

Have Psychologists Lost The Plot?
Analyzing the Quantity and Quality of Data Graphics in Psychology Papers
PSYC 6135
Gabriel Crone

Graphical displays abound in psychology, but it has not always been that way. Before the advent of modern computing (which would begin after 1950 and see maturity in the 1960s; Friendly, 2008), data visualizations were done by hand. The process—even for the simplest of plots—was labour-intensive, difficult, and expensive. To create a simple frequency polygon, for example, one needed to follow many steps: carefully consider the plot’s purpose; choose a plot type that would meet the purpose in mind; plan, design, and draft what the data visualization would look like; and construct it (which was 12 steps in itself; Schmid, 1954).¹ If one was not satisfied with their original plot, they would need to start the process over again, spending more of their finite resources. Given how difficult it was to construct plots, their usage in psychological literature was likely far less common than it is now.

In the modern computing age, plotting data is inexpensive, simple, and instant. With a computer, some data, and graphical software, one can create (nearly) any plot they desire.²

Given the explosion of new and accessible graphing technology, an important question arises: how has this technology affected the quality of plots in research? One might guess that it has improved, because technology has allowed researchers to “try out” an array of plots and select with the ones that work best. Still, others may argue that plot quality was better in the past, since researchers had to decide upon what their data graphic would look like carefully (because mistakes were more costly).

Before addressing the quality of plots in the literature, we must first ask ourselves: what does it mean for a data visualization to be effective? What differentiates a “poor” graphic from an “excellent” one?

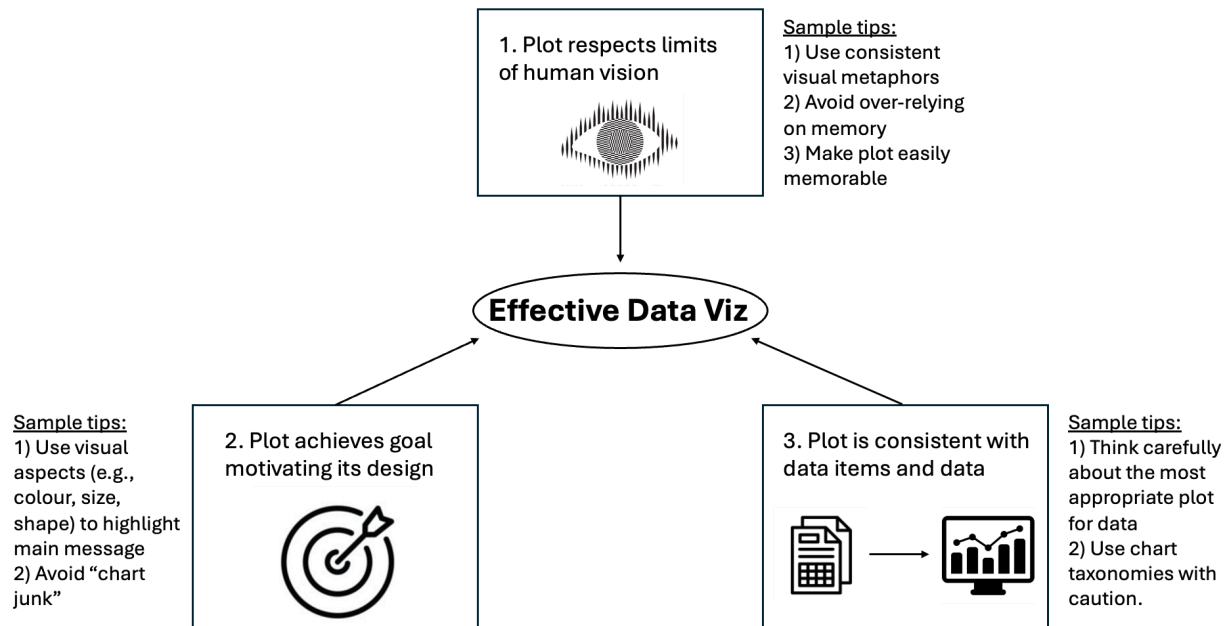
What makes an effective data visualization?

There is no simple answer to this question, because there is no universally accepted definition or framework to evaluate “effective” visualizations (Zhu, 2007). However, I would argue that the answer would ultimately depend on how one defined “effectiveness”. Across the extensive literature, there common, useful definitions have emerged. See Figure 1 for a graphical summary.

¹ Draftsmen were often commissioned to construct charts, but one needed to know how to draft in order to show the draftsmen what to draw (Schmid, 1954, Ch. 2).

² The only caveat here is that using point-and-click software inherently limits the possible chart types to those provided by the software designers. Syntax-based softwares such as R or python are limited in this regard, though inherently more flexible because extensions to the base software (such as libraries/packages) are regularly uploaded.

Figure 1
Three Definitions of Effective Data Visualization with Tips



Definition 1: Effective = Adherence to human visual system

The first definition of effectiveness is based on adherence to the human visual system: *Effective data graphics respect the human visual system*. Respecting the visual system entails designing plots in a way that takes into account the limitations of human visual perception so that they are not an issue when plots are viewed. Kosslyn (1985) provides an excellent overview of the process of visual perception and ways to respect its limitations. He describes visual perception as occurring in three phases: perceiving the plot (grouping visual elements together), short-term memory (recalling details of the plot), and long-term memory (understanding and taking meaning from the plot). Within each phase, Kosslyn (1985) provides examples of properties of a plot that respect it. Some examples include making different visual elements easily distinguishable, making the elements of the plot clear, respecting short-term memory limitations, and designing plots with an easily discoverable, clear, and unambiguous message.

Similarly, Franconeri et al. (2021) describe principles of designing plots so that they result in viewers accurately perceiving them. For example, based on previous human factors research, there exist particular visual channels that are more perceptually accurate than others (e.g., *position* is more effective than *length*; for a review, for a detailed overview, see Munzner, 2014). When using multiple channels, the viewer will take far more notice of more perceptually accurate visual channels than less accurate ones. Thus, designers should ensure that the most perceptually accurate visual channel shown illustrates the key purpose of the plot.

In all, when plots are designed to respect the human visual system, they are far easier to view, understand, and interpret. Interestingly, in Wainer's (1984) famous paper on the "dirty dozen" of data visualization, many of his do-notes (rules 3, 4, 6, 8, and 10) are based on violating the human visual system.³ However, this definition of effectiveness often excludes the intention

³ These rules Wainer (1984) mentioned were: 3 = ignore visual metaphor altogether, 4 = only order matters (e.g., distorting scales), 6 = change scales in mid-axis, 8 = jiggle the baseline, and 10 = label illegibly.

behind the plot's design. For instance, it is possible for plots of high perceptual clarity to lack substantive meaning (e.g., if axes are inappropriately labeled).

Definition 2: Effective = Plot Achieves the Goal Motivating its Design

The second definition of effectiveness has to do with the plot's purpose: *A plot is effective if it achieves the goal that originally motivated its design.* This definition respects the fact that plots have several distinct purposes, which vary by study and context. Usually, however, plots in research studies tend to have one of two (highly related) goals: a) give a richer sense of the underlying data in the study, or b) communicate one's results. Within these broader goals exist smaller goals, such as demonstrating the relationship between two variables or choosing an appropriate way to depict uncertainty in the data. One rather obvious way to meet the goal of a data visualization is to choose the type of visualization, and the appropriate visual metaphor, that best communicates what the designer of the plot wishes to communicate.

According to Wainer (1984), "the aim of good data graphics is to display data accurately and clearly" (p. 137). Thus, effective plots should depict more information than less and avoid needless window dressing (i.e., "chartjunk"; Tufte, 1983) wherever possible. Such a goal aligns most closely with the first purpose of a plot, since it gives a richer sense of the data at hand, as well as the second, since it helps to communicate one's results better.

More recently, Kelluher & Wagener (2011) discussed 10 guidelines to creating effective data graphics based upon classic books on the topic. Many of their guidelines were highly related to the goal of communicating one's own results. Their first rule emphasized this goal nicely: "Create the simplest graph that conveys the information you want to convey" (p. 822). They elaborated that plots should avoid redundancy or excessive detail, since it distracts the viewer from the plot's main intention. Software innovations complicate these matters, since they equip the user with ways to needlessly dress-up their plots in three dimensions—using two dimensions is (almost) always the better choice.⁴ Other recommendations they make also reinforce this idea; for example, guideline #4 says to select meaningful axis ranges, and #8 suggests selecting a plot that aligns well with different datasets in a meaningful way.

All in all, a data graphic is designed with a key purpose in mind. If the data graphic easily meets said key purpose, then it is said to be effective. There is one more definition of effectiveness that is common in the literature.

Definition 3: Effective = Graphic's visual elements coincide with the data items and the data

The third criteria has to do with choosing an appropriate graphic: *A plot is effective if its elements are consistent with the data items and the data itself.* This entails choosing the most effective graphic possible to depict the underlying data, and ensuring that the visual metaphors used effectively reflect the data items and dataset. For example, if one wants to show time-series data, often using a line plot is an effective means of displaying it, since the user could view how particular trends change over time (Kelleher & Wagener, 2011). Similarly, if one wanted to show the underlying distribution of a univariate distribution, there are many appropriate plots for the job, such as histograms, stem-and-leaf plots (if the sample size is small enough), and density plots. Indeed, there now exist several chart taxonomies which guide novices on deciding an

⁴ The only exception is if one wishes to code information in the third dimension, which can be effective when data is inherently three-dimensional (e.g., visualizing a multiple linear regression model as a plane). Still, in most cases, especially for simpler data graphics (e.g., bar charts, histograms), 3D introduces needless ambiguity and confusion.

appropriate means of visualizing their data, given factors such as the type of data, type of chart, and purpose of the visualization (e.g., see Evergreen, 2020; Ferdio, 2021; Holtz & Healy, 2018).

The third definition is also reflected in common suggestions to improve data graphics. For example, Kelluher & Wagener (2011) recommend considering the coding object and the attribute to create the plot (guideline #2) and to visualize data *patterns* or *details* depending on plot type (guideline #3). Wainer (1984) suggests that poor plots tend to ignore or misapply visual metaphors, and that poor plots tend to needlessly “reinvent the wheel” when effective plots exist. Both of these indicate that effective plots are those who use effective visual metaphors and who are an appropriate type for the underlying data.

It is important to emphasize that all three of the definitions for effective plots are *not* independent. Often, when a plot is designed according to one definition of effectiveness, it often indirectly satisfies the other definitions. For example, if a plot respects the limitations of the human visual system (definition one), it tends to be a highly appropriate visualization with respect to the data (definition three). Moreover, if the goal of the plots is to effectively communicate one’s underlying data, then meeting definition three necessitates meeting definition two (since the plot has achieved the goal underlying its creation). In this way, excellent plots can be considered those which meet all of the definitions of effective data visualization.

How do researchers code for the quality of graphics?

Each plot is uniquely different, and has a multitude of different components one can consider (e.g., axis labels, points, shapes, symbols, colors, font, scaling, transformations, shading, etc.). So, how does one quantify the quality of statistical graphics? A few researchers have attempted to create a system in which the quality of statistical graphics can be measured.

Kosslyn (1989) created a unique system in which any plot can be assessed for strengths and weaknesses. The system works by considering a plot on three levels of analysis (labels, framework, and specifiers) and across three levels of basic constituents (syntax, semantics, and pragmatics). In considering the plot, “accessibility principles”—rules underlying effective plot displays⁵—are assessed; whenever these principles are violated, the plot is no longer effective. Such a system is incredibly flexible (it can be used for practically any plot type), but is very rigorous and intensive, and lacks user-friendly options to learn it.

Zhu (2007) defined several ways to measure an effective data visualization. Based on their detailed survey of the literature, they arrived at three general principles of effective graphics: Efficiency, Utility, and Accuracy. Each one nicely parallels the aforementioned definitions for effectiveness. Efficiency entails plots reducing the cognitive load relative to a non-graphical alternative (definition #1), Utility entails plots meeting the goal of specific tasks (definition #2), and Accuracy entails plots having visual elements that are consistent with the underlying data and data structure (definition #3). With these principles, Zhu (2007) provided several ways to quantify plot quality, but most would require collecting data from participants (e.g., task completion time, eye movement, learning curve measurements, interviews). They, however, did not supply the reader with any concrete codified means of assessing plots from the literature, so their principles, while well-supported, lack accessibility.

Beyond the work on systems of assessing plots, others have quantified it by compiling lists of principles discussed through different studies and books and codifying them. For example, Gordon and Fitch (2015) drew primarily on advice from Cleveland (1984) and Tufte

⁵ These are based on the human visual system and how it processes information (similar to the first definition of effectiveness) and Goodman’s theory of symbols (see Goodman, 1968)

(1983) and created five rules of graphical excellence: *Show the data clearly* (p. 1211), *Use simplicity in design* (p. 1214), *use good alignment to a common scale* (p. 1214), *Keep the visual encoding transparent* (p. 1215), and *Use graphical forms consistent with principles 1 to 4* (p. 1216). From these principles, they created a meticulous list of 60 plot features!⁶ Such a list, while comprehensive, is unwieldy, especially if researchers would be interested in extending or replicating their results (which might explain why so few studies have done so).

Drawing recommendations from several sources (e.g., Gordon & Fitch, 2015; Wainer, 1984; Kelleher & Wagener, 2011), Astle (2023) distilled quality criteria into four discrete categories: *Labeling*, *Clear Understanding*, *Meaningful* (graphical elements add meaning to the information displayed), and *Scaling and Gridlines* (graph has adequate scaling and gridlines; Astle, 2023, p. 35). In all, there were a far more manageable number of items (24 compared to 60 from Gordon & Fitch, 2015) and they were all relatively easier to code, but they lacked a comprehensive overview of plotting principles.

This general approach had the benefit of covering several different aspects of plots, all which allude to several definitions of effective data graphics. However, it unfortunately had a major issue: there is a lack of a standardized metric across the literature, meaning that it is impossible for two research teams to assess the quality of data graphics in a consistent and replicable way. Regardless, several studies have examined the quality of plots in research papers.

Reviewing the Quality of Graphics in Research

What does the literature say when it comes to the quality of data graphics? Only two studies have specifically examined the plot quality within empirical research articles (Cleveland, 1984; Gordon & Fitch, 2015), and one study (a dissertation) has examined plot quality within Master's thesis reports (Astle, 2023).

Cleveland (1984) examined 377 graphs from a volume of *Science*, one of the most prestigious journals of all time. He focused on identifying four types of errors: construction (improper construction of plot), degraded image (quality of plot rendered is poor), explanation (plot lacks clear explanation of what it is depicting), and discrimination (difficult to discriminate visual elements of the plot). He found that 30% of the graphs had at least one error—which is startlingly high for a journal of such a high caliber.

Gordon and Fitch (2015) reviewed the quality of plots from several high-quality, prestigious journals across various (hard) scientific and social-scientific disciplines. In all, they sampled 97 graphs, and used a detailed, comprehensive rating system (for details, see *How do researchers code for the quality of graphics?*) to assess the quality of each plot. In all, ~39% of all plots reviewed were considered to have poor quality, and no plots were considered to have excellent quality.

The fact that the quality of plots was found to be poor in top-tier scientific journals is unsettling. If plot quality is so poor in top-tier journals, then less-prestigious journals might contain poorer data graphics. Overall, this indicates that the quality of plots in research papers is in need of improvement.

Astle (2023) examined the quality of data graphics within Master's thesis reports. Sampling from a broad range of reports (from 1930-2019), they examined the quality of 90 graphs using 4 graphical quality principles and a unique scoring system (for details, see *How do researchers code for the quality of graphics?*). Their primary goal was to examine general plot

⁶ Each plot feature was a binary item. More impressively, they coded these 60 features across 47 graphs, for a total of 60 x 47 (2820) individual assessments!

quality and determine if it changed over time. Across their analyses, they found that the quality of graphics in thesis reports were quite stable over time, but there was high variability with regard to the subscales of plot quality. Overall, the quality criteria ranged from 64% to 74% (out of 100%), indicating that most graphics had decent—but not excellent—quality.

In all, the (limited) research on graphical quality suggests that in both research papers and thesis reports, the quality of statistical graphics is wanting.

Reviewing the Quantity of Data Graphics

A related, equally important issue of data graphics is the *quantity* of particular graphics. Certain graphics, such as tables and bar charts, tend to be considered generally ineffective methods of summarizing one's data. If such less effective graphics are more frequently used, then it implies poorer communication of results.

A Tabular Tangent: Graphs vs. Tables

In any research article, researchers often have the choice of depicting their data with a table or with a graph. But which tends to be a more effective means of communicating key insights? The consensus is that plots tend to be more effective than tables (Feinberg & Wainer, 2011; Friendly & Kwan, 2011; Gelman, 2011). For example, while tables can show precise numerical values, it is difficult to extract key information from them (it is akin to extracting “sunbeams from cucumbers”; Feinberg & Wainer, 2011). There still exist some headstrong statisticians who insist that tables are superior (see Gelman, 2011, for a tongue-in-cheek satirical piece on these arguments), but most would agree that graphics are relatively better-suited for depicting statistical results.

Given this view, one would predict that the quantity of figures is greater than that of plots; however, they would be incorrect. Feinberg and Wainer (2011) surveyed the frequency of various displays (tables included) within the *Journal of Computational and Graphical Statistics* between 2005-2010. They observed that the most frequently displayed format (at 75% frequency) was a table. This was problematic because most tables were designed in a manner that made deriving crucial information from them difficult. Researchers have suggested ways to improve tabular displays, namely by making them more graphical, such as by emphasizing critical information with the use of size and bolding (Feinberg and Wainer, 2011) or using alternative tabular displays such as semi-graphics and tableplots (Friendly and Kwan, 2011).

Elevating Existing Plots

Particular types of plots, such as bar plots, tend to be simple yet uninformative means of summarizing data and communicating results. These plots are indeed quite prevalent: Weissberger et al. (2015) reviewed 703 research articles from high-impact psychology journals, finding that 85.6% of articles included at least one bar graph, and 61.3% of articles included figures depicting simple statistical results. According to the authors, “most figures provided little more information than a table” (p. 3).

This kind of finding indicates cause for concern regarding which plots are prevalently used. If, for example, the plots predominantly used in research settings tend to be uninformative, then their prevalent use could be hindering the crucial communication from empirical research. What does the literature say regarding the frequency of plots used in research?

Few studies have examined the frequency of particular plots in depth. Of the studies that have investigated it, they have generally found that the most frequently used plots have tended to

be bar plots and line graphs (Feinberg & Wainer, 2011; Weissberger, 2015), and that errors tended to frequently occur when bar charts were used (Nguyen et al., 2021). In her review of Masters theses, Astle (2023) found that line graphs were most frequently used. Lastly, Lane and Sándor (2009) investigated the use of graphs in psychology, finding that the types of charts typically selected rarely depicted distributional information beyond central tendency. In all, the literature indicates that the frequency of novel plots is limited, since it tends to be saturated with bar plots, line plots, and other simpler plots. This is not to say that simpler is not better; after all, a simpler plot that effectively captures the purpose of its design is ideal. Rather, what it indicates is that the choice of plots tends to be overly simplistic, to the point where key insights and nuances about the data are likely being missed.

Research Limitations

In all, the present research suggests that statistical graphics from a variety of different outlets is poor. The plots themselves were generally rated as being poor, and the types of plots found tended to be overly simplistic to adequately communicate the data and results.

It is important to note that this literature has many limitations. The most noticeable limitation is the lack of focus on psychology articles. Of all the present studies reviewed, only two focused exclusively on psychology (Lane & Sándor, 2009; Weissberger et al., 2015). While Weissberger et al. (2015) examined the quantity of plots in psychology (but not tables), Lane & Sándor did not directly examine quantity or quality. Gordon & Fitch's (2015) review included 4 psychology journals, but these were not analyzed separately from the aggregate sample, so conclusions regarding the quality of plots in psychology papers remains unknown.

An additional limitation was the lack of standardized measures to assess plot quality. All of the studies examining plot quality have used disparate systems that are based on different (albeit thematically similar) criteria, but no one has proposed a standardized coding metric that would flexibly allow any type of plot to be analyzed. The closest to proposing a standardized evaluation metric was Kosslyn (1989); however, Kosslyn's system lacked actionable, clear, and concrete guidelines to follow in order to evaluate the quality of any plot. Similarly, Zhu (2007) proposed a three-principle system for evaluating plot quantity and quality, but her focus was not on information visualization, and no definitive guidelines were provided. Of the two studies that used a coding system, one system was comprehensive but unwieldy (60 coded features per graph; Gordon & Fitch, 2015), while the other was more manageable but lacked comprehensiveness (24 criteria; Astle, 2023). A standardized, manageable number of coded items is wanting to enhance the literature further and to encourage others to verify the veracity of the discussed findings in other disciplines and contexts—especially within psychology.

A final limitation with the present literature is a lack of analysis of concurrent plot quality and quantity: most studies looked at plot quantity *or* quality, but never both. This is unfortunate, since plot quality and quantity are so intimately linked. For example, if effective plot types are plentiful, then it is likely (but not guaranteed) that they will be of higher quality, and vice versa. By only focusing on one, the researcher misses key information that could help provide a far clearer picture of the state of plots in research contexts far above that which has been discussed.

Present Research

In lieu of current limitations in the research literature, the purpose of the present research is to acquire a rich understanding of the state of data visualizations in psychology, both in terms of quality and quantity. Specifically, the current research would comprise three smaller studies.

Study one will involve creating and pilot-testing a new coding metric that is simple, intuitive, and comprehensive. Study two will randomly sample 100 psychology articles from a diverse selection of high-impact papers, and examine the quantity of plots and tables across the sampled articles. Study 3 will apply the metric from study 1 and the data from study 2 to evaluate the plot quality from across 100 psychology studies.

Study 1- Designing a Coding Rubric

The purpose of study 1 is to address the lack of a standardized tool to assess plot quality by creating a new one. This tool would need to be comprehensive enough so it captures rich data, while also not being too unwieldy. It also must be very intuitive, such that minimal training would be required for other researchers to adopt it for themselves. The new measure would consist of approximately 30 items.⁷ Earlier, we described the three definitions of effectiveness (echoed within Zhu's [2009] guidelines; see *What makes an effective data visualization?*), which entail three major, interrelated ways that researchers have defined plots' effectiveness. For each of these definitions, 10 items will be generated. To establish content validity for the measure, a sample of five data visualization experts will be surveyed and shown the items, and asked to provide feedback on how strong the items are at capturing the effectiveness of data graphics. Subsequently, the items will be revised until a set of finalized items will be found.

To ensure the measure has proper reliability, once the rubric is finalized, it will be pilot-tested on a sample of 10 randomly selected studies from the top-tier psychology journal, *American Psychologist*.⁸ 3 coders will be trained on using the criteria, and the Interrater reliability will be computed as the consistency between the raters' responses. After each round, the scale will be re-adjusted and 10 new random articles will be selected.⁹ This process will be repeated until sufficient inter-rater reliability is reached (>85% across all raters).

Study 2- Determining the Quantity of Plots in Psychology

The purpose of study 2 will be to conduct a random, representative sample of plots across various high-impact psychology journals and sub-disciplines. In doing so, we aim to replicate Feinberg & Wainer (2011) on a larger scale.

A total of 10 high-impact psychology journals will be sampled from, each specializing in a unique sub-discipline. All of these journals, as well as their corresponding sub-disciplines, are listed in Table 1. As the table shows, these APA journals will cover a wide range of psychological sub-disciplines and have moderate (2-4) or large (≥ 4) journal impact factors.

For each journal, 10 articles will be randomly selected from a recent (2022) volume. To do so, we will first download all 2022 volumes of each journal, filter those articles that are not empirical studies (e.g., review papers, comments, editorials), and assign a value to each study from $0-N_j$ (where N_j is the total number of empirical papers for the J th journal). Then, we will use a random number generator to select 10 random values between 0 and N_j for each journal. The articles that match the randomly selected values will be chosen for analysis.

Then, two coders will skim through each selected article and record every figure observed into its own row, along with data on the type of plot depicted. If a figure depicts

⁷ 30 was a good compromise between Gordon & Fitch's 60 comprehensive yet overly detailed criteria, and Astle's 24 manageable yet non-detailed criteria.

⁸ This journal was selected for its breadth of coverage and affiliation with the APA (since APA journals will be sampled from exclusively for Studies 2 and 3).

⁹ A natural follow-up study to study 1 would be to assess the quality of tables found, since no research to date has done so (likely because most tables are quite similar and do not follow Feinberg & Wainer's (2012) suggestions).

multiple plots, e.g., has multiple panels, each panel will be considered a unique plot (e.g., Figure 1a, 2c, etc...). Plots that do not depict data (e.g., flow-charts, illustrations of concepts being described, photographs of laboratory setups) will not be included. After all figures are recorded, coders will also count the number of tables included in each paper (excluding tables in appendices and supplemental materials), in order to get an estimate for the number of tables. A sample spreadsheet and codebook (with three filled-in examples) is available on our OSF page (linked [here](#)).

Table 1

List of Psychology journals, their sub-discipline, and their impact factors.

Journal	Sub-discipline	*2022 Journal Impact Factor
<i>Journal of Personality and Social Psychology</i>	Personality and Social Psychology	7.6
<i>Psychological Methods</i>	Quantitative Psychology	7.0
<i>Journal of Consulting and Clinical Psychology</i>	Clinical Psychology	5.9
<i>Journal of Educational Psychology</i>	Educational Psychology	4.9
<i>Journal of Applied Research in Memory and Cognition</i>	Cognitive Psychology	4.2
<i>Health Psychology</i>	Health Psychology	4.2
<i>Journal of Experimental Psychology: General</i>	Experimental Psychology	4.1
<i>Developmental Psychology</i>	Developmental Psychology	4.0
<i>Behavioural Neuroscience</i>	Neuroscience	2.5
<i>Neuropsychology</i>	Neuropsychology	2.4

*Note. Impact factors are from <https://www.apa.org/pubs/journals/resources/impact-factors>. They are organized in descending order.

Once all data are collected, coders will reconcile any differences by discussing them and coming to a consensus. Once that is done, we will aggregate through the data to generate frequency counts of the frequency of particular plots, as well as tables, and depict this information in a plot (similar to Figures 1-4 in Feinberg & Wainer, 2011), and differentiate results based on the journal.

Study 3- Determining the Quality of Plots in Psychology

Study 3 will draw upon the scoring metric from study 1 and the data from study 2. Using the data from study 2, all rows with tables will be removed, leaving only rows describing plots. Subsequently, 30 new columns will be added, one for each of the scoring criteria from the rubric. Using this modified spreadsheet, 4 coders will review plots from 25 studies each (approximately) until all 100 are coded. Once data are available, any inconsistencies will be resolved, and all of the aggregated scores (computed as percentages of items that were met) will be plotted as a whole and by journal.

Implications

This research, if complete, would provide other researchers with an invaluable tool that they could easily adopt for their own review purposes, and to generate new insights as to the quantity and quality of data visualizations within other fields. Additionally, if the tool is adopted widely, the results from various studies would be far more comparable, meaning that research literature could be compared in terms of the quantity and quality of the plots used. The rich data from this research alone would fill in a substantial gap in the research literature, since it will be the first to examine both quantity and quality of plots in high-impact, psychological research papers. The results will be invaluable to diagnosing potentially prevalent issues within plots used in psychology, and could help provide practical and actionable guidelines for researchers to follow to improve the types of plots selected and the quality of the plots used. In all, we hope this research inspires others to take notice of the plots they observe around them, and for new ways to improve the research literature for the better.

References

- Astle, R. (2023). *Statistical graph quality analysis of Utah State University Master of Science thesis reports*. [Master's Thesis, Utah State University].
<https://digitalcommons.usu.edu/etd/8815>
- Cleveland, W. S. (1984). Graphs in scientific publications. *The American Statistician*, 38(4), 261–269.
- Evergreen, S. D. H. (2020). *Effective data visualization: the right chart for the right data* (2nd ed.). SAGE Publications, Inc.
- Feinberg, R. A., & Wainer, H. (2012). Extracting sunbeams from cucumbers. *Journal of Computational and Graphical Statistics*, 20(4), 793–810.
<https://doi.org/10.1198/jcgs.2011.204a>
- Ferdio. (2022). *Data Viz Project*. <https://datavizproject.com/>
- Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. *Psychological Science in the Public Interest*, 23(3), 110–161. <https://doi.org/10.1177/15291006211051956>
- Friendly, M. (2008). A brief history of data visualization. In C. Chen, W. Härdle, & A. Unwin (Eds.), *Handbook of Data Visualization*. (pp. 16–48) Sprinkler-Verlag, Berlin Heidelberg.
<https://doi.org/10.1007/978-3-540-33037-0>
- Friendly, M., & Kwan, E. (2012). Comment. *Journal of Computational and Graphical Statistics*, 20(1), 18–27. <https://doi.org/10.1198/jcgs.2011.09166b>
- Gelman, A. (2011). Why tables are really much better than graphs. *Journal of Computational and Graphical Statistics*, 20(1), <https://doi.org/https://doi.org/10.1198/jcgs.2011.09166>
- Goodman, N. (1976). *Languages of art: An approach to a theory of symbols*. (2nd ed.). Hackett.
- Gordon, I., & Finch, S. (2015). Statistician heal thyself: Have we lost the plot? *Journal of Computational and Graphical Statistics*, 24(4), 1210–1229.
<https://doi.org/10.1080/10618600.2014.989324>
- Holtz, Y., & Healy, C. (2018). *From Data to Viz*. <https://www.data-to-viz.com/>
- Kelluher, C., & Wagener, T. (2011). Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software*, 26, 822–827.
<https://doi.org/10.1016/j.envsoft.2010.12.006>
- Kosslyn, S. M. (1985). Graphics and human information processing: A review of five books. *Journal of the American Statistical Association*, 80(391), 499–512.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185–226.
- Munzner, T. (2014). *Visualization Analysis and Design*. AK Peters Visualization Series, Routledge and CRC Press.
- Nguyen, V. T., Jung, K., & Gupta, V. (2021). Examining data visualization pitfalls in scientific publications. *Visual Computing for Industry, Biomedicine, and Art*, 4(27), 1–15.
<https://doi.org/10.1186/s42492-021-00092-y>
- Schmid, C. F. (1954). *Handbook of Graphic Presentation*. The Ronald Press Company, New York. <https://archive.org/details/HandbookOfGraphicPresentation>
- Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut.
- Wainer, H. (1984). How to display data badly. *American Statistician*, 38(2), 137–147.
- Zhu, Y. (2007). Measuring effective data visualization. *Advances in Visual Computing*, 652–661.
https://doi.org/10.1007/978-3-540-76856-2_64