DISCUSSION ARTICLE

Why Tables Are Really Much Better Than Graphs

Andrew GELMAN

The statistical community is divided when it comes to graphical methods and models. Graphics researchers tend to disparage models and to focus on direct representations of data, mediated perhaps by research on perceptions but certainly not by probability distributions. From the other side, modelers tend to think of graphics as a cute toy for exploring raw data but not much help when it comes to the serious business of modeling. In order to better understand the benefits and limitations of graphs in statistical analysis, this article presents a series of criticisms of graphical methods in the voice of a hypothetical old-school analytical statistician or social scientist. We hope to elicit elaborations and extensions of these and other arguments on the limitations of graphics, along with responses from graphical researchers who might have different perceptions of these issues.

Key Words: Statistical communication.

THE BENEFITS AND LIMITATIONS OF STATISTICAL GRAPHICS

My purpose in writing this article is to elicit lively discussion of the uses of graphical methods in statistical analysis.

Graphs tend to be ignored or underused in much of the literature of statistics and applied fields (see, e.g., Gelman, Dodhia, and Pasarica 2002; Kastellec and Leoni 2007), and the literature on graphical methods is small and is mostly separate from the rest of statistics. The related field of data visualization has become increasingly prominent in digital communication, and the arts, but there the focus is typically on eye-catching design rather than on conveying statistical information.[1]

Here I would like to stimulate considerations of the connections between graphics and more formal statistical analysis, along with a serious discussion of the drawbacks of visual

---

Andrew Gelman is Professor, Department of Statistics and Department of Political Science, Columbia University, New York, NY 10027 (E-mail: *gelman@stat.columbia.edu*; url: *http://www.stat.columbia.edu/~gelman/*).

[1]For example, when the influential (and interesting) Flowing Data blog published a list of the "5 Best Data Visualization Projects of the Year" (Yao 2008), a debate ensued over whether data visualizations should aim for transparent communication of information (Gelman 2009b, 2009c) or for visual novelty and beauty (Yao 2009).

Journal of Computational and Graphical Statistics, Volume 20, Number 1, Pages 3–7
DOI: 10.1198/jcgs.2011.09166

presentation of quantitative information: if graphs really are so wonderful and underused in applied statistics (as I in fact believe), what is holding people back from integrating them much more into data analysis?

Following the revolution begun by Tukey (1972, 1977) and continued by Chambers et al. (1983), Cleveland (1985), and Tufte (1983, 1990), graphical methods for exploratory data analysis have generally been recognized to be a useful first step in any statistical study. Beyond this, though, there is disagreement, with the dominant strain of applied researchers (at least in social science) feeling that, when the serious models come out, it is time to put the graphical toys away. From the other direction, researchers in statistical graphics often disparage models and focus on direct representations of data.

There is a lot of valuable work combining analytical modeling and graphical display (for a classical example, consider the article by Daniel (1959)), but in much of the published work in political science, economics, sociology, and other areas, graphics have little if any serious role, being used to display some simple data summaries and never seen again, with important findings displayed in tabular form.

Those of us who believe graphing to be important and even essential to research would be well advised to think hard about why visual displays are not used more extensively in serious applied research. To this end, this article presents a series of attacks on graphical methods in the voice of a hypothetical old-school analytical statistician or social scientist. Although this originated as an April Fool's blog entry (Gelman 2009a), I believe these are strong arguments to be taken seriously—and ultimately accepted in some settings and refuted in others. I welcome elaboration and discussion of these points by statisticians and statistically-minded researchers in applied fields. I have my own answers to some of these objections but do not present them here, in the interest of presenting an open forum for discussion.

The arguments I lay out are, briefly, that graphs are a distraction from more serious analysis; that graphs can mislead in displaying compelling patterns that are not statistically significant and that could easily enough be consistent with chance variation; that diagnostic plots could be useful in the development of a model but do not belong in final reports; that, when they take the place of tables, graphs place the careful reader one step further away from the numerical inferences that are the essence of rigorous scientific inquiry; and that the effort spent making flashy graphics would be better spent on the substance of the problem being studied.

SOME PROBLEMS WITH GRAPHS

Graphs are gimmicks, substituting fancy displays for careful analysis and rigorous reasoning. It is basically a trade-off: the snazzier your display, the more you can get away with a crappy underlying analysis. Conversely, a good analysis does not need a fancy graph to sell itself. The best quantitative research has an underlying clarity and a substantive importance whose results are best presented in a sober, serious tabular display. And the best quantitative researchers trust their peers enough to present their estimates and standard errors directly, with no tricks, for all to see and evaluate. Let us leave the dot plots, pie charts, moving zip charts, and all the rest to the folks in the marketing department and the art directors of *Newsweek* and *USA Today*. As scientists we are doing actual research and we want to see, and present, the hard numbers.

To get a sense of what is at stake here, consider two sorts of analyses. At one extreme are controlled experiments with clean estimate and p-value, and well-specified regressions with robust standard errors, where the p-values really mean something. At the other extreme are descriptive data summaries—often augmented with models such as multilevel regressions chock full of probability distributions that are not actually justified by any randomization, either in treatment assignment or data collection—with displays of all sorts of cross-classified model estimates. The problem with this latter analysis is not really the modeling—if you state your assumptions carefully, models are fine—but the display of all sorts of numbers and comparisons that in no way are statistically significant.

For example, consider a research article with a graph showing three lines with different slopes. It is natural for the reader to assume, if such a graph is featured prominently in the article, that the three slopes are statistically significantly different from each other. But what if no p-value is given? Worse, what if there are no point estimates or standard errors to be found, let alone the sort of multiple comparisons correction that might be needed, considering all the graphs that might have been displayed? Now, I am not implying any scientific misconduct—and, to keep personalities out of this, I have refrained from referring to the article that I am thinking about here—but it is sloppy at best and statistical malpractice at worst to foreground a comparison that has been presented with no rigorous—or even approximately rigorous—measure of uncertainty. And, no, it is not an excuse that the researchers actually "believe" their claim. Sincerity is no defense. There is a reason our forefathers developed p-values and all the rest, and let us remember those reasons.

To flip this discussion around, what sorts of graphs do we commonly see in statistics textbooks? Residual plots, influence diagrams, quantile-quantile plots... a bunch of cookbook routines that have little to do with exploratory data analysis as it might be practiced. In addition, it does not usually make sense to include diagnostic plots in published articles: if the plot reveals a problem, the model should be fixed, whereas if no problem is found, the graph is typically not very informative and so there is no point in displaying it.

THE POSITIVE CASE FOR TABLES

So far I have explained my aversion to graphs as an adornment to, or really a substitute for, scientific research. I have been bothered for a while by the trend of graphical displays in journal articles, but only in writing this piece right here have I realized the real problem, which is not so much that graphs are imprecise, or hard to read, or even that they encourage us to evaluate research by its "production values" (as embodied in fancy models and graphs) rather than its fundamental contributions, but rather that graphs are inherently a way of implying results that are often not statistically significant. (And all but the simplest graphs allow so many different visual comparisons, that even if certain main effects actually do pass the p-value test, many many more inevitably will not. Some techniques have been developed to display multiple-comparisons-corrected uncertainty bounds, but these are rarely included in graphs for the understandable reason that they magnify visual clutter.)

But enough about graphs. Now I would like to talk a bit about why tables are not merely a necessary evil but are actually a positive good.

A table lays down your results, unadorned, for the readers—and, most importantly, scientific peers—to judge. Good tables often have lots of numbers. That is fine—different readers may be interested in different things. A table is not meant to be read as a narrative, so do not obsess about clarity. It is much more important to put in the exact numbers, as these represent the most important summary of your results, estimated local average treatment effects, and all the rest.

It is also helpful in a table to have a *minimum* of four significant digits. A good choice is often to use the default provided by whatever software you have used to fit the model. Software designers have chosen their defaults for a good reason, and I would go with that. Unnecessary rounding is risky; who knows what information might be lost in the foolish pursuit of a "clean"-looking table?

There is also the question of what words should be used for the rows and columns of the table. In tables of regressions, most of the rows represent independent variables. Here, I recommend using the variable names provided by the computer program, which are typically abbreviations in all caps. Using these abbreviations gets the reader a little closer to your actual analysis and also has the benefit that, if he or she wants to replicate your study with the same dataset, it will be clearer how to do it. In addition, using these raw variable names makes it more clear that you did not do anything shifty such as transforming or combining your variables before putting them in your regression.

We would do well to take a lead from our most prominent social science colleagues—the economists—who have, by and large, held the line on graphics and have insisted on tabular presentations of results in their journals. One advantage of these norms is that, when you read an econ article, you can find the numbers that you want; the authors of these articles are laying it on the line and giving you their betas. Beyond this, the standardization is a benefit in itself: a patterned way of presenting results allows the expert readers—who, after all, represent the most important audience for journal articles—to find and evaluate the key results in an article without having to figure out new sorts of displays. Less form, more content: that is what tables are all about. If you have found something great and you want to share it with the world, sure, make a pretty graph and put it on a blog. But please, please, keep these abominations out of our scientific journals.

B-B-B-BUT...

Yes, you might reply, sure, graphics are manipulative tricks and tables are the best. But does not the ambitious researcher *need* to make graphs, just to keep up with everybody else, just to get his or her research noticed? It is the peacock's tail all over again—I do not want to waste my precious time making fancy 3-D color bar charts, but if I do not, my work will get lost in the nation's collective in-box.

To this I say, No! Stand firm! Do not bend your principles for short-term gain. We are all in this together and we all have to be strong, to resist the transformation of serious social science into a set of statistical bells and whistles. Everything up to and including ordered logistic regression is OK, and it is fine—nay, mandatory—to use heteroscedasticity-consistent standard errors. But No to inappropriate models and No to graphical displays that imply research findings where none necessarily exist.

Beyond this, even in the short term I think there are some gains from going graph-free. The time you save not agonizing over details of graphs can instead be used to think more seriously about your research. Undoubtedly there is a time substitution: effort spent tinkering with graphs (or, for that matter, blogging) is effort not spent collecting data, working out models, proving theorems, and all the rest. If you *must* make a graph, try only to graph unadorned raw data, so that you are not implying you have anything you do not. And I recommend using Excel, which has some really nice defaults as well as options such as those 3-D colored bar charts. If you are going to have a graph, you might as well make it pretty. I recommend a separate color for each bar—and if you want to throw in a line as well, use a separate y-axis on the right side of the graph.

I am sure there are a lot of other problems with statistical graphics that I have missed. You can take it from here.

[Received September 2009. Revised September 2010.]

REFERENCES

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983), *Graphical Methods for Data Analysis*, Pacific Grove, CA: Wadsworth. [4]

Cleveland, W. S. (1985), *The Elements of Graphing Data*, Pacific Grove, CA: Wadsworth. [4]

Daniel, C. (1959), "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments," *Technometrics*, 1, 311–341. [4]

Gelman, A. (2009a), "Why Tables Are Really Much Better Than Graphs," *Statistical Modeling, Causal Inference, and Social Science blog*, 1 April. [4]

——— (2009b), "More on Data Visualization, Beauty, etc.," *Statistical Modeling, Causal Inference, and Social Science blog*, 22 April. [3]

——— (2009c), "Doing Graphics the George Orwell Way," *Statistical Modeling, Causal Inference, and Social Science blog*, 17 July. [3]

Gelman, A., Pasarica, C., and Dodhia, R. (2002), "Let's Practice What We Preach: Turning Tables Into Graphs," *American Statistician*, 56, 121–130. [3]

Kastellec, J., and Leoni, E. (2007), "Using Graphs Instead of Tables in Political Science," *Perspectives on Politics*, 5, 755–771. [3]

Tufte, E. R. (1983), *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press. [4]

——— (1990), *Envisioning Information*, Cheshire, CT: Graphics Press. [4]

Tukey, J. W. (1972), "Some Graphic and Semigraphic Displays," in *Statistical Papers in Honor of George W. Snedecor*, ed. T. A. Bancroft, Ames, IA: Iowa State University Press. [4]

——— (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley. [4]

Yao, N. (2008), "5 Best Data Visualization Projects of the Year," *Flowing Data blog*, 19 December. [3]

——— (2009), "Narrow-Minded Data Visualization," *Flowing Data blog*, 22 April. [3]

DISCUSSION ARTICLE

Comment

William M. BRIGGS

Darn right, graphs are not serious. Any untrained, unsophisticated, non-degree-holding civilian can display data. Relying on plots is like admitting you do not need a statistician. Show pictures of the numbers and let people make their own judgments? That can be no better than airing your statistical dirty laundry.

People need guidance; they need to be shown what the data are supposed to say. Graphics cannot do that; models can. And it takes a trained expert—like myself—to devise a model that can support the conclusions I want.

For example, I had a client situation recently where displaying the data would have misled some of the client's audience because portions of that data did not appear to fit into the structure my client had envisioned. It took all the skills I had to find an analysis that suited his needs.

I had to rely on hard-core modeling that went way beyond showing simple pictures. The by-the-book stuff was not working: standard procedures kept spitting out p-values like 0.053, 0.060, 0.51; you get the idea. Finally, after a lot of effort and some pretty creative modeling—and through the use of an obscure test I found in a peer-reviewed journal—I was able to find a successful p-value of 0.049.

This meant a lot to my client. And to me, too, since I depend on repeat business. The moral—well familiar to readers—is that graphics can never help in situations like this.

I admit that our profession could do a better job teaching these concepts. We tend to send students off with just enough information so that they are a danger to themselves and others, numerically speaking.

Here is an anecdote to show the trouble inadequate training can cause. A client came to me claiming to "know statistics" but saying he did not have the time to do an analysis himself. Initially, all he wanted was some simple-to-understand plots. I used the fact that I had a Ph.D. in statistics, and therefore knew better, to talk him out of that mistake pronto.

What he needed—what all people need—was a model. So I built one and told him of the results. I explained to him that his 95% confidence interval was between this number and that, which meant it was likely that his theory of what explained the data was true.

William M. Briggs, New York, NY 10021 (E-mail: *matt@wmbriggs.com*).

Journal of Computational and Graphical Statistics, Volume 20, Number 1, Pages 16–17
DOI: 10.1198/jcgs.2011.09166a

He countered, “I thought individual confidence intervals were meaningless. That the best you could say was that some parameter was either in the interval or it wasn’t. I learned that the widths of confidence intervals can’t be used to infer the truth or falsity of any theory.”

See what I mean? Just enough knowledge to cause us experts grief. I explained to him that of course he was right, but the good news was that nobody ever remembered what a confidence interval was. People use them all the time to infer things that are forbidden to infer. Besides, I continued, there are alternative theories that transform those confidence intervals into objects which can be interpreted as giving evidence for or against a theory.

This satisfied him. Temporarily. Because no sooner had I completed the Bayesian analysis, and I told him that the credible interval was such that his theory was likely true, then he threw another ill-informed question my way.

He asked, “I understood that credible intervals just say something about the possible values of some parameter or parameters. They don’t provide any direct information about whether the model or theory behind that model are correct.”

I tried to explain to him that it was perception that is what counts. People will hear that a modern analysis was applied to his data—models with complex, impressive-looking equations were coded into computers—and that his audience would trust that he knew what he was talking about because not only was he an expert, but so was I. That is good enough for most people, so it should be good enough for him.

You will have already guessed that it was not good enough. He said, “Isn’t the only way to know whether—no matter how sophisticated it is—the model we used is any good is to use it to predict data that we have not yet seen? It might fit our present data well, but predict poorly on new data. Shouldn’t we be cautious and wait to announce our results until we could compare our model on independent data?”

I told him that *nobody* ever did this, so he need not worry. We went back and forth a few times like this, but my time was not inexhaustible, mostly because of my ever-growing importance. In the end, I let him win. I made the plots and sent him my bill. He was happy and I was happy. And in the end, that is all that matters.

DISCUSSION ARTICLE

Comment

Michael FRIENDLY and Ernest KWAN

We consider Gelman's claims about the relative merits of tables versus graphs from a psychological perspective that emphasizes the role of data displays in the communication of quantitative results from authors to readers or viewers. From this perspective, we consider these claims in relation to a cognitive distinction between graph people and table people.

1. GRAPH PEOPLE VERSUS TABLE PEOPLE

We are grateful to Andrew Gelman for what can best be described as thought-provoking *chutzpah*, in its most positive sense. To reply in like manner, we write this in the first-person, non-royal *I*.

Thus, I will begin with a bald assertion: there are two kinds of people in this world—graph people and table people. If you sit in your local Starbucks, or even in a departmental faculty meeting and gaze around, you will have trouble at first distinguishing them by sight. But trust me—I have a Ph.D. in quantitative and cognitive psychology, so I should know what I am talking about. With a little training, you can do this, too.

Establishing this assertion scientifically is similar to what psychologists have done for over 100 years, using techniques of principal components analysis, factor analysis, and more recently, structural equation modeling, almost all of which we developed. As a result (e.g., McCrae and Costa Jr. 1987) we now know that nearly all the aspects of your personality can be summarized along five dimensions: the so-called Big 5: *Openness* to experience (appreciation for art, adventure, curiosity, ...); *Conscientiousness* (self-discipline, act dutifully, ...); *Extraversion* (positive emotions, seeking the company of others, ...); *Agreeableness* (compassionate and cooperative toward others); *Neuroticism* (experience unpleasant emotions easily, such as anger, anxiety, depression, ...).

With only a little bit of training, you will easily be able to classify Aunt Bertha, cousin Charles, your department chair, and others along such dimensions. If pressed, I will even

Michael Friendly is Professor, Psychology Department, York University, Toronto, ON, Canada (E-mail: *friendly@yorku.ca*). Ernest Kwan is Assistant Professor, Sprott School of Business, Carleton University, Ottawa, ON, Canada (E-mail: *ernest_kwan@carleton.ca*).

Journal of Computational and Graphical Statistics, Volume 20, Number 1, Pages 18–27
DOI: 10.1198/jcgs.2011.09166b

admit that my academic forefathers established all this primarily with tables, though they did use graphic methods of factor rotation extensively before the advent of analytic methods. As psychologists, we even invented a convenient acronym for you to remember this: *OCEAN*.

So, my assertion is first that there is another underlying dimension, not of personality, but rather of cognition, underlying both the *presentation* of quantitative information in tables and graphs by authors and the *understanding* of this information by readers and viewers. What works best depends most strongly on the match between the requirements of a given task on the one hand, and the skills and orientation of reader or viewer on the other.

The second part of my assertion is that these distributions, in the general population, are at least strongly bimodal, if not fully two-point, discrete distributions—graph people versus table people. Thanks to the recent important developments in Bayesian computational statistics, I do not have to address the stronger, two-point claim here, given my prior.

My initial diagnostic impressions from reading Gelman's article are recorded in my case notes: "As clear a case of graphic-denial as I've ever observed; ask about tabular-tendencies of parents and mentors; should we try the penile-erectile test with brief visual presentations of tables and graphs? What is the *real* question?"

1.1 WHAT IS THE QUESTION?

Gelman raises the question of why tables might be better than graphs (or not) with tongue firmly planted in cheek to stimulate discussion, and this is a worthy goal. As he overstates his case, this tabular-centric view invites the conclusion that almost any form of tabular presentation will suffice, as long as it is factually correct.

But in any debate, it is useful to know exactly "what is the question?" In as fine an example of the shifting-sands school of rhetoric as I have seen in a while, Gelman frames the comparison of tables and graphs in different ways, each nicely illustrated with ad hominem arguments: Graphical methods (cute toys) versus statistical modeling (serious statistics); use of graphs in the statistical literature (ignored or under-used) versus data visualization (eye-catching fluff); use of graphs in applied social science (little serious role).

Part of this debate has a long history, largely centered on the nature of the *task* (look up a precise value? make comparisons? detect trends, differences or anomalies?); see the article by Gelman, Pasarica, and Dodhia (2002, sec. 2.1) for a brief summary. Here, I just want to call attention to a brief note by Karl M. Dallenbach (1963), the editor of the *American Journal of Psychology* from 1926–1967. Publication of graphs in this journal had always been difficult and deprecated (requiring expensive "line cuts"), and Dallenbach had been largely a table person. In this note, he reports an epiphany: a long-undetected error from an earlier article caused him to "deduce from that error some evidence regarding the relative value of tables and graphs in the presentation of experimental result." From his evaluation of this case he morphed to a graph person. He concluded:

> All the evidence obtained from the reproduction of the study mentioned here indicates that the graphic method is 'better' than the tabular. Tables, since graphs are based on them, are necessary, but they are like background rocks, heavy and uninteresting. Graphs, on the other hand, spice the reports; clarify them, and make them interesting and palatable. (Dallenbach 1963, p. 702)

I describe this here to give some comfort to Gelman and other table- or crypto-table people reading this. Although the evidence on the Big 5 traits of personality suggests that they are relatively immutable over one's lifetime and may even have some genetic component, cognitive capacities are more mutable, so even a predisposition as a table person *is* subject to change.

1.2 MODES OF COMMUNICATION: WORDS, NUMBERS, PICTURES

A good deal of confusion disappears when one considers a graph or table as an act of (or attempt at) communication, similar to using words. Then, more interesting questions arise:

- What is the communication goal?
- Who is the audience?
- Was the communication effective?

Thus, a given scientific result or statistical analysis can be conveyed in different forms—words, numbers (p-values, parameter estimates, or tables), or pictures (graphs or diagrams)—for different purposes (analysis or presentation), to achieve different communication goals (exploration, detection, comparison, aesthetics, or rhetoric). Moreover, this view suggests that communication is an activity directed from a source (author) to a target (reader or viewer) and therefore the communication *mode* should be tailored to the audience in order to achieve the desired goal.

In fact, we can take this further, and consider the proposition that the majority of human communication involves, to a fair approximation, different relative proportions of the three primary ingredients: words, numbers, and pictures. Figure 1 shows some examples that help to position graphs and tables in a wider context. I freely admit that this is just a cute toy and it is not based on any data. But it does show that (in my view) tables occupy a rather lonely position, and I would have been foolish to try to present this view in a table.

Most statisticians and applied researchers know implicitly how this works. In the analysis stage, you use a collection of statistical and graphical methods both to *summarize* the data (often in tables of numbers) and to *expose* it (in graphs); you make notes (in words) regarding what you have seen and come to believe. At this stage, you are both the author and viewer, and the communication goal is the "Aha!" experience—you have found something noteworthy. In a conference presentation, you have only 20 minutes to convince your audience that what you have found is indeed interesting, so you need to focus on the "Wow!" experience. If you are smart, you will use a larger proportion of incisive visual displays, not all of the words you could have used on your slides, and tables of numbers only as necessary. Finally, you want to publish your findings, so you need to think of the communication goals and audience anew, but particularly the editors and reviewers who will decide whether you publish or perish. This goal should help decide the mix of words, numbers, and pictures in what you write and submit. Of course, through all of this is another layer: are you a graph person, or a table person? What about your audience?

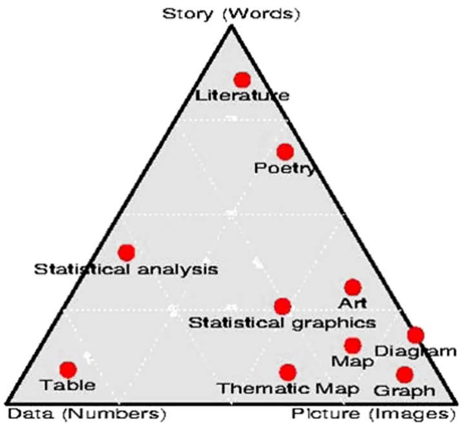

Figure 1. Modes of communication, as composed of words, numbers, and pictures, displayed in trilinear coordinates. Each point shows the (fictitious) composition of a given communication form, referred to the vertices representing 100%. The online version of this figure is in color.

2. USE OF GRAPHS AND TABLES IN SCIENTIFIC PUBLICATION

Gelman asserts that "graphs tend to be ignored or underused in much of the literature of statistics and applied fields," but this view is highly selective and ignores a growing body of research on the role of graphs (and of tables) in the construction and communication of science, as well as trends in the history of data visualization. From the previous section, it should be clear that use of graphs or tables in journal publication represents just a slice of the communication goal—intended audience tableau, but let us see where this takes us.

In a classic article Cleveland (1984) surveyed the use of graphs in 14 disciplines for the years 1980–1981, selecting 57 journals (4–5 in each area), with 50 articles selected randomly from each journal. Given that page space in journals is a limited resource, he measured the "fractional graph area" (FGA), or proportion of the total area of all journal pages devoted to graphs. Cleveland was careful in his tabulations, excluding figures such as apparatus illustrations, theoretical diagrams, etc.: "a figure was judged to be a graph if it had scales and conveyed quantitative information," so the FGA measure represented the amount of text displaced by graphs.

The results, presented in a dotplot (Cleveland 1984, fig. 3) compared journals in natural science, mathematical science, and social science. For example, the average graph use in natural science journals (chemistry: 0.18; physics: 0.17) vastly exceeded that in social science journals (economics: 0.025; sociology: 0.01).

More recently, other authors have taken up the more detailed study of the use of graphs versus tables across and within disciplines. Noteworthy here are articles by Smith, Best, and collaborators (Smith et al. 2000, 2002) where they asked respondents to rate each of the disciplines from the Cleveland study on a 1–10 scale, distinguishing "soft" science at the low end from "hard" science at the high end. As shown in Figure 2(A), use of graphs (in terms of FGA) across disciplines was nearly perfectly correlated with the rated hardness of the discipline.

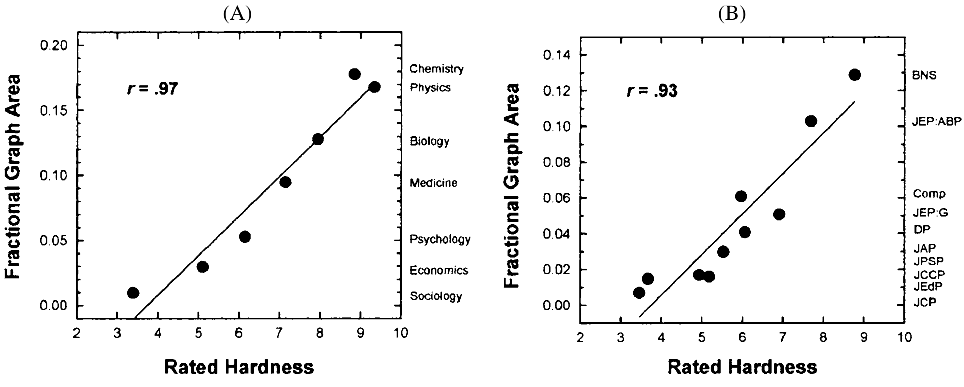

Figure 2. Proportion of journal page area devoted to graphs, in relation to rated hardness. (A) For seven scientific disciplines; (B) for 10 psychology journals. (Source: Smith et al. 2002, fig. 1.)

It should therefore not be surprising to Gelman that graph use in political science (somewhere between sociology and economics in hardness) is at the lower end of the continuum. If he is to "take a lead from our most prominent social science colleagues," he would do somewhat better to follow the exemplars set in psychology than in the dismal science of economics. As well, he (Gelman, Pasarica, and Dodhia 2002) and others in political science (Kastellec and Leoni 2007) have amply demonstrated some impressive ways in which tabular displays of even complex statistical models and model comparisons can be turned into graphic ones that preserve the essential information and make the results far more apparent.

What might be surprising is that this strong positive relation between graph use and "hardness" also applies *within* subfields of a given discipline. In psychology, journals range from the soft side (*J. Counseling Psychology*, *J. Educational Psychology*) to the harder side (*J. Experimental Psychology*, *Behavioral Neuroscience*). In a parallel study, Smith et al. (2000) obtained ratings of hardness for 10 psychology journals and also calculated graph use (FGA) for 156 articles distributed across these journals. Their results (Figure 2(B)) show nearly as strong a relation between hardness and graph use within psychology as the relation across disciplines.

The icing on this cake is shown in Figure 3, which shows the comparison between use of graphs and tables across these subfields of psychology. As rated hardness increases, area devoted to tabular displays decreases. This inverse relation is not unexpected, but the magnitude of the effect might be: in the two softest journals, the ratio of graph use to table use was about 1:10; among the hardest-rated journals, this ratio approached 10:1. It is also noteworthy that the total space devoted to data displays (tables and graphs) was more nearly constant, averaging about 14% of total page area; nevertheless, there was a smaller tendency for data display area to increase with rated hardness ($r = 0.35$).

As you will clearly see, all of this is easily explained by the cognitive distinction I introduced at the outset: The harder sciences (and harder subfields within a given area) are disproportionally populated by graph people—those whose significant phenomena are often so striking that they think it most natural to display their results in visual form. Those

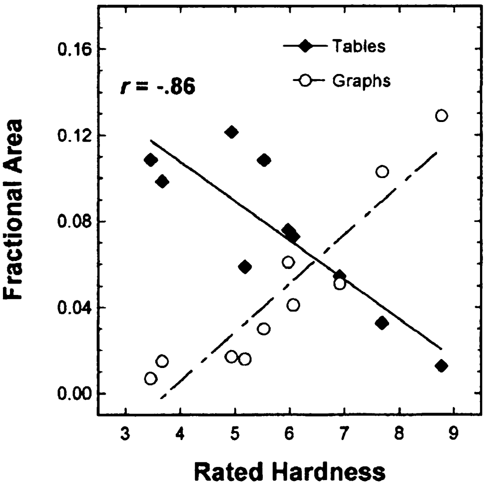

Figure 3. Proportion of page area devoted to tables and graphs in 10 psychology journals. (Source: Smith et al. 2002, fig. 3.)

who gravitate to the softer sciences often find that they have much smaller effects to deal with and so are often drawn to tabular displays, which conceal them nicely.

You can see from this description that there is a bit of a chicken-and-egg problem here: do graph people tend to gravitate to harder science, or does harder science shape researchers more as graph people? Are the data or phenomena studied in the softer sciences more complex or variable and less law-like that they resist portrayal in simple graphs? I leave these questions as open problems for future research.

2.1 When Tables Fail

One of Gelman's major premises is that the results of 'serious' modeling are better presented in tables of coefficients ("the authors are... giving you their betas") rather than in graphs ("less form, more content"). I referred above, in passing, to the article by Kastellec and Leoni (2007) illustrating how such tables of coefficients could be rendered more cogently as graphs.

There is one more important point to be made here: In many cases either the coefficients in fitted models are meaningless without graphical display or their interpretation is exceedingly difficult to understand. For examples, consider complex generalized regression models with transformations of predictors, polynomial terms, spline functions, not to mention interactions of the above, non-identity link functions, polytomous responses, generalized additive models, etc. Such models in tabular presentation often leave me gasping for air, but I am comforted that I can understand the terms in almost any model through an *effect display* (Fox 1987, 2003)—a plot of predicted values for a term, absorbing its lower-order relatives, and averaging over other terms in the model. Maybe it is just that I am a graph person, but I know for sure that even pure table people cannot extract any sunlight from such tabular cucumbers.

3. COMBINING TABLES AND GRAPHS: SEMI-GRAPHIC DISPLAYS AND TABLEPLOTS

Of course, it does not have to be either–or: tables *versus* graphs. There is a long history to what Tukey (1972) gave the name "semi-graphic" displays, integrating exact numbers into visual displays that also showed something more, or helped understand the numbers in a wider or more coherent way. Our historical ancestors (Playfair, Guerry, Minard, Nightingale, etc.) in the development of statistical graphics were usually cognizant of the fact that graph viewers might also want to know the numbers on which the graphs were based. Sometimes, they included the numbers directly on the graphs as annotations, sometimes in separate tables, a tip-of-the-hat to the table people among their readers. Sometimes, as in the case of Mendeleyev, simple quantitative data appeared unwieldy and resisted understanding until a method to represent them in a semi-graphic display was found.

It is no accident that Francis Galton (1886) developed the ideas of regression, starting from a table of the joint frequency distribution of characteristics of parents and their offspring, and overlaying lines connecting the means of $(Y|X)$ and $(X|Y)$ as well as contours of constant frequency. The remarkable visual insights he derived from this table turned into a graph became the foundation for correlation, regression, the bivariate normal distribution, and, ultimately, a huge chunk of modern statistical methods. Karl Pearson (1920, p. 37) would later say, "…that Galton should have evolved all this… is to my mind one of the most note-worthy scientific discoveries arising from analysis of pure observation." Clearly, Galton was a graph person.

Gelman clearly knows all the deficiencies of tables from the perspective of communication, and we will not belabor these here. He is also aware of the literature on how to make tables better, either by focusing more clearly on the message they should convey (Wainer 1993, 1997) or by reformulating them as graphs (Friendly and Kwan 2003; Gelman, Pasarica, and Dodhia, 2002; Kastellec and Leoni 2007).

3.1 TABLEPLOTS

Here, I would like to introduce another idea to help bridge the gap between table people and graph people: the tableplot (Kwan 2008; Friendly and Kwan 2009), designed as a semi-graphic display in the form of a table with numeric values, but supplemented by symbols with size proportional to cell value(s), and with other visual attributes (shape, color fill, background fill, etc.) that can be used to encode other information essential to direct visual understanding.

To illustrate, consider the statistical evidence that was used to develop the Big 5 dimensions of personality and to establish their stability over an individual's lifespan and cross-cultural replicability, and how the precise predictions of such theories can be exposed by tableplots (Kwan, Lu, and Friendly 2009). As commonly operationalized, the five dimensions are measured by 240 items grouped into 30 sub-scales ("facets") of the Revised NEO Personality Inventory (NEO PI-R; Costa Jr. and McCrae 1992), with six facets measuring each of the five dimensions. Using either exploratory factor analysis followed by rotation, or confirmatory factor analysis, researchers attempt to determine the extent to which the

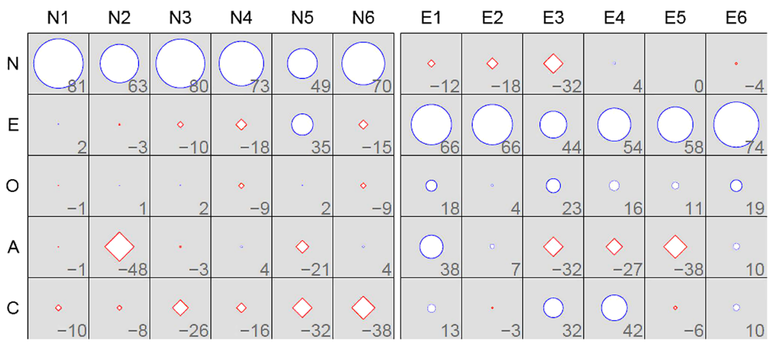

Figure 4. Tableplot of the first 12 facets from the normative NEO PI-R factor pattern measuring Neuroticism and Extraversion (Costa Jr. and McCrae 1992). Symbols (blue circles for positive loadings; red diamonds for negative loadings) scaled to |maximum| of 1; cell labels in hundredth decimal. (Source: Kwan, Lu, and Friendly 2009, fig. 2.) The online version of this figure is in color.

factor loadings conform to the five-factor theory or the extent to which two or more samples can be said to exhibit the same factor structure. Almost invariably, such results are presented in tables of factor loadings, sometimes accompanied by standard errors or other numerical comparison measures, but the model comprises 150 parameters (30 facets $\times$ 5 factors).

As an example of this graphic method, Figure 4 shows a portion of a tableplot of the facets for two factors from the normative study by Costa Jr. and McCrae (1992) to which other results are often compared. Each cell shows the factor loading ($\times$100) as a number, and as a circle (for positive loadings) or diamond (negative loadings), scaled to have a maximum size at |loading| = 1. Confirmation of the five-factor theory requires that all target loadings approach 1.0 and non-target loadings approach 0. The tableplot display could be supplemented by other annotations to indicate p-values or the results of significant tests, but the message from this example is clear enough: The Neuroticism factor appears to be relatively well-measured by its facets, while the Extraversion factor has smaller target loadings and some possibly troubling non-target ones.

This could be just another "cute toy," unnecessary in serious statistical modeling. A strength of the tableplot, however, is that it allows an easy detailed diagnosis of the fit between predicted and estimated factor patterns or between estimated results from multiple samples. Figure 5 shows one example, in which the normative results from Costa Jr.

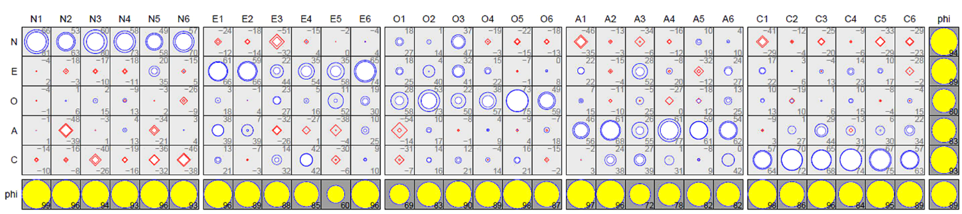

Figure 5. Superimposed tableplot of the normative (Costa Jr. and McCrae 1992) and Shona (Piedmont et al. 2002) NEO PI-R factor patterns, augmented by congruence coefficients (ϕ). Symbols scaled to |maximum| of 1; cell labels in hundredth decimal. (Source: Kwan, Lu, and Friendly 2009, fig. 4.) The online version of this figure is in color.

and McCrae (1992) are compared to those from a cross-cultural Shona-speaking sample from Zimbabwe (Piedmont et al. 2002). Such comparisons are extremely difficult in tabular displays, and so the similarities and differences are often summarized in congruence coefficients ($-1 \le \phi \le 1$) indicating the degree of similarity either within or across factors. Figure 5 shows these for the two samples on the tableplot margins. It is clear that overall, agreement is very strong, yet there are a few facets (E5, O1, A3) for which the evidence of identical factor structure is less compelling.

In case the use of tableplots or related semi-graphic displays is either unfamiliar in the context of factor analysis or unconvincing for closet table people who long to be trans-tabled, I provide examples in the supplementary materials illustrating how the results of more traditional models can be better understood and explained though tableplots than through the "background rocks, heavy and uninteresting" of tables described by Dallenbach. A **tableplot** package for R software is *http://cran.r-project.org/package=tableplot*.

4. WHY I LIKE (SOME) BAYESIAN STATISTICIANS TALKING ABOUT GRAPHS

What do Bayesian statisticians think about while they are doing an analysis and what do they think about when trying to communicate their findings to others? I do not think that those of the Bayesian or non-Bayesian persuasion are as fundamentally different as graph people versus table people in any way I could describe for your basic Starbucks test. Perhaps this is an area for future psychological research.

I would like to frame this part of the discussion in terms of this prescription for the central goal of applied statistical analysis: Tell an *A*ccurate, *C*redible, *U*nderstandable, and *I*nteresting story about some *R*eal problem (mnemonics: *ACUIR*, or *AURIC*, the gold standard).

By this standard, what Bayesian statisticians might lose on the understandability front they try to more than overcome on the side of credibility. In most cases, the point estimates from frequentist and Bayesian analyses are quite similar, or nearly identical with uninformative priors. What differs is how credibility is assessed and conveyed: by standard confidence intervals and p-values versus posterior distributions and their summaries.

Yet, perhaps because of the complexities of Bayesian modeling, particularly with complex models of significant practical importance, I find that I am often impressed with the pains that Gelman and some others (e.g., Gelman and Hill 2007) take to make their results comprehensible, not infrequently with graphic displays tuned to be responsive to the *AURIC* standard. The bottom line is this: however much Gelman professes otherwise, deep down both he and I know that he is really a closet graph person masquerading in table person garb. If he likes, we can always try the penile-erectile tests for tabular and graphic displays.

ADDITIONAL REFERENCES

Cleveland, W. S. (1984), "Graphs in Scientific Publications," *The American Statistician*, 38, 261–269. [21]

Costa Jr., P. T., and McCrae, R. R. (1992), *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEOFFI) Professional Manual*, Odessa, FL: Psychological Assessment Resources. [24-26]

Dallenbach, K. M. (1963), "Tables vs. Graphs as Means of Presenting Experimental Results," *The American Journal of Psychology*, 76 (4), 700–702. [19]

Fox, J. (1987), "Effect Displays for Generalized Linear Models," in *Sociological Methodology, 1987*, ed. C. C. Clogg, San Francisco: Jossey-Bass, pp. 347–361. [23]

——— (2003), "Effect Displays in R for Generalised Linear Models," *Journal of Statistical Software*, 8 (15), 1–27. [23]

Friendly, M., and Kwan, E. (2003), "Effect Ordering for Data Displays," *Computational Statistics and Data Analysis*, 43 (4), 509–539. [24]

——— (2009), "Where's Waldo: Visualizing Collinearity Diagnostics," *The American Statistician*, 63, 56–65. [24]

Galton, F. (1886), "Regression Towards Mediocrity in Hereditary Stature," *Journal of the Anthropological Institute*, 15, 246–263. [24]

Gelman, A., and Hill, J. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge, U.K.: Cambridge University Press. [26]

Kwan, E. (2008), "Improving Factor Analysis in Psychology: Innovations Based on the Null Hypothesis Significance Testing Controversy," unpublished doctoral dissertation, York University, Toronto, ON. [24]

Kwan, E., Lu, I. R. R., and Friendly, M. (2009), "Tableplot: A New Tool for Assessing Precise Predictions," *Zeitschrift für Psychologie/Journal of Psychology*, 217 (1), 38–48. [24,25]

McCrae, R. R., and Costa Jr., P. T. (1987), "Validation of the Five-Factor Model of Personality Across Instruments and Observers," *Journal of Personality and Social Psychology*, 52, 81–90. [18]

Pearson, K. (1920), "Notes on the History of Correlation," *Biometrika*, 13, 25–45. [24]

Piedmont, R. L., Bain, E., McCrae, R., and Costa Jr., P. T. (2002), "The Applicability of the Five-Factor Model in a Sub-Saharan Culture: The NEO PI-R in Shona," in *The Five-Factor Model of Personality Across Cultures*, eds. R. R. McCrae and J. Allik, New York: Kluwer Academic/Plenum, pp. 155–173. [25,26]

Smith, L. D., Best, L. A., Stubbs, D. A., Archibald, A. B., and Roberson-Nay, R. (2002), "Constructing Knowledge: The Role of Graphs and Tables in Hard and Soft Psychology," *American Psychologist*, 57 (10), 749–761. [21-23]

Smith, L. D., Best, L. A., Stubbs, D. A., Johnston, J., and Archibald, A. B. (2000), "Scientific Graphs and the Hierarchy of the Sciences: A Latourian Survey of Inscription Practices," *Social Studies of Science*, 30 (1), 73–94. [21,22]

Wainer, H. (1993), "Tabular Presentation," *Chance*, 6 (3), 52–56. [24]

——— (1997), "Improving Tabular Displays, With NAEP Tables as Examples and Inspirations," *Journal of Educational and Behavioral Statistics*, 22 (1), 1–30. [24]

DISCUSSION ARTICLE

Comment

Howard WAINER

Andrew Gelman's tongue-in-cheek advocacy of poor tabular presentation over more evocative visual/graphic displays has provided an opportunity to discuss some important issues that have arisen as part of modern data analysis. I am delighted to have been invited to participate in this discussion.

It seems to me that it would aid our discussion if we place the task of data display into a broader context. The context almost surely revolves around the use of evidence in science. Key contemporary issues are how and when should evidence be used. We hear the term 'evidenced-based decision making' in many fields; medicine, education, economics and political policy, to pick four, implying this is a new and modern way to try to solve modern problems. If what we are doing now is evidence-based, what were we doing previously[1]?

How can we consider the use of evidence in science new? Has not evidence been at the very core of science for millennia? The short answer is no. Making decisions evidence-based has always been a tough row to hoe, for once you commit to it, no idea, no matter how beautiful, no matter how desirable, can withstand an established contrary fact, regardless of how ugly that fact might be. The conflict between evidence and faith in the modern world is all around us, even in scientific issues for which faith is not required.[2] So it is not surprising that using evidence to make decisions took a long time to catch on. The origination of the formal idea of using evidence as a method for gaining knowledge is often dated, as are so many things, with Aristotle (384 BC–322 BC) but its pathway thereafter was not smooth, for once one commits to using evidence to make decisions, facts take precedence over opinion.[3] And not all supporters of an empirical approach had Alexander the Great

Howard Wainer is Distinguished Research Scientist, National Board of Medical Examiners, 3750 Market St., Philadelphia, PA 19104 (E-mail: *hwainer@nbme.org*).

[1]When I suggested that evidence-based medicine's predecessor must have been faith-based, my boss, Don Melnick, corrected me and said that he liked to think that medicine was intelligently designed.

[2]A story is told of a conversation between Napoleon and Laplace in which Napoleon congratulated Laplace on the publication of his masterwork, *Traité de Mécanique Céleste*, but then added that he was disappointed because "no where in this great work was the name of God mentioned even once." Laplace is said to have responded, "I did not need that hypothesis."

[3]Bertrand Russell reports that even Aristotle had trouble following the tenets of empiricism in all aspects of his own life, for he maintained that women have fewer teeth than men; although he was twice married, it never occurred to him to verify this statement by examining his wives' mouths.

Journal of Computational and Graphical Statistics, Volume 20, Number 1, Pages 8–15
DOI: 10.1198/jcgs.2011.09166c

to watch their backs. Hence it took almost 2000 years before Francis Bacon (1561–1626) repopularized the formal use of evidence, that was subsequently expanded and amplified by the British empiricists John Locke (1632–1704) and George Berkeley (1685–1753) and the Scot David Hume (1711–1776).

But having a formal epistemological basis for evidence-based science was not enough; looking at evidence required effective methods for doing so. Language, developed long before science, was not an ideal match for describing quantitative phenomena precisely. It was at this time that a new method of description, that substituted a visual description for a linguistic one, sprang into being. Although there were a few scattered examples of the graphic display of quantitative information in the 17th and early 18th century (see Wainer 2005, chaps. 1–3 and especially 7), the birth of graphic display is generally dated from 1786 with the publication of William Playfair's (1759–1823) *Commercial and Political Atlas* (the 1801 3rd edition of this path-breaking book has been republished and is widely available from Cambridge University Press). In this remarkable volume, an atlas that contained no maps, he invented the line, bar, and pie charts. Although the first two of these were in earlier documents (Christiaan Huygens in 1669 (reprinted in Huygens 1895) and Joseph Priestly in 1765, respectively), they were rough affairs, incompletely utilized. Playfair's graphic inventions were beautifully executed and used to perfection. These inventions took hold very quickly, although strangely, more quickly on the continent than in Playfair's native Britain, and soon found themselves widely used for purposes well beyond Playfair's economic applications.[4]

By 1978 the French physiologist Etienne Marey, whose graphic schedule of all the trains between Paris and Lyons provides a powerful illustration of the breadth of value of this approach, expressed the feelings of most natural scientists of the value of graphical representation.

> "There is no doubt that graphical expression will soon replace all others whenever one has at hand a movement or change of state—in a word, any phenomenon. Born before science, language is often inappropriate to express exact measures or definite relations."

Marey was also giving voice to the movement away from the sorts of subjectivity that had characterized prior science in support of the more modern drive toward objectivity. Although some cried out for the "insights of dialectic," "the power of arguments," and the "flowers of language," their protestations were lost on Marey, who dreamed of a wordless science that spoke instead in high-speed photographs and mechanically generated curves; in images that were, as he put it, in the "language of the phenomena themselves."

Historians have pointed out that "Let nature speak for itself" was the watchword of the new brand of scientific objectivity that emerged at the end of the 19th century. Daston and Galison (1992) pointed out that "at issue was not only accuracy but morality as well: the all-too-human scientists must, as a matter of duty, restrain themselves from imposing their hopes, expectations, generalizations, aesthetics, and even their ordinary language on the image of nature." Mechanically produced graphic images would take over

---

[4]Fletcher (1835) used statistical maps to display the locations of such social variables as ignorance, bastardy, and improvident marriages in England and Wales.

when human discipline failed. Marey and his contemporaries turned to mechanically produced images to eliminate human intervention between nature and representation. "They enlisted polygraphs, photographs, and a host of other devices in a near-fanatical effort to produce atlases—the bibles of the observational sciences"—documenting birds, fossils, human bodies, elementary particles, flowers, and economic and social trends that were certified free of human interference.

And so science advanced, now coupling the new tool of data display of quantitative phenomena with the older one of mathematical models. But despite the enthusiasm for an empirical approach to science supported by graphic displays of the evidence, progress was slow. In the 1920s there were debates even in the statistical literature about the legitimacy of graphics. Referring to Otto Neurath's (1882–1945) evocative isotype display, some complained that it was not the task of the scientist to be drawing 'little men.'

The development of the science of statistics in the first half of the 20th century was primarily mathematical, with a focus on fitting equations to data and making judgments of the legitimacy of the inferences one might draw from them. This changed in 1977 with the publication of John Tukey's (1915–2000) *Exploratory Data Analysis.* Tukey, a towering figure of 20th century science, legitimized the practice of the atheoretical plotting of points with the goal of finding suggestive patterns. He pointed out that "the greatest value of a graph is when it *forces* us to see what we never expected."

The simple, but powerful, tools that Tukey invented for looking at data quickly spread, and soon were even being taught in high schools. Although his methods could easily be implemented on the back of an envelope, they were quickly computerized. And with the computing revolution that moved high-speed computing from guarded sterile rooms to everyone's desktop, these methods began a revolution whose influence has only begun to be felt. The current enthusiasm for data mining is a direct lineal descendent of Tukey's ingenious, but humble stem-and-leaf diagrams.

But issues remain. Two which were alluded to in Gelman's instigating essay are most frequently seen within data mining, but exist whenever we plot points and look for suggestive patterns:

1. the problem of multiplicity, and
2. the problem of analyst ignorance and hubris.

Let me address them in turn.

MULTIPLICITY

When we look at large datasets we make many implicit comparisons. Typically we mouse around until we see something. It is hard to know whether what we have found is real or is an epiphenomenon yielded by chance. There is no obvious analog to the Bonferroni inequality to allow us to adjust our perceptions to this kind of fluid search. The most sensible approach is the gold standard for all experimental research, replication. If, after we discover some interesting phenomenon, we then turn to a parallel, but so far unexplored dataset, and find the same structure, we have confirmation. It seems logical that prior to doing any such investigation we should divide the existing dataset into two parts: exploratory and confirmatory, then dig away at the exploratory part until we are convinced that we have extracted from it all that we can. Then we use a more disciplined approach on the confirmatory portion. How the original dataset is divided will be determined by the specific circumstance. One approach is to make the confirmatory portion large enough for the power we need for a plausible number of hypothesis tests, and leave the balance for exploration. Of course there will be the usual concerns about wasting discovery power by not using all of the data for exploration, but when datasets are finite triage decisions must be made.

IGNORANCE AND HUBRIS

Too often there is a disconnect between the people who run a study and those who do the data analysis. This is as predictable as it is unfortunate. If data are gathered with particular hypotheses in mind, too often they (the data) are passed on to someone who is tasked with testing those hypotheses and who has only marginal knowledge of the subject matter. Graphical displays, if prepared at all, are just summaries or tests of the assumptions underlying the tests being done. Broader displays, that have the potential of showing us things that we had not expected, are either not done at all, or their message is not able to be fully appreciated by the data analyst.

Let me provide one dramatic example.

In the post-World War II world antibiotics were called "wonder drugs" for they provided quick and easy cures for what had previously been intractable diseases. Data were being gathered to aid in learning which drug worked best for what bacterial infection. Being able to see the structure of drug performance was an enormous aid for practitioners and scientists alike. In the Fall of 1951 the master designer Will Burtin published a graph showing the performance of the three most popular antibiotics on 16 different bacteria (see Wainer 2009). The data that were used in his display are shown in Table 1. The entries of the table are the Minimum Inhibitory Concentration (MIC), the amount of the antibiotic needed to stop the growth of the bacterium in question. The variable "Gram staining" describes the reaction of the bacteria to Gram staining. Gram-positive bacteria are those that are stained dark blue or violet; Gram-negative bacteria do not react that way.

Although there are many ways to display these data, some emphasizing one aspect, others another, any competent display tells us that if a bacterium is Gram negative (it does not take a Gram stain) we should treat it with Neomycin. And if it is Gram positive we should treat it with Penicillin. This can be inferred from the data within the table, although not without considerable effort, illustrating the truth of Farquahr and Farquahr's (1891) famous observation that "getting information from a table is like extracting sunbeams from cucumbers." Various alternative displays can make this point easily and clearly, as well as pointing out some minor variations; but this is the main story (see Wainer and Lysen 2009 for more details).

Table 1. The effectiveness of three antibiotics against 16 bacteria shown by the minimum concentration need to stop growth in vitro, plus the covariate 'gram staining.'

	Antibiotic			
Bacteria	Penicillin	Streptomycin	Neomycin	Gram staining
Aerobacter *aerogenes*	870	1	1.6	negative
Bacillus *anthracis*	0.001	0.01	0.007	positive
Brucella *abortus*	1	2	0.02	negative
Diplococcus *pneumoniae*	0.005	11	10	positive
Escherichia *coli*	100	0.4	0.1	negative
Klebsiella *pneumoniae*	850	1.2	1	negative
Mycobacterium *tuberculosis*	800	5	2	negative
Proteus *vulgaris*	3	0.1	0.1	negative
Pseudomonas *aeruginosa*	850	2	0.4	negative
Salmonella (Eberthella) *typhosa*	1	0.4	0.008	negative
Salmonella *schottmuelleri*	10	0.8	0.09	negative
Staphylococcus *albus*	0.007	0.1	0.001	positive
Staphylococcus *aureus*	0.03	0.03	0.001	positive
Streptococcus *fecalis*	1	1	0.1	positive
Streptococcus *hemolyticus*	0.001	14	10	positive
Streptococcus *viridans*	0.005	10	40	positive

An alternative display (Figure 1)[5] tells us more, but only if we know enough to see it. Here each bacterium is given its own icon with little bars indicating how much of it was needed. The horizontal line depicts what might be considered the maximum plausible dosage, and so bars going down from that line depict clinically efficacious drugs.

Looking at this display, specifically at the row of bacteria resistant to Streptomycin and Neomycin, we see something funny. The pattern of response to the antibiotics of all three bacteria is essentially identical—two of these bacteria are Streptococcus and one is not. That seems odd. What is *Diplococcus pneumoniae* doing there? And, why does the third Strep bacterium, *Streptococcus fecalis* (in the next row up), appear to be so different? One would think that the genus of a bacterium would reflect similarities in what kills other species that are members of that genus.

Because these oddities were not easily visible in Burtin's display, neither his, nor apparently anyone else's, curiosity was piqued. Had this odd structure been seen, perhaps it would not have taken until 1974 for *Diplococcus pneumoniae* to be recognized as a Streptococcus and to be renamed *Streptococcus pneumoniae*.

And why is *Streptococcus fecalis* so different? It would seem that its credentials as a member of the Strep family are impeccable; as Sherman, Mauer, and Porter (1937) described it:

> "In some respects Streptococcus fecalis (Andrewes & Horder, 1906) might be considered one of the better established species of the streptococci, and certainly some of the rather unique characteristics of this organism, or the general group to which it belongs, are commonly known by bacteriologists." (page 275)

[5] This is a variation on a display originally designed and prepared by Emory University biostatistician, Brian Schmotzer.

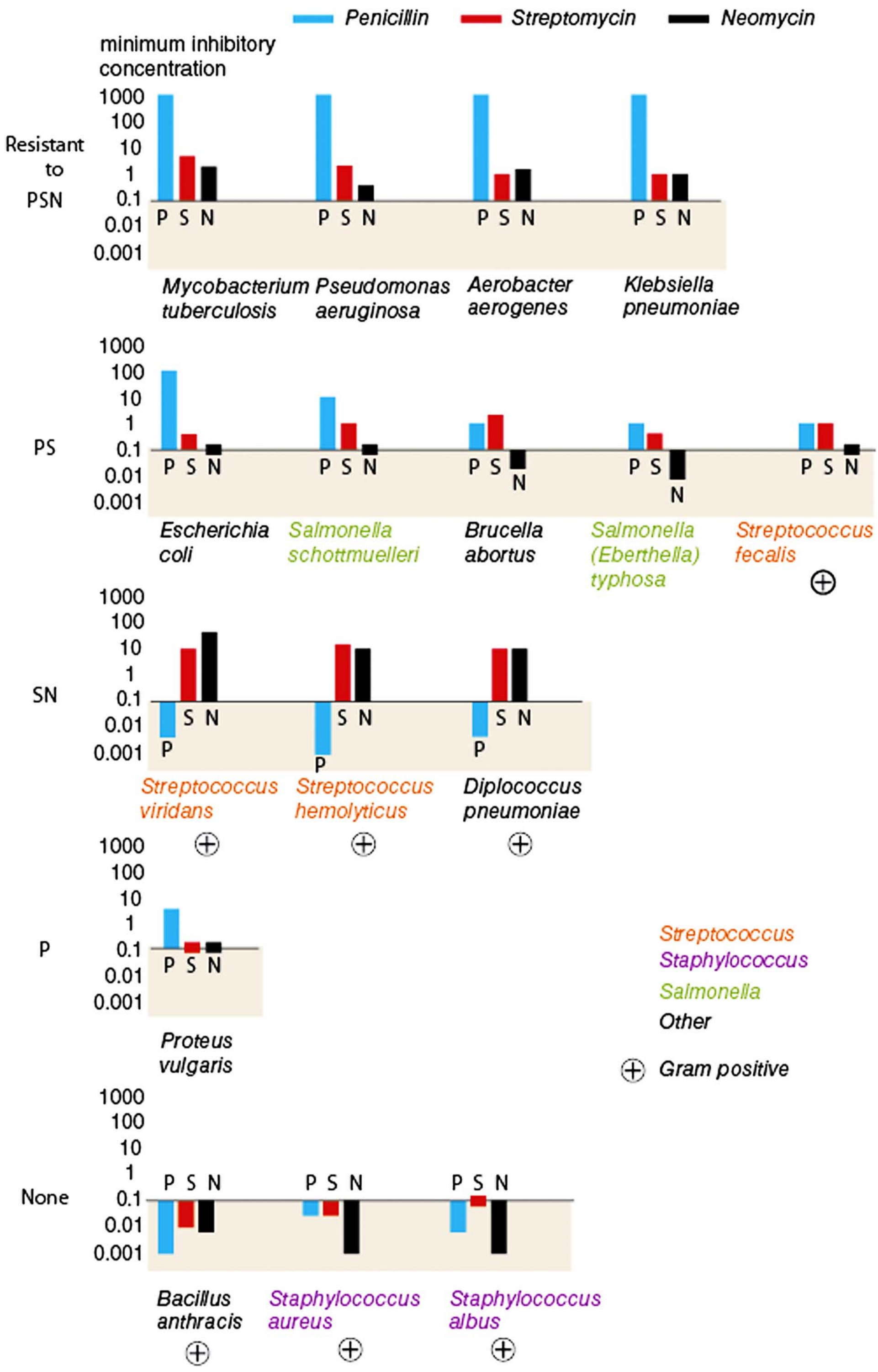

Figure 1. A graphic display of the data shown in Table 1. The online version of this figure is in color.

Yet, in 1984, its genus was changed and its name became *Enterococcus faecalis* (Schleifer and Kilpper-Balz 1984). Perhaps had these data been plotted in a way that allowed us to compare the profile of responses of these various bacteria to these antibiotics the classification of *Streptococcus fecalis* would have come under scrutiny sooner.

My point is this—if the 16×3 data matrix from which this display was constructed were handed off to someone with instructions to 'draw a nice graph of them' these anomalies would not have been found. It is an often insurmountable handicap for a data analyst to be given data without the associated science needed to understand them. For an analyst to willfully avoid learning about the science is akin to malfeasance. Of course, it is likely that a deep understanding both of the science and of data analytic methods does not reside in the same person. When it does not, data analysis should be done jointly. It is my understanding that data mining is not often done as a team. This is unfortunate, for then it is too easy to miss what might have been found.

There are many other issues that Professor Gelman raised that I did not address (rounding and other characteristics of table construction, to pick one juicy topic). I will leave those to the other discussants. Instead I chose, as my role, to lay out the historical reasons why data-based graphs entered so late into science's tool box. In short, we did not need ways of displaying data until we became convinced that data provide the evidence on which to base our science. Even today, as we view the ado initiated by the evidence-based advice about when women should have mammograms, we can see how easily other forces, not based on evidence, can hold sway.

ADDITIONAL REFERENCES

Andrewes, F. W., and Horder, T. J. (1906), "A Study of the Streptococci Pathogenic for Man," *Lancet*, 2, 708–713. [12]

Daston, L., and Galison, P. (1992), "The Image of Objectivity," *Representations*, 40, 81–128. [9]

Farquhar, A. B., and Farquhar, H. (1891), *Economic and Industrial Delusions: A Discourse of the Case for Protection*, New York: Putnam. [11]

Huygens, C. (1895), *Oeuvres Completes, Tome Sixieme Correspondance,* 515–518, 526–539. Den Hague: Martinus Nijhoff. [9]

Marey, E. J. (1878), *La Méthode graphique dans les sciences expérimentales et particuliérement en physiologie et en médecine*, Paris. [9]

Playfair, W. (1786), *The Commercial and Political Atlas*, London: Corry. [9]

——— (1801), *The Commercial and Political Atlas, Representing, by Means of Stained Copper-Plate Charts, The Progress of the Commerce, Revenues, Expenditure, and Debts of England, During the Whole of the Eighteenth Century*. Edited and introduced by Howard Wainer and Ian Spence. New York: Cambridge University Press, 2007. [9]

Priestley, J. (1765), *A Chart of Biography*, London: William Eyres. [9]

Schleifer, K. H., and Kilpper-Balz, R. (1984), "Transfer of *Streptococcus faecalis* and *Streptococcus faecium* to the Genus *Enterococcus* nom. rev. as *Enterococcus faecalis* comb. nov. and *Enterococcus faecium* comb. nov.," *International Journal of Systemic Bacteriology*, 34, 31–34. [14]

Sherman, J. M., Mauer, J. C., and Stark, P. (1937), "Streptococcus Fecalis," *Journal of Bacteriology*, 33 (3), 275–282. [12]

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley. [10]

Wainer, H. (2005), *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*, Princeton, NJ: Princeton University Press. [9]

——— (2009), "A Centenary Celebration for Will Burtin: A Pioneer of Scientific Visualization," *Chance*, 22 (1), 51–55. [11]

Wainer, H., and Lysen, S. (2009), "That's Funny...," *The American Scientist*, 97 (July–August), 272–275. [11]

DISCUSSION ARTICLE

Comment

Graham WILLS

CHARTS VERSUS TABLES: A REMATCH

Around 12000 BC, in the Navarra region of northern Spain, an engraver created a graphic representation of ibex herds and their locations relative to local landscape features. Around 2600 BC, in the Sumerian city of Shuruppag, a scribe created a 10×3 table giving the values for areas for a set of pairs of lengths. Charts or tables? Which work better for describing data? After 4000 years of contention it is time to settle this once and for all.

WHAT IS THE POINT OF EITHER?

A core assumption of the article is that the purpose of statistics is to support the process of science. This appears a good assumption. After all, as soon as we ask if something is "better" than something else, then we are trying to optimize something, and we need a way to measure "goodness." Therefore, if we take the purpose of statistics to be to support the scientific method, we should look at what that science actually consists of and what scientists do. Numerous definitions exist, but the following (*http://en.wikipedia.org/wiki/Science*) is a reasonable example:

> Science is a continuing effort to discover and increase human knowledge and understanding through disciplined research. Using controlled methods, scientists collect observable evidence of natural or social phenomena, record measurable data relating to the observations, and analyze this information to construct theoretical explanations of how things work.

The primary purpose of science, according to this, is to *discover*. So, at a basic level, what helps a scientist discover more about their data—a table or a graph?

To help answer this question, let us take a look at a recently published dataset where the purpose was to discover "important features of the data"—the 2009 Statistical Computing/Statistical Graphics Data Expo (*http://stat-computing.org/dataexpo/2009*). Although Gelman suggests that showing many numbers is good, trying to display *all* the data in tabular form would take a little over a million pages (the raw data contain over 120 million

Graham Wills is Visualization and Analytics Research/Architect, IBM, 233 S. Wacker, Chicago, IL 60606 (E-mail: *gwills@us.ibm.com*).

Journal of Computational and Graphical Statistics, Volume 20, Number 1, Pages 28–35
DOI: 10.1198/jcgs.2011.09166d

Table 1. Percentage of canceled flights by day of year.

	J	F	M	A	M	J	J	A	S	O	N	D
1	2.1%	2.3%	1.9%	1.5%	1.3%	1.4%	1.6%	1.8%	1.7%	1%	1.4%	2.1%
2	3.2%	2.5%	2.5%	1.3%	1.4%	1.6%	1.3%	1.8%	1.5%	1.3%	1.3%	1.5%
3	3%	2.7%	2.2%	1.6%	1.1%	1.3%	1%	1.7%	1.8%	1.1%	1.1%	1.6%
4	3.2%	3.2%	2.8%	1.5%	1.2%	1.4%	0.9%	1.7%	2.1%	1.4%	1.1%	1.5%
5	2.7%	3%	3.1%	1.1%	1.2%	1.3%	1.2%	1.6%	1.9%	1.6%	1.2%	2.3%
6	3.6%	3.4%	3.5%	1.2%	1.3%	1.6%	1.4%	1.2%	1.9%	1.3%	1.3%	2.1%
7	4.8%	2.1%	2.4%	1.5%	1.2%	1.4%	1.4%	1.4%	1.5%	1.2%	1.1%	1.9%
8	5.2%	2.6%	3%	1.3%	1.2%	1.6%	1.4%	1.3%	1.4%	1.3%	1.1%	2.2%
9	3.5%	2.4%	2.9%	1.6%	1.5%	1.6%	1.6%	1.4%	1.9%	1.2%	0.9%	2.5%
10	2.4%	2.3%	2%	1.7%	1.8%	1.7%	1.8%	1.8%	2%	1.3%	1.3%	2.3%
11	3.4%	3.6%	1.6%	1.9%	1.6%	1.6%	1.5%	1.7%	6.6%	1.3%	1.4%	3.2%
12	3.7%	3.8%	1.6%	1.2%	1.3%	1.9%	1.4%	1.3%	7%	1.1%	1.4%	2.3%
13	3%	3.3%	3.3%	1.1%	1.4%	2.1%	1.5%	1.6%	6.7%	1.2%	1.3%	2.7%
14	3.1%	3.6%	2.9%	0.9%	1.2%	1.9%	1.9%	1.9%	5.3%	1%	1.4%	2.7%
15	2.8%	2.3%	1.8%	1.2%	1.1%	1.8%	1.5%	1.6%	3.9%	1.2%	1.8%	2.9%
16	2.8%	3.1%	2.9%	1.2%	1.2%	1.6%	1.3%	1.4%	3.9%	1.2%	1.8%	2.4%
17	3.4%	3.1%	1.8%	0.9%	1.5%	1.7%	1.5%	1.5%	2.9%	1.3%	1.3%	2%
18	3.8%	2.9%	2.3%	0.9%	1.7%	1.5%	1.9%	1.2%	2.9%	1.3%	1.1%	1.9%
19	3.1%	1.6%	1.9%	1.2%	1.4%	1.7%	1.7%	1.6%	2.4%	1.3%	1.4%	2.5%
?0	2.8%	1.5%	1.9%	1.1%	1%	1.3%	1.6%	1.4%	2.1%	1.2%	1.3%	2.2%
21	2.5%	2.4%	1.8%	1.1%	1.1%	1.5%	1.6%	1.1%	2%	1%	1.2%	2.5%
22	3.2%	2.7%	1.3%	1.2%	1%	1.8%	1.8%	1.3%	2.4%	0.9%	1.3%	2.1%
23	3.2%	2.2%	1.1%	1.2%	0.8%	1.4%	2.4%	1.5%	2.3%	1%	1.3%	2.7%
24	1.8%	2.9%	1.1%	1.1%	1.4%	1.2%	1.4%	1.7%	1.7%	1.3%	1.1%	2%
25	3.7%	3.4%	1.2%	1.2%	1%	1.6%	1.4%	1.6%	2.1%	1.5%	0.8%	2.3%
26	3%	2.2%	1.2%	1.2%	0.8%	2%	1.7%	1.6%	2%	1.7%	0.8%	2.5%
27	3.1%	1.4%	1.4%	0.9%	0.9%	2.2%	2.1%	1.4%	1.8%	1.4%	0.8%	2.3%
28	2.7%	1.8%	1.3%	1.1%	0.8%	1.8%	2%	1.6%	1.4%	1.3%	0.9%	2.6%
29	3.3%	1.3%	1.1%	0.9%	0.9%	1.6%	1.7%	1.9%	1.4%	1.1%	1.1%	2.3%
30	2.8%		1.3%	1.3%	1.3%	1.4%	1.6%	2%	1.1%	1.2%	1.4%	2.9%
31	2.3%		1.7%		1.3%		1.7%	1.9%		1.5%		2.2%

records) and so summaries will be used. Even between proponents of graphical techniques and tabular reporting, there is consensus here—summarization is an important skill. One goal of the competition was to discover when flight delays occur. We would like to discover which days are good to fly on; so let us create displays that tell us about delays by day of the year. Table 1 is the result; the percent of canceled flights per day, arranged by month and day of month.

Unfortunately this table becomes too big to fit on a single page if we follow the guidelines suggested in the article: showing the absolute numbers as well as the percentages, and showing more precision. A graphical version that shows counts and percentages is more compact (Figure 1), and can also be oriented to read more naturally for standard western reading order (left to right, then top to bottom).

Our job is to compare the two ways of displaying information, with the goal of discovering theories that can be used to explain the world. A proper scientific approach would be to create an experiment to compare the two for different purposes, and then evaluate the

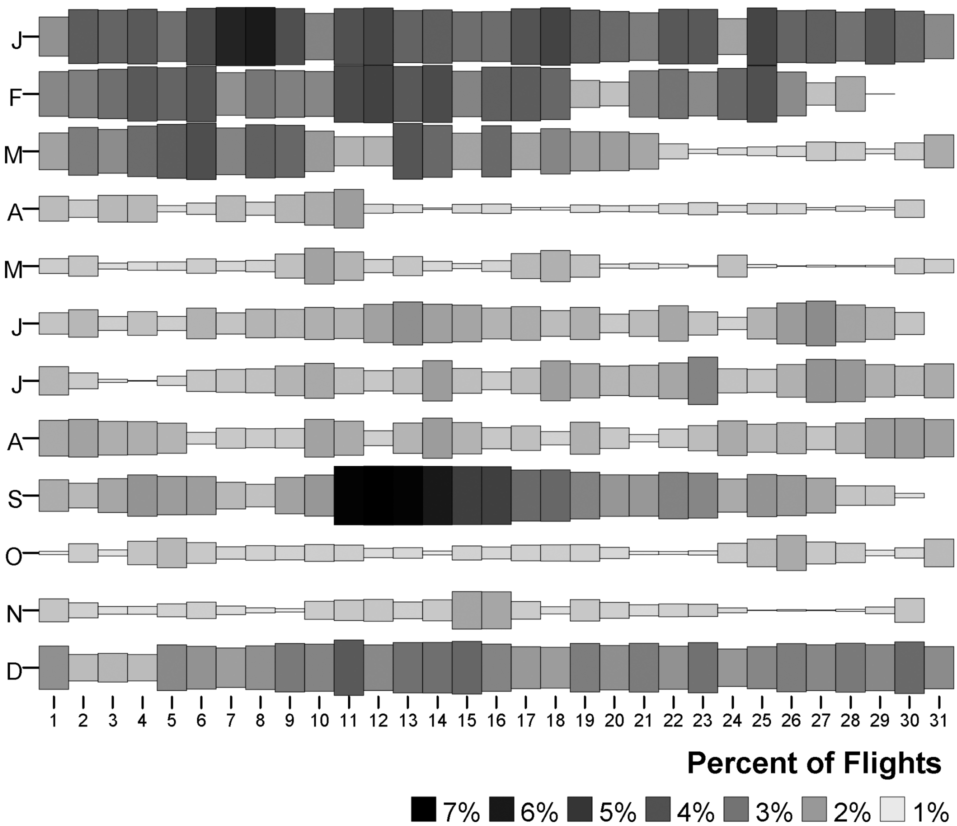

Figure 1. Percentage of flights canceled by day of year. The size is proportional to the absolute counts of flight canceled. Note the immediate outliers following September 11; one year of data where all flights were canceled is enough to make these dates look worse than they should.

results (presumably with another table or another chart?), but my initial guess as to what the results might be would include:

1. *Usability*. Although it is possible to look at the table rows and columns and compare them numerically, it is hard to do so. Our brains are designed to spot visual patterns; for many millions of years it has been much more important to tell broad differences of shape and color than to distinguish regular sequences of patterned squiggles. (Immediately seeing the difference between "1030105" and "1080105" is not as important as rapidly seeing the difference between the deer you are hunting and the tiger that is hunting you!) For the graph, not only does our visual processing power allow us immediately to compare months and discover that April, May, October, and November are great months to fly in, but it also zeroes in on interesting effects after September 11 and, with a little more scanning, around July 4 (about the only good time to fly in the summer). References such as Ware (2008) describe how our brains work and explain why (at least some) charts are far more effective conveyers of information. Because a graphical representation *works with the brain, not against it*, it is a clear winner.

2. *Cost*. Every page published costs money and effort. A simple model might be to estimate cost as proportional to the area needed to display the figure. For this, the graphical form wins—it provides the more space-efficient way to summarize data. However, cost in terms of dollars is not the only issue to be considered. Suppose I wanted to compare the data for canceled flights with that for delayed flights; or compare a few years? With a compact graphical representation, I can show multiple charts on the same page; attempting to add a dimension to Table 1 would make it impossible to discover anything from it. (If we have to flip pages to compare figures, our brains can only hold a few bits of information in them (Ware 2008), making comparisons hard.) In terms of *bang for the area-based buck*, the graphical representation wins again.

3. *Precision*. One argument against graphics is that they lack precision—it is hard to see if there is a difference between May 1 and September 25 in the chart, whereas the table gives us the information that there is a 0.04% difference between them. So is this a win for tables? I would actually argue that it is not a win, because there is no notion of significance attached to either representation. Boo to both representations for *not including measures of variability*. More about this in the next section.

4. *Reproducibility*. One often-touted advantage of tables is that they give actual numbers, and so can be used by other people to replicate experiments. This would indeed be a great advantage, but it suffers from one fatal flaw—the transcription from a published page to data to be used in someone's analysis is not a trivial or error-free step. Whether you scan Table 1 or copy it by hand, the chances of an error are quite high. With the graphical version, there is no such chance and the most error-free solution must be taken; the author should be emailed for a file containing the actual data. By *reducing the chances of error*, the graphical representation wins again.

JUST GIVE ME THE SIGNIFICANCE, MA'AM...

Gelman's article indicates that at the best end of the spectrum of analyses are experiments with good estimates and p-values, where the p-values "really mean something." The example of a graph is given that shows three different lines and does not show any measures that might help the reader know if the slopes really are different. I completely agree that such a graph is misleading. But a table that also omitted that information would be *equally* as misleading. As statisticians we have a responsibility always to show measures of variation. As an example, Figure 2 shows a 95% confidence interval for the mean line for a regression of "fraction of flights delayed" against "total number of flights." We have split the data into two groups: one for July/August, and one for the rest of the year.

We can see immediately that the confidence interval for the mean for the rest of the year does not include the summer month mean. There is a significant difference. An added advantage of the graphic over a table that gave the model estimates and significance test values is that when we look at the actual points, we can see suspicious patterns that should be investigated. Graphical representations *guard against model assumption violations*.

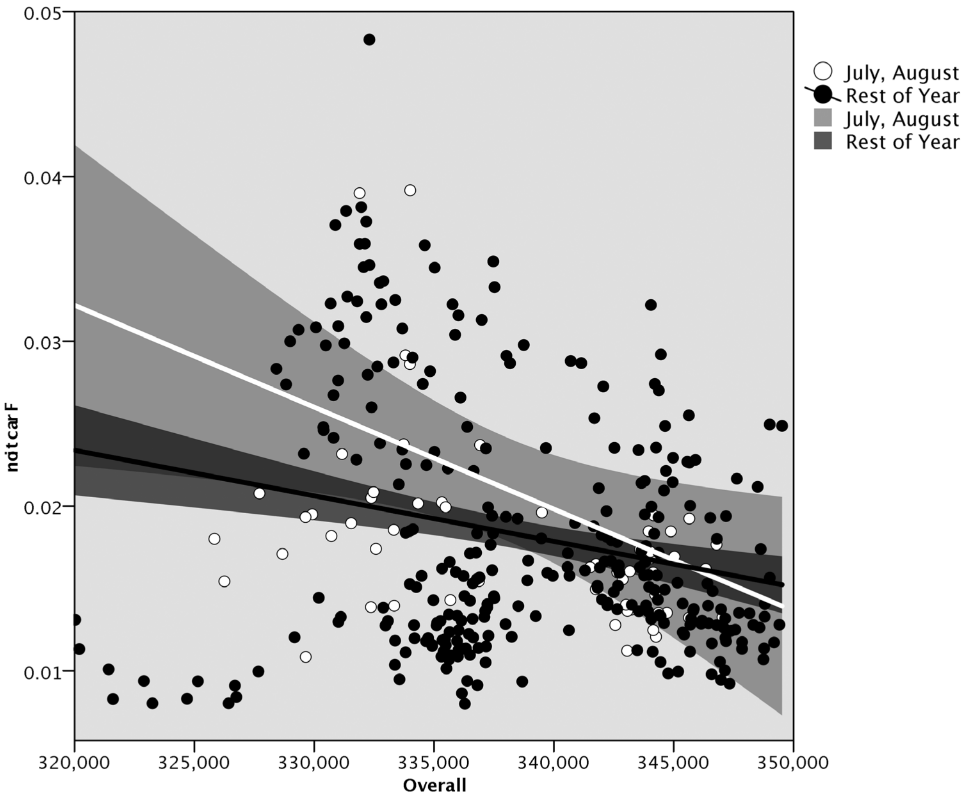

Figure 2. Regression of the fraction of canceled flights by the total number of flights in that month—answering the question "are cancellation probabilities related to the total number of flights?" The confidence interval for the line shows that there is a significant difference in the relationship during the summer. The raw data plotted show some issues with the assumptions used in the regression.

The above discussion is predicated on the basic assumption that p-values are a good idea, and that they are a good way to say what is real and what is not. Certainly, as the article suggests, tables seem to lead to the creation of p-values, whereas charts seem to lead to less definite comparisons. I am not going to argue that p-values are completely without merit, but it is important to understand their failings, which include:

1. vast numbers of people unable to understand what a p-value actually represents.
2. the difference between significance and importance. The airline dataset has 120 million records—pretty much any effect will be significant to several decimal places. But when you are planning a trip, do you care about an effect that is 99.9999% likely to be real, but makes an expected difference of less than three minutes to less than 0.01% of flights?
3. The use of p-values at fixed levels leads to paradoxical statements. It is not hard to find examples where group A is significantly different from group B, and group B is significantly different from group C, but groups A and C are not significantly different. How does knowing the p-values help you in this case? What you really

1	2.1%	2.3%	1.9%	1.5%	1.3%	1.4%	1.6%	1.8%	1.7%	1%	1.4%	2.1%
2	3.2%	2.5%	2.5%	1.3%	1.4%	1.6%	1.3%	1.8%	1.5%	1.3%	1.3%	1.5%
3	3%	2.7%	2.2%	1.6%	1.1%	1.3%	1%	1.7%	1.8%	1.1%	1.1%	1.6%
4	3.2%	3.2%	2.8%	1.5%	1.2%	1.4%	0.9%	1.7%	2.1%	1.4%	1.1%	1.5%
5	2.7%	3%	3.1%	1.1%	1.2%	1.3%	1.2%	1.6%	1.9%	1.6%	1.2%	2.3%
6	3.6%	3.4%	3.5%	1.2%	1.3%	1.6%	1.4%	1.2%	1.9%	1.3%	1.3%	2.1%
7	4.8%	2.1%	2.4%	1.5%	1.2%	1.4%	1.4%	1.4%	1.5%	1.2%	1.1%	1.9%
8	5.2%	2.6%	3%	1.3%	1.2%	1.6%	1.4%	1.3%	1.4%	1.3%	1.1%	2.2%
9	3.5%	2.4%	2.9%	1.6%	1.5%	1.6%	1.6%	1.4%	1.9%	1.2%	0.9%	2.5%
10	2.4%	2.3%	2%	1.7%	1.8%	1.7%	1.8%	1.8%	2%	1.3%	1.3%	2.3%
11	3.4%	3.6%	1.6%	1.9%	1.6%	1.6%	1.5%	1.7%	6.6%	1.3%	1.4%	3.2%
12	3.7%	3.8%	1.6%	1.2%	1.3%	1.9%	1.4%	1.3%	7%	1.1%	1.4%	2.3%
13	3%	3.3%	3.3%	1.1%	1.4%	2.1%	1.5%	1.6%	6.7%	1.2%	1.3%	2.7%
14	3.1%	3.6%	2.9%	0.9%	1.2%	1.9%	1.9%	1.9%	5.3%	1%	1.4%	2.7%
15	2.8%	2.3%	1.8%	1.2%	1.1%	1.8%	1.5%	1.6%	3.9%	1.2%	1.8%	2.9%
16	2.8%	3.1%	2.9%	1.2%	1.2%	1.6%	1.3%	1.4%	3.9%	1.2%	1.8%	2.4%
17	3.4%	3.1%	1.8%	0.9%	1.5%	1.7%	1.5%	1.5%	2.9%	1.3%	1.3%	2%
18	3.8%	2.9%	2.3%	0.9%	1.7%	1.5%	1.9%	1.2%	2.9%	1.3%	1.1%	1.9%
19	3.1%	1.6%	1.9%	1.2%	1.4%	1.7%	1.7%	1.6%	2.4%	1.3%	1.4%	2.5%
20	2.8%	1.5%	1.9%	1.1%	1%	1.3%	1.6%	1.4%	2.1%	1.2%	1.3%	2.2%
21	2.5%	2.4%	1.8%	1.1%	1.1%	1.5%	1.6%	1.1%	2%	1%	1.2%	2.5%
22	3.2%	2.7%	1.3%	1.2%	1%	1.8%	1.8%	1.3%	2.4%	0.9%	1.3%	2.1%
23	3.2%	2.2%	1.1%	1.2%	0.8%	1.4%	2.4%	1.5%	2.3%	1%	1.3%	2.7%
24	1.8%	2.9%	1.1%	1.1%	1.4%	1.2%	1.4%	1.7%	1.7%	1.3%	1.1%	2%
25	3.7%	3.4%	1.2%	1.2%	1%	1.6%	1.4%	1.6%	2.1%	1.5%	0.8%	2.3%
26	3%	2.2%	1.2%	1.2%	0.8%	2%	1.7%	1.6%	2%	1.7%	0.8%	2.5%
27	3.1%	1.4%	1.4%	0.9%	0.9%	2.2%	2.1%	1.4%	1.8%	1.4%	0.8%	2.3%
28	2.7%	1.8%	1.3%	1.1%	0.8%	1.8%	2%	1.6%	1.4%	1.3%	0.9%	2.6%
29	3.3%	1.3%	1.1%	0.9%	0.9%	1.6%	1.7%	1.9%	1.4%	1.1%	1.1%	2.3%
30	2.8%		1.3%	1.3%	1.3%	1.4%	1.6%	2%	1.1%	1.2%	1.4%	2.9%
31	2.3%		1.7%		1.3%		1.7%	1.9%		1.5%		2.2%
	J	F	M	A	M	J	J	A	S	O	N	D

Figure 3. Percentage of flights canceled by day of year. The background shading redundantly encodes the label data so that human visual systems can do their work more easily. Cells that are significant at the 5% level are outlined for those who prefer pass-fail grading to a continuum of scoring.

need is some way of describing how the groups are different. . . perhaps a chart would help?

A FLASHY WASTE OF TIME

A good debater knows when to concede a point, and this is an excellent place to do so. Too many charts are evaluated on coolness and eye-catching style. I would follow Keats's (1820) "Beauty is truth, truth beauty" and state that a chart that does not truthfully display data in a usable way is not beautiful. Edward Tufte's books have done a tremendous service to the graphical display of information, but people who try to follow his example often copy the form and fail to replicate the substance. Tables and charts are meant to aid discovery and further knowledge; if they are simply a pretty display from which little can be learned, they should be evaluated purely as art, not as science. Hang them on the wall (or on your blog), ooh and aah over them, but do not pretend they are useful.

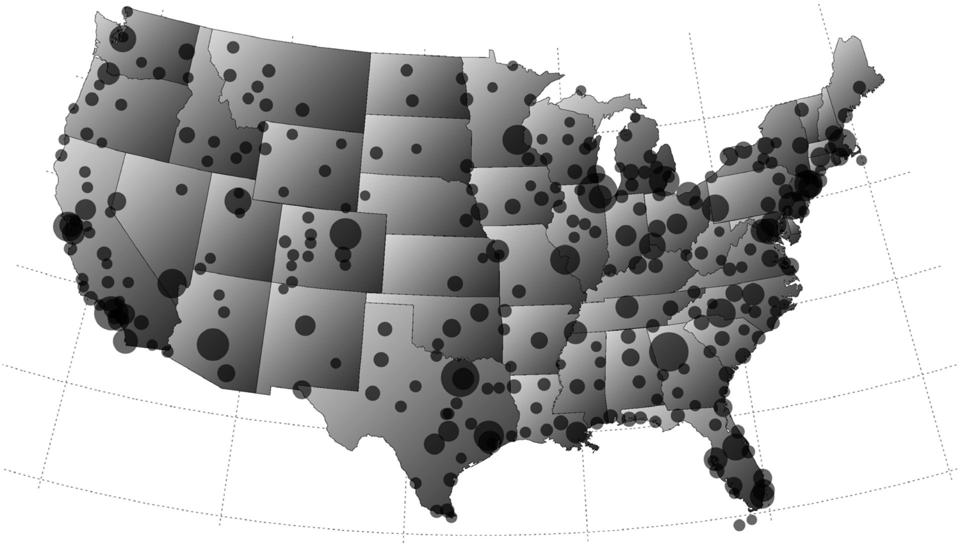

Figure 4. Busy airports. Locations of commercial airports, with the areas of circles proportional to the total number of flights originating from them over the entire time period.

Tangentially, this criticism highlights a difference between tables and charts: when doing an analysis, charts are used often in exploratory mode, but when the results are to be published, tables are more likely to be used. Broadly characterizing, people discover things with charts, but they share their discoveries with tables.

IS THERE A DIFFERENCE BETWEEN A TABLE AND A CHART?

The table presented first in this discussion was not actually produced by a table package; it was produced by the VizML-based engine used in SPSS/IBM that is based on the *Grammar of Graphics* (Wilkinson 1999). In fact that is true of all tables produced by SPSS products; the only important difference between the table and the graph shown earlier is the way the element at each cell location is drawn. For the table, it is shown as an invisible box with a label. For the chart, it is shown as a rectangle with size proportional to the count and color proportional to the fraction. The difference between a chart and a table lies only in the eyes of the user. Which is better? That depends how you classify a figure such as Figure 3.

In the first paragraph I mentioned a very early graphical representation. Figure 4 is a direct descendent of this—a geographical representation of information that, because the data are intrinsically spatial, works much better as a graphical representation than as a table. Time-based data are another form where the strong association with a dimension often makes graphical representations more compelling (Wills 2011). But the differences between tables and charts are not as great as they appear to be; a table can easily be thought of as a simplified chart. And if it adequately represents the data, why would you not use the simplest possible solution?

The original article asks an important question: “What is holding people back from integrating [visualization] much more into data analysis?” Perhaps the reason is that “standard" charts cannot show the depth of detail in many datasets and models, and so people resort to a table—a graphical representation which, if often not useful, does focus firmly on the data—whereas what they really want are visualizations that focus on *their* data. We need to stop telling people to make their data fit the chart, and start building charts that fit the data.

ADDITIONAL REFERENCES

Keats, J. (1820), "Ode on Grecian Urn," in *Lamia, Isabella, The Eve of St. Agnes, and Other Poems*, London: Taylor and Hessey. [33]

Ware, C. (2008), *Visual Thinking for Design*, San Mateo, CA: Kaufmann. [30,31]

Wilkinson, L. (1999), *The Grammar of Graphics. Statistics and Computing*, New York: Springer. [34]

Wills, G. (2011), *Visualizing Time. Statistics and Computing*, New York: Springer. [34]

DISCUSSION ARTICLE

Rejoinder

Andrew GELMAN

I appreciate the time and effort spent by the discussants of my article, and I hope the resulting collection of perspectives will be helpful in advancing the field of statistical communication. To this end, instead of responding to each point made in the discussion, I will focus my rejoinder on a single question: *If graphs are so great* (a sentiment with which we all evidently agree), *why are they not more popular? Why, in so many areas of quantitative social science, do tables still predominate?*

Before getting to this question, let me summarize the arguments in the discussion for the general superiority of graphs over tables. I agree with Wills and Wainer that the relevant goals here are *discovery* and *statistical communication*—to ourselves as well as to others. This is one reason that, in our own writing and research, my colleagues and I have emphasized the continuity between three areas that are often treated separately:

1. statistical modeling,
2. exploratory data analysis,
3. presentation-quality graphics.

Here are some connections between the above three ideas. Statistical modeling is most effective if you understand the data, for which graphics are a valuable tool. Exploratory data analysis has the goal of discovering patterns beyond what is expected under some implicitly or explicitly defined model; as such, improvements in models can lead to more effective exploratory analysis. Ideas such as clarity, focus, and awareness of cognitive processing that are relevant for graphical presentations—and specific techniques such as sensible ordering, direct labeling of points and lines, and descriptive plot titles—translate directly into exploratory analysis. As a statistician, your first client for any analysis is yourself, and you want to use presentation-quality statistical graphics for your own exploratory data analysis for the same reason you want it for others—to effectively connect the numbers to your substantive story. (Slipping for a moment into the ironic mode favored by

Andrew Gelman is Professor, Department of Statistics and Department of Political Science, Columbia University (E-mail: *gelman@stat.columbia.edu*; url: *http://www.stat.columbia.edu/~gelman/*).

Journal of Computational and Graphical Statistics, Volume 20, Number 1, Pages 36–40
DOI: 10.1198/jcgs.2011.09166rej

Briggs, we can rephrase this point as follows: The most effective way to fool others is to first fool yourself.)

As noted, ideas from graphical presentations are relevant to exploratory data analysis. The converse is also true: as has been emphasized by Edward Tufte, Howard Wainer, and others, statistical principles of conveying potentially unexpected information—the same ideas that are central to exploratory data analysis—also make for powerful presentations for non-statisticians. Finally, exploratory and presentation graphics are typically presented as tools for display of raw data, but both these sorts of displays can be more effective in the context of models (as demonstrated, I hope, in my own applied work). Tufte had it right when he wrote of the visual display of quantitative information—not just “data” but, more generally, “quantitative information,” which can include all sorts of derived quantities, ranging from the Consumer Price Index to logistic regression coefficients (generally best when divided by 4, to approximately transform to the probability scale). Much confusion has arisen because people think of statistical visualization to be about showing the raw data. No, it is more general than that.

In contrast, I would argue that the use of tables represents a form of pre-quantitative thinking—numbers as totems rather than quantities. This gets back to Wills’s useful formulation of statistical graphics as comparisons (an idea also discussed by Tukey and Cleveland in their classic books). Almost every result from any analysis or model we have fit is ultimately, in some way, a comparison, and one of the challenges and intellectual rewards of graphical display is to identify and isolate comparisons of interest, while it is still possible for unanticipated comparisons to be revealed.

TABLES ARE EASIER TO MAKE THAN GRAPHS

Why are graphs not more popular? Our short answer: Good statistical graphics are hard to do, much harder than running regressions and making tables. In Wills’s “bang for the buck” calculation, we must also include the cost to the author of creating and producing the graph.

Here is how I see things usually going in a work of applied statistics:

Step 1: Exploratory data analysis—some plots of raw data, possibly used to determine a transformation.

Step 2: The main analysis—maybe model-based, maybe nonparametric, whatever. It is typically focused, not exploratory.

Step 3: That’s it.

I have a big problem with Step 3 (as maybe you could tell already). Sometimes you will also see some conventional model checks such as chi-squared tests or qq plots, but rarely anything exploratory. Which is really too bad, considering that a good model can make exploratory data analysis much more effective. And, conversely, I will understand and trust a model a lot more after seeing it displayed graphically along with data.

But, again, here is the problem. Anyone can run a regression or an Anova! Regression and Anova are easy. Graphics are hard. Maybe things will change with the software and new media—various online tools such as Gapminder make graphs that are far far better than the Excel standard, and, with the advent of blogging, hot graphs are popular on the internet. We have come a long way from the days in which graphs were in drab black-and-white, when you had to fight to get them into journals, and when newspaper graphics were either ugly or (in the case of *USA Today*) of the notoriously trivial, "What Are We Snacking on Today?", style.

Even now, though, if you are doing research work, it is much easier to run a plausible regression or Anova than to make a clear and informative graph. I am an expert on this one. I have published thousands of graphs but created tens of thousands more that did not make the cut.

One problem, perhaps, is that statistics advice is typically given in terms of the one correct analysis that you should do in any particular setting. If you are in situation A, do a two-sample t-test. In situation B, it is Ancova; for C you should do differences-in-differences; for D the correct solution is weighted least squares, and so forth. If you are lucky, you will get to make a few choices regarding selection of predictors or choice of link function, but that is about it. And a lot of practical advice on statistics actually emphasizes how little choice you are supposed to have—the idea that you should decide on your data analysis before gathering any data, that it is cheating to do otherwise.

One of the difficulties with graphs is that it clearly does not work that way. Default regressions and default Anovas look like real regressions and Anovas, and in many cases they actually are! Default graphics may sometimes do a solid job at conveying information that you already have (see, e.g., the graphs of estimated effect sizes and odds ratios that are, I am glad to say, becoming standard adjuncts to regression analyses published in medical and public health journals), but it usually takes a bit more thought to really learn from a graph. Even the "superplot" of voting by income in Mississippi, Ohio, and Connecticut—a graph I envisioned in my head back in 2003 (!) back at the very start of our *Red State, Blue State* project, before doing any data analysis at all—even the superplot required a lot of tweaking to look just right.

Perhaps things will change. An important thread of current research ties graphics more closely to modeling, with one goal being to develop a default process for looking through lots of graphs in a useful way. Researchers were doing this back in the 1960s and 1970s—methods for rotating point clouds on the computer, and all that—but I am thinking of something slightly different, something more closely connected to fitted models. But right now, no, graphs are harder, not easier, than formal statistical analysis.

From the other direction, tables are easy, and there is a single standard way to, for example, tabulate a fitted regression model. I doubt that most economists or political scientists reflect upon what useful information, if any, is conveyed by such tables. It is just the default. I firmly believe in the superiority of graphs but it takes a lot to beat the default, especially when the alternative takes more effort.

SUMMARY

Graphs are not just for displaying raw data. Perhaps if researchers realize they can display their inferences graphically, building on basic summaries such as the so-called "secret

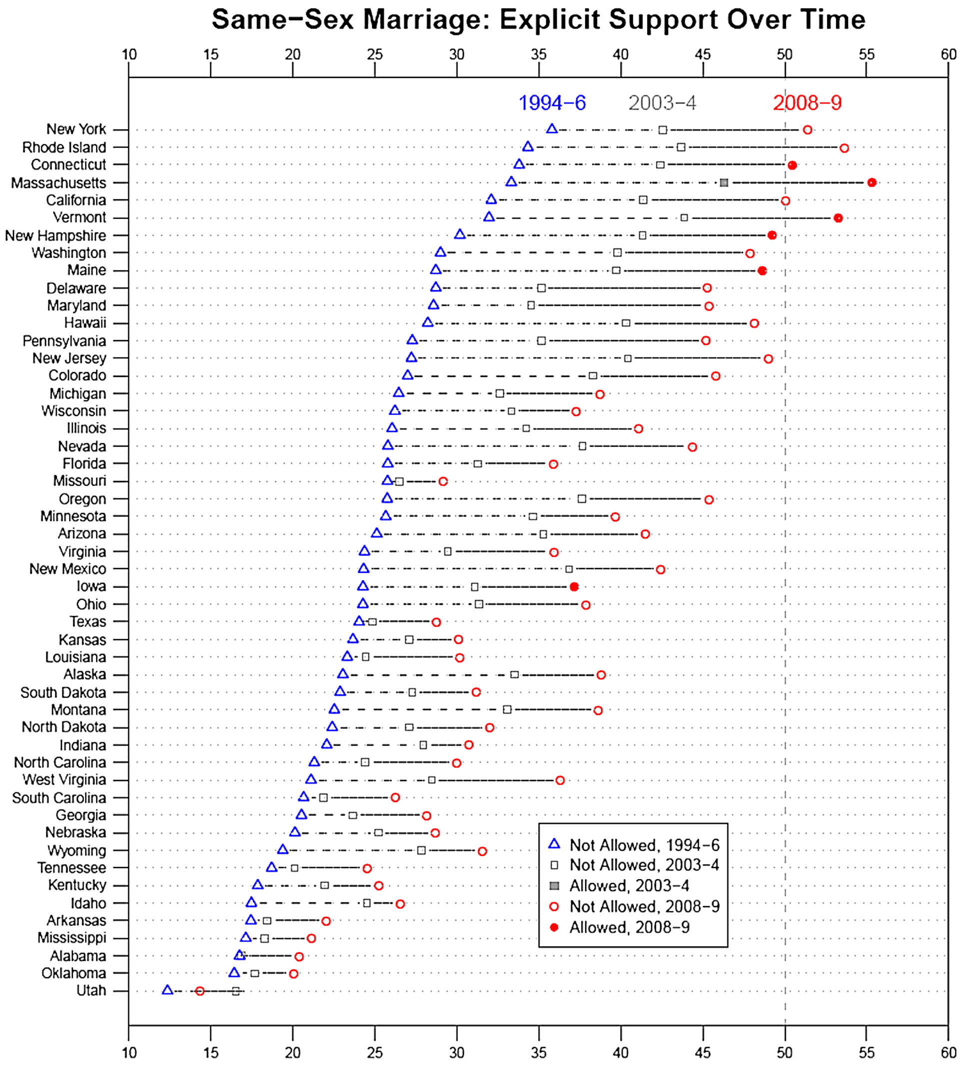

Figure 1. Estimates of support for gay marriage, as estimated using multilevel regression and post-stratification as applied to a collection of public opinion polls over several years. States are ordered by opinion in 1994–1996. Approximately as much change has occurred in the last four years (solid lines) as in the previous eight (dashed lines) and states with higher levels of early support changed the most (the opposite of what one might expect from a simple "regression to the mean" pattern, and a difference that persists if the probabilities are transformed on to the logit scale). Policy is as of June 2009. From Lax, J. R., and Phillips, J. (2009), "Gay Rights in the States: Public Opinion and Policy Responsiveness," *American Political Science Review*, 103 (3). The online version of this figure is in color.

weapon" (time-series or cross-sectional plots of estimates and standard errors) and automatic software such as coefplot() in R, they will learn the advantages of visual display of inferences, as illustrated in the beautiful graph created by Jeffrey Lax and Justin Phillips (and reproduced in Figure 1 here) of trends on opinions and policies on gay marriage in 50 states. I am sorry to report that the accompanying article published in the *American Political Science Review* also includes three ugly tables. Still, Figure 1 (which has some aspects of a “tableplot” as defined by Friendly and Kwan) represents a huge amount of progress, and I hope and expect that others will be motivated to follow its example. Wainer’s historical overview also is useful here, reminding us that the hard work required to make good graphs can pay off, but that change—even useful change—can be slow.

ACKNOWLEDGMENT

We thank Richard Levine for organizing the discussion and the Institute of Education Sciences for partial support of this work (ED-GRANTS-032309-005).