Invited Article

AMERICAN ACADEMY OF OTOLARYNGOLOGY–HEAD AND NECK SURGERY FOUNDATION

Otolaryngology–Head and Neck Surgery
145(6) 886–894

DOI: 10.1177/0194599811423608
http://otojournal.org
SAGE

Practical Guide to Choosing an Appropriate Data Display

Randal C. Paniello, MD[1], J. Gail Neely, MD[1], Jason T. Rich, MD[1], Eric L. Slattery, MD[1], and Courtney C. J. Voelker, MD, DPhil[1]

No sponsorships or competing interests have been disclosed for this article.

Abstract

The primary objective of graphing research data is to communicate key information visually in a rapid, accurate, and concise way. Graphs might be considered visual take-home lessons of the major point(s) of the manuscript. In choosing a graph, it is tempting to concentrate only on ways of illustrating summary statements characterizing the group(s). However, individual patients are unique, and their characteristics or outcomes may not be predicted by a group summary. Consequently, if possible, graphs should demonstrate individual responses as well as group summaries. "Graphical literacy," "graphical excellence," and "graphical acumen" are achievable with work and collaboration. To produce a well-designed graph, a combination of by-subject detail and overall results should be the goal within the same illustration. The practice gap addressed in this article is that little attention from authors, reviewers, editors, and publishers seems to be paid to graphical literacy. The purpose of this article is to present some practical guidelines for choosing or evaluating more appropriate data displays.

Keywords

graphing, graphs, illustrations, data display

Received August 11, 2011; accepted August 25, 2011.

The primary objective of graphing research data is to communicate key information visually in an accurate and concise way. It offers a rapidly understood snapshot of information, often better appreciated as an illustration than a wordy text or table. It might be considered a visual take-home lesson of the major point(s) of the manuscript. There are 2 uses of graphics: one to analyze data and the other to report data. Often more and different displays are involved in analysis than the final reporting. Here we will concentrate on graphics used in reporting.

The word *data* is plural, generally referring to information from a group of subjects. As such, it is tempting to concentrate on ways of illustrating summary statements characterizing the group(s). However, in the practice of medicine, individual patients are unique, and their characteristics or outcomes may not be predicted by a group summary. In the vernacular, we hear "he/she is not a statistic." Consequently, it is important to attempt to demonstrate how each individual fits into the whole, as well as to look at the group summary characteristics. Graphic illustrations can be uniquely capable of demonstrating individual responses as well as group summaries.

"Graphical literacy," "graphical excellence," and "graphical acumen" are achievable with work and collaboration.[1-4] To produce a well-designed graphic, a combination of by-subject detail and overall results should be the goal within the same illustration.[1,2] The practice gap addressed in this article is that little attention from authors, reviewers, editors, and publishers seems to be paid to graphical literacy. The purpose of this article is to present some practical guidelines for choosing or evaluating more appropriate data displays.

General Concepts of Graphing Data

The extensive writings and lectures of Edward Rolf Tufte, PhD, first at Princeton and later at Yale, significantly advanced the field of graphical excellence.[5] Below are some of his contributions that help clarify some general rules or guidelines in the construction of graphics.

Golden rectangle. A study of the esthetics and mathematics of graphs suggests that a landscape design is preferred and that the ratio of height to width should be 1:1.618. However, if

[1]Practical Guides Writing Group, Department of Otolaryngology–Head and Neck Surgery, Washington University School of Medicine, St Louis, Missouri, USA

Corresponding Author:
J. Gail Neely, MD, Department of Otolaryngology–Head and Neck Surgery, Washington University School of Medicine, 660 S Euclid Avenue, Box 8115, St. Louis, MO 63110
Email: neelyg@ent.wustl.edu

the proper display of the data requires a different configuration, then first go with the data; otherwise, the width should be about 50% greater than the height.

Table versus graphs. Tables generally work better for small data sets of 20 or fewer subjects. Graphs are particularly useful for larger data sets.

Relational graphics. The "greatest of all graphical designs [is the] scatter plot and its variants."[5] This allows the display and analysis of 2, or more, variables. The standard assumption is that X causes Y, and therefore these graphs are constructed so that the actual or presumed dependent (outcome) variable is plotted on the y-axis and the independent (predictor) variable on the x-axis. However, the actual relationship between the variables may be seen and questioned.

Graphical competence. The demands for graphical competence are (1) substance, (2) statistical integrity, and (3) design. Above all, the substantive data must be accurately and honestly reported. A statistically congruent display of descriptive and analytical data is a must; this often requires consultation and collaboration with a statistician to faithfully represent the full and accurate impact of the data. Keeping these 2 aspects in mind, the design should contain the appropriate aesthetic and be immediately understandable by the audience (more on this later).

Principles of graphical integrity. Six principles of graphical integrity include (1) graphically displayed representation of values should be directly proportional to the actual data; (2) clear labeling on the graph itself should avoid distortion and ambiguity; (3) data variations, not design variations, should be shown; (4) in time series, standardized units are usually better than nominal units; (5) graphical display of dimensions should not exceed the actual meaningful data dimensions; and (6) graphs must reliably represent the data in context.

Graphical excellence. Graphical excellence is fundamentally telling the truth about the data. Henry adds to this in his principles for achieving graphical competence, which includes (1) give primacy to data; (2) design information-rich graphs; (3) tailor design to reveal data by (a) avoiding distortion, (b) enhancing clarity, and (c) encouraging important comparisons; and (4) remember the audience.[6]

Words versus encoded keys. To avoid taxing the memory and/or the need to alternately refer to the code key for gray scale or cross-hatching encoded graphed data, it is better to clearly label the graph with words.

Chartjunk, data density, lie factor. Graphed non–data ink is "chartjunk" and should be removed. Data ink is the crucial portion of the graph that efficiently represents the important data and cannot be removed. Data density is the numbers of entries of individual data divided by the area of the graphic. For example, per Tufte, a bar graph with 2 named bars of differing heights in a graph 26.5 square inches has 4 individual entries in the data matrix (height of the bars and their labels). Thus, $4/26.5 = 0.15$ entries per square inch, which is a very sparse density. The lie factor is the size of the effect shown in the graph divided by the size of the effect in the data. A ratio of 1 suggests the graph may be representing the data accurately; however, if it is <1 or >1, distortion in the graph has occurred. This may occur if artistic license is used to make a point; such distortion tends to cast doubt as to the validity of the original data, not just its display.[5]

Specific Aspects of Creating Excellent Graphical Data

Audience. As mentioned earlier, graphics are participant in human communication of data. Consequently, it is important to remember that 2 major parties are involved: the presenter and the audience. The audience was likely uninvolved with the research design, data collection, and analysis; however, the audience is paramount in the selection of how the data are graphed.[6] Henry recapitulates some of the science differentiating short-term memory and long-term memory. Short-term memory involves understanding and recalling something just seen for the first time, whereas long-term memory refers to recognizing a familiar form and understanding new information that is presented in that form. If complex data are displayed in a form familiar to the audience, they are likely to grasp what is being portrayed quickly. If complex data are presented in an unfamiliar form, "it is a problem to be solved, not a display to be read." While data density is a worthwhile goal, it should not override the need to clearly illustrate data that the audience will quickly understand.[6]

Statistical rules of thumb. Gerald van Belle,[7] in online lecture notes, suggested to (1) chew words carefully, (2) watch table manners, (3) avoid pies, and (4) stay away from bars. This memorable mnemonic emphasizes the need to use text to describe 2 to 5 numbers, tables for more, and graphs for complex data. Tables should be self-contained and limit significant digits. Do not use pie charts in scientific publications, an admonition echoed by Schriger and Cooper.[3] Similarly, there are likely better ways to describe data than the use of bar charts. The limited information that might be in a bar chart may be conveyed in a sentence of text.[3]

By-subject data combined with group data. Schriger and Cooper[3] presented an excellent example of how graphing by-subject data with group data materially enhanced analysis and changed the conclusions. The initial graph of two 3-dimensional bars was replete with Tufte's chartjunk and a y-axis that distorted the appropriate range; when this was corrected, a simple bar chart showed the difference between 2 groups of 200 each to be statistically significant ($P < .05$). However, when the histograms of the 2 groups were displayed together, one group was seen to be bimodal, and the distributions of the groups made the parametric comparison of the groups meaningless. Further analysis of the reason behind the bimodal configuration of the one group showed that this group's findings were operator dependent. Consequently, the initial conclusion that one treatment was better than the other was not true.

Error Bars

Error bars are of 3 classic types: standard deviations (SD), standard error of the mean (SEM), and confidence intervals (CIs). To illustrate these values that flank the statistic (such as the mean), a ± sign precedes the entity; for example, ±SEM or

Table 1. Raw Data

Group A		Group B	
146	153	268	252
170	178	211	289
127	120	235	300
174	142	329	206
203	141	272	286
182	233	303	227
165	165	246	207
182	150	257	274
147	141	275	280
173	214	257	255
174	132	242	218
	163		245

Group Descriptive Summary

	Mean	SD	SE	CI of Mean	Median	25%	75%
Group A	163.975	27.607	5.757	11.938	164.5	142.625	176.854
Group B	257.957	32.409	6.758	14.015	257.0	236.833	278.375

CI or SD. The type of error bar depicted on the graph must be clearly defined. SDs describe the spread or dispersion of raw data in a group, the variability of individuals within the group. The SEM and the CI define the stability of the statistic of the sample group; it is the measure of the imprecision of the summary statistic, such as the mean, proportion, difference, or regression slopes.[8-10] It is usually better to graph the SD of a group than to graph the SEM of the group or the CI of the group if the intent of the graph is to show the central tendency and spread of each group. Conversely, if the intent is to demonstrate the inferential statistical comparison between groups focusing on the stability of the central tendency of the groups, then the 95% CI is better than the SEM. This is because the 95% CI participates in showing not only statistical significance but also the actual value range of the central tendency as if the experiments were done many times.[10]

Graphical Examples

From this point forward, a set of arbitrary data will be used to illustrate the use of graphs descriptive of the principles mentioned above (**Table 1**). A beautiful example of a data-rich graph is seen in figure 2.1, page 9, in Gardner and Altman's text.[8] In this graph, 2 histograms of 2 groups are displayed side by side onto which the mean of each group, the 95% CI of the differences and the range of differences in the measured entity, is displayed. We will explore our data set in light of this type of graphical display so consistent with the principles mentions at the beginning of this article.

Descriptive Statistics

Statistical programs, such as SigmaPlot (SigmaPlot for Windows, version 11.0, Systat Software Inc, Chicago, IL) or SPSS (PASW statistics 18, IBM Software, Armonk, NY), may offer preset graphics for a statistical result. For example, for the descriptive statistics in **Table 1**, SigmaPlot offers bar charts of means, scatter plot column means, point plot, point and column means, and box plot. We will discuss these further below.

Bar charts of means demonstrate the mean as the upper limit of the bar; in this case, the height of the bar, not the area, demarcates the mean as read on the *y*-axis. The error bars are the standard deviations above the mean; however, the error bars must be identified in the graph legend as such (**Figure 1A**). Scatter plot column means show the means and the SDs above and below the mean and use much less ink (**Figure 1B**). The point plot demonstrates each individual subject data point in each group (**Figure 1C**). The point and column means show a combination of each individual subject data and the summary data by demonstrating the means and SDs for each group (**Figure 1D**). The box plots (or box-whisker plots) show the distribution of the data equally divided into 4 parts, or quartiles (**Figure 1E**). The quartiles Q1, Q2, and Q3 correspond to the percentiles at the 25%, 50%, and 75% of the distribution. The median of the group data, or 50th percentile, is shown as a horizontal bar within the box. The box demarcates the middle 50% of the distribution, also known as the interquartile range. The top of the box demarcates the 75% percentile and the lower edge of the box the 25% percentile. The "whiskers," or error bars, show the upper 90th and lower 10th percentile, approximately 1.5 times the interquartile range above and below the box limits (but not below zero). Outliers, beyond the 3 SDs from the mean, are shown as small closed circles, dots.

Given the principles described at the beginning of this article, it becomes quite clear that the point plot and the point and column means plot are the two richest in demonstrating the individual data points and some summary statistics. The box-whisker plots are also quite good but fail to show the details of individual subject results.

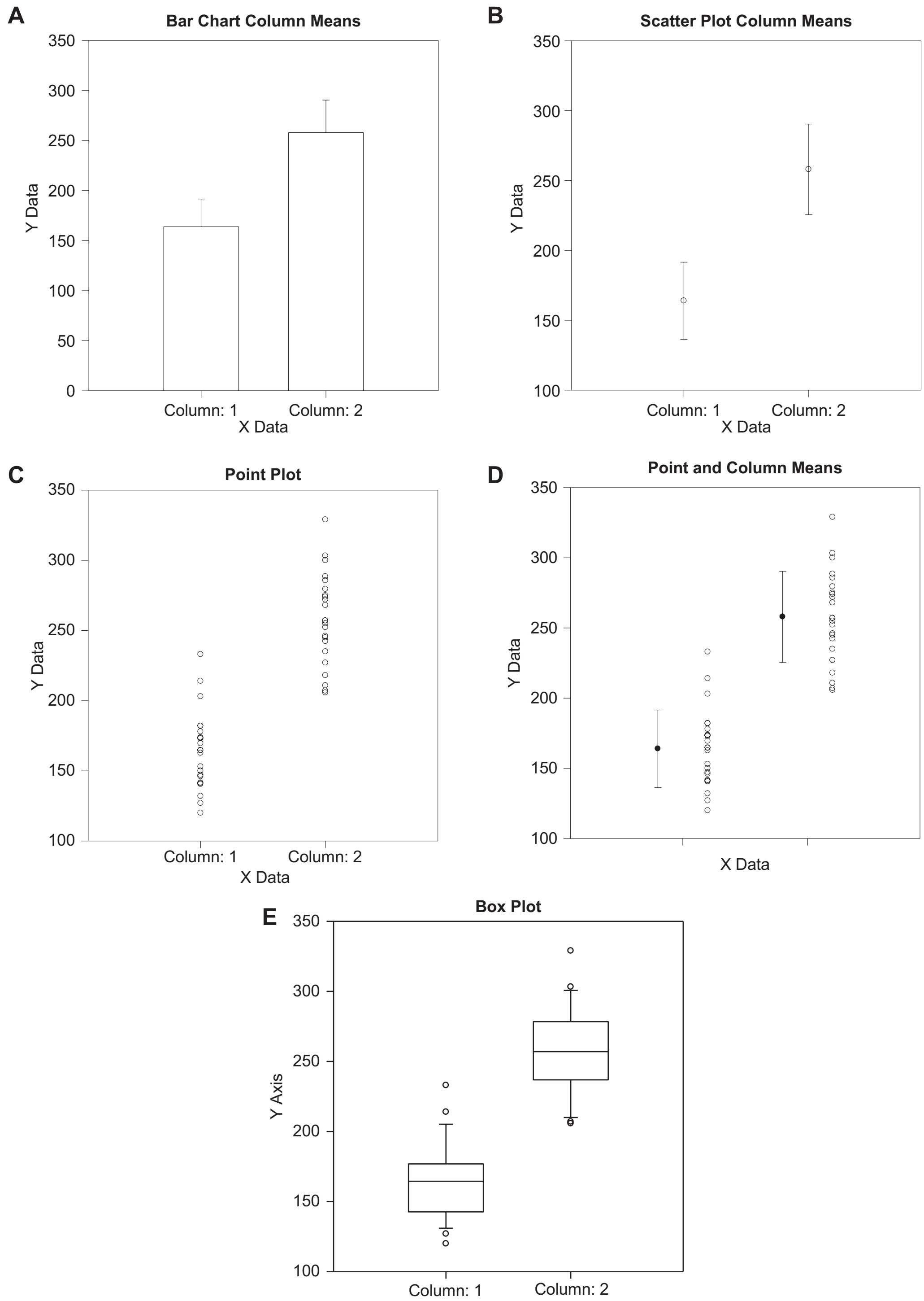

Figure 1. (A) Bar chart of means and standard deviations of the 2 groups seen in **Table 1**. Column 1 is group A, and column 2 is group B. Values on the *y*-axis are arbitrary values as they are in **Table 1**. (B) Scatter plot column means shows the means and standard deviations seen in A, but in a different format. (C) Point plot shows all of the raw data points in **Table 1** for both groups. The advantage of this is to show the value for each individual, which includes the full range from the very best patient value to the very worst for each group and how they overlap. It does not show summary values, such as means and standard deviations. (D) Point and column means is the best of both worlds. This graph shows all the point values of each subject and the summary values of means and standard deviations for each of the 2 groups in **Table 1**. (E) Box plot has much of the advantages of the point column means but adds summary values showing the distribution of individual values of both groups collapsed into the center 50% of values (boxes), the median, not mean (line in box), the top 90% and lower 10% of values ("whiskers"), and the extreme values beyond the whiskers as dots or small circles.

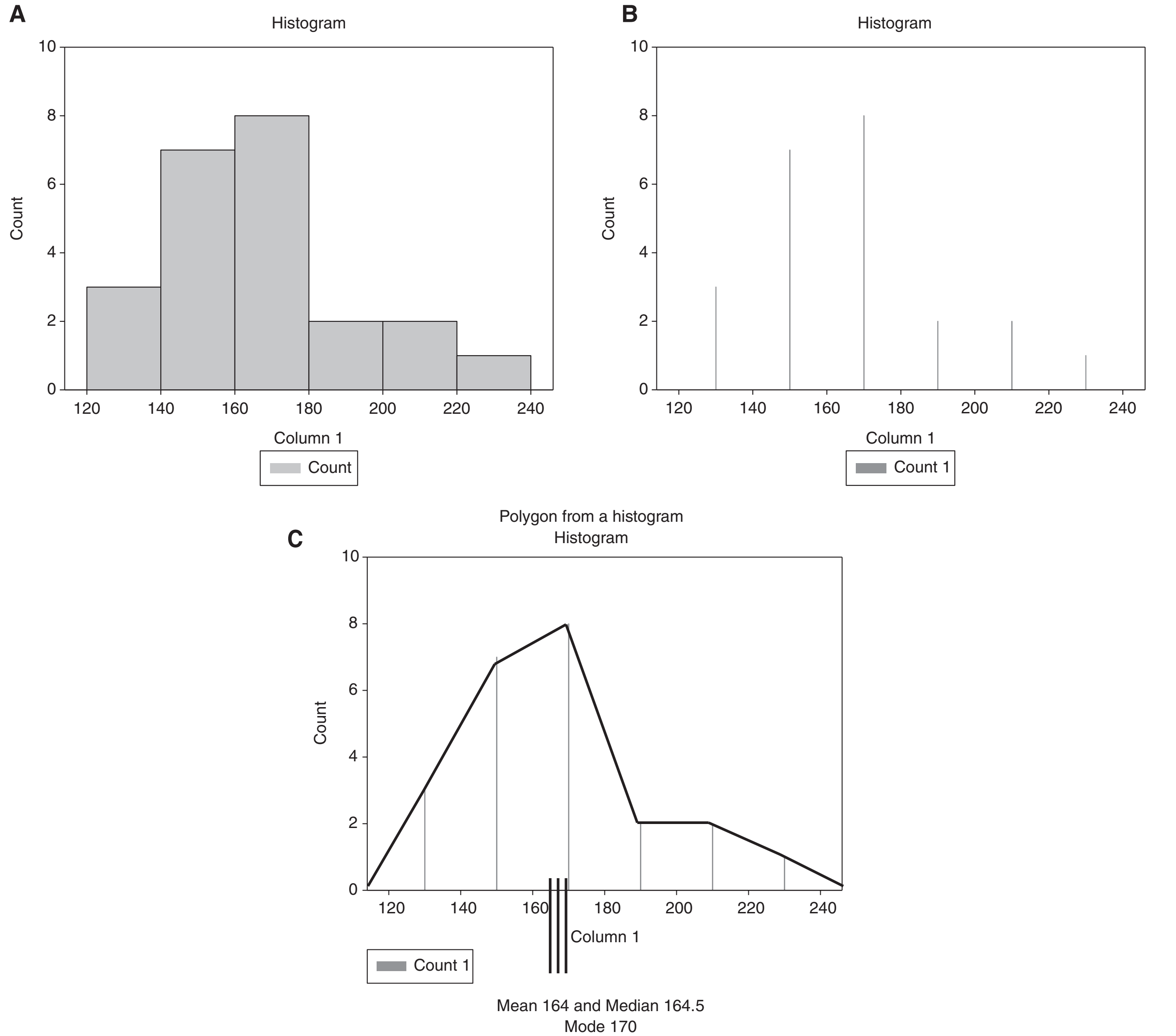

Figure 2. (A) Histogram of group A data in **Table 1**. (B) Needle histogram conversion of wide columns in A. (C) Polygon constructed using the needle histogram seen in B. In addition, the mode, median, and mean of the group have been added to the polygon illustration.

Histograms. Histograms are special bar graphs of contiguous bars, with each bar demarcated by bins or class intervals on the *x*-axis and frequency counts on the *y*-axis[9,11] (**Figure 2A**). The bar spans a uniform width, with the bar centered over the class interval that defines the "bin." Bin width is arbitrarily designated, although having either too wide or narrow of a bin width can be problematic. Having too wide of a bin width can overlook a bimodal characteristic, while having too narrow of a bin width can allow unimportant fluctuations in data to appear meaningful. Often data are inputted using different bin widths to see how the data appear before deciding on a bin width. For example, when the histogram was constructed for group A, the computer automatically calculated the bins most appropriate using a log formula built into the program. One can change the width of the bins manually. In this example, the bin width was 20.

A computer program can construct a "needle" histogram in which only the frequency count is graphed at the centermost point of the bin (**Figure 2B**). If a frequency polygon, a continuous line connecting the top of the histogram bars, is to be constructed, a point is placed at the center of each bar and the points connected with the continuous line (**Figure 2C**). A frequency polygon might have 2 advantages: (1) the shape of the distribution is easier seen than with the histogram and (2) a linear dose response may be more clearly seen than with a histogram.[11] Rather than trying to define the center of the bars on the histogram, the needle configuration makes that much easier.

Comparative Statistics

To compare the groups described in **Table 1**, a *t* test of the data shows $P < .001$ (statistically significant difference),

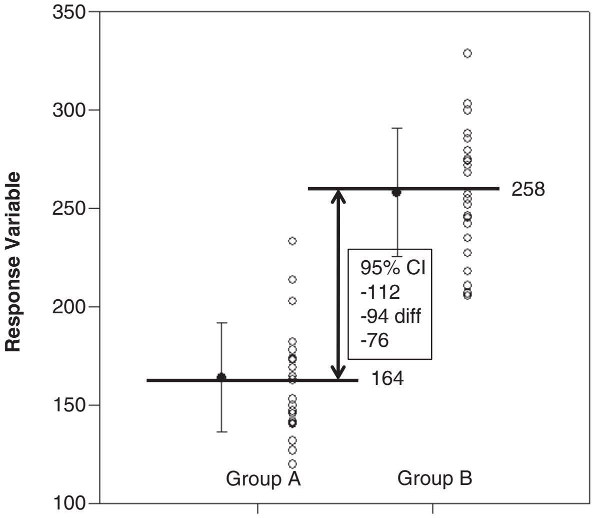

Figure 3. This graph shows the point and column means seen in **Figure 1D** and adds the inferential comparison values showing the difference between the means of group A minus group B (–94) and the 95% confidence interval of the difference between means (–76 to –112). This illustration maximizes the by-subject data for each group, 2 groups' summary data, and comparison statistics between the 2 groups.

difference between means (group A minus group B means) = –94, and 95% CI for difference of means –112 to –76. To be consistent with the principles mentioned in the first sections of this article, the task is to graph this summary data in such a way as to also describe the variance or spread of individual data within groups (**Figure 3**).

Graphing nominal, binary, or ordinal data leaves something to be desired. In most instances, a table serves better (**Table 2**). When the ordinal frequency counts by group are rearranged into a dichotomous data set by collapsing counts of grades I and II as satisfactory and counts of grades III to VI as unsatisfactory, the data get closer to being statistically significant but still do not meet the criterion of α set at $P < .05$. If the text were to mention that the group B intervention resulted in 57% satisfactory outcomes in comparison with group A with only 30% and a figure were used to illustrate these results without further statistical information, the reader might be misled as to the efficacy of the group B intervention. Again, the primary responsibility of high-quality graphic illustrations is to truthfully report data.

Additional Graphical Details

Line Plots

Connecting data points with a line implies that the data are continuous and that the *x*-axis intervals are equivalent. As seen in **Figure 4**, which is an incorrect use of a line plot, there is no particular relationship defining the spacing on the *x*-axis, but connecting the dots implies that there is and that the difference between apples and bananas is double that between apples and oranges (**Figure 4**).

When the data are continuous, multiple lines can illustrate comparisons between treatment arms in one plot, as seen in **Figure 5**. Of course, this plot would be improved if the spread of the data was added for each arm without overly confusing the graph.

Bar Graphs

Vertical. Bar graphs can be used for the same data as multiple line plots; the choice depends on which is easier to read. Using data from the 3-line comparison above, the data can be clustered in 2 ways. One is to group data by the date, as in **Figure 6**.

However, this does not really make sense if we are most interested in observing the month-to-month trend. In that case, we would have to choose a color and jump from one cluster to the next to see this trend. Printing in noncolor (gray scale) makes this even more difficult. It would be better to cluster by

Table 2. Frequency Counts of Grades by Groups

	Grade	I	II	III	IV	V	VI
Group A		2	5	6	3	4	3
Group B		3	10	8	2	0	0

Summary Statistics of Frequency Counts

	Mean	SD	CI of Mean	Median	25%	75%
Group A	3.833	1.472	1.545	3.5	3	5
Group B	3.833	4.215	4.423	2.5	0	8

Mann-Whitney *U P*(exact) .589 (not statistically significant).

Frequency Counts of Combined Grades

	Satisfactory (Grades I and II)	Unsatisfactory (Grades III to VI)	Number	% Satisfactory
Group A	7	16	23	30%
Group B	13	10	23	57%

χ^2, $P = .137$ (not statistically significant).

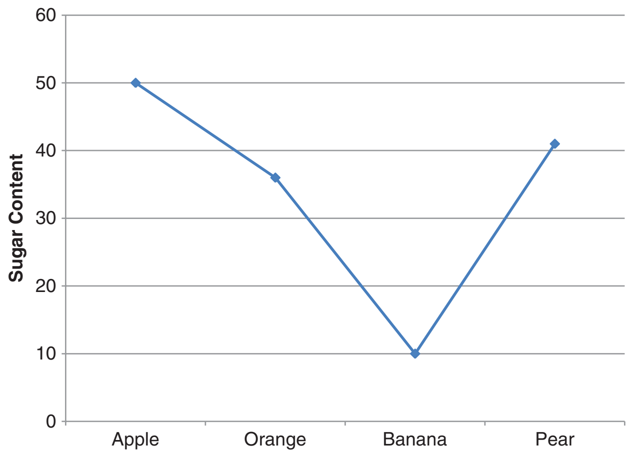

Figure 4. Wrong use of a line plot.

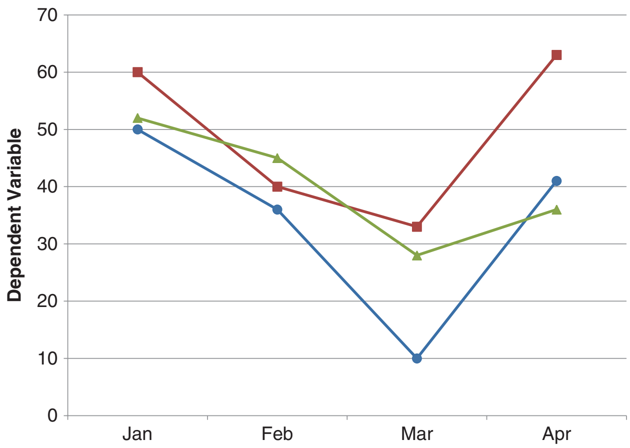

Figure 5. Multiple line plot showing changes over time in 3 groups of arbitrary data.

treatment group. Note that there is no advantage to adding extra effects, such as 3-dimensional plotting of a 2-dimensional graph. Three-dimensional effects add no information and actually make the graph less readable.

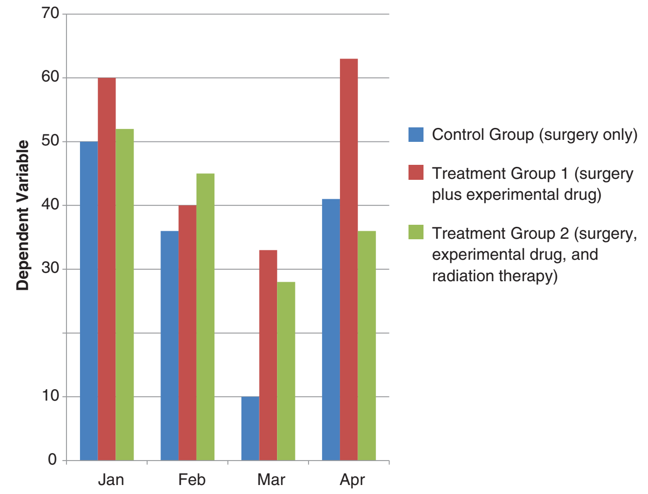

Figure 6. Vertical bar graph showing the same data in **Figure 5**.

Horizontal. Horizontal bar graphs may be useful if the labels on the x-axis are long or involved, making them hard to read when compressed into a small space (**Figure 7**). This orientation allows the cluster labels to take up whatever space they need without distorting the actual data plot.

Linear Correlations

For large sets of individual data pairs, illustrating the correlation between 2 variables, a line of best fit is usually better than connecting the dots (**Figure 8**). The trendline (best fit) shows the overall data trend, while the scatter plot still shows the range of measured values. The equation for this line can be included in the plot, if desired, along with the R^2 value, which shows a pretty good correlation for this set of data.

Scatter Plots

Scatter plots are good for a large set of paired variables with no apparent correlation. These can demonstrate clustering of data pairs (**Figure 9**). This scatter plot reveals 2 clusters of data points at the lower left and upper right.

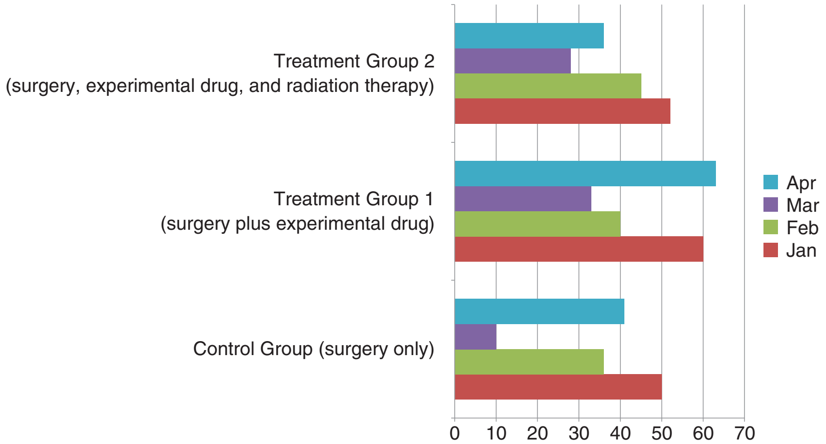

Figure 7. Horizontal bar graph showing reconfiguration of **Figure 6**.

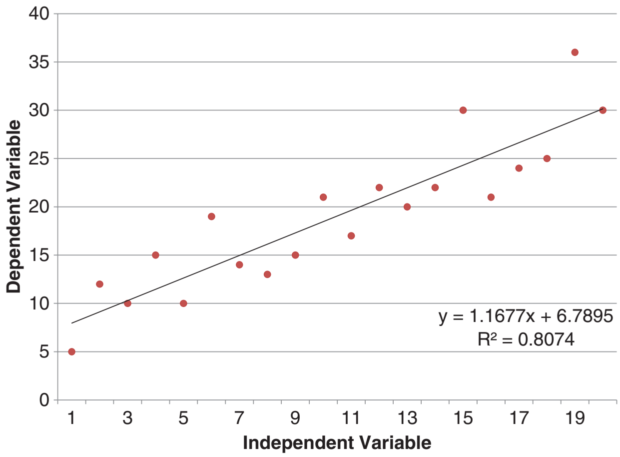

Figure 8. Line correlation of a set of arbitrary data.

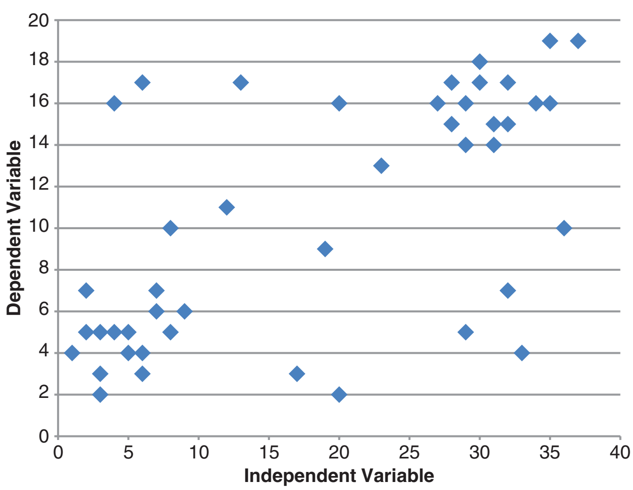

Figure 9. Scatter plot of a set of arbitrary data showing 2 clusters of data points at the lower left and upper right.

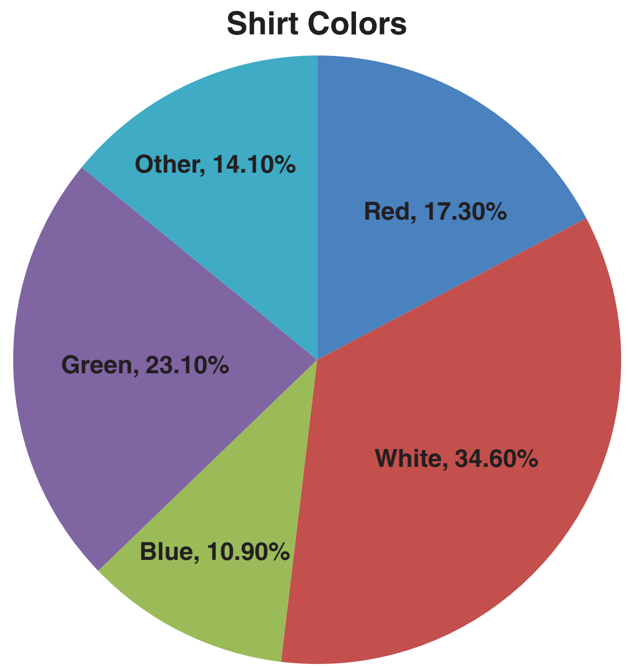

Figure 10. Pie chart of another set of arbitrary data.

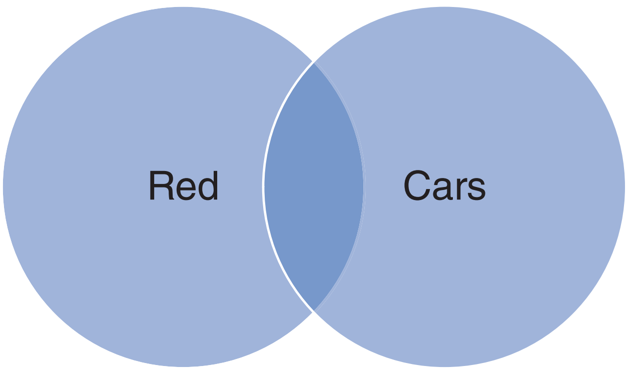

Figure 11. Venn diagram arbitrarily constructed to show the simplest form of a Venn diagram.

Pie Charts

Even though pie charts are seldom used in the scientific literature in lieu of tables, when used they can easily illustrate the complete population and the proportions belonging to each subgroup (**Figure 10**). At a glance, an alternative bar chart of subgroups does not demonstrate that the subgroups add to 100%. Inclusion of the actual subgroup values allows easy comparison of the relative sizes of pie slices.

Venn Diagrams

The Venn diagram was developed by John Venn, a Cambridge University mathematician, logican, and philosopher, in 1880. In the simplist form, a Venn diagram is 2 circles that partially overlap (**Figure 11**). Each circle contains members of a different data set, and the area in which they overlap contains members of both. For example, circle A contains red things and circle B contains specific cars; the overlapping area contains specific red cars. More complex shapes and numbers of shapes may constitue a Venn diagram. While these look simple, the logic and mathematics can be quite complex. Venn diagrams show similarities and differences between groups of data in the context of logic and probabilities and are useful in exploring the probabilities of relationships between variables.

Units on Graphs

When possible, it is important to include units on a graphic illustration. This increases the efficiency of the graph alone to communicate information.

Conclusion

The objectives of graphical illustrations are to present rapid, accurate, and concise take-home lessons of the major points in a manuscript. To produce a well-designed graph, a combination of by-subject detail and overall results should be the goal within the same illustration. The practice gap addressed

Table 3. Summary

General concepts of graphing data
Landscape best: height to width 1:1.6
Tables better for data sets 20 or less
Scatter plot or its variants are the greatest graphical designs
Statistical integrity and substance more important than design
Graphs must reliably represent actual data, be clearly labeled, and tell the truth
Concentrate on graphing only what cannot be removed; avoid "chartjunk" and make the graph as dense with data as possible, but still easily readable
Specific aspects of creating excellent graphical data
Remember the audience
Statistical rules of thumb
Strive to combine by-subject data and group data summaries into one graph
Error Bars
What they are and when to use them
Graphical examples
Descriptive statistics
Comparative statistics (inferential)
Additional graphical details
Line plots
Bar graphs
Line correlations
Scatter plots
Pie charts
Venn diagrams
Units on graphs

in this article is that little attention from authors, reviewers, editors, and publishers seems to be paid to graphical literacy. The purpose of this article is to present some practical guidelines for choosing or evaluating more appropriate data displays. A summary of these guidelines as they appear in the article are outlined in **Table 3**.

Author Contributions

Randal C. Paniello, primary author; substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published. **J. Gail Neely**, corresponding author; also primary author; substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published. **Jason T. Rich**, substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published. **Eric L. Slattery**, substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published. **Courtney C. J. Voelker**, substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published.

Disclosures

Competing interests: None.

Sponsorships: None.

Funding source: None.

References

1. Cooper RJ, Schriger DL, Close RJH. Graphical literacy: the quality of graphs in a large-circulation journal. *Ann Emerg Med*. 2002;40:317-322.
2. Cooper RJ, Schriger DL, Tashman DA. An evaluation of the graphical literacy of *Annals of Emergency Medicine. Ann Emerg Med*. 2001;37:13-19.
3. Schriger DL, Cooper RJ. Achieving graphical excellence: suggestions and methods for creating high-quality visual displays of experimental data. *Ann Emerg Med*. 2001;37:75-87.
4. Sonnad SS. Describing data: statistical and graphical methods. *Radiology*. 2002;225:622-628.
5. Tufte ER. *The Visual Display of Quantitative Information*. 2nd ed. Cheshire, CT: Graphics Press; 2001.
6. Henry GT. *Graphing Data: Techniques for Display and Analysis (Applied Social Research Methods)*. Thousand Oaks, CA: SAGE; 1995.
7. van Belle G. Statistical rules of thumb. In: http://designrulesus/presentations%5CCHSTalkNoPicturespdf [lecture notes]. Seattle, WA: University of Washington; 2003.
8. Gardner MJ, Altman DG. *Statistics With Confidence—Confidence Intervals and Statistical Guidelines*. London, UK: British Medical Journal; 1989.
9. Portney LG, Watkins MP. *Foundations of clinical research applications to practice*, 3 ed. Upper Saddle River, NJ: Prentice Hall; 2009.
10. Wang E, Ghogomu N, Voelker C, et al. A practical guide for understanding confidence intervals and *P* values. *Otolaryngol Head Neck Surg*. 2009;140:794-799.
11. Jekel JF, Katz DL, Elmore JG. *Epidemiology, Biostatistics, and Preventive Medicine*. 2nd ed. Philadelphia, PA: WB Saunders; 2001.