Achieving Graphical Excellence: Suggestions and Methods for Creating High-Quality Visual Displays of Experimental Data

From the University of California–Los Angeles Emergency Medicine Center, University of California–Los Angeles School of Medicine, Los Angeles, CA.

David L. Schriger, MD, MPH
Richelle J. Cooper, MD, MSHS

See related article, p. 13.

Graphics are an important means of communicating experimental data and results. There is evidence, however, that many of the graphics printed in scientific journals contain errors, redundancies, and lack clarity. Perhaps more important, many graphics fail to portray data at an appropriate level of detail, presenting summary statistics rather than underlying distributions.

We seek to aid investigators in the production of high-quality graphics that do their investigations justice by providing the reader with optimum access to the relevant aspects of the data. The depiction of by-subject data, the signification of pairing when present, and the use of symbolic dimensionality (graphing different symbols to identify relevant subgroups) and small multiples (the presentation of an array of similar graphics each depicting one group of subjects) to portray stratification are stressed. Step-by-step instructions for the construction of high-quality graphics are offered.

We hope that authors will incorporate these suggestions when developing graphics to accompany their manuscripts and that this process will lead to improvements in the graphical literacy of scientific journals. We also hope that journal editors will keep these principles in mind when refereeing manuscripts submitted for peer review.

[Schriger DL, Cooper RJ. Achieving graphical excellence: suggestions and methods for creating high-quality visual displays of experimental data. *Ann Emerg Med.* January 2001;37:75-87.]

INTRODUCTION

"Life is too short for *The New Yorker,*" an accomplished woman of letters quips in a recent play.[1] With the spread of computerization and digitization, images and sound are eroding the importance of the written word. Despite this, investigators and the journals that publish their work maintain an emphasis on the quality of prose while paying less attention to the quality of graphical displays. The *Annals of Emergency Medicine's* "Instructions to Authors" devotes less than 5% of its space to instructions regarding the preparation of tables and figures. In our review of the journal's graphical quality, we found that many of the graphics inadequately portrayed essential details of the data.[2] Other medical journals have similar deficiencies.[3-5]

Although there are abundant sources addressing empiric evidence,[6,7] theory,[6,8-11] style,[9-16] and practical considerations[12,16,17] regarding graphics development, it appears that too few authors, reviewers, and editors have availed themselves of these references. Perhaps, in the absence of standards for graphical style (the visual equivalent of the rules of grammar), authors feel no pressure to improve their graphics. A graphic can serve many purposes and audiences. It can convey quantitative concepts to persons with limited education, can be an eye-catching advertisement that overstates the sponsor's message through distortion of truth,[18] or can present detailed data to a sophisticated scientific audience. It is our contention that graphics in scientific manuscripts serve this final function and should be created solely with this intent. Furthermore, we believe that investigators who present detailed data empower readers to decide whether conclusions are justified. In an era when many studies are conducted or funded by groups that have a vested interest in the outcome, skepticism abounds. The presentation of detailed data is an important means of reassuring readers that conclusions are justified and that statistics have not been used to "sensationalize, inflate, confuse [or] oversimplify."[17]

In this article, we attempt to show how the kinds of graphics commonly found in our review of *Annals of Emergency Medicine* can be reconfigured so that they convey far more information. The underlying principle is that each graphic should tell the story of the investigation, showing as much of the relevant underlying data as possible in the most meaningful, unbiased way. We acknowledge that the dual goals of depicting detailed data and summarizing important results with clarity often produce contradictory demands on the graphic. The art of graphics creation is finding a format that resolves this apparent contradiction and meets both goals. The investigator who truly understands this concept will also understand that what follows is by no means a set of mandatory rules. The examples are an attempt to demonstrate the thought process that leads to high-quality graphics. In this short article, we cannot illustrate every type of graphic that has worthwhile properties. We encourage interested readers to expand their knowledge of graphical techniques by consulting the cited texts and articles.

Graphics created for the purpose of exploring data should differ from those created to present data. Although we concentrate here on the latter, we remind readers that graphical exploration of the data is the initial step in any data analysis. Although computers have greatly facilitated graphical exploratory data analysis, they have had a mixed

---

Table 1.
Steps in graphic preparation.

---

1. Carefully define what needs to be communicated.

2. Determine whether text, a table, or a graphic is the best way to communicate the information.

3. Determine the primary goal of the graphic. What is the story to be told? Is the message descriptive or comparative? Are the underlying data continuous, categorical, or summary statistics?

4. Determine the essential variables necessary to depict the study findings. Are there any additional variables that might be effect modifiers or confounders that are worthy of portrayal? (NB: This is a gestalt judgment that is not made through statistical analysis). It is sometimes worth portraying a variable to show that it is not a confounder.

5. Determine whether pairing exists and depict it when present.

6. Select a graphic format. What is the best graphical means of conveying these data? Consider whether you are fully exploiting the dimensionality of the data. Ask whether combined formats (the addition of box plots to univariate plots or either of these to the axes of scatter plots) are appropriate (see Table 2).

7. Label all axes clearly and use a consistent metric. Consider annotating key aspects of the graphic. Make sure that all abbreviations and symbols are defined. Write a legend that makes the graphic self-explanatory.

8. Edit meticulously. Does each component convey unique, essential information? Is each graphical element maximally exploited? Erase all nonessential ink.

---

effect on the construction of presentation graphics. Before graphics software was widely available, researchers worked with illustrators and artists and often found inventive methods for effectively depicting scientific data. Today most researchers create their own graphics—often using lowest-common-denominator spreadsheet software ubiquitous on personal computers. Such software may be a convenient way to examine data, but it may produce highly unsatisfactory graphics when the default choices are accepted.

How then does one approach the task of developing a graphic for the presentation of data? By determining what is to be communicated, whether this is best done with a graphic and, if so, what specific concepts must be conveyed (Table 1, Steps 1 to 3). Once this vision is established, the careful execution of steps 4 through 8 should lead to a high-quality graphic. Step 6 can be carried out in consultation with Table 2, which provides tips regarding the indications, pros, and cons of common graphic formats.

To illustrate our contention that a graphic can more effectively convey information than prose, and to demonstrate how using the steps suggested in Table 1 can enhance graphical quality, we provide examples from 3 hypothetical experiments. For each experiment, we offer a visual progression that demonstrates how suboptimal graphics can be transformed to outstanding ones by following the steps in Table 1. For each graphic, we demonstrate that qualitative improvements are accompanied by quantitative improvements in a statistic—the data den-

sity index (DDI)—that measures graphical efficiency by counting the average amount of information per square centimeter of graphic.[11]

DEPICTING DISTRIBUTIONS

In the first scenario, we consider an experiment that used a visual analog scale (VAS) to measure the pain experienced by subjects randomly assigned to 2 alternate methods of performing a procedure and reported that "method A was less painful than method B ($P<.05$, t test)" (Figure 1). Several comments can be made regarding this depiction. First, this graphic is cluttered with every imaginable type of "chartjunk."[4] Unneeded gridlines; meaningless 3-dimensional bars that create ambiguity regarding whether height, area, or volume is being compared; and shading that creates op-art–like moire patterns all distract and detract from the main purpose of the graphic. Second, because of the birds'-eye–view perspective, the heights of the bars do not correspond to the numbers they portray. If the bars were the correct height, the numeric labels above the bars could be removed because they can be read directly from the y-axis. Third, the y-axis scale, which spans only a small portion of the universe of VAS scores (which can range from 0 to 100), creates visual distortion and makes the difference between groups seem much larger than it actually is. Although not every graphic need use a zero origin (think how absurd this would be for pH), the selected scale and

Figure 1.

Mean pain score as measured on a visual analog scale (VAS) by treatment group. Some of the problems with this intentionally flawed graphic are the large amount of nondata ink (<5% of the ink conveys information), the inappropriate use of a 3-dimensional format, and the magnified y-axis scale. The data density index (DDI) is 0.05 cm^{-2}. For details, see text.
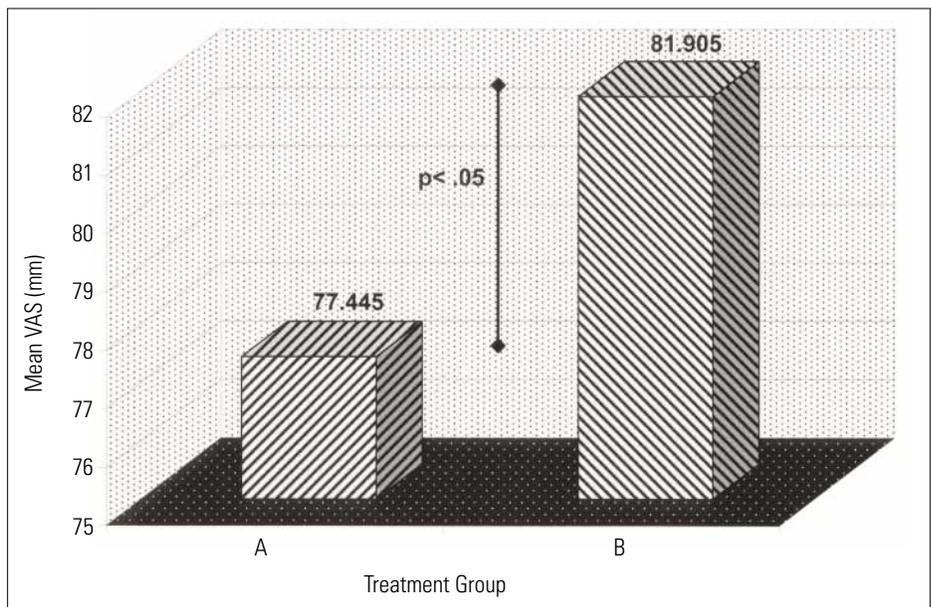
origin should be chosen to place observed differences on a scale that acknowledges the range of possible differences.

Removing the chartjunk and distortion results in a simple bar graph (Figure 2) that retains all of the meaningful content of the first figure. However, a single sentence in the article stating "mean VAS for group 1 was 77 (SD 30) and for group 2 was 82 (SD 7)" could convey equivalent information in far less space. Twenty-seven percent of the graphics we encountered in our review of *Annals of Emergency Medicine* depicted the mean of a population in this way.[2] This is a poor choice as only one or two statistics

Figure 2.

The mean pain score (VAS) with SD, by treatment group. Many of the problems with Figure 1 have been fixed, and the inclusion of SD ticks allows the reader to see that the VAS scores were more widely dispersed in treatment A. The inclusion of the number of subjects per group (N) allows the reader to understand how much data each mean is based on. Despite these improvements, the DDI of 0.09 cm^{-2} is very low.
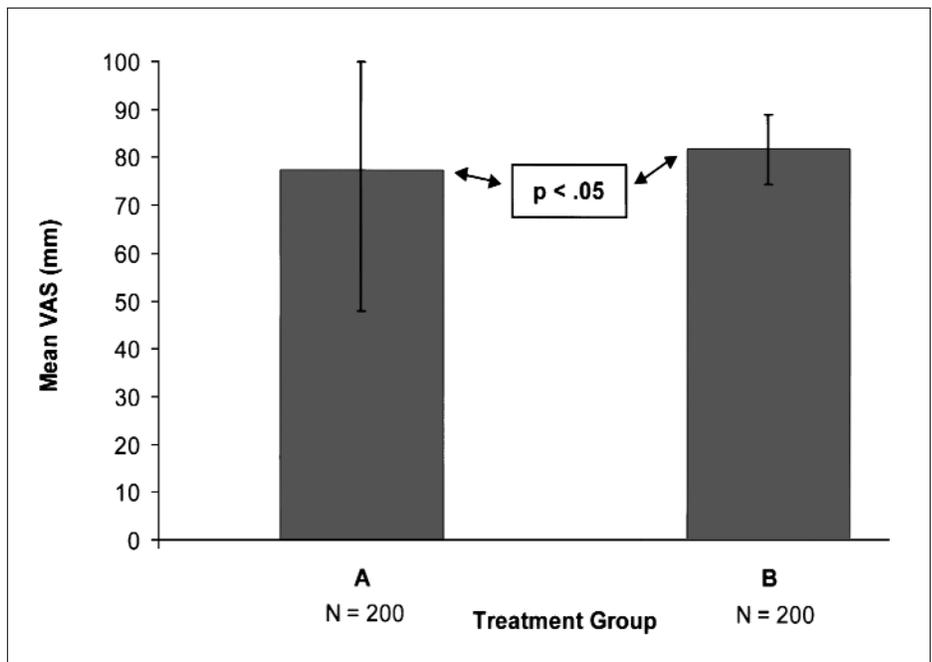
Figure 3.

VAS pain score by treatment group. This graphic depicts 2 adjacent histograms, 1 for each group. The number of subjects that reported each VAS value is shown. The x-axis is the VAS score, which ranges from 0 mm to 100 mm. The DDI of 1.5 cm^{-2} quantitatively demonstrates that Figure 3 provides far more information than Figures 1 and 2.
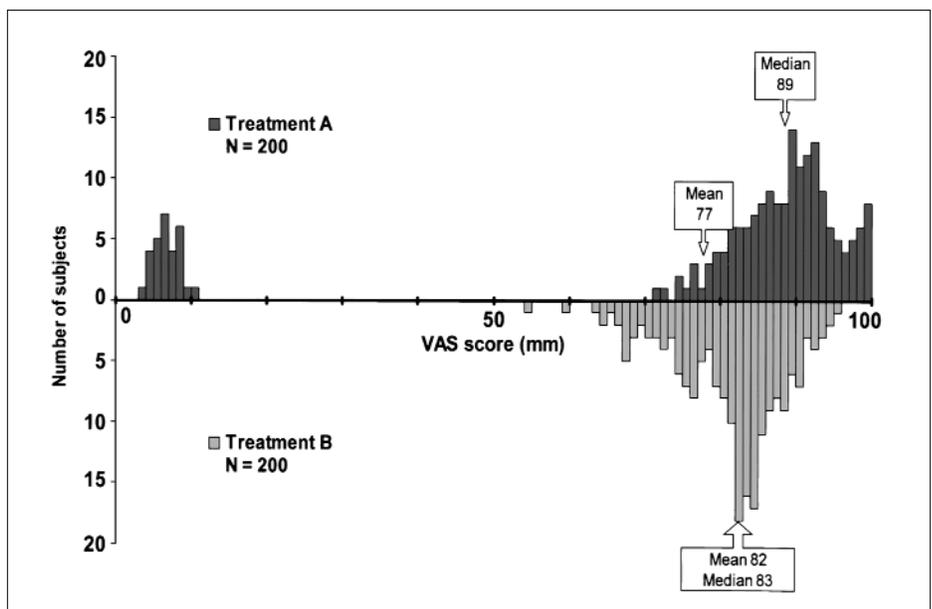
(the mean and SD) are conveyed, when the entire distribution could be shown in the same space.[19]

Better methods for displaying and comparing distributions include quantile-quantile (Q-Q plots),[20,21] probability-probability (P-P) plots,[20] adjacent stem-and-leaf plots and adjacent histograms. For this example, we chose adjacent histograms (Figure 3). This new graphic functions at 2 levels. First, we see the experiment's story in exquisite detail. The entire data set could be reverse-engineered from the picture. Second, the graphic provides a compelling overview of the experiment, far surpassing what could be conveyed with text or tables. We learn that group A's pain experience is bimodal. The mean difference between groups A and B and the statistical significance of this difference, as calculated by parametric statistics, is meaningless given these distributions. Roughly 15% of subjects in group A experience little pain—and these subjects shift the Group A mean lower than the Group B mean. On the other hand, the most common occurrence for Group A subjects is to experience more pain than the typical group B subject. Figure 3 tells this story, eliminating all of the underlying ambiguity inherent in the bar graph or the reporting of a *P* value. We no longer have to imagine what a distribution with a mean of 77 and SD of 30 might look like; we see it. Figure 3 outperforms Figures 1 and 2 in all respects and is a more accurate portrayal of the data than reporting findings in

the text as means and SDs. There is no downside. It serves well the casual reader, the interested researcher, and the investigators. It is detailed, focused, and unbiased. The DDI for graphs 1, 2 and 3 are 0.05, 0.09, and 1.5 cm^{-2}, respectively, quantitatively documenting our qualitative argument that more information is conveyed when principles of graphical excellence are followed.

For the investigator, Figure 3 is a source of new questions. Why is the distribution of group A's VAS scores bimodal? Do qualities of the subjects or providers explain why there is a "low pain" and "high pain" group? To investigate the bimodal nature of the group A VAS scores, the investigator chose to incorporate symbolic dimensionality—the use of plotting symbols to convey another dimension of detail in the data—into the graphic to aid in the search for an explanation. In this case, a number was assigned to each of the 9 treating physicians and these numbers were used as symbols in the histogram of VAS scores for the 2 group A peaks (Figure 4). It is apparent that the majority of patients who reported little pain were treated by physicians 1 or 2, and that these physicians saw very few of the patients in the "high pain" peak. This association suggests that the procedure may be operator-dependent and points the investigator toward a new hypothesis and further experiments. This graphic was not designed for publication. It is an example of how exploratory graphics should be used to understand one's data.

Figure 4.

Histogram of Group A's low pain and high pain peaks for 9 physicians. The histogram is built with numeric symbols, each signifying the identity of the physician who performed the procedure. The vast majority of less painful procedures were undertaken by physicians 1 and 2. The DDI is 2.1 cm^{-2}.
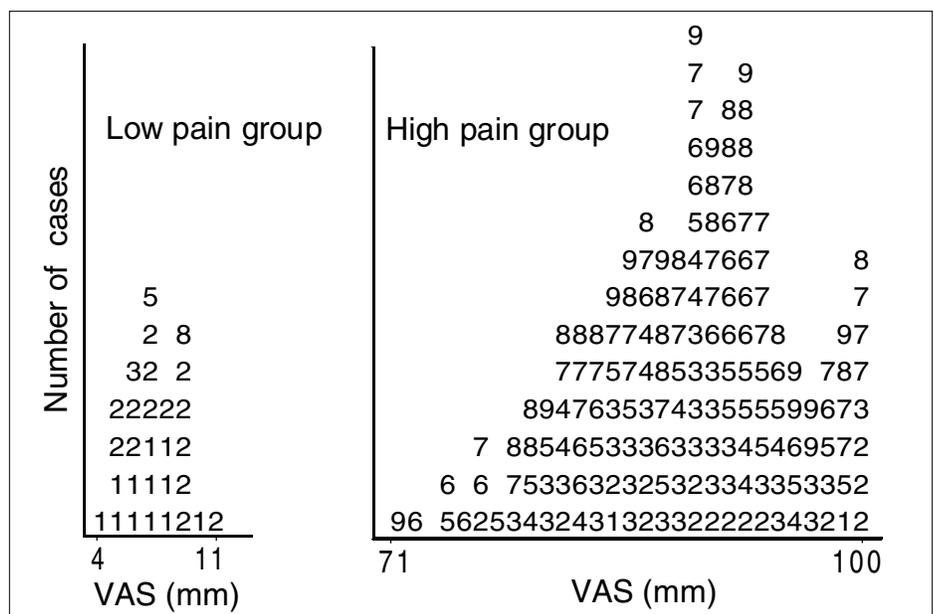
Table 2.

Indications, advantages, and disadvantages of some common graphic formats.

Graphic Type and Uses	Indications and Advantages	Disadvantages
Univariate		
Pie charts **Use:** Depiction of percent distribution among categories	**Indication:** Short answer, none. Effective for communicating percentage concepts to quantitatively unsophisticated audiences. Exception—multiple pie charts in the same graphic may allow patterns to emerge that cannot easily be gleaned from a table. (eg, using 50 pie charts to show the distribution of a nonbinary categorical variable in each United State). **Advantages:** None.	Solitary pies have no role in scientific publications since readers should be able to generate the picture from tabular data, making the picture redundant with text and tables. Pie charts are not easily modified to communicate concepts of distribution, uncertainty, and sample size, and do not handle nested data well.
Bar and point charts of categorical data **Use:** Depiction of categorical data by group.	**Indication:** To permit comparison of binary (percentage) or categorical data among groups. **Advantages:** Stacked bars or internally shaded bars can be used to further characterize different strata within the data.	Convey little information, especially when binary data for only 2 groups are being compared, since a single sentence can convey all information more efficiently.
Bar and point charts of summary statistics **Use:** To depict the mean and standard deviation of a group.	**Indication:** To trend data over time when the summary statistic is known and the underlying distribution is not. **Advantages:** Useful when many points are to be graphed and the portrayal of the underlying distributions is impossible or impractical.	When underlying distributions are available, this format needlessly condenses the distribution into 1 or 2 statistics when a box plot, one-way plot, or scatter plot would convey far more information.
One-way plots **Use:** A plot that uses 1 tick mark per observation along a solitary axis to convey the distribution of a univariate parameter.	**Indication:** To show by-subject observations of a single parameter. **Advantages:** Nicely portrays exact occurrences of events along an axis. Can be aligned with other plots for comparison purposes and, when placed side by side with a box plot conveys individual observations and distribution percentiles. Can be linked to show pairing (see Figure 6).	They are unidimensional (except when used in pairing) and therefore fail to take advantage of the 2 dimensions available on the printed page. Authors should be able to justify their decision to ignore the information potential of the orthogonal axis. Can get overly congested when the N is large. A stem-and-leaf plot or histogram may be preferred in this case.
Stem-and-leaf plots and histograms **Use:** Same as one-way plots, but use the orthogonal axis to emphasize the density of each region of the distribution.	**Indication:** To show the shape of a univariate distribution. Two can be placed back to back (see Figure 3) or interweaved for comparison sake. Can include symbolic dimensionality to show stratification (see Figure 4). **Advantages:** Excellent for showing univariate continuous data when there is no need for comparison with additional variables.	In contradistinction to one-way plots, cannot be linked to show pairing. Must choose bin size (the number of groups into which the data are divided) wisely so data are not overly compressed or dispersed.
Box-and-whisker plots (box plots) **Use:** To depict selected percentiles for continuous univariate data.	**Indication:** To convey the distribution of larger data sets in a small area, especially when multiple groups are to be portrayed (see Figure 5). Can be drawn along the axes of scatter plots to highlight the univariate distributions of the individual variables (see Figure 11). **Advantages:** Outperforms bar graphs and point charts of means in all respects. Multiple box plots are easily juxtaposed for comparative purposes.	Do not depict by-subject data (except for outliers), cannot depict pairing, and are incompatible with techniques that use symbolic dimensionality. Should not be used alone when the N is small enough to use a one-way plot or histogram.
Survival curve **Use:** A cumulative histogram that demonstrates how a population changes with respect to a single dichotomous variable over time.	**Indication:** To convey the cumulative experience of a population over time for such binary outcomes as death, cure, presence of symptoms, and so forth. **Advantages:** Easy to understand, and useful when the goal is to compare cumulative experience over time rather than the differences at discrete time points.	Although survival curves do an excellent job of conveying univariate change over time, they are not well suited for depicting additional stratifying variables, though this limitation can be overcome through small multiples–repetitive survival curves each representing one stratum of the data.
Bivariate		
Scatter plots **Use:** To portray 2 observations or calculated values both derived from the same subject. Includes plots of raw data, and Bland-Altman, Tukey, Q-Q, P-P, and residual plots.	**Indication:** An outstanding format for both diagnostic and presentation graphics. **Advantages:** Depicts the relationship of 2 measures made in each individual, thereby conveying univariate information for each variable and the linkage between the variables. Can be easily augmented with special features such as pairing, symbolic dimensionality, and small multiples.	Points may be obscured if many fall on the same coordinates. Can feasibly portray 3 (or at most 4) characteristics of each data point, which is still too few, but is the best we have.

DEPICTING PAIRED DATA

Our second hypothetical example is a randomized trial of a new inhaled medication for asthma. In this trial, peak exploratory flow rate (PEFR) is measured just before and 30 minutes after administration of a new medication (Drug N) or standard inhalant therapy (Drug S). Each drug is given to 15 asthmatic patients. Setting aside the fact that short-term change in PEFR is an intermediate outcome and a poor measure of drug performance, let us consider how the results of such an experiment might be reported. The investigators might have chosen the poor, but commonly used, method of before-and-after bar graphs of mean PEFR for each group. This graphic would have looked similar to 2 copies of Figure 2, one for each treatment group. Instead these scientists, aware that bar graphs of means are substandard, chose to use box-and-whisker plots (or box plots) (Figure 5).

The box plot is intended to provide more information about the shape of a distribution than a bar graph with SD ticks. The horizontal line within the rectangle demarcates the median (50th percentile), while the upper and lower limits of the rectangle surrounding the median line represent the 75th and 25th percentiles. This central part of the distribution contains the middle 50% of patients. The difference between the 75th and 25th percentile is called the interquartile range. The upper and lower ticks outside the rectangle, called the upper and lower adjacent values, represent the most extreme data points that are within 1.5 times the interquartile range from the end of the rectan-gle.[*] Points beyond this range are considered "outside" values and are individually signified.[8] Depending on the shape of the distribution, these outside values could be considered outliers. These conventions regarding the components of the box plot should be followed so that it is clear what is being displayed. In some box plots, the mean is also portrayed as a broken horizontal line within or beside the rectangle. Similarly, one can vary the width of the rectangle in adjacent box plots to indicate the number of subjects from which each plot is derived.[22]

Figure 5 provides considerably more information than a bar graph. We learn that the 2 groups have similar median PEFRs at the outset, although the subjects who received Drug N were more tightly clustered about their median. The median PEFR for the Drug S group increased from 220 to 430 L.min, and the distribution of subjects' post-therapy PEFRs coalesced about the median as a result of therapy. There is one outside value (signified by an ○), in this case an outlier, whose posttreatment PEFR of 200 L/min falls well below the group experience. The median improvement in PEFR in the Drug N group is smaller, from 220 to 360 L/min, and results are more disparate. Some subjects attained PEFRs higher than those achieved by the Drug S group, whereas others remained quite obstructed.

---

[*]There have been a host of alternative definitions for these ticks, most commonly the points just within the 2.5th and 97.5th percentiles. Although use of standard definitions is encouraged, until these definitions are universally adopted, it is best that each author state his or her conventions in the figure's legend.

Figure 5.
Peak expiratory flow rate (PEFR) before and after standard (Drug S) or new (Drug N) therapy. The parallel box plots illustrate the distribution of each group's baseline and posttreatment peak flow rates and are easily compared. The DDI of 0.3 cm⁻² is low.
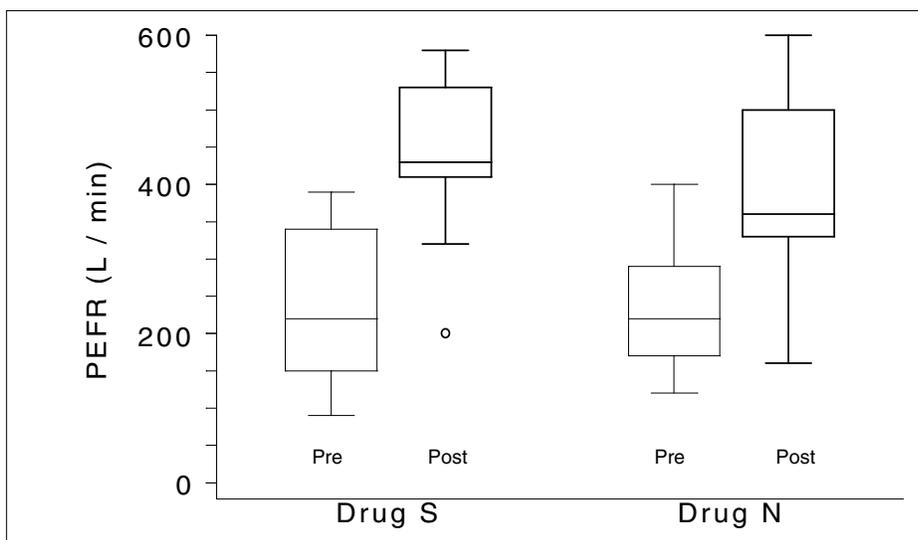
Although box plots represent a considerable improvement over bar graphs, they fail to convey what happened to individual subjects. One commonly used method of depicting paired data is to create 2 parallel aligned vertical axes (one-way plots) and use one for pretreatment values and the other for posttreatment values. By drawing a line segment between each pair of pretreatment and posttreatment points, subject-specific data can be portrayed (Figure 6, a parallel-coordinate–line-segment plot). From Figure 6a, we learn that the subjects in the Drug S group

had a fairly homogeneous response except for 3 subjects who appear to have flatter slopes (less improvement). Two versions of the Drug N response are provided (6b and 6c). Both are completely consistent with the box plots of the Drug N data presented in Figure 5; notice that the one-way plots of pretreatment values in Figures 6b and 6c are identical, as are the posttreatment plots. The only difference between 6b and 6c is which posttreatment value is linked with each pretreatment value. Figure 6b is similar to 6a; it suggests that the 2 drugs have similar effects. In

Figure 6.
PEFR before and 30 minutes after treatment, by subject. Each parallel-coordinate–line-segment plot consists of two one-way plots depicting by-subject pretreatment and posttreatment PEFR, with each subject's values linked by a line. **6a** *shows the response to Drug S. In contrast to Figure 5, the experience of each individual can be seen. The importance of depicting by-subject data is stressed by* **6b** *and* **6c,** *which provide 2 versions of response to Drug N, both of which are completely consistent with the box plots for Drug N shown in Figure 5. Note that the pretreatment values are identical for* **6b** *and* **6c** *as are the posttreatment values. Without by-subject data, it is impossible to distinguish which pattern is occurring.* **6d** *shows how this method fails when the number of subjects exceeds 20 or 25.* **6a, 6b,** *and* **6c** *each have a DDI of 5.1 cm^{-2}, which is markedly higher than in Figure 5, even though half as many subjects are portrayed. Despite depicting many more subjects, the DDI for* **6d** *is 1.4 cm^{-2} because the poor choice of format obscures most of the points, making these data unintelligible.*
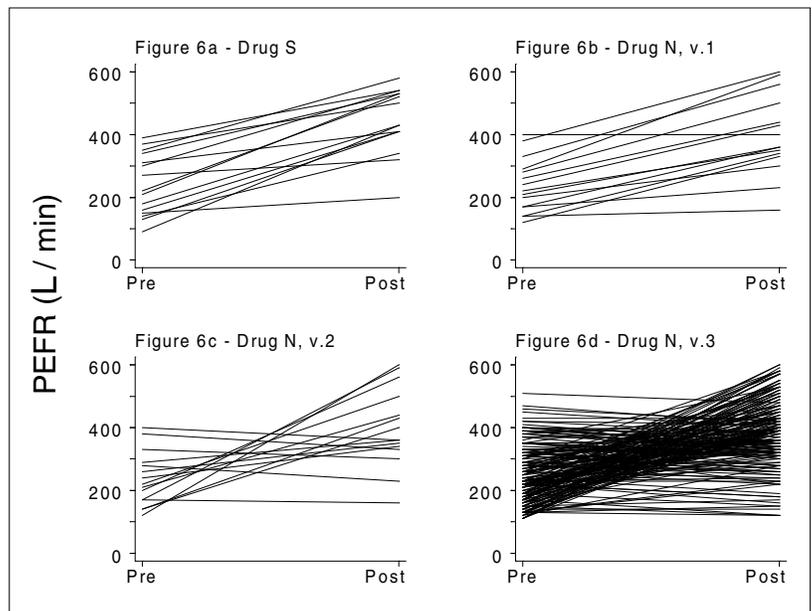
Figure 7.
Change in PEFR by subject. This is a parallel line plot of the subjects who received Drug N who were depicted in Figure 6d. Each subject is represented by a vertical line starting at the pretreatment PEFR (○) and running to the posttreatment PEFR. The length of the line portrays the magnitude of the change. The subjects are ordered by pretreatment PEFR to facilitate comparison. The graphic demonstrates that those with lower pretreatment PEFRs experience greater improvement than those with higher initial values. The graphic uses symbolic dimensionality to further stratify these data. Subjects who were already taking steroids when they presented to the ED are depicted with bold lines. This feature helps us interpret these data by making clear that steroid use is (1) associated with higher initial PEFR, and (2) associated with less consistent response to treatment. The DDI is 6.9 cm^{-2}.
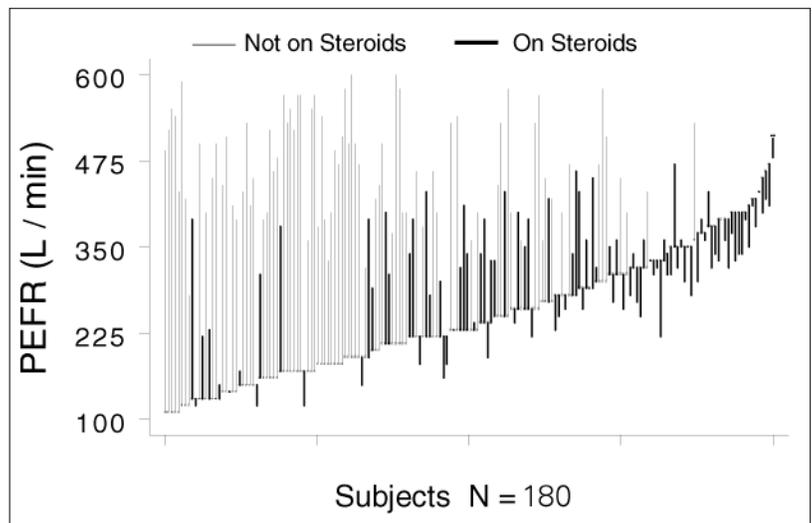
the second version of the Drug N data (6c), subjects with more airway obstruction before treatment respond superbly to the agent and achieve PEFRs that exceed those achieved by subjects with higher PEFRs at the outset. We will explore this seemingly biologically implausible finding shortly. The 2 versions of the Drug N effect are very different and lead to different conclusions regarding the drug's potential utility. If the first version (6b) is true, then Drugs S and N produce similar changes in PEFR. If the second version (6c) is true, this experiment may have established an important therapy for use in substantially impaired asthmatics. This important distinction cannot be made with Figure 5; it requires the by-subject data presented in Figure 6.

But what if the study was larger (Figure 6d)? The parallel-coordinate–line-segment technique fails and we must seek an alternative method for effectively depicting these data. There are a number of alternatives; including parallel line plots[23] and scatter plots that graph each pair's difference against the pretreatment value of the pair, the sum of the pair, or the mean of the pair.[24,25] In Figure 7, we present the data from Figure 6d (which examines only those patients assigned to receive Drug N) as a parallel line plot. We organize the subjects from lowest to highest initial PEFR, graph each subject's initial PEFR (the open circle) and draw a vertical line to the subject's posttreatment PEFR. The length of the line indicates the pretreatment-posttreatment difference. From this graphic, we can easily see that in the leftmost 5 subjects (initial PEFR 110 L/min), PEFR improved to about 475 L/min (a change of 365 L/min) after treatment, and that in the sub-

ject with the highest initial PEFR, the PEFR decreased from about 500 to 475 L/min after therapy.

In Figure 8, we graph each subject's change in PEFR on the y-axis and pretreatment PEFR on the x-axis. The horizontal line through zero helps us differentiate those subjects whose PEFR improved with Drug N from those whose PEFR did not improve. Note also that the extreme ticks on the axes in Figure 8 are labeled with the values of the highest and lowest data points rather than rounded values. Although this technique may seem nonstandard, it is another way of increasing the information content of the graphic. Interested readers may consult Chapters 4 and 6 of Tufte's *The Visual Display of Quantitative Information* for other innovative ways of improving on standard formats.[4] In contrast to the parallel line plot (Figure 7), the scatter plot (Figure 8) has the advantage of directly comparing the change in PEFR with the initial PEFR, although it has the disadvantage of not portraying each subject's posttreatment PEFR and hence, each subject's raw experience.[23]

Figures 7 and 8, by summarizing the experiment's results while portraying detailed information about each subject, approach our definition of graphical excellence. From either figure we can see that the improvement in PEFR is greatest for those with low initial PEFR, and that subjects with the highest initial PEFRs had little response to the Drug N. Parallel-coordinate–line-segment plots (Figure 6) are more commonly found in medical journals than parallel line plots (Figure 7) or scatter plots of change versus reference (Figure 8). Cleveland,[6] based on experimental findings regarding human visual perception, has

Figure 8.

Change in PEFR versus initial PEFR, by prior steroid use. This graphic demonstrates an alternate presentation of the data in Figure 7. The change in PEFR (post-PEFR minus pre-PEFR) is graphed against the pretreatment PEFR. Points above the horizontal line at zero indicate that the patient improved. The extreme ticks on the axes are labeled with the values of the highest and lowest data points rather than rounded values, another means of improving the data content of a graphic. The DDI of 5.8 cm^{-2} is lower than that of Figure 7 since the post-PEFR cannot be read directly off the graph.
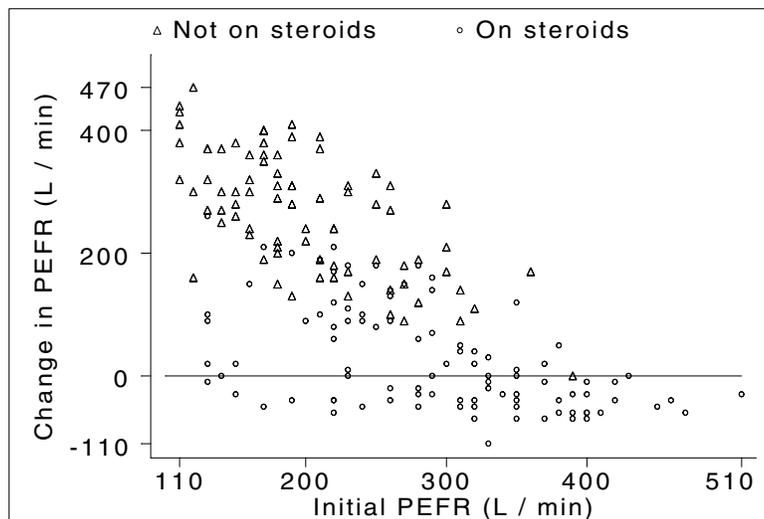
convincingly argued that there is a hierarchy of methods for comparing data which, from best to worst, proceeds as follows: (1) position along a common scale, (2) position along identical nonaligned scales, (3) length, (4) slope/angle, (5) area, (6) volume, (7) color hue/saturation/density. Because the parallel coordinate plots require that we assess slopes to compare differences (the fourth level of the hierarchy), whereas the parallel line plots and scatter plots permit comparisons of differences along the same axis (the first level), it can be argued that the latter techniques are preferred, regardless of sample size.

Because they depict by-subject data, the formats shown in Figures 6 through 8 can be further embellished to show an additional layer of stratification. It turns out that our investigators wisely noted whether each subject had been taking oral or inhaled steroids before presenting for emergency care. In Figure 7, those who were taking steroids are denoted with bold lines; in Figure 8 they are identified with Δs. This additional dimension of information enhances (and in this case corrects) our interpretation of this experiment. From either Figure 7 or Figure 8, we see that (1) those taking steroids presented with higher initial PEFRs, and (2) those already taking steroids did not improve with drug N. With this knowledge, we see the crude observation—that those with low initial PEFRs had great improvement with Drug N, whereas those with high PEFRs had no response or a negative response—makes more sense now that we understand that steroid use is an effect modifier and confounder. Because steroid use is strongly associated with both the likelihood of having a high initial PEFR and the likelihood of having little response to the drug, it confounds the interpretation of the unstratified data. Figures 7 and 8, through their depiction of paired data stratified on steroid use, increase our understanding of these results beyond that achieved in Figures 5 and 6. The DDIs for Figures 5 (0.3 cm^{-2}), 6a through 6c (5.1 cm^{-2}), 6d (1.4 cm^{-2}) (because most points are too cluttered to decipher), 7 (6.9 cm^{-2}), and 8 (5.8 cm^{-2}) support this. We again stress that the point of this example is not to find the one best graphic among Figures 6, 7, and 8, but to demonstrate how better graphics and clearer thinking about experimental results are entwined.

DEPICTING STRATIFICATION

Our final example considers an observational study relating emergency department length of stay (LOS) to patient satisfaction as reported on a continuous scale from 1 to 10. The investigators dropped patients whose LOS was more than 8 hours (although by analyzing log [LOS], which pulls in the right tail caused by patients with long ED stays, they could have achieved their purpose without discarding information). They were disappointed when their statistician informed them that linear regression analysis revealed that "longer LOSs were associated with an increase in satisfaction ($P<.001$)." However, when they graphed the data, the scatter plot of 6,063 cases with linear and lowess regression lines demonstrated no clinically

Figure 9.
Patient satisfaction versus length of stay (LOS), all patients. Scatter plot of patient satisfaction as a function of ED LOS for 6,063 subjects. As indicated by linear and lowess (see text for explanation) regression lines, there is no apparent relation between the 2 variables. In theory, the DDI could be as high as 200 cm^{-2}, but because we score only 100 observations per group, it is 3.3 cm^{-2}.
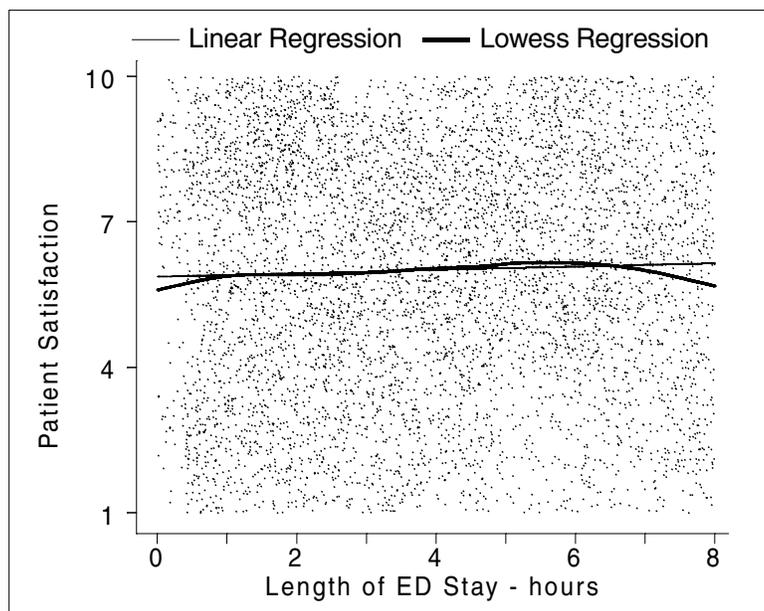
important relationship between LOS and satisfaction (Figure 9). Lowess (<u>lo</u>cally <u>we</u>ighted <u>s</u>catter plot <u>s</u>moothing) is a form of robust regression that, in contrast to simple least-squares linear regression (which assumes linearity over the universe of values), assumes only that the true regression curve has some smooth (differentiable) form.[27] Through this experience, the investigators were reminded that the use of regression models and the assessment of statistical significance must always be relegated to a second tier far below visual data analysis.[28,29]

The investigators then hypothesized that the relationship between LOS and satisfaction might be stronger for patients who are discharged than for those who are admitted. By highlighting the symbol for discharged patients (○), deemphasizing that for admitted ones, and graphing results of a lowess regression, they observed that there seems to be an appreciable inverse relationship between LOS and satisfaction for those who are sent home (Figure 10). This is another example of the importance of stratification—in this case, through the use of symbolic dimen-

Figure 10.

Scatter plot of patient satisfaction versus ED LOS, stratified by disposition. By adding the additional dimension of admission versus discharge and a lowess plot for discharged patients, we appreciate the inverse relationship of LOS and satisfaction. The DDI increases to 5.0 cm^{-2} because each point now conveys an additional piece of information.
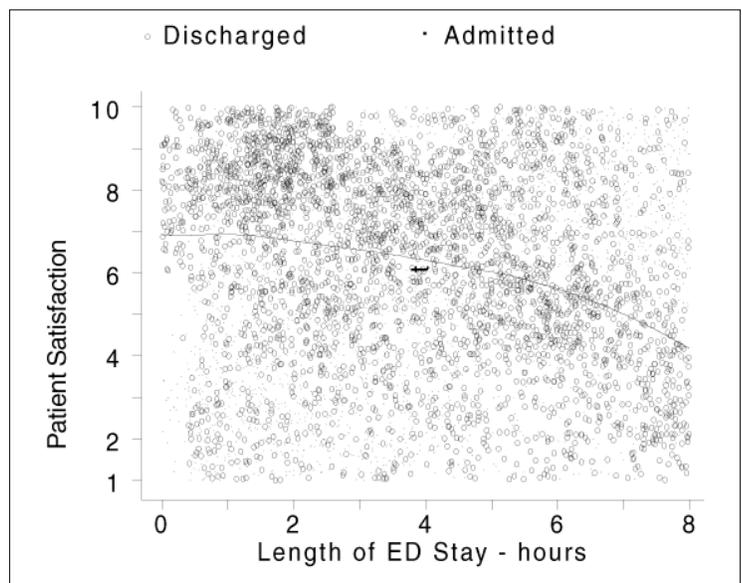
Figure 11.

Small multiple display of satisfaction as a function of LOS, for discharged patients, stratified by chief complaint. The array of small graphical elements of similar design allows the investigator to depict a great degree of detail in a small space. Many patterns emerge that cannot be appreciated in Figures 9 and 10. The box plots along the axes support univariate comparison across chief complaints (eg, "Which chief complaint was associated with the shortest median LOS?"). The DDI of 41 cm^{-2} demonstrates the power of the small multiple array to convey information succinctly.
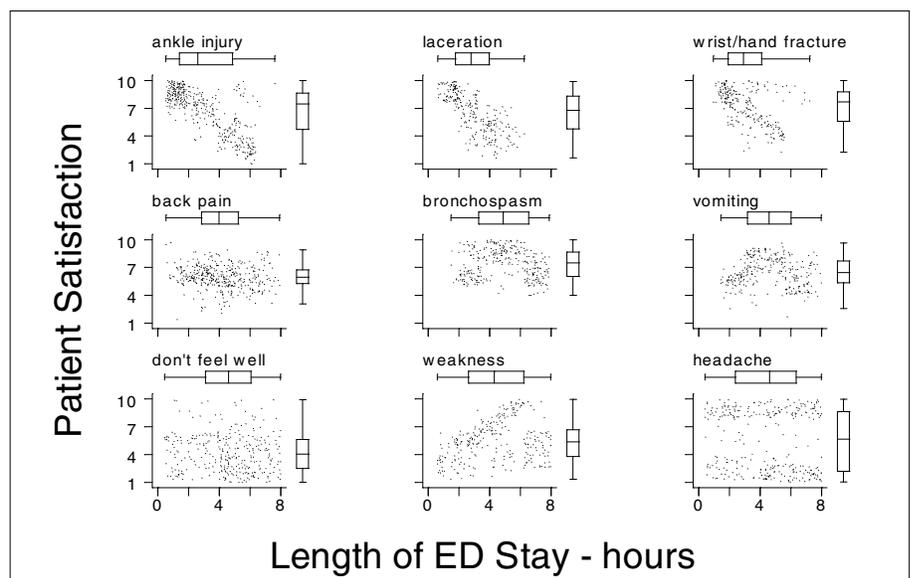
sionality—in the analysis and presentation of experimental data.

The authors then considered whether the relationship of satisfaction to LOS would differ among subjects with different chief complaints. Assigning patients with each complaint a different colored symbol in Figure 10 might be effective, but not in a journal with black-and-white figures. Symbolic dimensionality, as used in Figure 10, would fail here because the number of symbols required to portray the chief complaints of the 3,398 subjects who were discharged from the ED would produce unintelligible clutter. The alternative is the use of small multiples—repetitive graphical elements that reveal patterns within each element and between elements.

Figure 11 shows the relationship of satisfaction and LOS for discharged patients with 9 different chief complaints, each depicted by an identically structured scatter plot. Alongside the axes of each scatter plot are box plots that illustrate the univariate distributions. This feature makes the graphic multifunctional, allowing for isolated consideration of each variable (eg, "Which complaints have the shortest median LOS?"), as well as consideration of the relationships among them. Quick perusal reveals a few predominant patterns. For simple outpatient conditions such as ankle injury, laceration, and wrist fracture, there is a strong inverse linear relationship between LOS and satisfaction. For conditions that may require serial evaluation and therapy such as vomiting and bronchospasm, a ∩-shaped pattern is observed. We leave it to the reader to contemplate explanations for the various patterns but emphasize that without the use of some visual method of data stratification, these relationships would go unnoticed. Blind multivariate modeling techniques, many of which look only for linear relationships and make strong and often implausible assumptions about the interactions among variables, will also fail to appreciate these relationships.[30]

In each of these examples, the better graphic, with its higher DDI, used the same space as the less informative alternative. It was made in about the same amount of time, at the same marginal cost (pennies). The only obstacles to the achievement of graphical excellence are failure to identify those data (if any) that should be portrayed, lack of knowledge of the best display techniques, and unfamiliarity with the required software. Each of these can be overcome with education or through collaboration with knowledgeable colleagues.

The examples did not include every important graphic type. Most notably omitted are survival curves,[31] receiver operating characteristic curves,[32] dot graphs,[33] and stem-and-leaf,[7] Q-Q,[20-21] P-P,[20] Tukey,[34] and Bland-Altman plots.[25] Descriptions of these and many other techniques can be found in the cited references.

We improve as writers by reading. Likewise, we improve our capacity to achieve graphical excellence by critically appraising graphics created by others. Each reading of a scientific paper is an opportunity to increase one's graphical acumen. Consider the heart of the experiment and test the graphics with the steps in Table 1. Is a graphic truly needed? Is the essence of the study portrayed? Have the essential variables, including potential confounders been included? Has pairing been depicted? Was the correct graphic format selected, and was dimensionality fully exploited? Has the editing process eliminated nondata ink, distortion, and errors? By asking these questions of each graphic encountered, scientists will enhance their capacity to create graphics that optimally convey the results of their work.

REFERENCES

1. Margulies D. *Collected Stories.* New York, NY: Theatre Communications Group; 1998.

2. Cooper RJ, Schriger DS, Tashman DA. An evaluation of the graphical literacy of *Annals of Emergency Medicine. Ann Emerg Med.* 2001;37:13-19.

3. De Amici D, Klersy C, Tinelli C. Graphic data representation in anaesthesiological journals: a proposed methodology for assessment of appropriateness. *Anaesth Intensive Care.* 1997;25:659-664.

4. Tufte ER. Chartjunk: vibrations, grids, and ducks. In: Tufte ER, ed. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press; 1983:107-121.

5. Cleveland WS. Graphs in scientific publications. *Am Stat.* 1984;38:261-269.

6. Cleveland WS. *Elements of Graphing Data.* Monterey, CA: Wadsworth Advanced Books and Software; 1985:229-294.

7. Wainer H, Thissen D. Graphical data analysis. *Ann Rev Psychol.* 1981;32:192-241.

8. Cleveland WS. *Visualizing Data.* Summit, NJ: Hobart Press; 1993.

9. Tufte ER. *Envisioning Information.* Cheshire, CT: Graphics Press; 1990.

10. Tufte ER. *Visual Explanations: Images and Quantities, Evidence and Narrative.* Cheshire, CT: Graphics Press; 1997.

11. Tufte ER. Data density and small multiples. In: Tufte ER, ed. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press; 1983:161-175.

12. Briscoe MH. *A Researcher's Guide to Scientific and Medical Illustrations.* New York, NY: Springer-Verlag; 1990:7-14, 75-107.

13. Briscoe MH. *Preparing Scientific Illustrations: A Guide to Better Posters, Presentations, and Publications.* 2nd ed. New York, NY: Springer-Verlag; 1996.

14. Scientific Illustration Committee. *Illustrating Science: Standards for Publication.* Bethesda, MD: Council of Biology Editors; 1988.

15. Cox DR. Some remarks on the role in statistics of graphical methods. *Appl. Statist.* 1978;27:4-9.

16. Wainer H. How to display data badly. *Am Stat.* 1984;38:137-147.

17. Huff D. *How to Lie With Statistics.* New York, NY: WW Norton; 1954:8.

18. Wilkes MS, Doblin BH, Shapiro MF. Pharmaceutical advertisements in leading medical journals: experts' assessments. *Ann Intern Med.* 1992;116:912-919.

19. Rashcalf P, et al. Advanced applications of the theory of the bar chart. In: Scherr GH, ed. *The Journal of Irreproducible Results.* 3rd ed. New York, NY: Dorset Press; 1986:243-244.

20. Wilk M, Gnanadesikan R. Probability plotting methods for the analysis of data. *Biometrika.* 1968;55:1-17.

21. Cleveland WS. *Visualizing Data.* Summit, NJ: Hobart Press; 1993:21-24.

22. McGill R, Tukey JW, Larsen W. Variations of box plots. *Am. Stat.* 1978;32:1-12.

23. McNeil D. On graphing paired data. *Am Stat.* 1992;46;307-311.

24. Cleveland WS. *Visualizing Data.* Summit, NJ: Hobart Press; 1993:110-125.

25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;8476:307-310.

26. Cleveland WS. *Elements of Graphing Data.* Monterey, CA: Wadsworth Advanced Books and Software, 1985:254.

27. Cleveland WS. *Visualizing Data.* Summit, NJ: Hobart Press; 1993:167-178.

28. Gallagher EJ. p<0.05: threshold for decerebrate genuflection. *Acad Emerg Med.* 1999;6:1084-1087.

29. Rothman KJ, Greenland S. *Modern Epidemiology.* Philadelphia, PA: Lippincott Williams & Wilkins; 1998:183-199.

30. Rothman KJ, Greenland S. *Modern Epidemiology.* Philadelphia, PA: Lippincott Williams & Wilkins; 1998:359-399, 401-432.

31. Cox DR, Oakes D. *Analysis of Survival Data.* New York, NY: Chapman and Hall; 1984.

32. Metz, CE. Basic Principals of ROC analysis. *Semin Nucl Med.* 1978;8:283-298.

33. Cleveland, WS. Graphical methods for data presentation: full scale breaks, dot charts, and multibased logging. *Am Stat.* 1984;38:270-280.

34. Cleveland WS. *Visualizing Data.* Summit, NJ: Hobart Press; 1993:23.