Circulation

PRIMER

Reveal, Don't Conceal

Transforming Data Visualization to Improve Transparency

ABSTRACT: Reports highlighting the problems with the standard practice of using bar graphs to show continuous data have prompted many journals to adopt new visualization policies. These policies encourage authors to avoid bar graphs and use graphics that show the data distribution; however, they provide little guidance on how to effectively display data. We conducted a systematic review of studies published in top peripheral vascular disease journals to determine what types of figures are used, and to assess the prevalence of suboptimal data visualization practices. Among papers with data figures, 47.7% of papers used bar graphs to present continuous data. This primer provides a detailed overview of strategies for addressing this issue by (1) outlining strategies for selecting the correct type of figure depending on the study design, sample size, and the type of variable; (2) examining techniques for making effective dot plots, box plots, and violin plots; and (3) illustrating how to avoid sending mixed messages by aligning the figure structure with the study design and statistical analysis. We also present solutions to other common problems identified in the systematic review. Resources include a list of free tools and templates that authors can use to create more informative figures and an online simulator that illustrates why summary statistics are meaningful only when there are enough data to summarize. Last, we consider steps that investigators can take to improve figures in the scientific literature.

Tracey L. Weissgerber, PhD
Stacey J. Winham, PhD
Ethan P. Heinzen
Jelena S. Milin-Lazovic, MD
Oscar Garcia-Valencia, MD
Zoran Bukumiric, MD, PhD
Marko D. Savic, MPH
Vesna D. Garovic, MD, PhD
Natasa M. Milic, MD, PhD

Key Words: bar graphs ■ basic science ■ continuous data ■ data visualization

https://www.ahajournals.org/journal/circ

Informative figures that allow readers to critically evaluate data are essential; however, a systematic review demonstrated that the figures commonly used in basic biomedical science obscure the data.[1] Figures highlight the most important findings in publications, posters, and talks. The data underlying published studies typically are not available[2] or reusable.[3] Success rates for requesting data from authors decline precipitously with time since publication.[4] Until open science becomes more widespread, our knowledge of the data underlying study conclusions will be limited by what the authors choose to show.

Many journals and publishers have recently introduced policies that encourage or require authors to use more informative figures that show the data points or distribution, including *PLOS Biology*,[5] *eLife*,[6] *Journal of Biological Chemistry*,[7] and Nature Publishing Group.[8] We conducted a systematic review of studies published in top peripheral vascular disease journals to determine what types of figures are commonly used to present data and to assess the prevalence of suboptimal data visualization practices. The results show that the inappropriate use of bar graphs to present continuous data is a major problem in peripheral vascular disease journals. This primer provides a detailed overview of strategies for addressing this problem by (1) outlining techniques for selecting the correct type of figure depending on the study design, sample size, and type of outcome variable; (2) examining techniques for creating effective dot plots, box plots, and violin plots; and (3) showing how to avoid sending mixed messages by aligning the structure of the figure with the study design and goals of the analysis. We also present solutions to other common visualization problems identified in the systematic review. Last, we provide links to free tools and templates that make it easy for authors to prepare informative figures.

Visualization Problems in Peripheral Vascular Disease Journals

The systematic review included all original research articles published in September 2018 in the top 25% of peripheral vascular disease journals, ranked according to 2017 impact factor (n=206 papers from 13 journals, Figure I and Table I in the online-only Data Supplement). Detailed methods and results are presented in online-only Data Supplement. The protocol and data were deposited in a public repository.[9] We found that 180 papers (87.4%) included a data figure. Table 1[10-14] provides a brief overview of findings and presents steps that authors can take to correct common visualization problems.

The inappropriate use of bar graphs to display continuous data was the most common visualization problem

Table 1. Correcting Common Visualization Problems in Peripheral Vascular Disease Journals

Solution	Rationale
Replace bar graphs with figures that show the data distribution	Of papers reviewed with data figures, 47.7% used bar graphs to present continuous data, typically for small datasets. The median sample size for the smallest and largest groups shown in a bar graph were 3 (interquartile range, 3–5) and 10 (interquartile range, 7–17), respectively. This suggests that most bar graphs should be replaced with dot plots (Figures 1 and 2).
Consider adding dots to box plots	The median sample size for the smallest and largest groups shown in a box plot were 20 (interquartile range, 11.5–56) and 66 (interquartile range, 17.5–302.5), respectively. There are no clear guidelines for the sample size at which a dataset becomes large enough to use box plots; however, these data show that box plots are sometimes used for small samples. In these cases, authors should combine the box plot with a dot plot and emphasize the box plot, as shown in Figure 4.
Change journal policies	Only 3 of the 13 peripheral vascular disease journals examined had policies requesting that authors replace bar graphs with more transparent figures that show the data points or distribution.
Use symmetric jittering in dot plots to make all data points visible	Among papers with dot plots (n=39), 46.2% included overlapping data points that were not clearly visible. Figure 3 illustrates strategies for making all points visible.
Use semi-transparency or show gradients to make overlapping points visible in scatter plots and flow-cytometry figures	Scatter plots in 89.2% of papers (33 of 37) had overlapping points; however, only 12.1% of papers (4 of 33) used techniques such as semitransparency, shaded color gradients, or gradient lines to make overlapping points visible. Of papers with flow cytometry plots, 76.5% (13 of 17) used gradients to identify regions with many overlapping points. Semitransparency works best for small datasets with few overlapping points (Figure II in the online-only Data Supplement). Gradients should be shown for large datasets with many overlapping points, including flow cytometry plots.
Consider adding a flow chart or study design diagram	Only 63 of 206 (30.6%) of the papers had a flow chart or a study design diagram. These figures make it easy to understand the study design and to follow the flow of participants or animals through the experiment. The Experimental Design Assistant tool is extremely useful for creating flow charts when planning animal studies.[11] This tool also provides extensive feedback to facilitate rigorous experimental design and transparent reporting of key features such as blinding and randomization. The STROBE guidelines recommend using flow charts for observational studies.[10] The CONSORT guidelines[12] include sample flow charts for randomized controlled trials.
Use colorblind-safe color maps	The most common form of colorblindness affects up to 8% of men and 0.5% of women of northern European ancestry.[14] Although 12.6% of papers (26 of 206) used a color map on a graph or clinical image, only 15.4% of these papers (4/26) used color maps in which key features were visible to someone with deuteranopia. Most papers with heat maps used color palettes that were not colorblind safe (7 of 12; 58.3%). Free tools such as Color Oracle[13] allow investigators to quickly see how figures might look to a colorblind person (Figure 5). Investigators should avoid non–colorblind safe color maps (ie, rainbow, red/green) and choose colorblind-safe alternatives. This may require working with device manufacturers, as many rainbow color maps appeared to have been generated by flow cytometry or ultrasound software.

CONSORT indicates Consolidated Standards of Reporting Trials; and STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

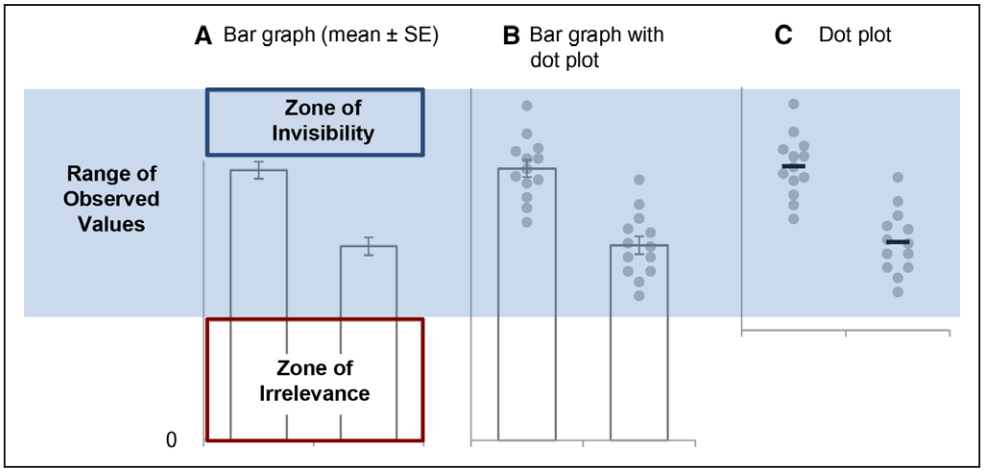


Figure 1. Why one shouldn't use a bar graph, even if the data are normally distributed.
Bar graphs arbitrarily assign importance to the height of the bar, rather than focusing attention on how the difference between means compares to the range of observed values. **A**, The bar height represents the mean. Error bars represent 1 standard error. The y-axis starts at zero and ends just above the highest error bar. **B**, Adding data points reveals that the bar graph in **A** includes low values that never occurred in the sample ("Zone of Irrelevance") and excludes observed values above the highest error bar ("Zone of Invisibility"). **C**, The dot plot emphasizes how the difference between means compares to the range of observed values. The y-axis includes all observed values. Reprinted from Weissgerber et al[22] under a CC-BY license. SE indicates standard error.

in peripheral vascular disease journals. Bar graphs showing continuous data were the most common figure type, appearing in 47.7% of papers (Table II in the online-only Data Supplement). These figures do not allow readers to critically evaluate the data.[1] This is problematic because many different data distributions can lead to the same

Figure Types	Example	Type of Variable	What the Plot Shows	Sample Size	Data Distribution	Best Practices
Dot plot		Continuous	Individual data points & mean or median line Other summary statistics (i.e. error bars) can be added for larger samples	Very small OR small; can also be useful with medium samples	Sample size is too small to determine data distribution OR Any data distribution	• Make all data points visible - use symmetric jittering • Many groups: Increase white space between groups, emphasize summary statistics & de-emphasize points • Only add error bars if the sample size is large enough to avoid creating a false sense of certainty • Avoid "histograms with dots"
Dot plot with box plot or violin plot		Continuous	Combination of dot plot & box plot or violin plot (see descriptions above and below)	Medium	Any	• Make all data points visible (symmetric jittering) • Smaller n: Emphasize data points and de-emphasize box plot, delete box plot and show only median line for groups with very small n • Larger n: Emphasize box plot and de-emphasize points
Box plot		Continuous	Horizontal lines on box: 75th, 50th (median) and 25th percentile Whiskers: varies; often most extreme data points that are not outliers Dots above or below whiskers: outliers	Large	Do not use for bimodal data	• List sample size below group name on x-axis • Specify what whiskers represent in legend
Violin plot		Continuous	Gives an estimated outline of the data distribution. The precision of the outline increases with increasing sample size.	Large	Any	• List sample size below group name on x-axis • The violin plot should not include biologically impossible values
Bar graph		Counts or proportions	Bar height shows the value of the count or proportion	Any	Any	• **Do not use for continuous data**

Figure 2. Figures for comparing groups in cross-sectional or experimental studies.
When choosing among different types of graphs, it is important to consider the study design, sample size, and data distribution. This figure provides a detailed overview of different types of graphs, describes when to use each graph, and lists best practices for clear data presentation.

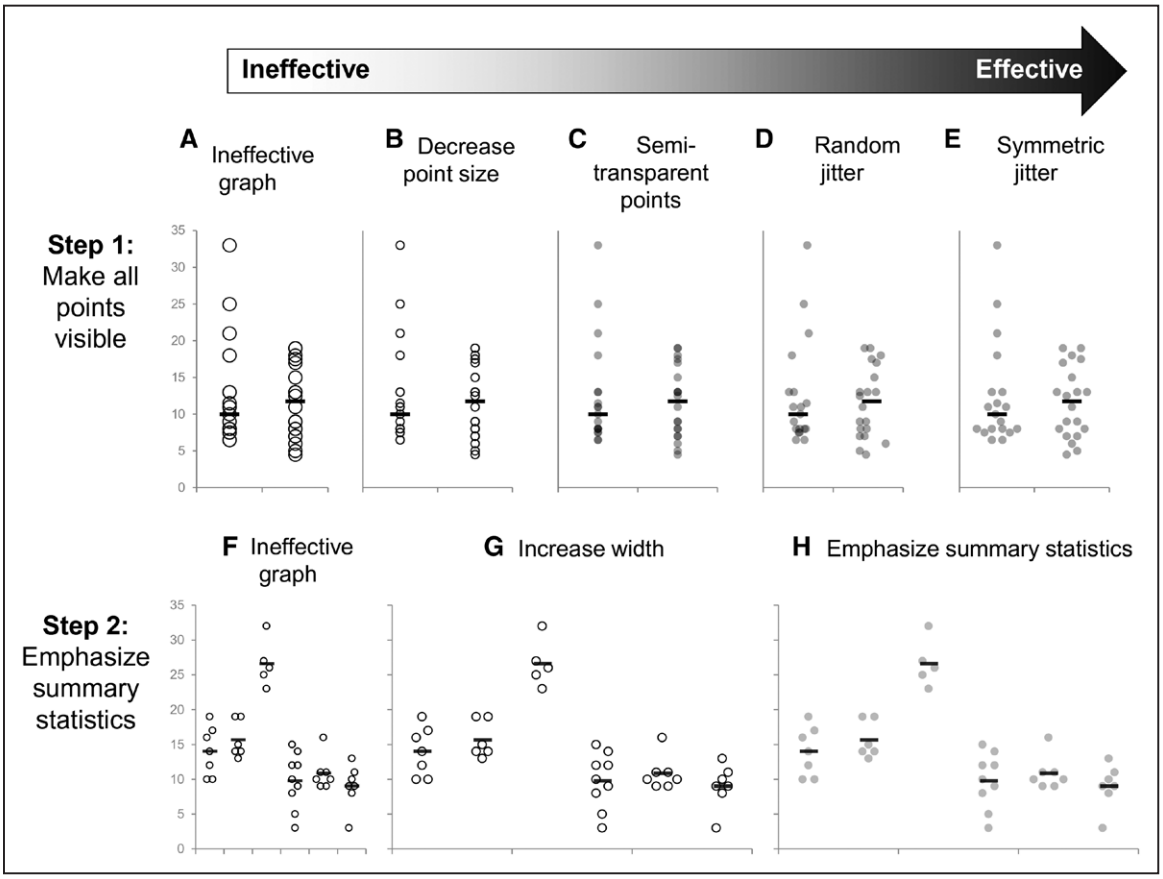


Figure 3. Strategies for making effective dot plots.
The initial graph is hard to interpret because it has many overlapping data points (**A**). Strategies for making all points visible include decreasing the size of the data points (**B**), making the data points semi-transparent (**C**), and using random (**D**) or symmetric (**E**) jittering. The bottom row illustrates how to clearly show the main finding, while allowing readers to critically evaluate the data. Increasing the white space between groups (**G**) and emphasizing the summary statistics (**H**) makes the graph (**F**) much easier to interpret.

bar or line graph and because the actual data may suggest different conclusions from the summary statistics alone.[1] Bar graphs also misleadingly assign importance to the bar height and distort our perception of how the differences between means compare to the variability in the data.[15] Given that the y-axis typically starts at zero and ends above the highest error bar, bars often include biologically impossible values (zone of irrelevance; Figure 1) while omitting high values that were observed in the sample (zone of invisibility). Bars also introduce within-the-bar bias, where viewers erroneously believe that data points are more likely to fall inside the bar instead of outside (above) the bar.[16] Investigators should use more informative alternatives to bar graphs—such as dot plots, box plots, and violin plots—especially for the situations outlined in Table 2.[17–21]

Selecting the Best Graphic for the Data

Effective figures in scientific papers should (1) immediately convey information about the study design and statistical analysis, (2) illustrate important findings, and (3) allow the reader to critically evaluate the data.[22] The type of figure that is selected will depend on the study design. When working with continuous data, dot plots, box plots, and violin plots may be used to compare independent groups in cross-sectional or experimental studies, whereas line graphs are commonly used to present data from longitudinal studies or studies with matched participants. Table 3 lists free tools and resources for creating these graphs.[1,15,22–31]

Comparing Independent Groups

More informative alternatives to the bar graph include dot plots, box plots, and violin plots. Figure 2 provides detailed information about how to determine which type of figure to use and describes best practices. When creating dot plots, avoid programs that create histograms with dots by grouping the data points into bins and then moving all data points to the center of their respective bins.[32] These fake dot plots can be misleading, especially for small datasets. Every point should be visible in dot plots (Figure 3, step 1). Emphasizing mean or median lines and de-emphasizing data points conveys a clear message, while allowing readers to critically evaluate the data

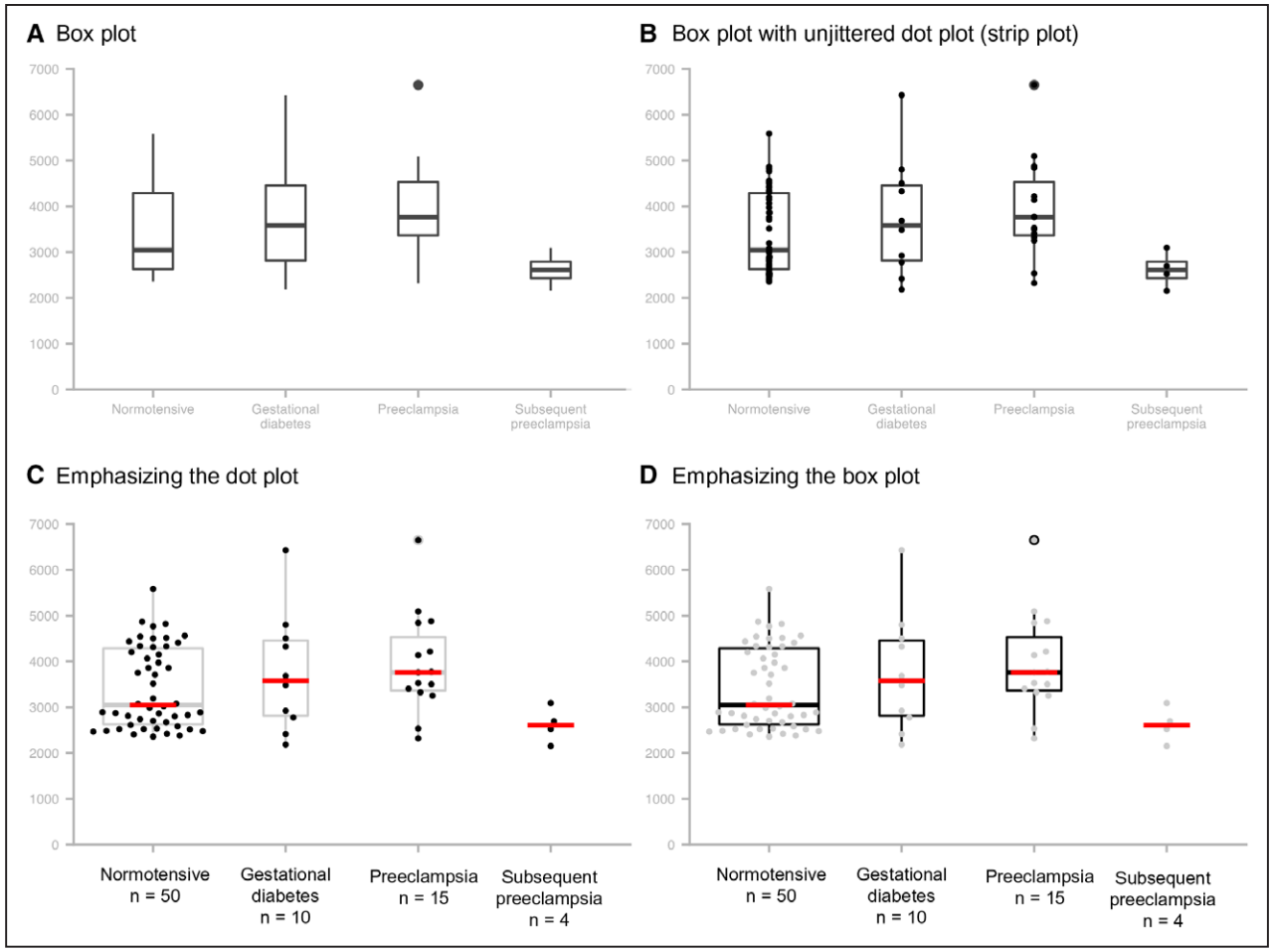


Figure 4. Combining dot plots with box or violin plots.
A, Data distribution shown but no information about sample size. **B**, The overlapping points offer little additional insight. **C** and **D,** Readers can evaluate the data by including symmetrically jittered data points, making the box plot width proportional to the sample size and listing the sample size on the x-axis. Only the dot plot is shown for the last group because the sample size was too small for a box plot. This dataset includes small groups (n=10–15); therefore, it would be better to emphasize the dot plot (**C**). If all groups have larger sample sizes, investigators can choose whether to emphasize the dot plot (**C**) or the box plot (**D**).

(Figure 3, step 2). This is especially useful for complex graphs with many groups. Similar principles can be applied when combining dot plots with box or violin plots (Figure 4C and 4D). Avoid adding dot plots to bar graphs. The bar does not add information and often obscures data points.

Longitudinal, Repeated Measures, or Matched Data

Line graphs were the second most common type of graph in peripheral vascular disease journals, appearing in 34.4% of papers. These graphs are used for longitudinal or repeated measures studies with 2 or more time points or conditions. The lines indicate that the same measurement was repeated on each participant, specimen, or sample. Lines can also show matched or related observations. The simplest design is paired data, in which measurements are performed on the same participants at 2 time points or under 2 experimental conditions, or 2 different participants are matched for important characteristics. Previous publications provide advice on visualizing paired data.[33,34]

As with bar graphs, many datasets can lead to the same line graph. There are some alternatives for small datasets that allow readers to go beyond the summary statistics. We recently introduced a free web-based tool for creating interactive line graphs for publications.[29] Users can gain additional insight by examining different summary statistics, viewing lines for any individual in the dataset, focusing on groups, time points or conditions where important changes are occurring, and examining change scores for any two time points or conditions. These features make it easier to assess overlap between groups and determine whether all individuals respond the same way. Alternatives to the line graph are useful for certain types of datasets, as described in the supplement of a recent article.[29] These include lasagna plots,[35] small multiples,[36] spaghetti plots,[37] and showing lines for selected individuals on a dot plot.[38]

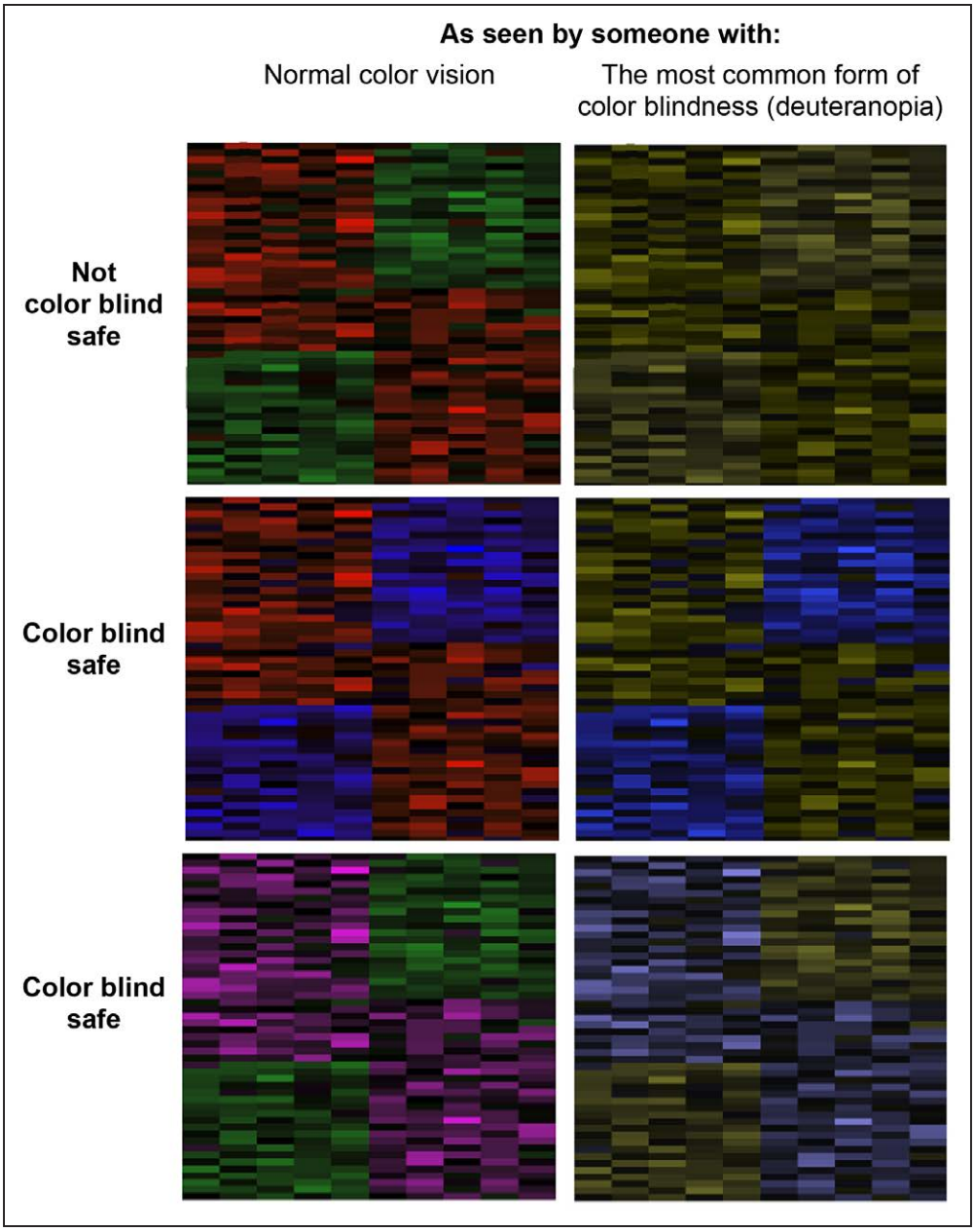


Figure 5. Select color blind safe color maps.
This figure illustrates how heat maps created using different color palettes would appear to someone with normal color vision (left column) versus someone with the most common form of color blindness (right column). Color blindness was simulated using Color Oracle.[13]

All Means Are Not Equally Meaningful: When to Show Summary Statistics

Bar and line graphs create the illusion of certainty by focusing readers' attention on differences in summary statistics, while erroneously suggesting that all means are equally meaningful. This assumption is problematic particularly for small datasets because summary statistics for small samples can be very different from the true population values (Figure 6). Preclinical and basic biomedical science studies typically have fewer than 15 observations per group, and 3 to 6 observations per group are common (RESULTS in the online-only Data Supplement).[1,39] Figure 6A shows how the summary statistics might change if we repeated the same experiment 100 times, with 5 or 20 observations per group. The smaller the sample size, the greater the variability in the summary statistics and the higher the probability that a sample will yield estimates that are very different from the true population value. Figure 6B illustrates that increasing the sample size allows one to move from the turbulent "Seas of Uncertainty,"[40] where summary statistics can be wildly imprecise, into the "Corridor of Stability."[41] Investigators can explore the effects of sample size on summary statistics by using a free online simulator with interactive versions of these figures (https://rtools.mayo.edu/size_matters/, Simulator 1).[42]

Table 2. Situations Where Showing the Data Points or Distribution Is Particularly Important

Situation	Reasons Transparent Figures are Particularly Important
Small sample size studies	Dot plots are crucial for small studies, as summary statistics for small samples can be quite different from those of the population from which the sample was drawn.[17]
Responses are highly variable	When an article asserts that responses are highly variable, figures showing the data distribution are necessary to support this conclusion. Understanding the factors that contribute to variability between individuals can provide critical insight into pathophysiology as well as the potential usefulness of tests and treatments. A blog post[18] based on a study[19] examining highly variable weight regain patterns among participants of the *Biggest Loser* television series illustrates how examining individual-level data may provide additional insight.
Heterogeneity or subgroups are expected	Bar and line graphs mask heterogeneity and conceal subgroups. Data points must be visible in order to understand heterogeneity or identify subgroups. The hypertensive pregnancy disorder, preeclampsia, is one example. Many pathways can lead to the diagnostic signs of hypertension and proteinuria, for example, and the relative contribution of each pathway likely varies from woman to woman.[20] Markers for any pathway are likely to be normal in some preeclamptic women and abnormal in others.[21] Showing the data distribution allows others to examine the overlap between groups and estimate the proportion of patients with abnormal values.
The SD is larger than the mean, and the variable cannot be negative	This indicates that the data are skewed, and the mean and SD are misleading.
Previous study indicators	Previous studies suggest that the variable is not normally distributed in the study population or in related populations.

An additional problem with showing mean and standard error or standard deviation for small samples is that these summary statistics are most appropriate for normally distributed data. Very small samples do not provide enough data to distinguish among normal, skewed, and bimodal distributions (Figure 7). Normality tests are underpowered when the sample size (n) is small and often fail to detect nonnormal distributions. Interactive versions of this figure, and a table showing the percentage of samples that would fail a normality test, are available in the online simulator (https://rtools.mayo.edu/size_matters/, Simulator 1).[42] Investigators can use this tool to actively explore potential effects, instead of relying on arbitrary thresholds. Results may vary from those shown in the simulator based on the degree of skewness or the separation between bimodal peaks.

The simulator illustrates that dot plots of very small samples help readers to evaluate the data, but don't

Table 3. Free Resources to Create Figures for Small Datasets

Figure	Program	Resource
Static or interactive dot plots, box plots, and violin plots	Web-based tool; features for showing subgroups and clusters of nonindependent data*	http://statistika.mfub.bg.ac.rs/interactive-dotplot/[22]
Dot plots, box plots, and violin plots	Web-based tool (Shiny app)	https://huygens.science.uva.nl/PlotsOfData/[27]
Combination of dot plots, box plots, and kernel density plots	Tutorial for R, Python, MATLAB—this visualization is appropriate only for large datasets	https://wellcomeopenresearch.org/articles/4-63/v1[30]
Dot plots	Excel templates, GraphPad PRISM instructions	Supplemental files of Weissgerber et al. (2015)[1]
	SPSS code	https://www.ctspedia.org/wiki/pub/CTSpedia/TemplateTesting/Dotplot_SPSS.pdf[23]
	R code	Blog post by Jamie Ashander[25] Blog post by Ben Marwick[26]
Box plots	Web-based tool (for independent or clustered/grouped data; Shiny app)*	https://lancs.shinyapps.io/ToxBox[28]
	Web-based tool (for independent data; Shiny app)	http://boxplot.tyerslab.com/[15]
	R code	Blog post by Jamie Ashander[25]
Violin plots	Web-based tool (for independent data; Shiny app)	https://interactive-graphics.shinyapps.io/violin/[24]
Paired data: spaghetti plots and dot plots of change scores	Excel templates, GraphPad PRISM instructions	Supplemental files of Weissgerber et al. (2015)[1]
	R code	Blog post by Jamie Ashander[25] Blog post by Ben Marwick[26]
Interactive line graphs	Web-based tool; features for showing lines for any individual, focusing on groups or time points of interest, viewing change scores for any two conditions	http://statistika.mfub.bg.ac.rs/interactive-graph/[29]
Various types of graphs	Excel, R code, Stata	http://faculty.washington.edu/kenrice/heartgraphs[31]

Shiny indicates R package in statistical software.

*Grouped or clustered data refers to measurements performed in subjects, specimens, or samples that are related to each other. This might include replicates or animals from the same litter.

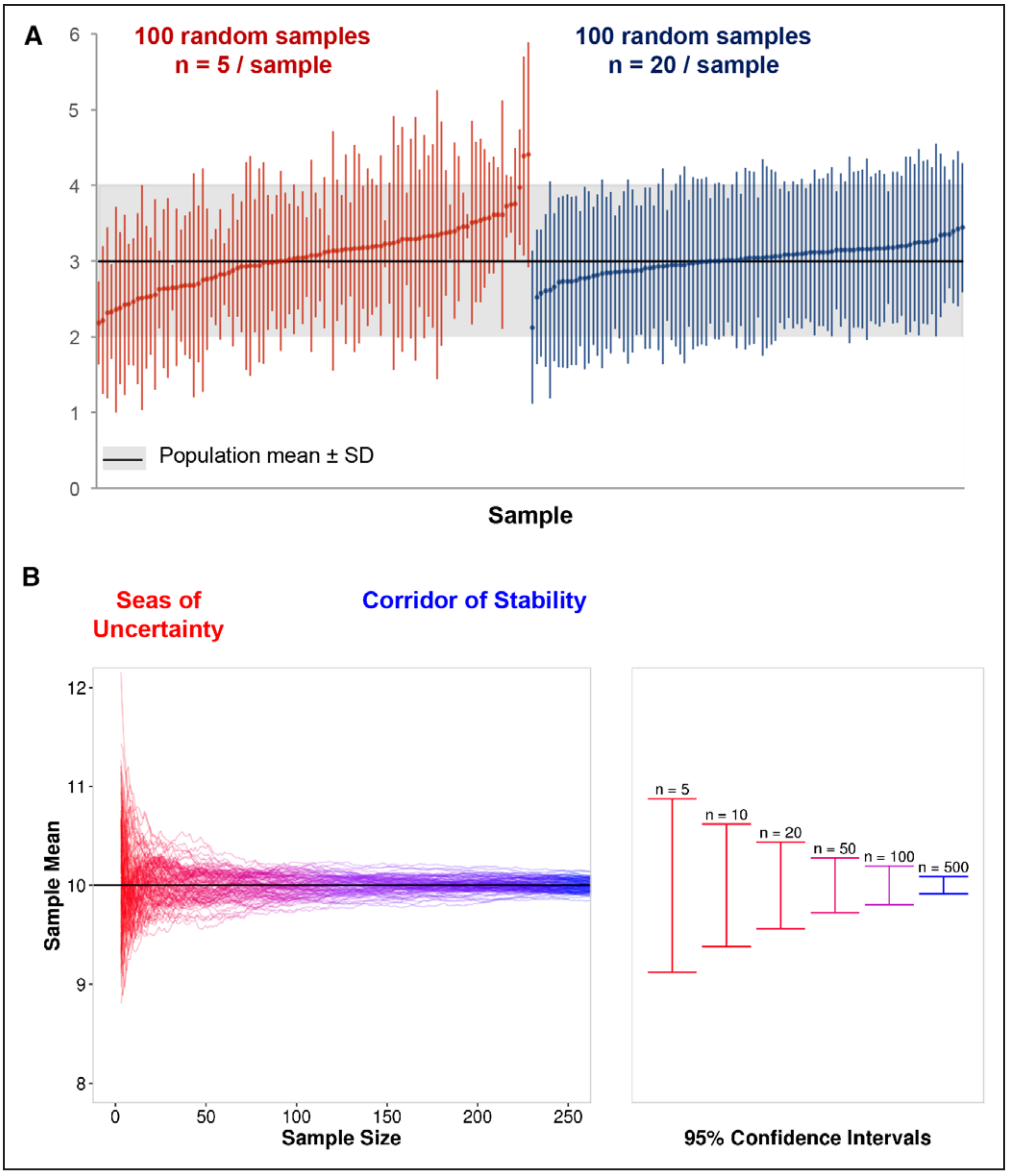


Figure 6. Summary statistics are only meaningful when there are enough data to summarize.
A, How the sample mean (red and blue dots) and standard deviation (red and blue error bars) might change if we repeated the same experiment 100 times, with n=5 or n=20. The black line and gray shaded region show the population mean and standard deviation. If all samples gave precise estimates, each sample mean would be on the black line and the error bars would fill the gray region. **B,** How cumulative means change with increasing sample size. The sample mean is calculated for 3 participants. New participants are added 1 at a time. The colored lines illustrate how the sample mean changes as each new participant is added. The experiment is repeated 100 times. The black line shows the true population mean. When n is small, the sample means are often quite different from the population mean ("Seas of Uncertainty"). As n increases, the sample means converge on the population mean ("Corridor of Stability"). Interactive version: https://rtools.mayo.edu/size_matters/, Simulator 1.[42] The terms "Seas of Chaos/Uncertainty" and "Corridor of Stability" were used in articles that examined the effects of sample size on correlation coefficients[41] and effect sizes.[40]

reveal the data distribution. As sample size increases, data distributions can be identified, and summary statistics become more precise. This explains why dot plots with mean or median lines are the best choice for very small datasets—error bars, box plots, and violin plots are meaningful only when there are enough data to summarize. These features can be added when the sample size is large enough to give reasonably precise estimates. In most bar graphs in our sample, error bars show the standard error (66.3%) rather than the standard deviation (20.9%) or 95% confidence interval (2.3%). As described previously, standard errors are sample size dependent and provide information about the precision of the mean, rather than the variability in the data.[43] If the sample size varies among groups, avoid creating a false sense of certainty by showing only error bars, box plots, or violin plots for larger groups. In Figure 4, the box plot was removed for the group with 4 observations. When combining plots for small samples, investigators should emphasize what is known (the data points) and deemphasize what is uncertain (the box/violin plot). For larger samples, in-

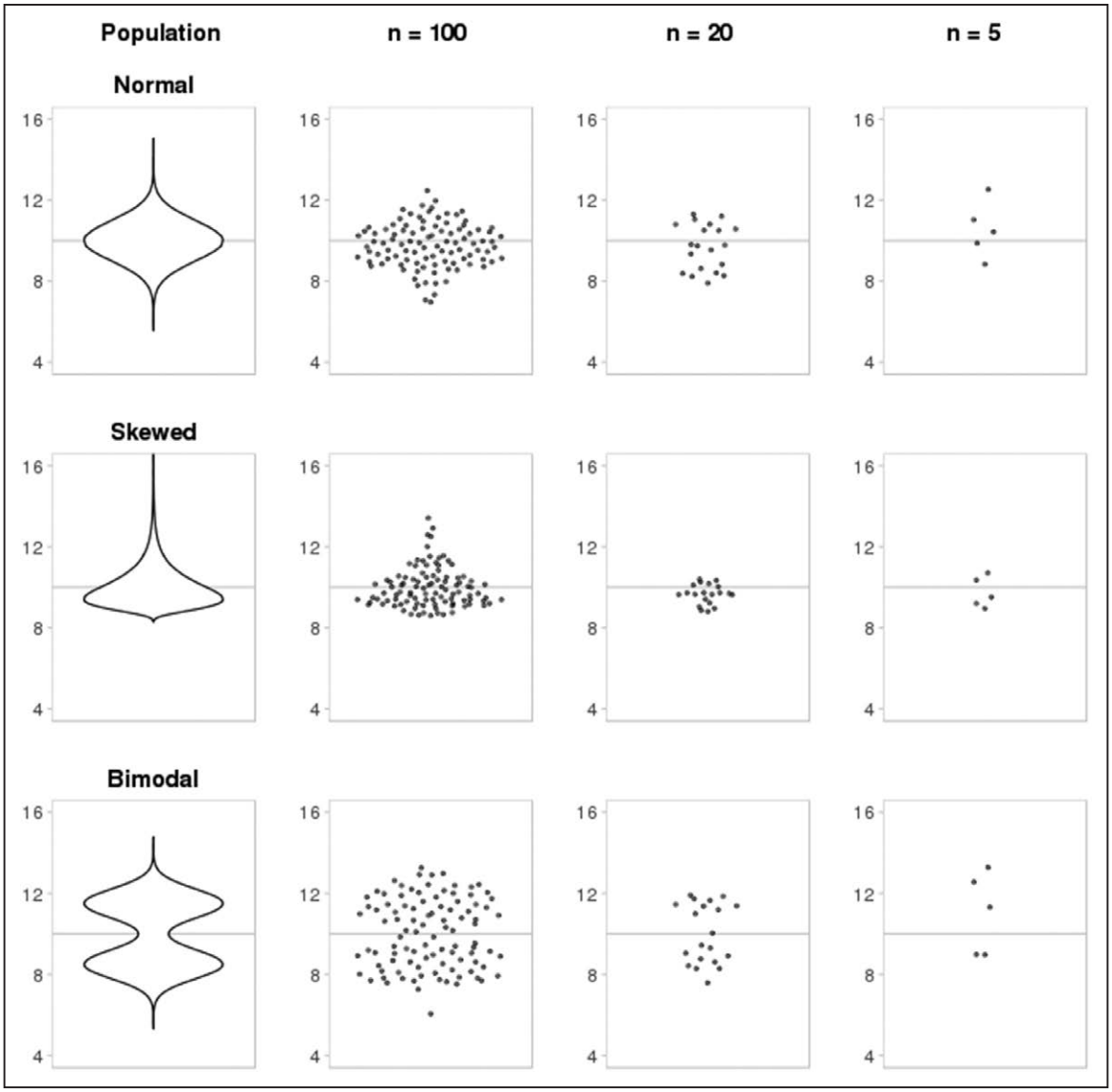


Figure 7. Small samples do not contain enough information to determine the distribution.
Random samples of different sizes (n=100, n=20, and n=5) were drawn from populations with a normal, skewed, or bimodal distribution (black violin plots). One can clearly identify the different data distributions when n=100. Determining the data distribution becomes more difficult when n=20 and is impossible when n=5. Interactive version: https://rtools.mayo.edu/size_matters/, Simulator 1.[42]

vestigators can choose whether to emphasize the dot plot (Figure 4C) or the box/violin plot (4D).

Practice Good Statistical Hygiene

The figure structure provides visual cues about the study design and statistical analysis. Misleading structures confuse readers (Figure 8). The experiment shown in Figure 8 was designed to compare normotensive and hypertensive patients. Concentrations of 3 different vascular biomarkers (dependent variables) were each compared using a *t*-test (normotensive vs hypertensive). Sometimes published articles present different dependent variables in the same graph (Figure 8A). This erroneously suggests that *t*-tests would not be appropriate given that the authors wanted to compare biomarkers (A vs B vs C) in addition to examining the effects of hypertension. In contrast, presenting each biomarker in a separate panel (Figure 8B) shows that the goal was to compare hypertensive and normotensive patients, and not to compare biomarkers. For small studies, each graph should present 1 statistical analysis and include all groups that were part of that analysis. Figure 9 illustrates how to structure figures and panels for common analyses.

The statistical methods, figure legend, and results should contain the information needed to reproduce the analysis. Specifying the statistical test in the figure legend makes it easy for readers to determine what the authors were comparing and to confirm that the

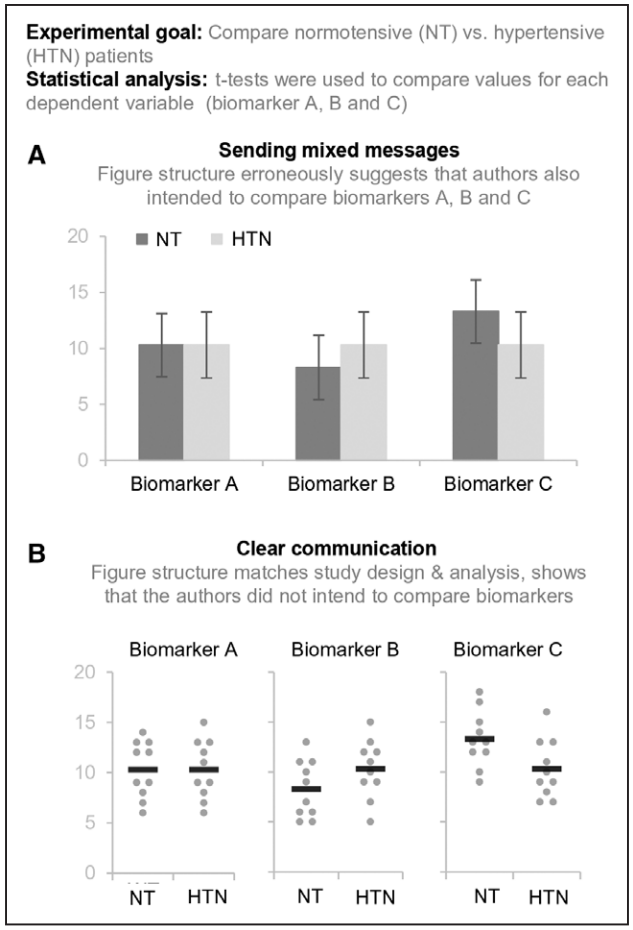


Figure 8. Avoid sending mixed messages—why the figure structure should match the study design and statistical analysis.
The experiment was designed to compare normotensive and hypertensive patients. Separate analyses were performed for each dependent variable (biomarkers A, B, and C). **A,** A common strategy for presenting this type of data. Including all dependent variables on the same graph erroneously suggests that the authors intended to compare biomarkers A, B, and C. **B,** Avoids confusion by presenting each biomarker separately. HTN indicates hypertensive; and NT, normotensive.

statistical methods match the design. When this is not possible, list table or figure numbers next to each technique in the statistical methods. If data are normally distributed but means and standard deviations are not evident from the figure, report these statistics in the legend or in the results. Readers may need exact values to confirm the statistical results, perform power calculations or conduct meta-analyses. Reporting test statistics, degrees of freedom and exact *P* values (*P*=0.43), instead of thresholds (*P*<0.05), is essential to allow others to confirm the test results.[44] Exact sample sizes for each group should be reported in the figure or legend, for example, by listing sample sizes below the group name on the x-axis. Do not report a range of sample sizes. This conceals information needed to reproduce the analysis and raises questions about whether unequal sample sizes were planned or may be due to the unexplained exclusion of participants or samples.

It is necessary to show outliers or excluded observations in figures, provide explanations (if known), and state how outliers were handled in the analysis. Excluding 1 or 2 outliers that oppose the expected effect in small studies can dramatically increase the odds of finding a significant effect when none exists.[39] If showing extreme outliers alters the scale such that the remaining data are not clearly visible, omit the outlier from the graph and report the value in the legend. Flow charts that illustrate the planned number of participants, animals, or samples for each experiment and show the reasons for attrition or exclusion of each subject or specimen are underused in preclinical and observational studies. This information is needed to confirm that exclusions were unbiased.

Articles that show data points or provide open data allow others to reproduce the analysis and determine whether using different analytic techniques would have yielded different results. Transparent reporting and open data are becoming increasingly important as scientists, funding agencies, and journals implement new strategies to improve scientific rigor and reproducibility.[45] These include reducing our reliance on hypothesis testing and the widespread use of the *P*<0.05 threshold for statistical significance, and training investigators in the strengths and weaknesses of this approach compared to other techniques (eg, effect sizes, Bayesian analysis).[17,46,47]

Additional Graphs and Resources

This section briefly reviews visualization techniques for other common situations in the basic biomedical sciences and introduces some new visualization tools.

Two-Way ANOVA

A 2-way ANOVA could refer to several different tests, each of which is appropriate for a different study design and requires a different figure structure.[48] Authors should avoid confusion by specifying whether a repeated measures ANOVA was used and whether each independent variable was analyzed as a between-subjects or a within-subjects factor. The figure structure should match the study design and analysis (Figure 3 in the online-only Data Supplement).

Static Graphs Comparing Effect Sizes

P values indicate whether groups are significantly different, but they do not assess the size of the difference or determine whether this difference is biologically meaningful. A new web-based tool creates graphs that highlight the size of the difference between groups.[49]

Comparing Differences in Variability

Common statistical tests compare differences in means or ranks, however sometimes scientists want to compare differences in variability (ie, is the range of observed values larger

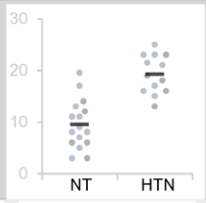
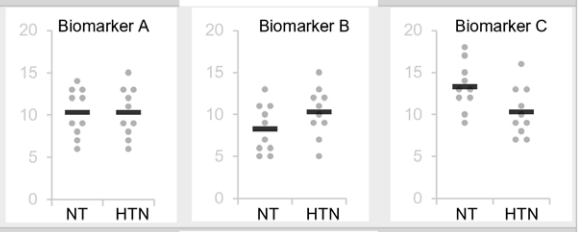
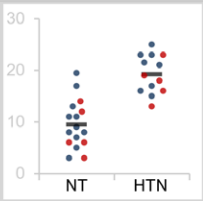
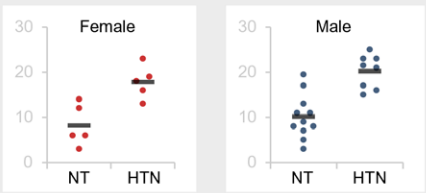
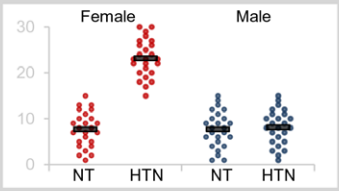
Analysis Strategy	Example	Figure Structure	Illustration
Comparing groups	Figure compares normotensive vs. hypertensive patients	One figure showing all groups that were included in the analysis	
Repeating the same analysis on different dependent (outcome) variables	Figure compares normotensive vs. hypertensive patients. Three different tests are performed on different biomarkers.	Separate panels for each analysis (i.e. dependent variable)	
Comparing groups with pooled subgroups	Figure compares normotensive vs. hypertensive patients. Men and women are pooled.	One figure showing all groups that were included in the analysis; data points for different subgroups are shown in different colors	
Stratified analysis	Figure compares normotensive vs. hypertensive patients. Separate analyses are performed for men and women.	Separate panels for each analysis When possible, using the same scales can facilitate visual comparisons	
Testing for an interaction	Figure compares four different groups of patients (normotensive women, hypertensive women, normotensive men, hypertensive men). The analysis tests for an interaction between hypertension and sex.	One figure showing all groups included in the analysis	

Figure 9. How to structure figures for common analyses.
This figure illustrates how to structure figure panels and groups for common types of analyses, including comparing groups, repeating the same analysis on different dependent variables, comparing groups with pooled subgroups, stratified analyses, and testing for an interaction. HTN indicates hypertensive; and NT, normotensive.

in males than in females?). Investigators with larger samples can compare variability by graphing shift functions.[50]

Interactive Graphics

Free online tools allow scientists with no programming expertise to quickly make interactive dot plots, box plots, and violin plots[22] or interactive line graphs[29] for publications. Investigators who use R statistical software can create customized interactive graphics using Shiny.[51]

What Can Scientists Do to Improve Data Presentation in the Scientific Literature?

Scientists, journal editors, and funding agencies can use several strategies to promote better data visualization.

1. Use transparent figures for papers, posters and talks. Avoid the common errors outlined in Table 1.
2. Support data presentation choices, as one would any other aspect of the study. Provide references that illustrate the importance of showing the data distribution when asked.[1,28,52]
3. When reviewing articles, request informative figures that allow others to critically evaluate the data.
4. Talk to journal editors about strategies for improving data presentation. Encourage them to consult editorials,[7,53] presentations,[54] and policies[5,8,55,56] of journals that have implemented policy changes. Policy changes are most effective when they are integrated into the review process. Decision letters should include comments requesting data

visualization changes, with citations to show why this is important and how to create more informative graphics.

5. Organize data visualization training for researchers at all career stages.[1,57] Record sessions for later viewing.

CONCLUSIONS

Better data visualization practices are needed to promote transparency in small sample size studies.[1] The strategies and resources outlined in this review are designed to promote transparency by improving the quality of figures, while assisting scientists, academic journals, and funding agencies in making lasting improvements to data visualization in the scientific literature.

ARTICLE INFORMATION

The online-only Data Supplement is available with this article at https://www.ahajournals.org/doi/suppl/10.1161/circulationaha.118.037777.

Correspondence

Tracey L. Weissgerber, PhD, Division of Nephrology and Hypertension, Mayo Clinic, 200 First Street SW, RO-HA-06-675-11, Rochester, MN 55905 USA. Email weissgerber.tracey@mayo.edu

Affiliations

Division of Biomedical Statistics and Informatics (S.J.W., E.P.H.), Division of Nephrology and Hypertension (T.L.W., O.G.V., V.D.G., N.M.M.), Mayo Clinic, Rochester, MN. QUEST – Quality | Ethics | Open Science | Translation, Charité - Universitätsmedizin Berlin, Berlin Institutes of Health (BIH), Germany (T.L.W.). Department of Medical Statistics and Informatics, Medical Faculty, University of Belgrade, Serbia (Z.B., M.D.S., N.M.M.).

Sources of Funding

TLW was funded by American Heart Association grant 16GRNT30950002, a Robert W. Fulk Career Development Award (Mayo Clinic Division of Nephrology and Hypertension) and the Office of Research on Women's Health (Building Interdisciplinary Careers in Women's Health, K12HD065987). SJW was funded by a Walter and Evelyn Simmers Career Development Award for Ovarian Cancer Research and National Cancer Institute (R03-CA212127). This study was made possible by Clinical and Translational Science Awards Grant Number UL1 TR000135 from the National Center for Advancing Translational Sciences, a component of the National Institutes of Health. The content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health. The writing of the manuscript and the decision to submit it for publication were solely the authors' responsibilities.

Disclosures

None.

REFERENCES

1. Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol.* 2015;13:e1002128. doi: 10.1371/journal.pbio.1002128
2. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JP. Public availability of published research data in high-impact journals. *PLoS One.* 2011;6:e24357. doi: 10.1371/journal.pone.0024357
3. Hardwicke TE, Mathur MB, MacDonald K, Nilsonne G, Banks GC, Kidwell MC, Hofelich Mohr A, Clayton E, Yoon EJ, Henry Tessler M, et al. Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*. *R Soc Open Sci.* 2018;5:180448. doi: 10.1098/rsos.180448
4. Vines TH, Albert AYK, Andrew RL, Débarre F, Bock DG, Franklin MT, Gilbert KJ, Moore JS, Renaut S, Rennison DJ. The availability of research data declines rapidly with article age. *Curr Biol.* 2014;24:94–97. doi: 10.1016/j.cub.2013.11.014
5. PLOS Biology. Submission Guidelines: Data Presentation in Graphs. 2016. http://journals.plos.org/plosbiology/s/submission-guidelines - loc-data-presentation-in-graphs. Accessed July 1, 2019.
6. Teare MD. Transparent reporting of research results in eLife. *Elife.* 2016;5:e21070. doi: 10.7554/eLife.21070
7. Fosang AJ, Colbran RJ. Transparency is the key to quality. *J Biol Chem.* 2015;290:29692–29694. doi: 10.1074/jbc.E115.000002
8. Towards greater reproducibility for life-sciences research in Nature. *Nature.* 2017;546:8. doi:10.1038/546008a
9. Weissgerber T, Milin-Lazovic J, Garcia-Valencia O, Milic N; Open Science Framework. Data visualization practices in peripheral vascular disease journals: How can we improve? 2019. https://osf.io/pm3vn/. Accessed July 1, 2019. doi:10.17605/OSF.IO/PM3VN
10. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med.* 2007;4:e296. doi: 10.1371/journal.pmed.0040296
11. Percie du Sert N, Bamsey I, Bate ST, Berdoy M, Clark RA, Cuthill I, Fry D, Karp NA, Macleod M, Moon L, et al. The experimental design assistant. *PLoS Biol.* 2017;15:e2003779. doi:10.1371/journal.pbio.2003779
12. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med.* 2010;7:e1000251. doi: 10.1371/journal.pmed.1000251
13. Jenny B, Kelso NV. Monash University. Color Oracle. 2006. https://colororacle.org/index.html. Accessed July 10, 2019.
14. National Eye Institute. Facts about color blindness. 2015. https://nei.nih.gov/health/color_blindness/facts_about. Accessed March 13, 2019.
15. Spitzer M, Wildenhain J, Rappsilber J, Tyers M. BoxPlotR: a web tool for generation of box plots. *Nat Methods.* 2014;11:121–122. doi: 10.1038/nmeth.2811
16. Newman GE, Scholl BJ. Bar graphs depicting averages are perceptually misinterpreted: the within-the-bar bias. *Psychon Bull Rev.* 2012;19:601–607. doi: 10.3758/s13423-012-0247-5
17. Halsey LG, Curran-Everett D, Vowler SL, Drummond GB. The fickle P value generates irreproducible results. *Nat Methods.* 2015;12:179–185. doi: 10.1038/nmeth.3288
18. Joyner M. Behind the "Biggest Loser" study headlines—a lost opportunity to educate about weight loss options. Health News Review.org. May 9, 2016. http://www.healthnewsreview.org/2016/05/behind-the-biggest-loser-study-headlines-a-lost-opportunity-to-educate-about-weight-loss-options/. Accessed January 5, 2019.
19. Fothergill E, Guo J, Howard L, Kerns JC, Knuth ND, Brychta R, Chen KY, Skarulis MC, Walter M, Walter PJ, et al. Persistent metabolic adaptation 6 years after "The Biggest Loser" competition. *Obesity (Silver Spring).* 2016;24:1612–1619. doi: 10.1002/oby.21538
20. Myatt L, Redman CW, Staff AC, Hansson S, Wilson ML, Laivuori H, Poston L, Roberts JM; Global Pregnancy CoLaboratory. Strategy for standardization of preeclampsia research study design. *Hypertension.* 2014;63:1293–1301. doi: 10.1161/HYPERTENSIONAHA.113.02664
21. Powers RW, Roberts JM, Plymire DA, Pucci D, Datwyler SA, Laird DM, Sogin DC, Jeyabalan A, Hubel CA, Gandley RE. Low placental growth factor across pregnancy identifies a subset of women with preterm preeclampsia: type 1 versus type 2 preeclampsia? *Hypertension.* 2012;60:239–246. doi: 10.1161/HYPERTENSIONAHA.112.191213
22. Weissgerber TL, Savic M, Winham SJ, Stanisavljevic D, Garovic VD, Milic NM. Data visualization, bar naked: A free tool for creating interactive graphics. *J Biol Chem.* 2017;292:20592–20598. doi: 10.1074/jbc.RA117.000147
23. Savic M; CTSpedia. SPSS instructions for creating univariate scatterplots. 2016. https://www.ctspedia.org/wiki/pub/CTSpedia/TemplateTesting/Dotplot_SPSS.pdf. Accessed July 10, 2019.
24. Savic M, Weissgerber TL, Bukumiric Z, Milic NM; Department of Medical Statistics and Informatics, Medical Faculty, University of Belgrade. Violin plot app. 2016. https://interactive-graphics.shinyapps.io/violin/. Accessed March 10, 2019.
25. Ashander J. Easy alternatives to bar charts in native R graphics. Rapid evolution: Theory, computation and inference. April 28, 2015. http://www.ashander.info/posts/2015/04/barchart-alternatives-in-base-r/. Accessed July 10, 2019.
26. Marwick B. Using R for the examples in "Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm." 2015. https://cdn.rawgit.

com/benmarwick/new-data-presentation-paradigm-using-r/582a80eaba654237231fe4b06d3eda5a61587d73/Weissgerber_et_al_supplementary_plots.html. Accessed July 1, 2019.
27. Postma M, Goedhart J. PlotsOfData—A web app for visualizing data together with their summaries. *PLoS Biol.* 2019;17:e3000202. doi: 10.1371/journal.pbio.3000202
28. Pallmann P, Hothorn LA. Boxplots for grouped and clustered data in toxicology. *Arch Toxicol.* 2016;90:1631–1638. doi: 10.1007/s00204-015-1608-4
29. Weissgerber TL, Garovic VD, Savic M, Winham SJ, Milic NM. From static to interactive: transforming data visualization to improve transparency. *PLoS Biol.* 2016;14:e1002484. doi: 10.1371/journal.pbio.1002484
30. Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res.* 2019;4:63. doi: 10.12688/wellcomeopenres.15191.1
31. Rice K, Lumley T. Graphics and statistics for cardiology: comparing categorical and continuous variables. *Heart.* 2016;102:349–355. doi: 10.1136/heartjnl-2015-308104
32. Dang TN, Wilkinson L, Anand A. Stacking graphic elements to avoid over-plotting. *IEEE Trans Vis Comput Graph.* 2010;16:1044–1052. doi: 10.1109/TVCG.2010.197
33. Schriger DL. Graphic portrayal of studies with paired data: a tutorial. *Ann EmergMed.*2018;71:239–246.doi:10.1016/j.annemergmed.2017.05.033
34. Rousselet GA, Foxe JJ, Bolam JP. A few simple steps to improve the description of group results in neuroscience. *Eur J Neurosci.* 2016;44:2647–2651. doi: 10.1111/ejn.13400
35. Swihart BJ, Caffo B, James BD, Strand M, Schwartz BS, Punjabi NM. Lasagna plots: a saucy alternative to spaghetti plots. *Epidemiology.* 2010;21:621–625. doi: 10.1097/EDE.0b013e3181e5b06a
36. Tufte E. *The visual display of quantitative information*. Cheshire, CT: Graphics Press; 1983.
37. Knaflic CN. Strategies for avoiding the spaghetti graph. Storytelling With Data. March 14, 2013. http://www.storytellingwithdata.com/blog/2013/03/avoiding-spaghetti-graph. Accessed January 5, 2019.
38. Diggle PJ, Heagerty PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. 2nd ed. Oxford: Clarendon Press; 2002.
39. Holman C, Piper SK, Grittner U, Diamantaras AA, Kimmelman J, Siegerink B, Dirnagl U. Where have all the rodents gone? The effects of attrition in experimental research on cancer and stroke. *PLoS Biol.* 2016;14:e1002331. doi: 10.1371/journal.pbio.1002331
40. Lakens D, Evers ER. Sailing from the seas of chaos into the corridor of stability: practical recommendations to increase the informational value of studies. *Perspect Psychol Sci.* 2014;9:278–292. doi: 10.1177/1745691614528520
41. Schoenbrodt FD, Perugini M. At what sample size do correlations stabilize? *J Res Pers*. 2013;47:609-612.
42. Heinzen EP, Weissgerber TL, Winham SJ. Mayo Clinic. Sample size matters. 2019. https://rtools.mayo.edu/size_matters/. Accessed March 30, 2019.
43. Davies HT. Describing and estimating: use and abuse of standard deviations and standard errors. *Hosp Med.* 1998;59:327–328.
44. Nuijten MB, Hartgerink CH, van Assen MA, Epskamp S, Wicherts JM. The prevalence of statistical reporting errors in psychology (1985–2013). *Behav Res Methods.* 2016;48:1205–1226. doi: 10.3758/s13428-015-0664-2
45. Announcement: Reducing our irreproducibility. *Nature*. 2013;496:398. doi:10.1038/496398a
46. Goodman SN. STATISTICS. Aligning statistical and scientific reasoning. *Science.* 2016;352:1180–1181. doi: 10.1126/science.aaf5406
47. Wasserstein R, Lazar N. The ASA's statement on p-values: context, process, and purpose. *Am Stat*. 2016;70:129-133. doi:10.1080/00031305.2016.1154108
48. Weissgerber TL, Garcia-Valencia O, Garovic VD, Milic NM, Winham SJ. Why we need to report more than "data were analyzed by t-tests or ANOVA." *Elife*. 2018;7:e361363. doi:10.7554/eLife.36163
49. Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A. Moving beyond p values: everyday data analysis with estimation plots. *bioRxiv*. 2018. doi: 10.1101/377978. Accessed May 1, 2019.
50. Rousselet GA, Pernet CR, Wilcox RR. Beyond differences in means: robust graphical methods to compare two groups in neuroscience. *Eur J Neurosci.* 2017;46:1738–1748. doi: 10.1111/ejn.13610
51. Ellis DA, Merdian HL. Thinking outside the box: developing dynamic data visualizations for psychology with shiny. *Front Psychol.* 2015;6:1782. doi: 10.3389/fpsyg.2015.01782
52. Saxon E. Beyond bar charts. *BMC Biol.* 2015;13:60. doi: 10.1186/s12915-015-0169-6
53. Weissgerber TL, Garovic VD, Winham SJ, Milic NM, Prager EM. Transparent reporting for reproducible science. *J Neurosci Res.* 2016;94:859–864. doi: 10.1002/jnr.23785
54. Colbran RJ, Fosang A. Data presentation: the importance of transparency. 2016. http://ht.ly/NyOF30005Ly. Accessed January 5, 2019.
55. International Journal of Primatology. Instructions for authors. 2015. https://beta.springer.com/journal/10764/submission-guidelines. Accessed July 10, 2019.
56. Journal of Biological Chemistry. Collecting and presenting data. 2015. http://jbcresources.asbmb.org/collecting-and-presenting-data. Accessed July 10, 2019.
57. Angra A, Gardner SM. Development of a framework for graph choice and construction. *Adv Physiol Educ.* 2016;40:123–128. doi: 10.1152/advan.00152.2015