suggest a difference in dispersion of the data for the two groups. However, it needs to be noted that the similarity of the medians (4.921 for group 1 and 4.712 for group 2) could lead to misguided conclusions.

**Figure 5**  
*Boxplots for Simulated Dataset 1*



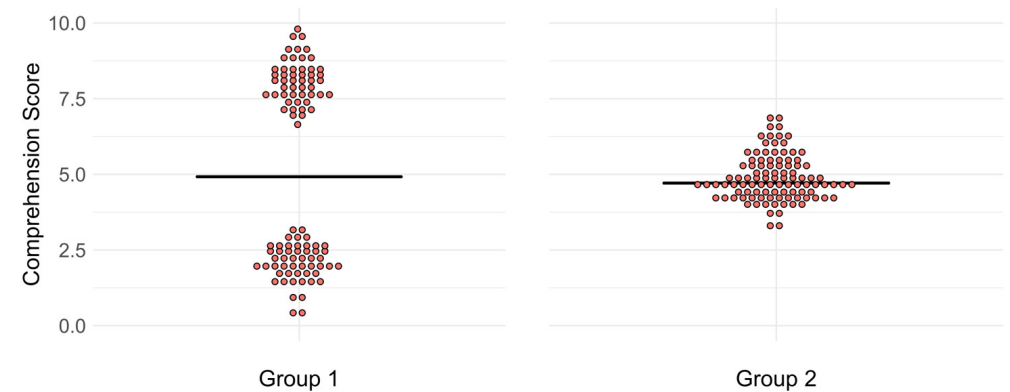
One way to increase the explanatory richness of a boxplot is to augment it with the individual data points from the samples. This can be done quite easily in R, and with the use of jittering (spreading out the points) to avoid superimposing points with the same value over each other (overplotting) it is possible to show an entire dataset on a boxplot (Figure 6). With our example where each sample size is  $n = 100$ , this works reasonably well, clarifying the difference in distributions of the two groups.

**Figure 6**  
*Boxplots for Simulated Dataset 1 Augmented with Jittered Data*



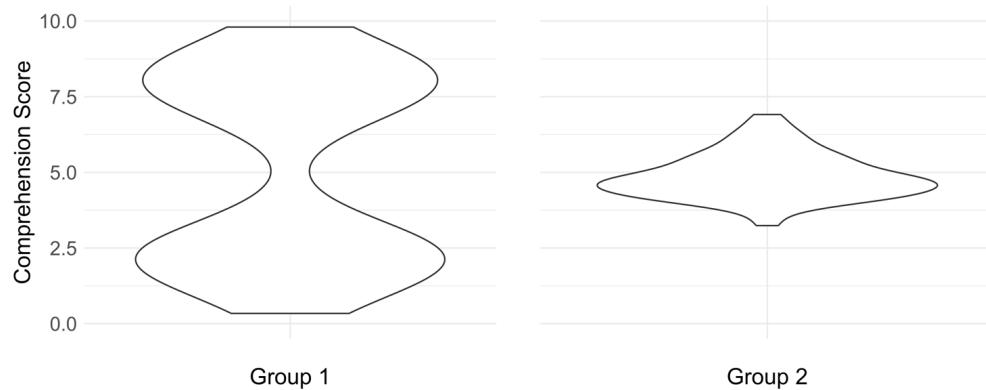
A somewhat cleaner variation of the above is to use a dotplot. This shows similar information about the data, but with a clearer indication of the distribution of the data points. Each individual point is plotted, making this kind of graph an especially good choice for smaller samples as it can show the location of every response. This is illustrated in Figure 7.

**Figure 7**  
*Dotplots for Simulated Dataset 1*



As mentioned, when plotting data points, overplotting can be an issue. This is especially true for large samples; superimposing data points over a boxplot or making a dotplot is probably not going to be very helpful if you are working with samples of hundreds or thousands. An alternative graph that works well here is a violin plot. They show the distribution of the data as well as indicating differences in the data (Wickham & Stryjewski, 2011). This kind of graph works well for both exploratory and explanatory purposes. Figure 8 shows violin plots for our simulated data.

**Figure 8**  
*Violin Plots for Simulated Dataset 1*

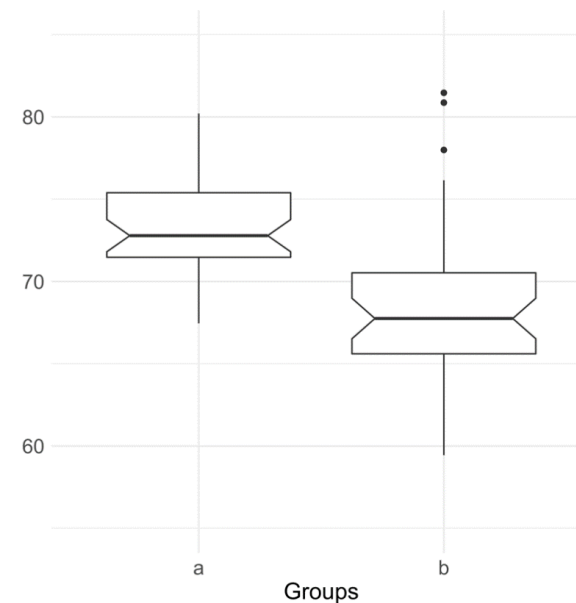


With boxplots, it is possible to get an alternative measure to a  $p$ -value by including notches. The notch shows the 95% confidence interval around the median. As such, if the notches do not overlap, it provides strong evidence that the medians differ (Wickham & Stryjewski, 2011). To provide an example here, the data used for the previous graphs will not work here, so it is necessary to create a new randomly-generated set of simulated data, representing two independent samples, group a and group b, each of  $n = 40$ . In this case, the data represents an outcome  $x$  on a test with a maximum possible score of 100. We want to check if there is any difference between the outcomes of the two groups. Prior to carrying out the analysis, the data should be checked with histograms and Q-Q plots. While not displayed here, these indicated group b deviates from normality, and has some outliers. To compare the two sets, descriptive statistics (Table 1) and a notched boxplot (Figure 9) are used.

**Table 1**  
*Descriptive Statistics for Simulated Dataset 2*

Group	N	Mean	20% Trimmed Means	SD	Median	95% CI Lower (Median)	95% CI Upper (Median)	Skew	Kurtosis
a	40	73.47	73.15	3.05	72.79	71.92	74.02	0.36	-0.62
b	40	68.58	67.73	4.55	67.75	66	68.73	1.09	1.2

**Figure 9**  
*Notched Boxplot for Simulated Dataset 2*



Note the skew for group b; an oft-quoted rule of thumb from Bulmer (1979) is that skew greater than  $+ / - 1$  should be considered as high, which points to another indication of the violation of the normality assumption for the sample. The notches of the boxplots do



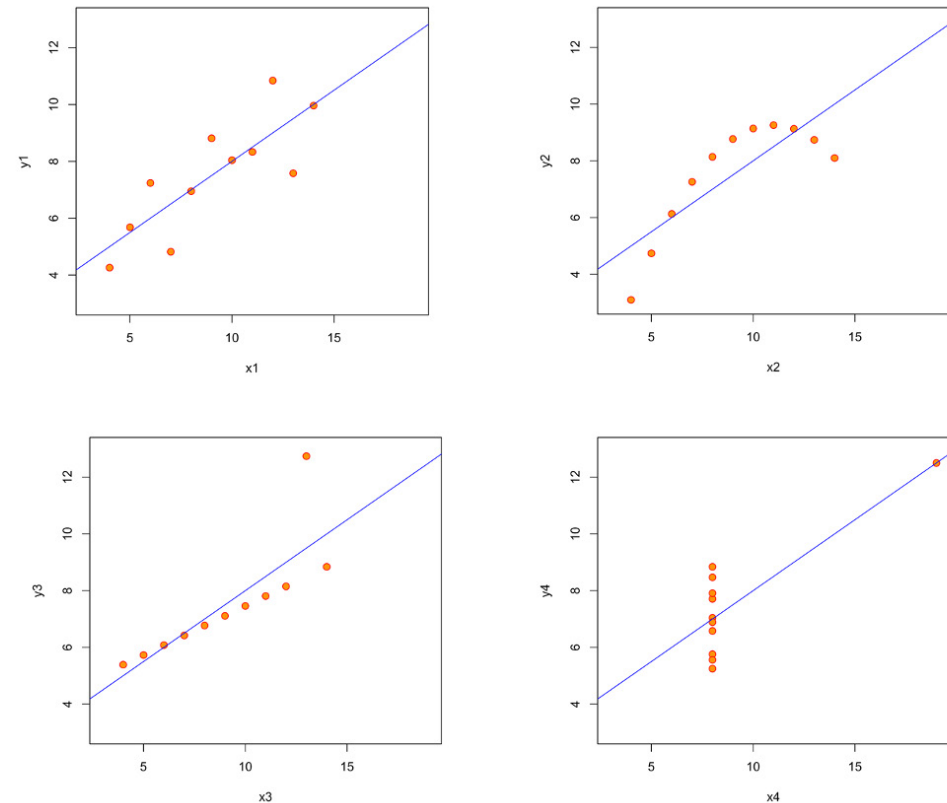
not overlap, so we have a strong indication that there is a true difference between the two samples. Here, the next step would be to then calculate an effect size for the difference between the two groups, and interpret the findings in terms of whatever is assumed to have led to the difference. However, if final confirmation of the statistical significance of the difference is required, a robust independent samples *t*-test, Yuen's modified *t*-test for independent means (Field & Wilcox, 2017) can be applied to the data. This is used due to the violation of the normality assumption in group b, and also the relatively small sample size. Yuen's test calculates the *t* statistic based on trimmed means. A trimmed mean is one where *x*% of the top and bottom scores are removed from the sample and the mean calculated based on this adjustment. This is a statistically-justified approach to dealing with outliers. Note that an alternative here would be to apply a non-parametric test, the Mann-Whitney *U* test, however in certain situations non-parametric tests are more susceptible to type II errors, and as such, robust tests should be generally preferred (Wilcox, 2017). As expected, based on the boxplots, the *t*-test gives a significant difference between the 20% trimmed means,  $M_{diff} = 5.42$ ,  $CI_{95\%} [3.78, 7.09]$ ,  $t_{yuen}(44.01) = 6.66$ ,  $p < .001$ .

When carrying out a statistical analysis such as a *t*-test, ANOVA, or a robust or nonparametric alternative, boxplots or one of the variations covered above should be plotted and included with confidence intervals and effect sizes alongside the calculated test statistic. This will provide a richer source of information about the differences in measures between your samples, to help make clear whatever conclusions you draw from the data.

### Scatterplots

Moving on to cases where one is interested in analyzing the relationships between groups rather than the difference between them, we can next look at scatterplots. This type of graph is used to show the relationship between variables, such as the correlation of data, and are appropriate for cases like regression analysis. Scatterplots are very useful in the exploratory stage of data analysis. An illustration here involves four datasets created by Anscombe (1973). Each set consists of 11 pairs, with identical means,  $\bar{x} = 9$ ,  $\bar{y} = 7.5$ . Carrying out a regression of the four sets (x1, y1; x2, y2; x3, y3; x4, y4) gives basically equivalent results,  $R^2 = .667$ ,  $F(1, 9) = 17.97$ ,  $p = .002$  with intercept = 3,  $b = 0.5$ ,  $t = 4.24$ ,  $p = .002$ . This all suggests that the data should be very similar across the four sets. However, plotting scatterplots of the four sets (Figure 10) shows the difference between them, and indicates that quite different conclusions can be made about the relationships in each set. The message should be clear – examine your data carefully to check it tells you what you think it does.

**Figure 10**  
*Anscombe's Scatterplots*

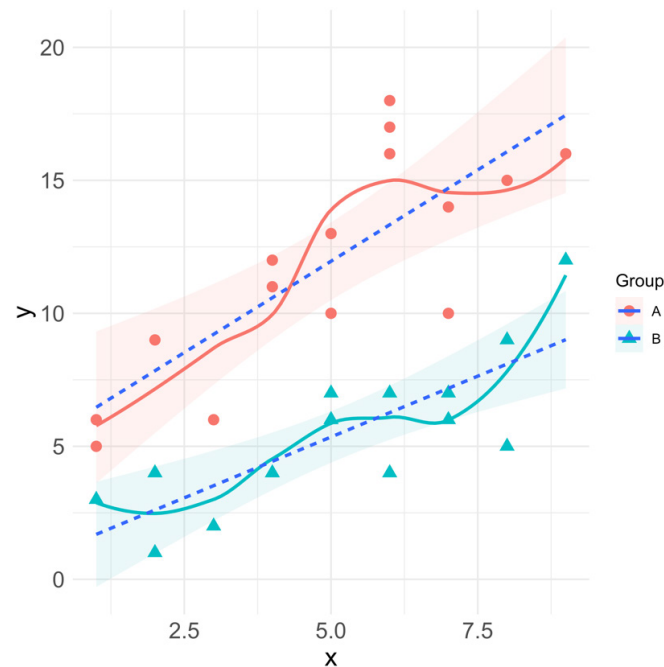


At their most basic, scatterplots show two dimensions of data. By adjusting some of the options in the R code, it is quite easy to add extra layers of information, for example increasing the size of data points or using color to clarify different groups. It is common to include a smoother curve in scatter plots. Often this is a plot of the regression curve which predicts the direction of the data based on an aggregate of the dataset, as in Figure 10 above. Larson-Hall and Herrington (2010) propose also including a loess line; this is basically a curve plotting the trends of the individual data points, rather than the

overall regression. The relationship between the loess line and regression curve reflects the assumption of linearity between variables, the closer the lines are, the better the likelihood of obtaining a meaningful correlation between variables.

Figure 11 shows a scatterplot for a third simulated dataset with two groups, A and B, each  $n = 30$ . The solid curved lines are the loess lines, while the dashed straight lines show the linear regression with a 95% confidence interval band. The trends are easily apparent here; using color and different shapes for the two different groups of data can make the patterns in the data clearer. When plotting grouped data like this, it may be better to plot each group in its own panel (as in Figure 10) to reduce clutter, overplotting and other overlap of the plot elements. As the online code should demonstrate, this is something that is handled quite easily in the R package used here.

**Figure 11**  
*Scatterplot with Loess and Regression Lines for Simulated Dataset 3*

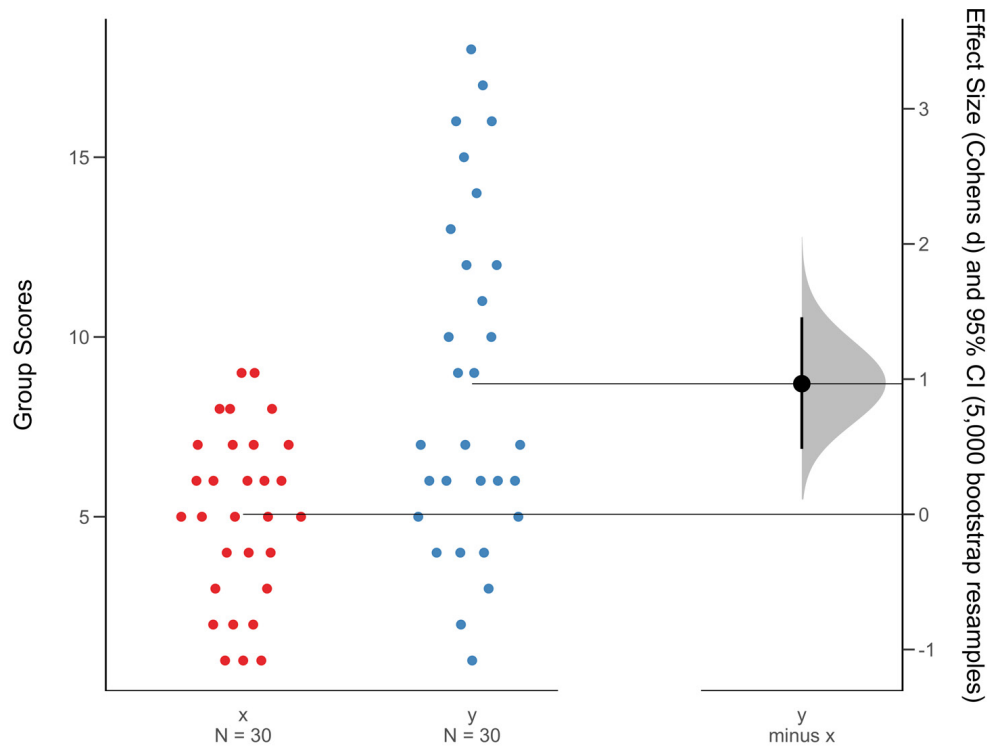


## Estimation Plots

Estimation plots are a relatively new kind of graph, which show the complete dataset as well as the effect size and confidence interval for the difference between groups. Gardner and Altman (1986) argued that presenting confidence intervals was more helpful than reporting  $p$ -values due to the limitations of the latter. This kind of graph includes all data points along with a confidence interval for the population difference between means. The mean difference is one possible measure of effect size (Cumming, 2012). Larson-Hall (2017) gives an example of such a graph, created using Cumming's (2012) ESCI Excel program. Ho et al. (2019) have since developed software to generate estimation plots in R using a bootstrapping technique which can be used for paired data or multiple groups. Bootstrapping is a robust statistical method where a dataset to be tested is randomly resampled hundreds or thousands of times to create a new theoretical sample which is then tested (see Field & Wilcox, 2017; LaFlair et al., 2015). The software has the option to generate different measures of effect size calculated from the bootstrapped sample. This provides a strong alternative to reporting  $p$ -values, which fits within a New Statistics framework (Calin-Jageman & Cumming, 2019; Cumming, 2012). An example of an estimation plot is provided in Figure 12, using the same dataset as for the scatterplot. The data is plotted on the left side of the graph, with the right side showing the effect size, indicated by the black dot, along with the CI for the effect size,  $d = 0.966$ ,  $CI_{95}[0.484, 1.46]$ . The horizontal lines indicate the means for the two datasets; 5.1 for group x, and 8.7 for y.



**Figure 12**  
*Estimation Plot for Simulated Dataset 3*

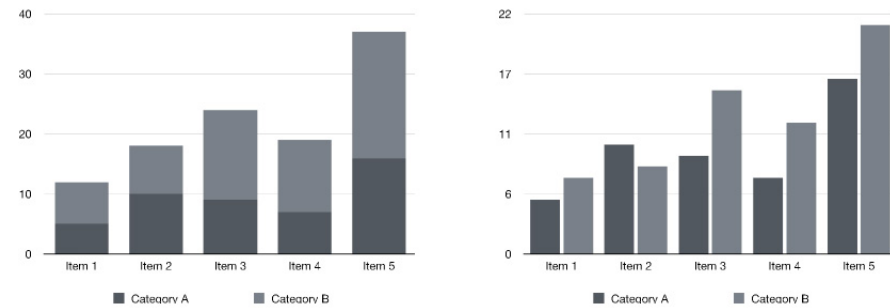


### Closing Considerations

As with any kind of data analytic technique, presenting data graphically can misrepresent findings. Healy (2019) and Hudson (2015) cover numerous issues here, including the following.

Bar graphs can be helpful if presenting categorical data, however it is best to avoid stacked bar graphs where group data is layered, in favor of grouped as the latter are easier to interpret. Examples are given in Figure 13.

**Figure 13**  
*Examples of a Stacked (left) and Grouped (right) Bar Graph*



Pie charts should generally be avoided when the number of categories exceeds four or five, as they are difficult to interpret in these situations.

One point to be careful of when presenting data in graphs is to avoid exaggerating or minimizing trends in the data. This can happen when choosing how to set your axis range, or in setting the plot dimensions to overemphasize the height or width. Trying different size settings when producing graphs can give a good idea of what is appropriate.

Plotting in color is a good technique for showing multiple dimensions of your data. If your intention is to share the resulting graph publicly, for example online or on a presentation slide, choosing a color-blind-friendly color palette is advisable. Such functionality is provided by the viridis package in R (Garnier et al, 2021). This can be used in conjunction with plotting packages such as ggplot2 to provide a suitable range of colors for plots which will also print well in grayscale.

It is important to ensure any graphs are suitably annotated with meaningful axis labels, clear scales, and legends. If including error bars, indicate if they represent CIs, standard errors, or standard deviations.

Plotting too much on one grid can distract from the relationships in your data. Using faceting to spread data across a lattice of plots is one easy way to work with complex data. Combining individual plots into compound figures is another possibility—look to the R cowplot package (Wilke, 2020) for one way to do this.

## Conclusion

As has been illustrated, graphical analysis of data is an important aspect of any quantitative research project. It serves as both an exploratory and explanatory process, identifying issues with data that may be problematic for the analysis; supplying insights into data trends; and summarizing outcomes in a clearly-understandable fashion. The R statistical framework provides numerous options to visualize data, and is strongly recommended as a tool for use in research.

Learning to use R to produce the types of plots outlined here is admittedly complex. However, if there is to be meaningful development in any discipline, complexity and challenge need to be embraced for substantive change to occur. It is hoped that the techniques and justifications presented here will encourage some to look at these alternative approaches to analyzing and reporting research, and to understand the benefits to be gained for professional development.

## Bio Data

**Paul Collett** teaches at Shimonoseki City University, Yamaguchi Prefecture. His interests include teacher and learner psychology, along with research methodology and epistemology. <collett@shimonoseki-cu.ac.jp>

## References

- Al-Hoorie, A. H., & Vitta, J. P. (2019). The seven sins of L2 research: A review of 30 journals' statistical quality and their CiteScore, SJR, SNIP, JCR Impact Factors. *Language Teaching Research*, 23(6), 727-744. <https://doi.org/10.1177/1362168818767191>
- Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1), 17-21. <https://doi.org/10.2307/2682899>
- Bulmer, M. G. (1979). *Principles of statistics*. Courier Corporation.
- Calin-Jageman, R. J., & Cumming, G. (2019). The new statistics for better science: Ask how much, how uncertain, and what else is known. *The American Statistician*, 73:sup1, 271-280. <https://doi.org/10.1080/00031305.2018.1518266>
- Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, 49(12), 997-1003. <https://doi.org/10.1037/0003-066X.49.12.997>
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.
- Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy*, 98, 19-38. <https://doi.org/10.1016/j.brat.2017.05.013>
- Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than p values: estimation rather than hypothesis testing. *British Medical Journal (Clinical research ed.)*, 292(6522), 746-750. <https://doi.org/10.1136/bmj.292.6522.746>
- Garnier, S., Ross, N., Rudis, R., Camargo, A. P., Sciaini, M., & Scherer, C. (2021). viridis - Colorblind-friendly color maps for R. <https://doi.org/10.5281/zenodo.4679423>
- Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 391-408). Sage.
- Haig, B. D. (2013). The philosophy of quantitative methods. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods volume 1: Foundations* (pp. 7-31). Oxford University Press.
- Haig, B. D. (2018). *The philosophy of quantitative methods*. Oxford University Press.
- Healy, K. (2019). *Data visualization*. Princeton University Press.
- Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond  $p$  values: data analysis with estimation graphics. *Nature Methods*, 16, 565-566. <https://doi.org/10.1038/s41592-019-0470-3>
- Hudson, T. (2015). Presenting quantitative data visually. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 78-105). Routledge/Taylor & Francis.
- LaFlair, G. T., Egbert, J., & Plonsky, L. (2015). A practical guide to bootstrapping descriptive statistics, correlations, t tests, and ANOVAs. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 46-77). Routledge/Taylor & Francis.
- Larson-Hall, J. (2017). Moving beyond the bar plot and the line graph to create informative and attractive graphics. *The Modern Language Journal*, 101, 244-270. <https://doi.org/10.1111/modl.12386>
- Larson-Hall, J., & Herrington, R. (2010). Improving data analysis in second language acquisition by utilizing modern developments in applied statistics. *Applied Linguistics*, 31(3), 368-390. <https://doi.org/10.1093/applin/amp038>
- Mizumoto, A., & Plonsky, L. (2016). R as a lingua franca: Advantages of using R for quantitative research in applied linguistics. *Applied Linguistics*, 37(2), 284-291. <https://doi.org/10.1093/applin/amv025>
- Norris, J.M., Plonsky, L., Ross, S.J., & Schoonen, R. (2015). Guidelines for reporting quantitative methods and results in primary research. *Language Learning*, 65, 470-476. <https://doi.org/10.1111/lang.12104>

- RStudio Team (2021). RStudio: Integrated development environment for R. <https://www.rstudio.com/>
- Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician*, 34(1), 23-25. <https://doi.org/10.1080/00031305.1980.10482706>
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on *p*-values: Context, process, and purpose. *The American Statistician*, 70(2), 129-133. <https://doi.org/10.1080/00031305.2016.1154108>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- Wickham, H., & Stryjewski, L. (2011). *40 years of boxplots*. <https://vita.had.co.nz/papers/boxplots.pdf>
- Wilcox, R. R. (2017). *Understanding and applying basic statistical methods using R*. Wiley.
- Wilke, C. O. (2020). cowplot: Streamlined plot theme and plot annotations for 'ggplot2'. <https://CRAN.R-project.org/package=cowplot>
- Wilkinson, L., & Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604. <https://doi.org/10.1037/0003-066X.54.8.594>