METHODOLOGY

Statistical Graphics in *AJG*: Save the Ink for the Information

Jason T. Connor, PhD, FACG[1,2]

Am J Gastroenterol 2009;104:1624–1630; doi:10.1038/ajg.2009.259

Often when reviewing papers for the *American Journal of Gastroenterology* (*AJG*), I find myself thinking how the essence of a manuscript could be summarized with just one or two informative statistical graphics. Sometimes the authors include this key plot, but often it is a missed opportunity. The mind is much better at remembering graphics than tables. Feliciano *et al.* demonstrated that journal readers were faster and more accurate at understanding comparisons when data displays complemented textual descriptions (1). Furthermore, a well-constructed plot can be more convincing to a reader than a statistical test (2).

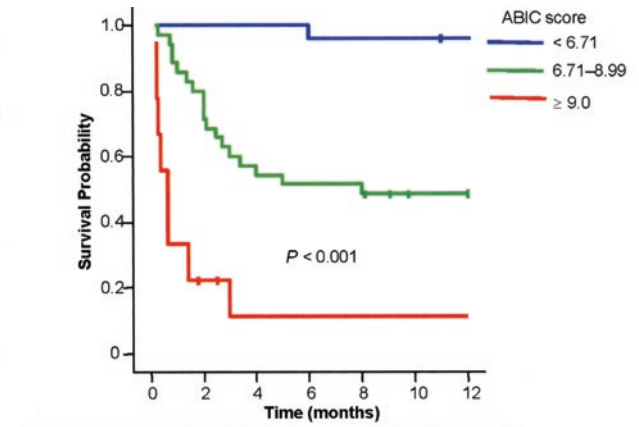
Our task as authors is to translate data into knowledge honestly, clearly, and without bias. But too often we (and I am guilty too) report our data via tables. This simple method of writing may sometimes fail to reach the point of translating our data into knowledge for the reader.

There is a growing statistical/psychological literature on the theory of statistical graphics. The simplest conclusion is that tables are best if you want your readers to be able to look up specific information and graphics are best for illustrating trends and making comparisons (3–5). Typically, in the medical literature, the focus is less on precise numerical estimates (patient populations tend to be quite variable from practice to practice) and more on relative comparisons of one treatment with another, or the magnitude of a risk due to a risk factor. Therefore, plots frequently highlight these comparisons better than tables.

During our editorial board discussions, a common question asked of individual papers being considered for publication is "Will this change clinical practice?" Even as a statistical reviewer, it is a question I keep in mind when reviewing papers. Some papers are packed so full of information it is difficult to determine which parts the clinician should remember and incorporate into her practice. I often comment in my reviews that the more numbers are included in a manuscript, the fewer a reader is likely to remember. When translating knowledge from *AJG* to his practice, a gastroenterologist mostly cares about the direction, the magnitude, and the certainty of a relationship. All three of these points can be succinctly displayed in a clever plot more memorably than in a table.

In addition to being more memorable than tables, graphics can also illustrate multiple dimensions and important relationships far better than tables. Edward Tufte says it best (6): "Modern data graphics can do much more than simply substitute for small statistical tables. At their best, graphics are instruments for reasoning about quantitative information. Often the most effective way to describe, explore, and summarize a set of numbers—even a very large set—is to look at pictures of those numbers. Furthermore, of all methods for analyzing and communicating statistical information, well-designed data graphics are usually the simplest and at the same time the most powerful."

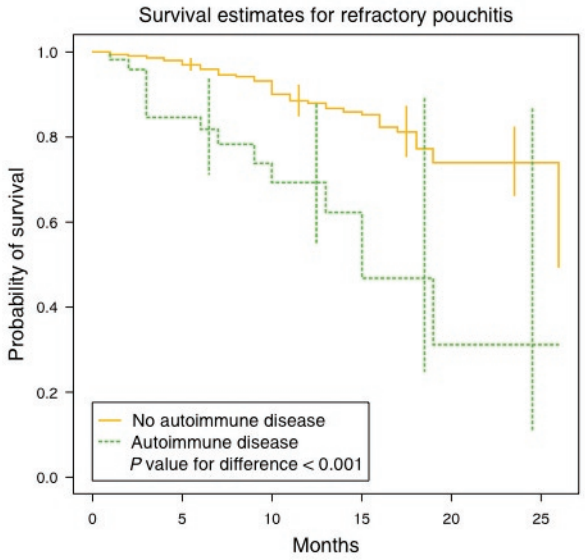
In this paper I highlight some of the best plots published in *AJG* in the past year, show some graphics that could be improved, and illustrate how perhaps one plot could convey information more directly than a table.



Kaplan-Meier curve showing 1-yr survival analysis in the study cohort using a cutoff of the ABIC score of 6.71 and 9.0.

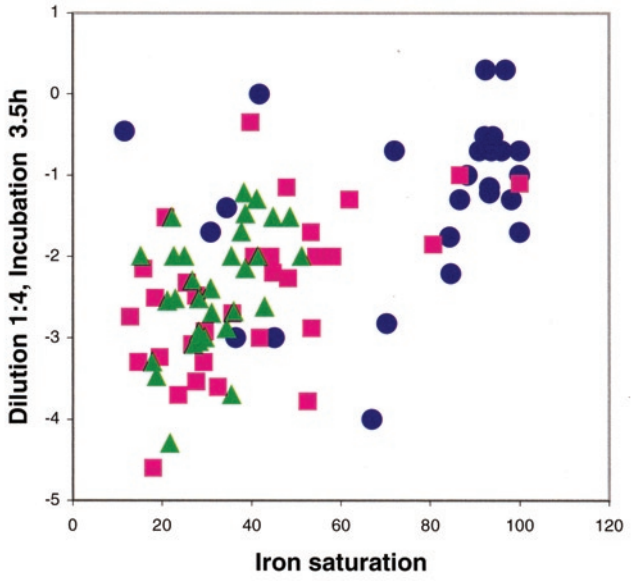
Figure 1. Kaplan–Meier plot showing all data and illustrating precision by showing sample size at each 2 months of follow-up.

[1]Berry Consultants, Orlando, Florida, USA. [2]University of Central Florida College of Medicine, Orlando, Florida 32826, USA. **Correspondence:** Jason T. Connor, PhD, FACG, Berry Consultants, 9757 Cypress Pine Street, Orlando, Florida 32827, USA. E-mail: jason@berryconsultants.com



Kaplan–Meier plots for the comparison of survival estimates for refractory pouchitis in patients with or without autoimmune disorders.

Figure 2. Kaplan–Meier plot showing precision via 95% confidence intervals.



Relationship between transferrin saturation and antibacterial activity level (serum dilution 1:4, incubation 3.5 h) in the whole population. Control (C; triangles), iron-depleted hemochromatosis patients (DHH; squares), and iron-overloaded hemochromatotic patients (HH; circles) ($P < 0.0001$).

Figure 3. Plot effectively showing three dimensions on the two-dimensional page.

Rules of statistical graphics

Tufte has written multiple books on the topic of effective graphical communication (6–9), and the *New York Times* called him "the da Vinci of Data" (10). He states a set of rules (including obvious points such as "graphical excellence requires telling the truth about the data"), but we can reduce his set of rules to one key principle: "graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space" (6).

The latter points in this principle are key, though sometimes Tufte takes it to extremes in his own plots. The eye is drawn to the color—the ink—within plots. Therefore the ink should be used to show the important parts of the plot, not merely to decorate the page. Tufte suggests that just as we prune unnecessary words when editing our own papers before submission, we should similarly edit our graphics and prune unnecessary ink.

If we may add one more rule, it is to show all the data whenever possible. We need not show just a mean and standard error or a proportion and confidence intervals when we could instead show all the data points in a study and let the reader study the whole data set for himself.

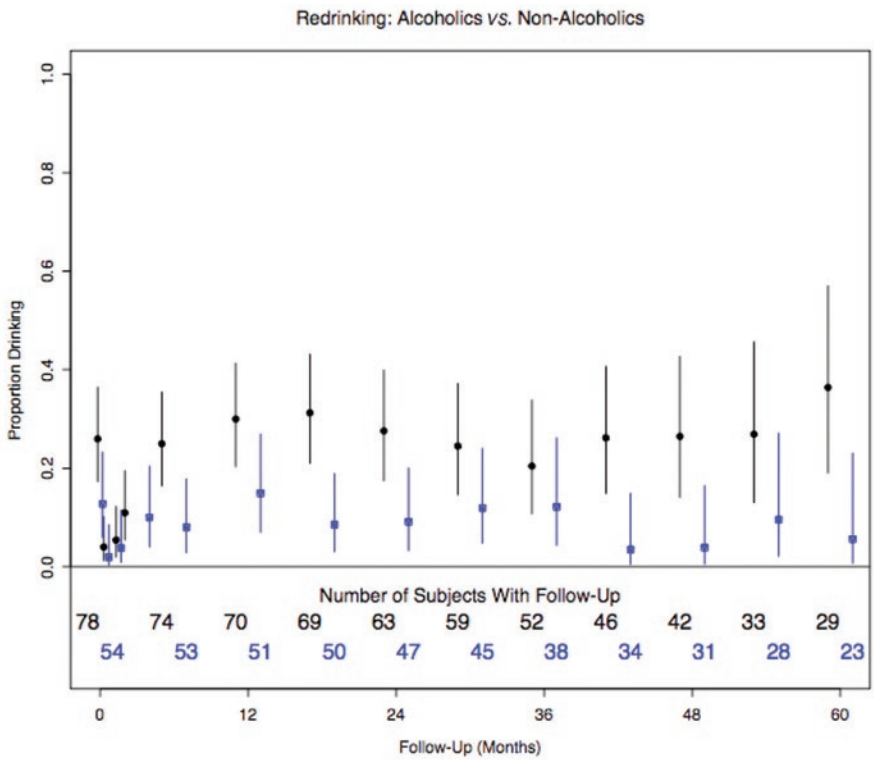
Great plots in *AJG*

Just as a discussion section should detail study limitations, the best statistical graphics illustrate both what we should conclude and what we cannot conclude from our data.

For example, Kaplan–Meier curves are ubiquitous in *AJG* and can be fine examples of effective plots: they are simple and memorable, they show all the data, and the majority of the ink is used in the valuable information. But each curve is merely an estimate of the true survival function the authors are trying to estimate. Therefore each Kaplan–Meier curve has variability associated with its estimate. Kaplan–Meier curves published in *AJG* in the past year ranged from summarizing very few patients and thus having large variability (11) to summarizing several thousand patients and therefore providing much more precise estimates (12).

Figure 1, from Dominguez *et al.* (13), shows survival up to 1 year based on a liver function score devised by the authors. The plot shows all the relevant data: time of death by vertical drops and the times patient data are censored by small ticks. Furthermore, the bottom of the figure shows the number of patients being tracked in 2-month intervals. This is vital information and gives the viewer an indication of the precision surrounding the estimated survival curves. Approximately 25% of Kaplan–Meier curves in *AJG* show this additional information, though all could benefit, since it paints a clear picture for the reader.

An alternative is to show the 95% confidence intervals for the survival curves, as in **Figure 2**, from Shen *et al.* (14). This method can become messy when there are three or more survival curves on a plot but is very effective here. Note, however, that this plot fails to show all of the data, since it fails to show censoring times for surviving patients.



Reported alcohol use during the DIVERT study stratified according to a prestudy diagnosis of alcoholic liver disease (circles) and nonalcoholics (squares). Shown as estimated proportion with 95% confidence interval.

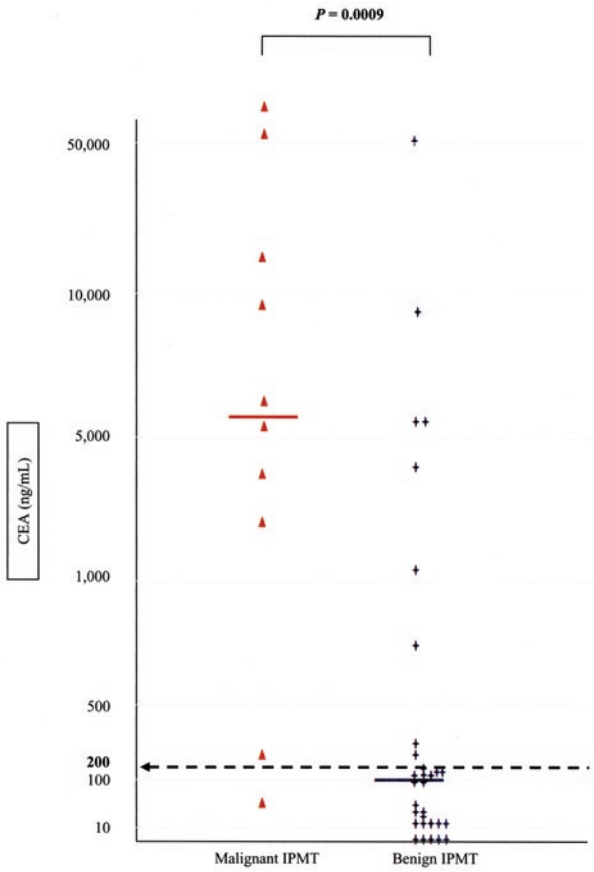
Figure 4. Plot showing longitudinal estimates of drinking in alcoholic vs. non-alcoholic liver disease patients.

Effective statistical graphics also illustrate multiple data dimensions on the two-dimensional page. One of the first famous examples of three-dimensional data on the printed page is John Snow's famous "ghost map" illustrating London's 1854 cholera deaths centered around the Broad Street pump (15), which has been reprinted often (16). A recent *AJG* example is **Figure 3**, from Jolivet-Gougeon *et al.* (17). These authors clearly illustrate that their three study groups have three different transferrin saturation–antibacterial activity level relationships. The plot shows all of the data (not results of a model or summary statistics), and the vast majority of the ink is used descriptively, not decoratively. Such plots are extremely valuable for comparing the relationship of two continuous variables across a set of groups.

Figure 4, from Lucey *et al.* (18), is another example of showing multiple data dimensions and reserving the ink for the information. We studied alcohol consumption by patients after either transjugular intrahepatic portosystemic shunt or distal splenorenal shunt. This plot compares re-drinking in alcoholic versus non-alcoholic liver disease patients over time. Most patients stop drinking shortly after the procedure, but 30% of alcoholics rapidly revert to drinking, a proportion that remains steady through 5 years of follow-up. Like the Kaplan–Meier plots shown above, the plot also shows the number of patients being tracked. I drew this original graphic, and in hindsight I would add a legend in the white space at the top so the viewer could tell which two groups (alcoholics and non-alcoholics) are displayed by the blue and black bars.

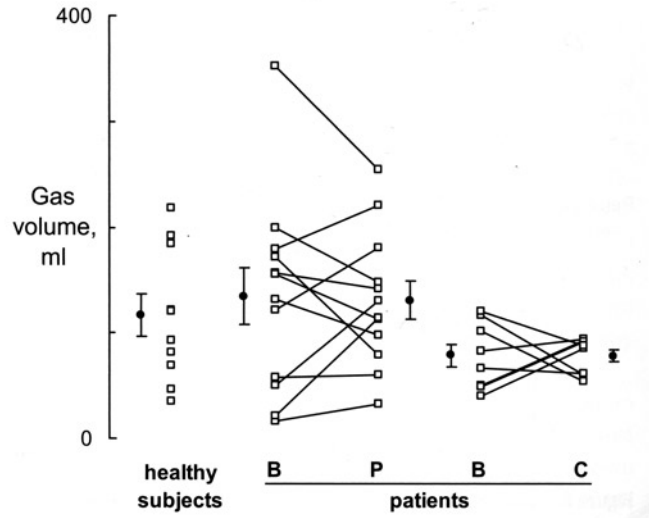
A common shortcoming of plots is that they fail to show all the data although they easily could. Many studies in *AJG* contain groups of less than 100 or just a few hundred patients. But rather than showing all the data, authors show plots that simply summarize the mean or some other summary statistics. For example, Accarino *et al.* had only 20 patients and 10 healthy controls in their study (19). But when showing gas distribution in the gut, they show the means over their 20 or 10 patients instead of showing all the data, which could easily be shown for their small data set. (To be fair, I cite a positive graphical example from the same article below).

A better example is in an article by Maire *et al.*, who studied pancreatic fluid to predict malignancy (20). They show their complete data set for the 41 patients they studied. For example, in **Figure 5**, they show the entire distribution of carcinoembryonic antigen (CEA) levels in patients with both malignant and benign intraductal papillary mucinous neoplasms (IPMNs). A quick study of the plot reveals how much information is contained in this simple graphic. This is one example of a plot that nearly summarizes the entire study. I believe it is more informative than their likely alternative: showing a receiver operating characteristic (ROC) curve. The plot shows CEA levels



Distribution of CEA levels in fluid obtained by EUS-FNA in 41 patients with IPMN (31 benign IPMN and 10 malignant IPMN). The medians are represented by the horizontal bars. The comparison of CEA levels was performed by the nonparametric Mann-Whitney test. The dotted arrow represents the discriminate cutoff value.

Figure 5. Extremely informative plot illustrates sensitivity, specificity, positive predictive value, and negative predictive value. It also lets the reader see how these metrics would change if the cutoff were changed.



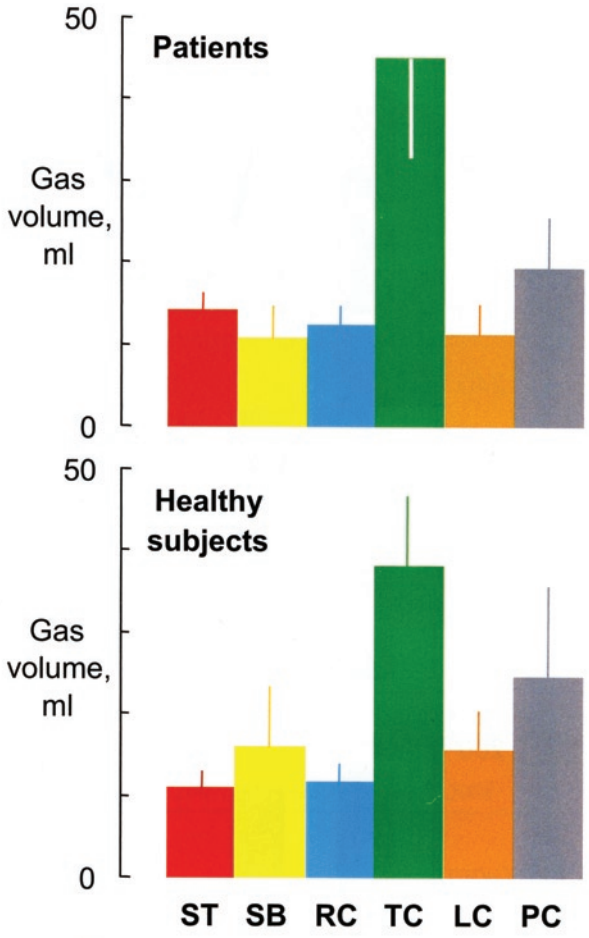
Total volume of gas in the gut. Figure shows individual data and mean values (± SE) in each group. No differences between groups were observed. Furthermore, volumes during administration of pyridostigmine (P) or placebo (C), as control, were similar as during basal conditions (B) in each group of patients.

Figure 6. An example of a plot that completely describes the study. The plot shows a failure in randomization but nevertheless illustrates the lack of effectiveness of the treatment compared with placebo.

for all 41 patients and lets the reader see how sensitivity (the proportion of red triangles above the horizontal line at CEA = 200, 90%) and specificity (the proportion of blue crosses below the line at CEA = 200, 71%) would change if the cutoff changed. Furthermore, it allows the reader to easily visualize the positive predictive value (the proportion of points above the line that are malignancies, 50%) and the negative predictive value (the proportion of points below the line that are benign IPMNs, 96%). One can even visualize the 2 × 2 table in which we would classify patients as diseased or not vs. predicted diseased or not.

Although there have been many more excellent plots in *AJG* in the past year, the final one I cite here was drawn by Accarino *et al.* (19). In this small study, the authors randomized 20 patients complaining of abdominal bloating and studied their intra-abdominal gas volume using computed tomographic images. They randomized 12 patients to pyridostigmine and 8 to placebo and also studied gas volume in 10 healthy controls. **Figure 6** effectively shows all the data and succinctly describes the whole study.

The first thing we see is that the randomization may have failed in this small study. Therefore this plot effectively and fairly highlights a shortcoming of the study (due in no part to the authors). Basal gas volumes tended to be higher in patients randomized to pyridostigmine than to control. In fact, control patients tended to have lower gas volumes than asymptomatic healthy controls. (If I could make one change, I'd let the error bars show means and 95% confidence intervals rather than standard errors.) Even though the randomization was less than ideal, the (typically) diagonal lines illustrate every patient's response to treatment. Downward-sloping lines illustrate improvements, and upward-sloping lines illustrate increased gas volumes. A quick look at the plot shows that six pyridostigmine patients improved, five worsened, and one had essentially no change in gas volume after treatment. Similarly, in the control group, four patients improved (one just slightly) and four worsened. The plot would also allow the reader to see an interaction between baseline gas volume and effectiveness if one existed. Here there is none, but, for example, if the treatment had tended to be highly effective only in patients with high baseline gas volumes, it would clearly be seen by the steepest slopes coming from patients with high baseline values. This important result would be obscured by a *t*-test alone. These are very informative plots that I would like to see more often in studies that involve comparing pre-to-post changes between treatment groups.



Gas distribution within the gut. The content of gas in stomach (ST), small bowel (SB), as well as right (RC), transverse (TC), left (LC), and pelvic (PC) colon was similar in patients with bloating and in healthy subjects.

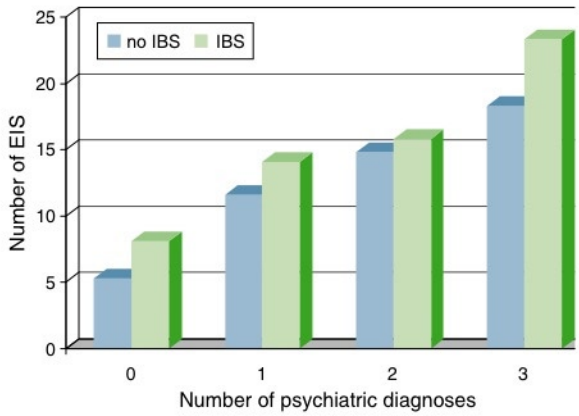
Figure 7. The superfluous color overwhelms the plot and obscures the more important standard errors that allow gastrointestinal segments to be compared.

The paper by Accarino *et al.* is a rare and laudable one because the authors use no tables and present all of their data in graphical form (19). Besides the less-than-ideal bar plots that I'll discuss below and these highly effective longitudinal plots, the authors also show some fantastic computed tomographic images of gas in the abdomen.

Not-so-great plots in *AJG*

In this section I highlight just a few counterexamples. I am certain I could populate this section solely with examples from my own manuscripts, so I hope the authors whom I discuss here do not take offense.

Quoting Tufte, I have stressed how ink should be reserved for information, and, other than their axes and labels, all of the ink in the plots I've shown is used to convey information. Plots that violate this principle are ubiquitous not only in *AJG* and other medical literature but everywhere graphics can be seen.



The relationship between number of extra-intestinal symptoms (EIS) and number of psychiatric diagnoses. IBS, irritable bowel syndrome.

Figure 8. Three-dimensional graphics make the *y*-axis more difficult to read.

One of the hundreds of examples in *AJG* is from Accarino *et al.* (**Figure 7**) (19). While the plot is very colorful, the vast majority of the ink is superfluous. In fact, the solid bars obscure a key component of the plot. There are three pieces of information per group in this plot: mean gas volume and the upper and lower bounds of the 68% confidence interval. The authors presumably (it is not detailed in the caption) show the standard error rather than letting the vertical line represent the 95% confidence interval; thus one standard error in each direction creates a 68% confidence interval. The solid bars cover the lower half of the standard error bars, making comparisons between groups more difficult because we are left to imagine the downward half of the bar. In fact, since the gut is never gasless, shading the plot all the way to zero is misleading—zero is not a relevant figure here, but the eye is more drawn to zero than to the more important lower bound of the estimate of gas per segment.

An improved plot could show all the data, as there are just 30 subjects measured repeatedly. If, however, the means are the most important information to convey, we could improve the plot by showing just the means and confidence intervals. An example is shown in **Figure 4** above, where it is much simpler to compare drinking rates between alcoholic and nonalcoholic patients and compare alcoholics' rates over time because the ink is used only to convey important information, not to obscure it.

Another way that style complicates plots is by adding false depth. For example, **Figure 8** is one of the many plots that are "three-dimensional" (21). Not only is there superfluous ink on the bar graphs that could be better used to convey additional information (e.g., 95% confidence intervals or all of the data points), but the superfluous ink added to provide pseudo-depth requires diagonal lines at the *y*-axis that make the *y*-axis more difficult to read. The horizontal lines in the background guide the eye leftward but do not line up with the scale on the *y*-axis. The plot makes another common error with a vague *y*-axis label.

Proportion of Patients With Resolution of Investigator-Assessed Upper Gastrointestinal Symptoms at 26 wk (Intention-to-Treat Population)

Symptom	Esomeprazole 20 mg n/N (%)	Placebo n/N (%)	*P* Value (CMH)
Epigastric pain	47/56 (83.9)	28/42 (66.7)	0.047
Epigastric burning	32/44 (72.7)	18/31 (58.1)	0.1876
Epigastric discomfort	43/63 (68.3)	29/57 (50.9)	0.0533
Heartburn	35/39 (89.7)	28/42 (66.7)	0.0131
Acid regurgitation	19/22 (86.4)	13/23 (56.5)	0.0290
Nausea	25/27 (92.6)	11/14 (78.6)	0.1988
Bloating	67/86 (77.9)	41/62 (66.1)	0.1126

CMH = Cochran-Mantel-Haenszel χ^2 test (stratified by baseline severity).

Figure 9. Table comparing symptom resolution in treatment vs. control group may be better illustrated with a graphic.

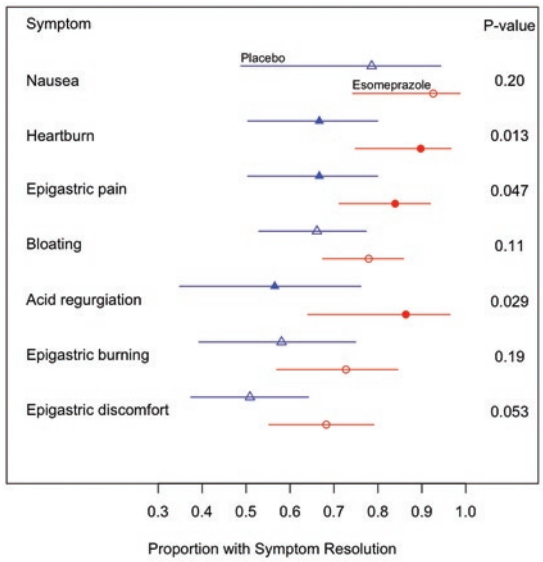


Figure 10. Graphical version of the table shown in **Figure 9**. Symptom resolution. Observed proportions are shown by blue triangles for the placebo and red circles for esomeprazole. Horizontal lines show the 95% confidence interval. Solid symbols represent statistically significant differences at $\alpha = 0.05$.

Here the authors presumably show the mean number of extra-intestinal symptoms in IBS and non-IBS patients. But both the label and the caption are unclear. Furthermore, replacing the simple boxes with either box plots or, better yet, all the data (as in **Figure 5**) would convey more information. While color can be very informative, the authors here use three shades of green to convey just five numbers. They are to be complimented for adding a legend on the plot, not merely in the caption, something I failed to do in my own plot (**Figure 4**).

A final example of a not-so-great plot is when plots are not used at all. Again there are many examples, but I chose a table from Yeomans *et al.* (22) (**Figure 9**). They show a table comparing symptom resolution for patients randomized to esomeprazole and placebo. There is nothing wrong with this table, and it conveys substantial information, but the layout makes it suboptimal for a quick comparison. By studying it we see that symptom resolution is higher for all seven symptoms and the difference is statistically significant at the 5% level for three of the seven symptoms. The ordering of the seven symptoms seems to contain no information to guide the reader's comprehension.

In **Figure 10**, I use the same data and present the information in graphical form. We can see much more readily that symptom resolution is higher for all seven symptoms in patients randomized to esomeprazole and that the relationship is significant in three of seven symptoms (you'll probably quickly notice without even reading the caption that symbols are solid when the relationship is statistically significant at the $\alpha = 0.05$ level and hollow otherwise). Also, rather than a seemingly random order, the symptoms are arranged from most often to least often resolved.

Summary

Nature Publishing Group allows *AJG* authors to produce color plots in the printed journal at no additional cost to authors. In addition to full-color pathology slides or colonoscopic images, authors can take advantage of this by producing informative color plots that complement their manuscripts.

Here I illustrate a few of the most informative plots from recent issues of *AJG*. While these plots convey different types of information, they all reserve their ink for the information, and as a result, readers quickly and efficiently gain valuable information when studying the plots. Conversely, I show two plots that could be improved by following this simple principle.

With today's software it is very easy to produce graphics. But much like good writing, good graphics are not produced merely by the click of a button. They require substantial forethought, meticulous craftsmanship, and repeated revision until the final product is clear, concise, and informative.

The chances are that your plots may be the first impression that you, as an author, make on the reader who is flipping through the latest issue of *AJG*. Informative plots free of clutter and gratuitous decoration can catch a reader's attention, and even if he does not read your whole article, a carefully considered and crafted graphic can convey both the strengths and the weaknesses of your data and help him incorporate your research into his clinical practice.

CONFLICT OF INTEREST

Guarantor of the article: Jason T. Connor, PhD, FACG.
Financial support: None.
Potential competing interests: None.

REFERENCES

1. Feliciano GD, Powers RD, Kearl BE. The presentation of statistical information. Av Commun Rev 1963;11:32–9.
2. Tukey JW. The future of data analysis. Ann Math Stat 1962;33:1–67.
3. Dibble E. The Interpretation of Tables and Graphs. University of Washington: Seattle, 1997.
4. Gelman A, Pasarica C, Dodhia R. Let's practice what we preach: turning tables into graphics. Am Stat 2002;56:121–30.
5. Meyer DE, Kuskin SM, Gopher D. Information structure and the relative efficacy of tables and graphs. Hum Factors 1999;41:570–87.
6. Tufte ER. The Visual Display of Quantitative Information. Graphics Press: Cheshire, CT, 2001.
7. Tufte ER. Envisioning Information. Graphics Press: Cheshire, CT, 1990.
8. Tufte ER. Visual and Statistical Thinking: Displays of Evidence for Decision Making. Graphics Press: Cheshire, CT, 1997.
9. Tufte ER. Beautiful Evidence. Graphics Press: Cheshire, CT, 2006.
10. Shapley D. The da Vinci of data. New York Times, 30 March 1998.
11. Imamura J, Tateishi R, Shiina S. Neoplastic seeding after radiofrequency ablation for hepatocellular carcinoma. Am J Gastroenterol 2008;103:3057–62.
12. Talley NJ, Locke GR, McNally M *et al.* Impact of gastroesophageal reflux on survival in the community. Am J Gastroenterol 2008;103:12–9.
13. Dominguez M, Rincón D, Abraldes JG *et al.* A new scoring system for prognostic stratification of patients with alcoholic hepatitis. Am J Gastroenterol 2008;103:2747–56.
14. Shen B, Remzi FH, Nutter B *et al.* Association between immune-associated disorders and adverse outcomes of ileal pouch-anal anastomosis. Am J Gastroenterol 2009;104:655–64.
15. Snow J. On the Mode of Communication of Cholera. London, 1855.
16. Johnson S. The Ghost Map. Riverhead Books: New York, 2006.
17. Jolivet-Gougeon A, Loréal O, Ingels A *et al.* Serum transferrin saturation increase is associated with decrease of antibacterial activity of serum in patients with HFE-related genetic hemochromatosis. Am J Gastroenterol 2008;103:2502–8.
18. Lucey MR, Connor JT, Boyer TD *et al.* Alcohol consumption by cirrhotic subjects: patterns of use and effects on liver function. Am J Gastroenterol 2008;103:1698–706.
19. Accarino A, Perez F, Azpiroz F *et al.* Intestinal gas and bloating: effect of prokinetic stimulation. Am J Gastroenterol 2008;103:2036–42.
20. Maire F, Voitot H, Aubert A *et al.* Intraductal papillary mucinous neoplasms of the pancreas: performance of pancreatic fluid analysis for positive diagnosis and the prediction of malignancy. Am J Gastroenterol 2008;103:2871–7.
21. Lembo AJ, Zaman M, Krueger RF *et al.* Psychiatric disorder, irritable bowel syndrome and extra-intestinal symptoms in a population-based sample of twins. Am J Gastroenterol 2009;104:686–94.
22. Yeomans N, Lanas A, Labenz J *et al.* Efficacy of esomeprazole (20mg once daily) for reducing the risk of gastroduodenal ulcers associated with continuous use of low-dose aspirin. Am J Gastroenterol 2008;103:2465–73.