[Vanderbilt Law Review](https://scholarship.law.vanderbilt.edu/vlr)

Volume 59 | Issue 6 Article 1

11-2006

On the Effective Communication of the Results of Empirical Studies, Part II

Lee Epstein

Andrew D. Martin

Matthew M. Schneider

On the Effective Communication of the Results of Empirical Studies, Part I*

Lee Epstein, Andrew D. Martin, & Matthew M. Schneider+

---

* Prepared for presentation at the Vanderbilt Law Review's Symposium on Empirical Legal Scholarship, February 17, 2006.

+ Lee Epstein (http://epstein.law.northwestern.edu/) is the Beatrice Kuhn Professor of Law at Northwestern University School of Law; Andrew D. Martin (http://adm.wustl.edu/) is Professor of Law and Professor of Political Science at Washington University in St. Louis; Matthew M. Schneider is a Ph.D. student at Washington University in St. Louis. We are grateful to the National Science Foundation for supporting our research and to Jessica Flanigan and Kate Jensen for their research assistance. We also thank Gary Klass for his helpful comments. We used Stata and R to conduct the analyses and to produce the graphs presented in this article. The project's web site (http://epstein.law.northwestern.edu/research/communicating.html) houses replication materials. Please send all correspondence to Lee Epstein. Email: lee-epstein@northwestern.edu. Post: Northwestern University School of Law, 357 E Chicago Ave, Chicago, IL, 60611-3069.

I. INTRODUCTION

In an important and certainly timely article published in the *N.Y.U. Law Review*, Nancy C. Staudt demonstrates that, in taxpayer standing cases, judges are motivated by politics but can be constrained when the law is clear and oversight exists.[1] As part of that demonstration, Professor Staudt offers an empirical analysis of the decision to grant standing to federal taxpayers—the results of which we reproduce in Table 1.[2]

Table 1

	Federal Taxpayer N = 120
Spending	β = -1.345 S.E. = .544 p = .013*
Spending and Establishment Clause	β = 1.212 S.E. = .501 p = .015*
Party of the Appointing President	β = .148 S.E. = .485 p = .760
Plaintiff Politics	β = 1.141 S.E. = .488 p = .019*
Plaintiff Sought Standing on More Than One Ground	β =-.240 S.E. = .447 p = .591
Court Discussed Standing	β =-.973 S.E. = .583 p = .095
Case Decided After *Flast v Cohen*	β =.098 S.E. = .574 p = .941
Constant	β =.806 S.E. = .978 p = .409

Table 1: Results of Nancy C. Staudt's analysis of the decision to grant standing (coded 1) or not (coded 0) to federal taxpayers. Each cell contains coefficient estimates from a logistic regression model, (asymptotic) standard errors, and p-values. * indicates that $p < .05$.

1. Nancy C. Staudt, *Modeling Standing*, 79 N.Y.U. L. REV. 612, 617 (2004).

2. *Id.* at 657. Using data generously provided to us by Professor Staudt, Table 1 parallels the "Federal Taxpayer" model in Table 3 of her article. We have reinserted the negative signs that were left out of the original publication of the article.

What are we to make of this rather ominous-looking table? Professor Staudt suggests two key takeaways. First, the analysis, she reports, shows that doctrine helps explain standing decisions even when political factors are taken into account. Both legal variables ("Spending" and "Spending and Establishment Clause") are statistically significant, controlling for all other factors listed in the table. Second, she finds an important role for the politics of the plaintiff: Judges are more likely to grant standing to a liberal plaintiff, regardless of their own political leanings.

No doubt, the data support Professor Staudt's claim about the importance of politics. The asterisk on the Plaintiff Politics variable, for example, tells us that a "statistically significant" relationship exists between a plaintiff's political ideology and the decision to grant standing.[3]

Moreover, the positively signed coefficient (1.141, and not -1.141) conveys information about the direction of that relationship: Given Staudt's coding rules, and accounting for the other six variables in her analysis, liberal plaintiffs are significantly more likely than conservative plaintiffs to receive a favorable decision on their standing claim.

On the other hand, Table 1 is not just ominous-looking and off-putting to most readers; it communicates information of little value either to its audience or author. For starters, while we know that liberal plaintiffs are more likely to be granted standing, we do not learn how much more likely. 0.2 times more likely? Two times? Or perhaps even four times? Certainly, this is the quantity of interest that matters most to readers.[4] But it is not one that we can readily learn from a tabular display of coefficients; in fact, all we learn from the coefficient of 1.141 is that, controlling for all other factors, as we move from a conservative plaintiff to a liberal plaintiff, we move up about 1 on a logit scale—an esoteric statement at best.[5]

Moreover, even if we could calculate a "best guess" about the likelihood of standing for a plaintiff based on political ideology, Table 1 conveys no useful information about the error surrounding that guess. Suppose Professor Staudt reported that the probability of a

3. A relationsbip is said to be "statistically significant" if its existence cannot be explained by chance alone. The term is shorthand for rejecting a null hypothesis of no effect. *See* ALAN AGRESTI & BARBARA FINLAY, STATISTICAL METHODS FOR THE SOCIAL SCIENCES 174 (3d ed. 1997).

4. Indeed, to assess any conclusions authors draw, we need to know whether "statistically significant" has any substantive effect. In Professor Staudt's case, her conclusion that judges with different attributes reach different decisions would have a good deal more punch if states were, say, twice as likely to commute the sentences of women, rather than 0.2 times more likely.

5. To make matters worse, because the logit scale is non-linear, moving up 1.141 units will result in different probabilities of clemency depending on where we start on the scale.

judge granting standing to a conservative plaintiff is about 0.40, while it is 0.60 for a liberal plaintiff. That would seem to constitute a big substantive difference. But this difference is not necessarily very important if the 95% confidence interval surrounding the 0.40 was [0.19-0.59], meaning that "[o]ur best guess about the probability of standing for conservatives is 0.40 but it could be as low as 0.19 and as high as 0.59 (which is very close to 0.60, our best guess for liberals)." Such is the reality of the statistical world: We can never be certain about our best guesses (i.e., inferences) because they themselves are based on estimates. We can, however, report our level of uncertainty (e.g., a confidence interval) about those guesses.[6]

In making these points, we do not mean to pick on Professor Staudt. Quite the opposite: In many other ways, her article is exemplary empirical scholarship.[7] But her work also demonstrates the problems legal scholars confront when attempting to convey their research results. Indeed, based on a survey we conducted of law review articles making use of quantitative data, we can safely say that nearly each and every one would have benefited from greater attention to the communication of their results.[8]

Indeed the benefits are many. Most crucially, it seems nearly incontrovertible that moving towards more appropriate and accessible presentations of data will heighten the impact of empirical legal scholarship on its intended audience—be that audience other academics, students, policy makers, lawyers, or judges—not to mention raise the level of intellectual discourse among scholars themselves.[9] When analysts write that "the coefficient on Plaintiff Politics is statistically significant at the 0.05 level," they likely immediately turn off many potential their readers. But if they were to translate their findings into a visual display, as we do in Figure 1, they would he able to supplant sterile statistical claims with the far more pleasing: "Other things being equal, the estimated probability of a liberal plaintiff being granted standing in a federal taxpayer lawsuit is .41 (though it could be as low as .25 or as high as .60).[10] That

6. Most scholars, Professor Staudt included, appreciate this fact and supply the error surrounding their estimated coefficients. But, as even the brief example in the text shows, this is far less useful than conveying uncertainty about the *substantive effect* of the results.

7. We also should note that Professor Staudt was one of the few scholars out of the many we contacted who was willing to share her data—yet again underscoring the need for replication policies in the law reviews.

8. For more on our survey, see *infra* Part II.

9. This paragraph adopts and parallels sentiments expressed in Gary King, Michael Tomz & Jason Wittenburg, *Making the Most of Statistical Analyses: Improving Interpretation and Presentation*, 44 AM. J. POL. SCI. 347, 347-48 (2000).

10. The figures .25 and .60 represent the 95% confidence interval.

probability decreases for conservative plaintiffs, to only .18 (though it could be low as .08 or as high as .38)."[11] Unlike the terms "coefficient" or "0.05 level," this statement is easy to understand even by the most statistically challenged among us.[12]

Figure 1

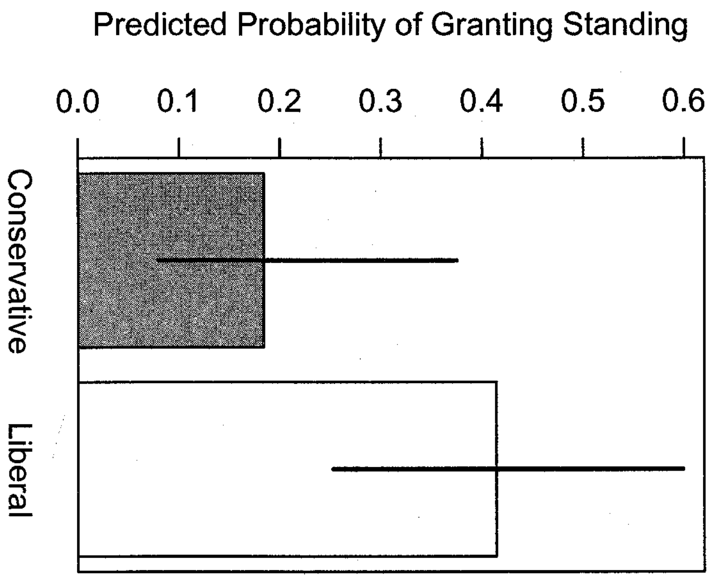


Figure 1: Predicted probabilities of being granted standing for conservative and liberal plaintiffs, from the results in Table 1. We held all dichotomous covariates at their sample modes and all others at their sample means. The vertical (error) bars represent 95% confidence intervals for the predicted probabilities.

How can legal academics develop such claims from their data? How might they go about visually depicting these claims? In what follows we draw on a growing literature in the social and statistical sciences to delineate, first, general principles for communicating the fruits of empirical legal research (Part III) and for graphing quantitative information (Part IV). In the next installment of this Article, we outline specific strategies for effectively (and accessibly) presenting data and statistical results. From these strategies, we devise a set of protocols for implementation by legal publications. We

11. The differential precision of these estimates is caused by the non-linearity of the model. The predictions would differ in magnitude if we were to hold the other covariates at different values.

12. It is important to note that while the confidence intervals for these two predictions overlap, there is still a statistically significant difference between the predictions. *See* Peter C. Austin & Janet E. Hux, *A Brief Note on Overlapping Confidence Intervals*, 36 J. VASCULAR SURGERY 194, 194-95 (2002) (explaining that two means can have confidence intervals that abut or overlap and still be significantly different from one another).

start, though, at the beginning—with our motivation for writing this paper.

II. MOTIVATION

Why an article on the presentation of data and results? We were driven by two separate forces: recent developments 1) in and 2) beyond the legal academy. As we explain below, these coalesce. Precisely at a moment when law professors desire to convey their research to diverse audiences, other disciplines have made a good deal of progress toward reaching that end.

A. Developments in the Legal Academy

To claim that empirical work is now a fundamental part of legal scholarship borders on the boring.[13] It has been said (and documented) too many times for us to recount here; even the Association of American Law Schools ("AALS") acknowledged the increasing centrality of empirical work when it devoted its 2006 annual meeting to the topic.[14] More important for our purposes is that while law professors are increasingly making use of data in their scholarship and while the data work housed in their studies is (generally) of a high quality, its presentation could be improved.

This much we learned from a systematic analysis of articles, essays, and notes published in twenty leading law reviews between 2000 and 2004.[15] It will come as no great surprise that (quantitative)

13. We can differentiate between two types of evidence employed in empirical work: quantitative (numerical) and qualitative (non-numerical). Though neither is any more "empirical" than the other, our focus here is (generally) on the presentation of quantitative data. *See, e.g.*, Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 2 (2002) ("The word 'empirical' denotes evidence about the world based on observation or experience. That evidence can be numerical (quantitative) or nonnumerical (qualitative); neither is any more empirical than the other.").

14. "There is a long tradition of empirical scholarship in law and there has recently been a burgeoning of interest in conducting empirical research in America's law schools." ASS'N OF AM. LAW SCHOOLS, EMPIRICAL SCHOLARSHIP: WHAT SHOULD WE STUDY AND HOW SHOULD WE STUDY IT? 50 (2006), *available at* http://www.aals.org/am2006/program/finalprogrammain2006.pdf.

15. We included the twenty flagship, student-edited journals with tbe highest "impact" from 1997-2004. Law Journals: Submissions and Ranking, http://lawlib.wlu.edu/LJ/index.aspx (last visited Nov. 22, 2004). They are (in order of impact): YALE LAW JOURNAL, COLUMBIA LAW REVIEW, VIRGINIA LAW REVIEW, NEW YORK UNIVERSITY LAW REVIEW, STANFORD LAW REVIEW, CORNELL LAW REVIEW, HARVARD LAW REVIEW, UCLA LAW REVIEW, MINNESOTA LAW REVIEW, UNIVERSITY OF PENNSYLVANIA LAW REVIEW, VANDERBILT LAW REVIEW, TEXAS LAW REVIEW, DUKE LAW JOURNAL, NORTHWESTERN UNIVERSITY LAW REVIEW, UNIVERSITY OF CHICAGO LAW REVIEW, SOUTHERN CALIFORNIA LAW REVIEW, CALIFORNIA LAW REVIEW, WILLIAM AND MARY LAW REVIEW, IOWA LAW REVIEW, AND MICHIGAN LAW REVIEW. For the figures that follow in the text, we counted all articles, notes, and essays.

evidence—whether in the text, tables, or figures—appears in a non-trivial number: 245 of the 3695 total articles, to be precise. Also likely to no one's surprise, methods for depicting data varied considerably, from simple bar charts designed to summarize information collected by the author to detailed tables meant to convey inferences about information the author has not collected. Finally, the range of substantive topics under scrutiny impressed us. From constitutional law to commercial law, from statutory interpretation in the tax context to the use of scientific evidence in criminal cases, from the appointment through the retirement of judges, no subject now seems beyond the reach of empirical analysis.

And yet it was the commonalities among the articles—and not their differences—that struck us. On the one hand, as we suggest above, the data work was competently executed. In only 18.8% of the 245 articles were we able to identify a statistical issue that required some remedying. In a few instances, the problems were rather severe, but in the great majority, they were far more minor: the occasional missing N, the neglect of model fit, and so on. Clearly, law professors have become adept data handlers.

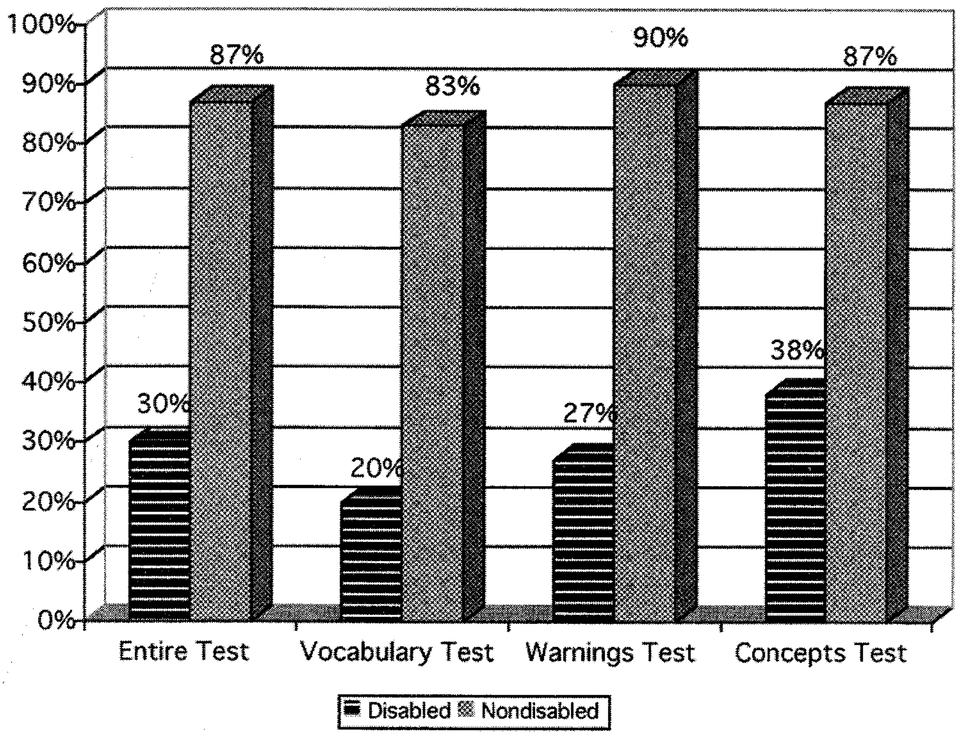
On the other hand, as judged by emerging standards for communicating data summaries and results, they are far less adept at conveying information. So, for example, while most scientists have "declared a war" on tables—or, at the least, have expressed a strong preference for graphical displays—law professors seem to have embraced them.[16] The 245 articles all made use of quantitative evidence, but fewer than 40% (N = 92) employed figures; even when conveying the results of multivariate statistical analyses, authors commonly eschewed graphical displays for tables (i.e., Table 1 was the rule, not the exception). Moreover, when figures did make their way into the articles, they were often so "busy" or otherwise marred that

16. For more on the general principle of "graphs, not tables," see *infra* Part III.C. Worth noting here, though, is that the preference for figures over tahular displays is hardly new. As early as 1801, William Playfair, a "key figure in the history of quantitative graphics," wrote that information "obtained [in charts] in five minutes . . . would require whole days to imprint on the memory, in a lasting manner, hy a table . . . ." Patricia Costigan-Eaves & Michael Macdonald-Ross, *William Playfair (1759-1823)*, 5 STAT. SCI. 318, 318, 323 (1990) (emphasis omitted). With new developments in the social and statistical sciences, see *infra* Part II.B, Playfair's sentiment has become a near hattle cry. This is not to say, however, that scientists always practice what they preach. "Statisticians recommend graphical displays but often do not follow this recommendation in presenting their own research." Andrew Gelman et al., *Let's Practice What We Preach: Turning Tables into Graphs*, 56 AM. STATISTICIAN 121, 121 (2002). After reviewing tables published in the March 2000 issue of the JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, Gelman and his colleagues show that using well-designed graphs is actually superior to using tables. *Id.*

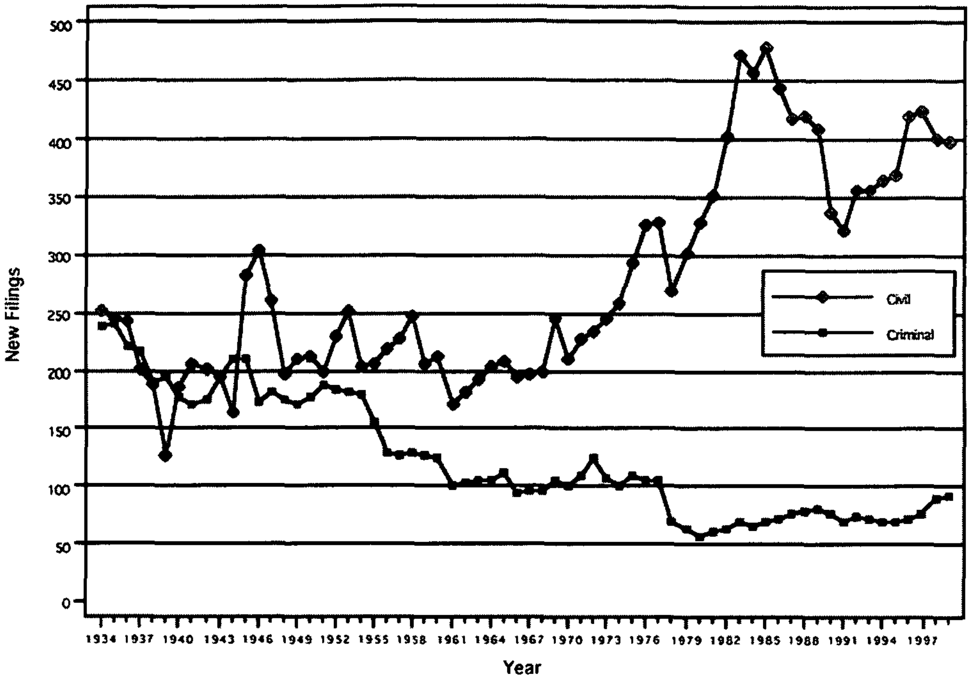
their impact was all but lost.[17] This is the case, in varying degrees, for the trio of pictures we reproduce in Figure 2—all of which appeared in law publications.

These are but three examples; the list of ills is longer, and we devote most of this article to offering antidotes. But the basic point should not be missed: A very large percentage of empirical work could be improved if authors would pay as much attention to how they present information as to how they collect and analyze it.

Figure 2



17. For more on the concepts of "busyness" and "impact," see *infra* Part IV.A.



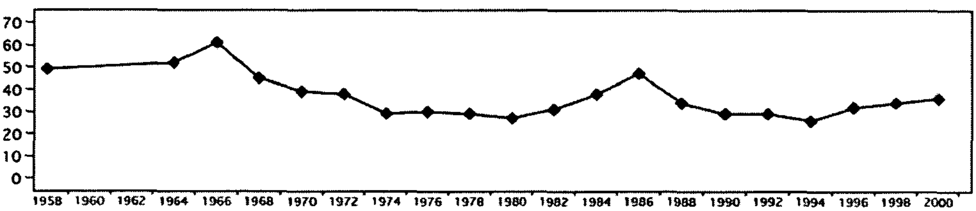


Figure 2: Each panel displays a figure (reproduced from a law review article) that obscures the data and thus interferes with visualization. The graph in the top panel is cluttered with irrelevant elements, including the depth cue, internal data labels, tick marks, and legend. The cross hatching and the number of tick marks is also distracting. In the middle panel, the dark grid and legend obscure visualization of the data, as do the non-circular subelements. In the bottom panel, the inclusion of zero obscures patterns in the data, and the connected lines make it difficult to observe missing years in the dataset. We offer correctives in Figures 9, 10, and 13.

The existence of this gap is not too surprising. While a spate of books and articles provides guidance to law professors seeking to undertake empirical work, we cannot identify one—not one—explicitly geared toward supplying counsel on how to present data and results.[18]

18. *E.g.*, Epstein & King, *supra* note 13, at 1 (delineating the "rules of inference" but having very little to say about conveying data and research results). *But see Id.* at 26-27 (discussing histograms). We can say the same of primers on social science methods, statistics, or both. *E.g.*, DAVID L. FAIGMAN ET AL., SCIENCE IN THE LAW (2002); MICHAEL O. FINKELSTEIN & BRUCE LEVIN, STATISTICS FOR LAWYERS (2001); JOHN MONAHAN & LAURENS WALKER, SOCIAL SCIENCE IN LAW

By the same token, various legal style guides are silent with regard to the communication of empirical evidence. As we might expect, THE BLUEBOOK speaks only to the question of how to cite others' figures and tables, not to matters of original presentation. Ditto for the flagship law journals (or at least the ten we consulted).[19] Less expected, though, is that relative to major journals in the sciences, important peer-reviewed journals in law provide only skeletal instructions to authors on how to convey numerical data.[20]

(2006). An exception comes in the latter's claim that while "graphs are useful for revealing key characteristics of a batch of numbers," readers must take care to examine their "markings" to ensure that they are not conveying deceptive information. FAIGMAN ET AL., *supra*, at 177. For more on the assumption that readers are "dumb" when it comes to decoding figures (but apparently not in reading tables), see *infra* Part IV.

19. We visited the websites of the COLUMBIA LAW REVIEW, HARVARD LAW REVIEW, MICHIGAN LAW REVIEW, NEW YORK UNIVERSITY LAW REVIEW, NORTHWESTERN LAW REVIEW, STANFORD LAW REVIEW, UNIVERSITY OF CHICAGO LAW REVIEW, UNIVERSITY OF PENNSYLVANIA LAW REVIEW, VANDERBILT LAW REVIEW, and YALE LAW JOURNAL.

20. The instructions for preparing figures and tables in the JOURNAL OF EMPIRICAL LEGAL STUDIES ("JELS"), most of which deal with matters of format, are as follows:

> Initial caps and italics are used for each major word in table column headings.
>
> Only the first word is capitalized in table rows.
>
> Horizontal lines separate the table heading from the table body and are also used at the end of a table.
>
> Initial caps are used for each major word in the title of tables: E.g., "Table 1: Number of Verdicts in Tort Trials by Jurisdiction by Decade"
>
> Only the first word of a figure title is capitalized. E.g., "Figure 1: Average and median damage awards in tort verdicts by year."
>
> Authors should strive to use a font in figures that is compatible with JELS's text font. Times Roman is acceptable.
>
> Tables and figures should stand on their own. When appropriate, authors should include an explanatory note for a table or figure. The goal is to have the table or figure "stand on its own" so that a busy reader can understand the table without reading the whole article. JELS, Author Guidelines, http://www.blackwellpublishing.com/submit.asp?ref=1740-1453&site=1.

The JOURNAL OF LAW, ECONOMICS, & ORGANIZATION'S ("JLEO") instructions are even skimpier. JLEO, Information for Authors, http://www.oxfordjournals.org/our_journals/jleorg/for_authors/ms_preparation.html. But JELS and JLEO pale in comparison to, say, NATURE, *see* NATURE, Formatting Guide: Manuscript Preparation and Submission, http://www.nature.com/nature/authors/gta/index.html, or even POLITICAL ANALYSIS, the journal of the Political Methodology Section of the American Political Science Association and the Society for Political Methodology, which also contains detailed instructions for the preparation of tables (though not graphs):

> Numbers in the text of articles and in tables should be reported with no more precision than they are measured and are substantively meaningful. In general, the number of places to the right of the decimal point for a measure should be one more than the number of zeros to the right of the decimal point on the standard error of this measure.
>
> Variables in tables should be rescaled so the entire table (or portion of the table) has a uniform number of digits reported. A table should not have regressions coefficients reported at, say, 77000 in one line and .000046 in another. By appropriate rescaling (e.g., from thousands to millions of dollars, or population in millions per square mile

In short, while law professors face no shortage of sources from which to learn about research design, data analysis, and citation practices, they lack a standard reference or even basic guidance on how to convey the fruits of their labors. This would be less than optimal for researchers working in any discipline, but it is uniquely problematic for legal academics. Perhaps more than most, law professors have a strong interest in conveying their results to both the statistically informed and uninformed. Moving away from rather meaningless statements—meaningless, at least, to most lawyers, policy makers, judges, and even fellow law professors—about "regression coefficients," "statistical significance," and "p-values," towards displays housing estimates of key quantities of substantive interest may therefore provide especially large payoffs for legal researchers and their audiences.

If this is so, then King et al.'s wise words—that the "raw results of any statistical procedure . . . [should] require little specialized knowledge to understand"[21]—are particularly apt for law professors. Only by following them will they and their communities in and outside the faculty commons reap the full benefits of empirical work.

> to population in thousands per square mile), it should be possible to provide regression coefficients that are easily comprehensible numbers. The table should clearly note the rescaled units. Rescaled units should be intuitively meaningful, so that, for example, dollar figures would be reported in thousands or millions of dollars. The rescaling of variables should aid, not impede, the clarity of a table.
>
> In most cases, the uncertainty of numerical estimates is better conveyed by confidence intervals or standard errors (or complete likelihood functions or posterior distributions), rather than by hypothesis tests and p-values. However, for those authors who wish to report "statistical significance," statistics with probability levels of less than .001, .01, and .05 may be flagged with 3, 2, and 1 asterisks, respectively, with notes that they are significant at the given levels. Exact probability values may always be given. POLITICAL ANALYSIS follows the conventional usage that the unmodified term "significant" implies statistical significance at the 5% level. Authors should not depart from this convention without good reason and without clearly indicating to readers the departure from convention.
>
> It cannot be stressed too much that all articles should strive for maximal clarity. Choices about figures, tables, and mathematics should be made so as to increase clarity. In the end all decisions about clarity must be made by the author (with some help from referees and editors).

POL. ANALYSIS, Information for Authors, http://www.oxfordjournals.org/polana/for_authors/general.html.

21. King et al., *supra* note 9, at 347. Daniel B. Wright puts it this way: "A second year undergraduate should be able to read the results section, on its own, and know what the main findings are." Daniel B. Wright, *Making Friends with Your Data*, 73 BRIT. J. EDUC. PSYCHOL. 123, 132 (2003).

B Developments Outside the Legal Academy

Surely the existing state of the legal literature was a key motivating factor for this Article, but it is hardly the only one. Actually, we would have been unable to offer a corrective in the absence of developments external to the legal academy.[22] Three are worthy of mention.[23]

First, a burgeoning literature in the statistical and social sciences focused on the presentation of data has arisen in response to the growing prevalence of data displays, and especially graphs in scholarly works.[24] Some of this amounts to little more than "armchair," though valuable, guidelines for graphic design.[25] But other work—and there is a good deal of it—invokes observational and experimental evidence to learn about how accurately and quickly people perceive and process information presented to them in prose versus tables versus graphs.[26]

---

22. We focus here on relatively recent developments. For discussions of the evolution of data graphs, see generally, EDWARD R. TUFTE, THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION (2d ed. 2001); Howard Wainer & Paul F. Velleman, *Statistical Graphics: Mapping the Pathways of Science*, 52 ANN. REV. PSYCHOL. 305 (2001).

23. We are not the first to take note of recent developments in the visualization of data and results. *E.g.*, WILLIAM G. JACOBY, STATISTICAL GRAPHICS FOR UNIVARIATE AND BIVARIATE DATA 1 (1997) (pointing to "advances in graphical methodologies," new research on graphical perception, and "the rapid evolution and widespread availability of powerful computing equipment"); Wainer & Velleman, *supra* note 22, at 314 (discussing the importance of "cheap powerful computing" and advances in software).

24. "Graphing data as a means of communicating information has become increasingly prevalent. [One study reports] an estimated increase from 900 billion statistical graphs published in 1983 to 2.2 trillion published in 1994." Martin H. Fischer, *Do Irrelevant Depth Cues Affect the Comprehension of Bar Graphs?*, 14 APPLIED COGNITIVE PSYCHOL. 151, 151 (2000). Growth has come not only in the natural and social sciences, but also in popular magazines and corporate reports. *See, e.g.*, Vivien A. Beattie & Michael John Jones, *Changing Graph Use in Corporate Annual Reports,* 17 CONTEMP. ACCT. RES. 213 (2000); Vivien Beattie & Michael John Jones, *The Use and Abuse of Graphs in Annual Reports: Theoretical Framework and Empirical Study*, 22 ACCT. & BUS. RES. 291 (1992).

25. Fischer uses the term "armchair" to describe principles for graphic design that are based on intuition rather than empirical evidence. Fischer, *supra* note 24, at 152. By way of example, he points to Tufte's "data-ink ratio" principle. *Id.*; *see generally* TUFTE, *supra* note 22, at 91-105. *See also* Jerry Lohse, *A Cognitive Model of the Perception and Understanding of Graphs*, 8 HUMAN-COMPUTER INTERACTION 353 (1993) ("[I]t is essential that graphic design [has] a sound scientific foundation"). While we think Tufte has made important contributions to the visualization of quantitative information, we generally agree with Fischer. *See infra* Part IV.

26. There are many articles and books devoted to this topic. A few prominent (or interesting) examples include: WILLIAM S. CLEVELAND, THE ELEMENTS OF GRAPHING DATA (2d ed. 1994) [hereinafter CLEVELAND, ELEMENTS]; WILLIAM S. CLEVELAND, VISUALIZING DATA (1993); STEPHEN M. KOSSLYN, ELEMENTS OF GRAPHIC DESIGN (1993); William S. Cleveland & Robert McGill, *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods,* 79 J. AM. STAT. ASS'N 531 (1984); Fischer, *supra* note 24, at 161 (demonstrating that "irrelevant depth cues" slow down graph comprehension); Stephan

Because we integrate many lessons of this literature throughout, one example serves to make the point here: the poor performance of certain "pop charts"—but most notoriously pie charts.[27] No less than William S. Cleveland, a towering figure in the field of data visualization, has demonstrated that pie charts so often mask important patterns and other properties of the data that researchers should outright reject them.[28] In their place (and based on his research on pattern perception), Cleveland suggests dot plots—a category of charts underdeployed in the law reviews but used prominently in many other disciplines.[29]

We wholeheartedly endorse Cleveland's advice, and Figure 3 shows why. There we provide two visual depictions of the same data—juxtaposed pie charts (with six slices or sectors in the form of percentages) and a dot plot (with each dot representing a percentage).[30]

Lewandowsky & Ian Spence, *Discriminating Strata in Scatterplots*, 84 J. AM. STAT. ASS'N 682 (1989); Ian Spence & Stephan Lewandowsky, *Displaying Proportions and Percentages*, 5 APPLIED COGNITIVE PSYCHOL. 61 (1991) (presenting experimental results which indicate that charts are superior to tables).

27. CLEVELAND, ELEMENTS, *supra* note 26, at 262-63, uses the term "pop chart" to refer to three graphs (pie, divided bar, and area charts) that often appear in media and business publications. While Cleveland claims that these are rare in scientific journals, we found an ample presence in the law reviews.

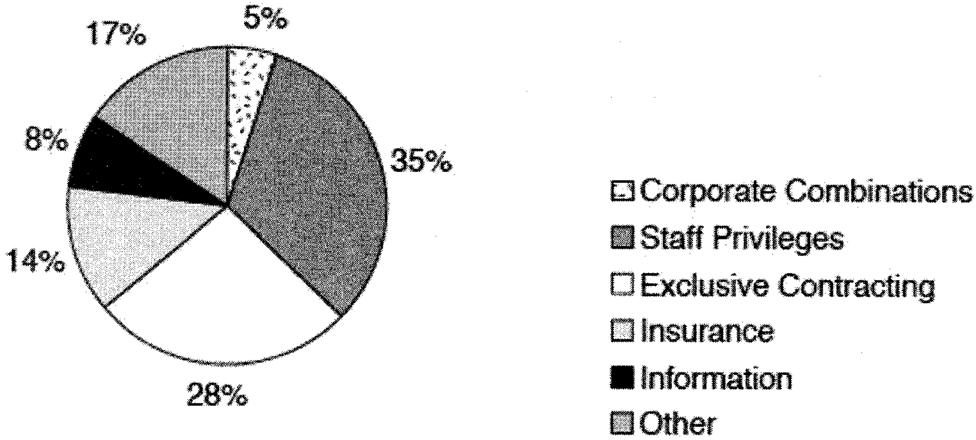
28. CLEVELAND, ELEMENTS, *supra* note 26, at 263; Cleveland & McGill, *supra* note 26, at 545. Many others agree. *See, e.g.*, JACQUES BERLIN, GRAPHICS AND GRAPHIC INFORMATION 111 (1981) (describing multiple pie charts, in particular, as "completely useless"); TUFTE, *supra* note 22, at 178 ("[P]ie charts should never be used"). *But see* Spence & Lewandowsky, *supra* note 26, at 61 (suggesting that, unless the judgment is a "complicated one," pie charts are not as inferior to bar charts as most of the literature suggests). However, some of the Spence & Lewandowsky experiments compared pie charts to stacked bar charts, an equally problematic display.

29. Cleveland & McGill, *supra* note 26, at 545 ("A pie chart can always be replaced by a bar chart . . . [but] we prefer dot charts . . . .").

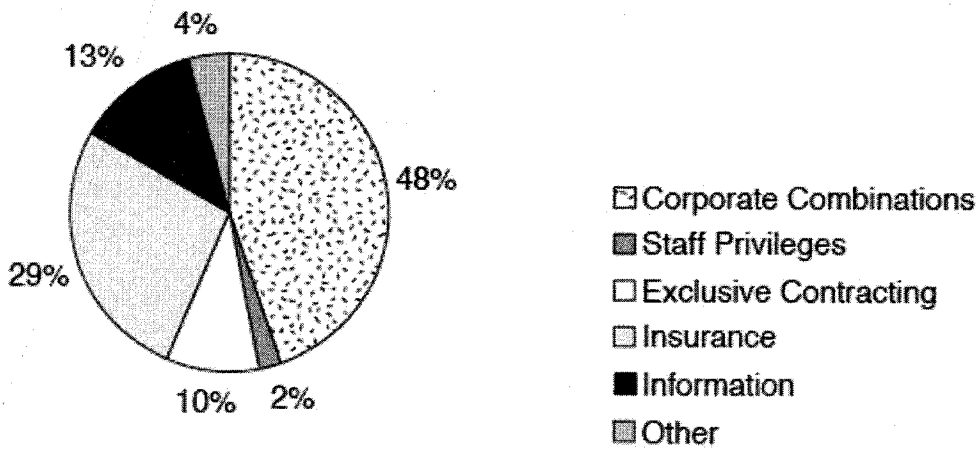
30. Peter J. Hammer & William M. Sage, *Antitrust, Health Care Quality, and the Courts*, 102 COLUM. L. REV. 545, 569 (2002) (depicting the pie charts).

Figure 3

Private Enforcement



Public Enforcement



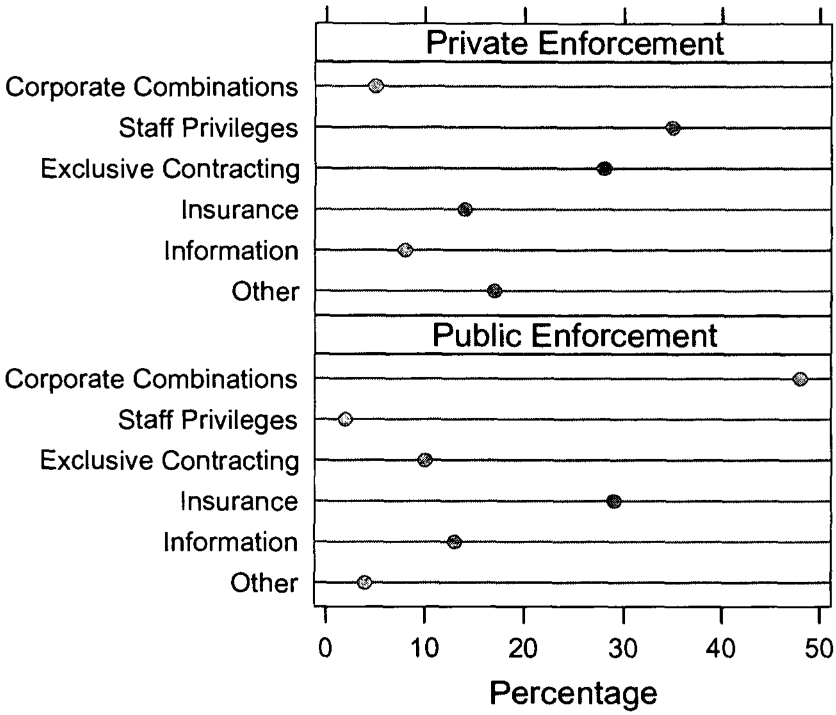


Figure 3: The stacked pie charts on the top depict the types of business conduct at issue in antitrust litigation in health care. We present the same data on the bottom with stacked dot plots. The dot plots are superior because they facilitate look-up of the data patterns; they also, by providing a common baseline, ease comparisons within and between the juxtaposed panels.[31]

The pie charts are designed to depict the types of business conduct at issue in antitrust litigation in health care, but two general problems are immediately apparent. First, decoding data patterns (the so-called "lookup") becomes an unnecessarily demanding cognitive task (with systematic biases) because it requires visually estimating an angle (rather than estimating a length).[32] Second, it is hard, visually, to detect differences in the data values because the pieces of the pie do not share a common baseline.[33] Looking at the top pie panel in Figure 3, for example, can you discern whether "Insurance" or "Other" are the bigger pieces? How about the difference between "Exclusive Contracting" and "Information" in the bottom pie? Now try comparing pieces of the pie across the two panels. Owing to the lack of a common baseline, this too is no easy task. The dot plots, in contrast,

31. *Id.*

32. *See* CLEVELAND, ELEMENTS, *supra* note 26, at 245, 264 (providing basis from which this discussion is adapted).

33. *See* Cleveland & McGill, *supra* note 26, at 545 (providing basis from which this discussion is adapted).

have the advantage of common baseline, which facilitates comparisons among the "pieces."[34]

Until recently, lessons of this sort pertained mostly to the communication of quantitative information, rather than statistical results. In other words, Cleveland, among other important figures in the field,[35] focused almost exclusively on developing principles for the display of raw data or summaries of those data (as in all the panels of Figures 2 and 3). But now—and this brings us to a second prominent development—scholars have moved beyond articulating theories for displaying *data* and towards developing techniques for presenting the *results* of statistical analyses.[36] As even our brief demonstration in the Introduction suggests (see Figure 1), these emerging strategies have the potential to transform the field of empirical legal scholarship, and we make use of them throughout the article.

In so doing, we employ sophisticated software—yet a third development motivating our work here. Just a decade ago, scholars hoping to present their results or even raw data faced a problem of no small proportion: a dearth of software able to meet their needs. The ever-popular Microsoft Excel was no more up to the task then than it is now, but neither were existing statistical packages (such as SAS and SPSS). No longer. These days, scholars can deploy Stata (along with modules that they can import into it[37]) or R, an open source implementation of the S language (also available as the commercial S-Plus package), to create sophisticated, yet accessible, data presentations.[38] What both Stata and R have in common is the ability

34. These are but two problems with pie charts. Other issues with those in Figure 3 are, first, the internal data labels are unnecessary, and the shading patterns are difficult to discern (especially between "Insurance" and "Other"). Moreover, the raw percentages in each pie chart do not sum to 100%, which may raise questions about the integrity of the data, and the manner in which the authors constructed the graphs.

35. *See, e.g.*, JOHN W. TUKEY, EXPLORATORY DATA ANALYSIS 99-101 (1977) (exploring alternative forms of displaying summaries of data).

36. *See, e.g.*, Gelman et al., *supra* note 16, at 121 (examining the fact that while statisticians recommend graphs, they often do not use them in their own work, and recommending ways to improve presentations through the use of graphs); King et al., *supra* note 9, at 347 (describing techniques to "extract the currently overlooked information and present it in a reader-friendly manner").

37. *See* King et al., *supra* note 9, app. at 360 (describing Clarify, a program which is not a graphing package but rather a freely available software program that converts statistical results into estimates of quantities of interest, along with estimates of the uncertainty surrounding those estimates, and the ability to depict these numbers, including confidence intervals, visually); *see also* http://gking.harvard.edu/clarify/docs/clarify.html (providing links to instructions for installing Clarify, along with the documentation necessary to use it).

38. There are a number of software packages dedicated to statistical computing. The most popular commercial packages are Stata, http://www.stata.com, SPSS, http://www.spss.com, S-PLUS, http://www.insightful.com, and SAS, http://www.sas.com. R, http://www.r-project.org, is

to customize all the fine features of a particular figure, and to create (and, more importantly, re-create) figures rapidly using scripts. (By the same token, the advent of fast computers on the desktop is another technological development that makes possible some of the strategies we describe below. Indeed, many of these methods require computationally intensive simulations that would have been prohibitively expensive in the 1990s.)

III. GENERAL PRINCIPLES FOR COMMUNICATING DATA AND RESULTS

If legal scholars have the tools for communicating more effectively and are motivated to do so, what they lack is guidance on how to do so. We hope to change this by offering sets of strategies for presenting data and results. We begin, though, with three general principles that all data work should strive to follow: (1) Communicate *substance*, not statistics; (2) When communicating substantive results (that is, when performing inference), also communicate *uncertainty*; (3) Regardless of whether the communication is of results or data, generally use *graphical*, not tabular, displays.

A. Communicate Substance, not Statistics

In perusing the data work in law-related articles, we were struck by the authors' emphasis on statistics over substance. More often than not they claim that their result is "statistically significant at the .05 level," without ever communicating very much, if anything, about the effect of that result.[39]

Staudt's study of taxpayer standing provides an example,[40] as does an article by Epstein and her colleagues on dissent in the U.S. Supreme Court.[41] The overall goal of the Epstein, et al. project was to assess the long-held belief that dissent rates remained low during the 19th century because the Court followed a "norm of consensus."[42]

---

an open source implementation of the S language, and has become the *linguq franca* of statistical computing. When managing data, creating graphs, or performing analyses, it is important to use a dedicated statistical package.

39. A search in LEXIS's US & Canadian Law Review file for the last two years (1/1/2004-12/31/2005) on the term "statistically significant" brings up 853 documents. And while we cannot say that many or even most of these articles fail to probe the substantive importance of the "statistically significant" finding (or, for that matter, even present original results), we can say that the majority of the 245 articles we examined neglected to do so.

40. Staudt, *supra* note 1, at 657-58.

41. Lee Epstein, Jeffrey A. Segal & Harold J. Spaeth, *The Norm of Consensus on the U.S. Supreme Court*, 45 AM. J. POL. SCI. 362, 362 (2001), *available at* http://epstein.law.northwestern.edu/research/norm.html (follow "click here for the article" link) (rounded to hundredths).

42. *Id.*

That is, the justices may have privately disagreed over the outcomes of cases but masked their disagreement from the public by producing consensual opinions.

To evaluate this claim, Epstein, et al. turned to Chief Justice Morrison L. Waite's docket books, in which the chief recorded votes cast during the Court's private conferences (1874-1888 terms). With the private and public votes in hand, the Epstein team estimated several multivariate logistic regression models, one of which we reproduce in Table 2.[43]

Table 2

Variable	Coefficient	(Std. Err.)
Justice was in Conference Minority	7.588**	(0.591)
Majority at Conference Affirmed Lower Court's Decision	-0.812**	(0.251)
Number of Years Justice Had Served on the Court	0.059**	(0.020)
Ideological Distance between the Justice and the Conference Majority	0.018	(0.039)
Conference Minority * Affirmed	0.903**	(0.252)
Conference Minority * Years	-0.058	(0.030)
Conference Minority * Ideological Distance	-0.014	(0.041)
Intercept	-6.301**	(0.459)
N	22001	
Log-likelihood	-1442.303	
$X^2_{(7)}$	53981.23	

Table 2: Results of the Epstein, et al. analysis of vote shifts on the Waite Court, 1874-1888 terms. This is a logistic regression with robust standard errors (clustered on justice) in parentheses. ** indicates p $\leq$.01. The dependent variable is whether the justice switched his vote (coded 1) or not (coded 0) between the time of the Court's (private) initial conference and publication of the decision.44

In this model, the dependent variable is whether a justice shifted his vote between the time of the Court's (private) initial conference and publication of the decision (regardless of whether the shift was in the direction of the minority or the majority). The independent variable of primary interest is "Conference Minority": If, according to Epstein et al., a justice was in the minority and switched to the majority (controlling for all other possible explanations of vote switching), that would provide evidence of a norm of consensus. Note that the authors also take into account other possibilities for shifts; for example, whether the majority initially affirmed the lower court's decision. The expectation here is that justices are less likely to exhibit

43. *Id.* at 374-75.

44. *Id.*

fluidity in "easy" cases, which the researchers take to be those the Court affirmed at its private conference.[45]

A mere glance at the results in Table 2 seems to confirm Epstein and her colleagues' conclusion that these hypotheses bear out. The two variables, "Conference Minority" and "Affirm," produce, as the authors note, "significant coefficients."[46]

Hence, Epstein, et al. are not wrong (note the ** indicating p > .01) but neither should we be impressed with the emphasis on statistical significance inherent in tabular displays like theirs (see also Table 1). First, no doubt some of the "statistically significant" estimates are of course substantively uninteresting. Take the "Affirmed" variable in Table 2. Given the authors' coding, it is in fact the case that justices were less likely to switch their votes if the majority initially affirmed the lower court's decision. But the effect of "Affirmed" is truly trivial. All other things being equal, the predicted probability of a first-year justice changing his vote is .0020;[47] that figure drops to .0008 when the Court voted to affirm.[48] Certainly a difference, but one we cannot imagine being of interest to any reader or, for that matter, the researchers themselves.

On the other hand, by focusing readers' attention solely on matters of statistical significance, analysts can *understate* the importance of their results. This holds for the Epstein et al. project on dissent[49] and Staudt's article on standing,[50] along with many of the others we examined. Consider Mark J. Roe's important paper in the STANFORD LAW REVIEW, which developed a simple but elegant theory: "[S]ocial democracies widen the natural gap between managers and distant stockholders, and impede firms from developing the tools that

45. *Id.* at 371 ("[G]iven the time period under analysis here—an era during which the justices were flooded with cases that *the lower courts had decided 'correctly'* but that they were forced to hear—[it is not a stretch to measure] easy cases [as] those the majority affirmed at conference . . . ." (emphasis in original) (citation omitted)).

46. *Id.* at 373.

47. The 95% confidence interval is [.0008, .0043].

48. The 95% confidence interval is [.00043, .00154].

49. For example, while it is difficult to glean from Table 2, the substantive effect of Conference Minority is substantial. All other things being equal, the predicted probability of a justice switching his vote when he is in the minority at conference is a substantial .79 [.707, .859].

50. *See supra* Part 1 (finding that, controlling for all other variables, the estimated probability of a liberal plaintiff being granted standing is .41; that figure for a conservative plaintiff was only .18).

would close that gap."[51] When the gap widens sufficiently, it renders "the large American-style public firm . . . unstable."[52]

One observable implication of this account, according to Roe, is that the more to the left the politics of a nation, the less diffuse ownership. To assess it, he estimated a linear regression model with the proportion of firms under diffuse ownership in the sixteen richest nations as the dependent variable, and experts' assessment of the nations' politics (from most left to most right) as the sole independent variable.

Roe communicates his results in two forms: Table 3, which we have replicated from his data, and the following statement: "these results are statistically significant."[53] Yes, that is true, but all readers should want to know "by how much?" That is, what is the practical effect of politics on diffuse ownership? It turns out that had Roe provided an answer to this question—the key question—his conclusion would have been far more dramatic because the effect seems substantial indeed: As we move from the most liberal societies to those in the middle of the political spectrum, diffusion of ownership nearly triples, from 16%[54] to 47%.[55]

Table 3

Variable	Coefficient	(Std. Err.)
Nation's Politics	0.329**	(0.090)
Intercept	-0.571	(0.289)
N	16	
R^2	0.489	
SEE	0.212	

Table 3: Mark J. Roe's regression of ownership on politics. ** indicates p ≤ .01. The dependent variable is the proportion of firms under diffuse ownership in the sixteen richest nations; the independent variable is experts' assessment of the nations' politics (from most left to most rigbt).[56]

Yet a second set of problems emerging when authors focus on statistical significance to the neglect of substantive effect is that they disservice their audiences in the short and long terms. In the short term, and this returns us to a point we made in the Introduction, no

51. Mark J. Roe, *Political Preconditions to Separating Ownership from Corporate Control*, 53 STAN. L. REV. 539, 561 (2000).

52. *Id.* at 543.

53. *Id.* at 562.

54. The 95% confidence interval is [-3.5, 36.0].

55. The 95% confidence interval is [36.5, 57.2].

56. *See* Roe, *supra* note 51, at 562 (providing the data from which we produced this table).

reader can possibly make (substantive) sense of the estimates in Table 2 (the Epstein, et al. table). All we can really say about the coefficient of 7.59 (on the "Conference Minority" variable), to provide but one example, is this: "Setting all other variables at 0, the log odds ratio of switching by a justice in the conference minority increases (-6.30 [the constant] + 7.59 =) 1.29"—a rather meaningless statement to readers regardless of their statistical prowess. The particular ohstacle here is that the model is nonlinear, meaning that one needs to know a good deal about the explanatory variables to decipher the magnitude of the coefficients. But even in the case of Roe's regression analysis, in which the coefficients are easier to interpret because the model is linear, why compel others to make the calculations (especially since many will not know how)? More generally, in both instances the authors are missing a prime opportunity to communicate with their readers. They are failing to present results in a way that "require[s] little specialized knowledge to understand"[57] and thus to reach the widest possible audience; they are instead limiting themselves to a very small sector of the community.

In the longer term, the emphasis on statistical significance, unaccompanied by any substantive message, can work to perpetuate questionable interpretations. Under present practice in the law reviews, it is entirely conceivable—in the absence of any statements or displays about key quantities of interest—that ensuing articles will report that "Affirmed" has a "statistically significant" effect on vote shifts despite its trivial substantive impact. Actually, this is more than a possibility. In perusing the law reviews, we are struck by the extent to which subsequent authors convey a finding as "statistically significant" (and sometimes third hand at that) without ever denoting the substantive punch of the result.[58] Of course, we can hardly lay the blame with the consumers of data work; it is the original researchers'[59] responsibility to provide estimates of the key quantities

57. King et al., *supra* note 9, at 347.

58. *E.g.*, Note, *The Case for Federal Threats in Corporate Governance*, 118 HARV. L. REV. 2726 (2005) (citations omitted). Judge Winter contends that a regime facilitating takeovers would enhance shareholder value. To reconcile this belief with his view that state competition benefits shareholders, Judge Winter argues that Delaware's antitakeover statute is relatively innocuous. Professor Romano holds a similar position. Reviewing empirical studies on the effects of takeover laws on shareholder wealth, she notes that "[e]vent studies find either statistically significant negative stock price effect or no effect." *Id.* at 2735 n.43.

59. *E.g.*, Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1489, 1491-92 (2005). Kang describes a study by political scientists Frank Gilliam and Shanto Iyengar which communicated substance and not merely statistical significance:

> [The researchers] created variations of a local newscast: a control version with no crime story, a crime story with no mugshot, a crime story with a Black-suspect mugshot, and a crime story with a White-suspect mugshot.

of interest.[60] More importantly, they should want to communicate substantive effects, for when they do, benefits accrue: it is those effects, perhaps in addition to (or even in lieu of) claims about statistical significance, that tend to get transported from study to study—thus generating more precise knowledge about the work's key findings.

In short, to begin reaping the benefits of empirical research, researchers ought to move away from an emphasis on statistics and towards substance (and readers ought demand that they do). At least conceptually, this is not an especially difficult move. All it requires is that analysts identify the quantity (or more likely *quantities*) of the greatest interest or relevance to their project. Returning to the example of Epstein and her colleagues, they were most concerned with the extent to which the "norm of consensus" led justices in the conference minority to switch their vote to the majority—meaning that the "Conference Minority" variable was of the greatest interest to them, though they had secondary interests in the others.

Once researchers have identified the variable(s) of interest it is just a matter of computation to estimate a quantity of interest (along with an assessment of uncertainty about that estimate). In contemplating the 7.59 coefficient on the "Conference Minority" variable, for example, why follow the crowd and write, "it is statistically significant at the .01 level" when it is nearly as easy to answer "how much effect does it exert?"[61] Not only do various statistical software packages enable researchers to estimate the quantity of interest, but also because contemporary software packages use simulations (repeated sampling of the model parameters from

> The Black and White suspects were represented by the same morphed photograph, with the only difference being skin hue—thus controlling for facial expression and features. The suspect appeared for only five seconds in a ten-minute newscast; nonetheless, the suspect's race produced statistically significant differences in a criminal law survey completed after the viewing. Having seen the Black suspect, White participants showed *6%* more support for punitive remedies than did the control group, which saw no crime story. When participants were instead exposed to the White suspect, their support for punitive remedies increased by only *1%*, which was not statistically significant.

Id. (emphasis added).

60. *But see* MOD. LANG. J., Frequently Asked Questions, http://polyglot.lss.wisc.edu/mlj/suhmission.htm#faq6 ("Manuscripts often benefit from reporting effect sizes of both the new study and studies cited in the literature review. To do the latter, it may be necessary to calculate the effect size measures from data summaries or test statistics given in those cited studies. In some cases, of course, such calculation will not be possible. Ideally, the author will relate the effect size of the new study to those of previous studies.").

61. *See supra* note 49 and accompanying text (noting the dramatic effect of the Conference Minority).

their sampling distribution) to produce the estimate, researchers are also capable of generating assessments of error (e.g., confidence intervals.).[62]

In the next Section we have more to say about the importance of conveying error (uncertainty). For now it is worth noting that while contemporary software is geared toward estimating quantities of interest (and error) for the results of rather fancy models (as in the cases of Staudt, Epstein, et al. and Roe), the same general principle of "Communicate Substance, Not Statistics" applies with equal force to simpler tools. Take the omnipresent Pearson chi-square (χ^2) statistic. Just as too many authors merely note whether a regression coefficient is "statistically significant" without providing a substantive interpretation, those reporting χ^2 do likewise.

Moore's work on judges and juries in patent cases provides a classic example.[63] The author is interested in determining whether many in the legal community are right to be skeptical about the ability of jurors to reach informed decisions in patent litigation. As she puts it, "[T]here is a popular perception that the increasing complexity of technology being patented . . . has made patent trials extremely difficult for lay juries to understand."[64] To assess this perception, she collected data on all patent cases that went to trial, whether before a judge or jury, between 1983-1999 (N = 1209).[65] The top panel of Table 4, which we reproduce from the article, displays Moore's overall results.

62. King, *supra* note 9, at 352 (describing the Monte Carlo algorithm used for these simulations; it is implemented in the Clarify plug-in for Stata). To use these methods, investigators need make no additional assumptions; that is, beyond those they make to perform statistical inference. *Id.*

63. Kimberly A. Moore, *Judges, Juries, and Patent Cases*, 99 MICH. L. REV. 365 (2000).

64. *Id.* at 365.

65. *Id.* at 367.

Table 4

	Jury Decisions	Judge Decisions	Total
Patentee Prevails	363 (68.1%)	343 (50.7%)	706
Infringee Prevails	170 (31.9%)	333 (49.3%)	503
Total	553	676	1,209

	Jury Decisions	Judge Decisions	Total
Patentee Prevails	363 (311)	343 (395)	706
Infringee Prevails	170 (222)	333 (281)	503
Total	553	676	1,209

$X^2_{(1)} = 36.28$, $P \leq .01$

Table 4: Results of Kimberly A. Moore's examination of the outcome in patent trial decisions, by party and adjudicator. The top panel shows the observed frequencies (with column percentages in parentheses); the bottom panel shows the observed and expected frequencies.[66]

In attempting to convey her findings, Moore writes:

> There is a significant difference in win rate when tbe jury decides patent claims. Hence, the null hypothesis that "when a jury decides a patent claim there is an equal chance of success for the patent holder and the infringer" can be rejected. There is not a significant difference in the win rate, however, when the patent case is decided by a judge. The null hypothesis that "when a judge decided a patent claim there is an equal chance of success for the patent bolder and the infringer" cannot be rejected. The identity of the adjudicator, therefore, is a statistically significant predictor of wbo wins the claims in the lawsuit.[67]

To us this statement is as uninformative as "the coefficient on 'Conference Minority' is statistically significant at the .01 level." Far better and certainly of more use to Moore and her audience is the information conveyed in the bottom panel of Table 4. There we display the observed and expected counts—that is, the counts we would expect in each cell if no difference existed between judges and jurors.[68] Now the contrast in "win rates" just pops: In Moore's sample, patentees won fifty-two more cases before juries than we would have anticipated if no association existed between the decision maker and outcome—or about a 17% increase in success.

66. *Id.* at 386 (providing the data from which these tables are produced).

67. *Id.* at 385.

68. Of course, the first step in computing the Pearson R2 is to calculate the expected count within each cell.

B. When Performing Inference, Convey Uncertainty

In a poll released on December 21, 2005, ABC News/Washington Post reported that 54% of Americans (95% CI [±3]) thought the Senate should confirm Samuel Alito for a seat on the Supreme Court.[69] Of course, the pollsters did not survey every American; they rather made a statistical inference about all Americans (the "population") by drawing a "sample" (N = 1003). In other words, via an examination of a small piece of the world (1003 Americans) they attempted to learn about the entire world (all Americans), and, assuming that they drew a random probability sample, we can have some faith in the information they convey.

But there is more. When researchers draw random probability samples, they also can communicate their degree of uncertainty about the sample statistic (e.g., the 54% favoring Alito). Surveys reported in the press, for example, typically convey this degree of uncertainty as "the margin of error," which is usually a 95% confidence interval (or, as we have written above, 95% CI). So when the press reports that 54% of respondents support Alito's confirmation with a ±3 "margin of error," they are supplying the level of uncertainty they have about the sample statistic of 54%: here, that the true fraction of American supporting Alito will be captured in the stated confidence interval (of 54 [±3] or 51-57%) in ninety-five of one hundred applications of the same sampling procedure. Note that this information does not tell us whether the population (parameter) lies within this range. What is critical, though, is that if we continue to draw samples from the population of Americans, the mean of the samples will eventually equal the mean of the population, and if we graphed the mean of each sample, we would see a shape resembling a normal distribution. This is what enables us to make an inference—here, in the form of a sample statistic and a margin of error—about how all Americans (the population) feel about Alito by observing a single sample statistic (54%).

When we read the results of surveys reported in the press—such as the Alito survey—we have come to expect pollsters to convey the level of uncertainty about their sample statistics. (If they did not, we would be unable to judge their results. For example, consider a

69. Jon Cohen, *Majority Wants Alito on Court; Most Also Want Roe to Remain*, ABC NEWS, Dec. 18, 2005, http://abcnews.go.com/images/Politics/1001a2AlitoandAbortion.pdf (last visited on Nov 5, 2006) (explaining the results to the question: "As you may know, Bush has nominated federal judge Samuel A. Alito to serve on the U.S. Supreme Court. Do you think the U.S. Senate should or should not confirm Alito's nomination to the Supreme Court?" in which 28 percent said no and 19 percent had no opinion).

survey reporting that 54% of Americans supported Alito, ±20(!)—meaning it is entirely possible that a majority did not approve of his candidacy.) We should expect the same of scholarship. That is, when scholars perform statistical inference—as did Staudt, Epstein, et al., Roe, and Moore in the examples above—they too have an obligation to convey their level of uncertainty. After all, because statistics are only estimates developed from a sample of the world (and not from the entire world) we can never be 100% certain that we got it right. Most analysts take this obligation seriously but tend to rely on p values and standard errors to do the work. This is unfortunate since neither conveys information of much value. Take the omnipresent standard error, for example the ".090" on the estimate of .329 in Table 3; all that this error value supplies is an estimate of the standard deviation of the estimated slope—which, standing alone, is of interest to almost no one, readers and scientists alike.[70]

It is thus hardly surprising that reporting the far-more-meaningful 95% (or even 99%) confidence interval rather than (or in addition to) p values and standard errors has become de rigeur in other disciplines,[71] and we recommend the same for law. Hence, in the case of the Roe project, we suggest supplanting the standard error of .090 with the 95% confidence interval; here, those values are a lower bound of .136 and an upper bound of .522.

This interval, we believe, comes far closer than the standard error to conveying what Roe wants: that his best guess about the coefficient is .328 but he is "95 percent certain" that it is in the range of .136 to .522. Because zero is not in this range (the confidence

70. Its value, rather, lies in its role in computing 95% confidence intervals.

71. *See, e.g.*, POLITICAL ANALYSIS, *supra* note 20 (applying the confidence interval standard in politics). *See also, e.g.*, THE AMERICAN PSYCHOLOGICAL ASSOCIATION, PUBLICATION MANUAL 21 (American Psychological Association 2001) (suggesting that authors include estimates of the "effect size"); CLINICAL BIOMECHANICS, Author Gateway Guide for Authors, http://authors.elsevier.com/GuideForAuthors.html?PubID= 30397&dc=GFA (last visited Aug. 31, 2006) ("Confidence intervals are preferred over just P values."); DEV. PSYCH., Instructions to Authors, http://www.apa.org/journals/dev/submission.html (last visited Aug. 31, 2006) (stating that "for all study results, measures of both practical and statistical significance should be reported; the latter can involve either a standard error or an appropriate confidence interval," but emphasizing "practical significance" or "effects" which "can be reported using an effect size, a standardized regression coefficient, a factor loading, or an odds ratio"); J. OF VISUAL COMM. IN MED., Instructions for Authors, http://www.tandf.co.uk/journals/authors/cjauauth.asp (last visited Aug. 31, 2006) ("Statistical advice is available in STATISTICAL GUIDELINES FOR CONTRIBUTORS TO MEDICAL JOURNALS. Confidence intervals are preferred over just P values."); MOD. LANG. J., *supra* note 60 ("Ideally, the author will relate the effect size of the new study to those of previous studies. Confidence intervals are generally appropriate for this purpose."); THE NEW ENG. J. MED., Author Center, http://authors.nejm.org/Misc/NewMs.asp (last visited Aug. 31, 2006) ("Measures of uncertainty, such as confidence intervals, should be used consistently, including in figures that present aggregated results.").

interval), he (and we) can safely reject the null hypothesis of no relationship between a society's politics and ownership (without ever providing a p value!). Likewise in the Epstein et al. study, rather than reporting the standard error of .591 on the Conference Minority variable, we suggest denoting the confidence interval around the coefficient of 7.59 [6.43, 8.75]. (Hereinafter we use this notation [x,x] to indicate the lower and upper bounds of the 95% confidence interval around the estimate of the quantity of interest.)

But researchers would not be making the most of their analyses. This is especially true in the case of Epstein and her colleagues who estimated their model using logistic regression. If they were to write that they are "95 percent certain" that the true logit coefficient lies between 6.43 and 8.75, they would fail to speak clearly and accessibly to their audience. What we commend instead is combining the lesson here of relating uncertainty with the general principle of conveying substantive information. In Epstein et al.'s case, this translates into the following claim: All other things being equal, the predicted probability of a justice switching his vote when he is in the minority at conference is a whopping .79 [.71, .86]. For Roe it is this: As we move from societies in the middle of the political spectrum to those on the far right, diffusion of ownership increases by nearly 60%, from 47% [36.5, 57.2] to 75% [56.7, 92.4]. Now in both instances, readers need no specialized knowledge to understand the results of the study or the researchers' assessment of their uncertainty about those results. As a result, they are in a far better position to evaluate the study's findings.

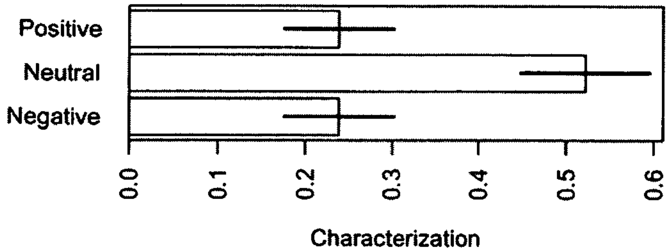
By offering this advice, we do not mean to suggest that all data projects must house assessments of uncertainty. Indeed, the principle is "When *Performing Inference*, Convey Uncertainty," not "When Describing Data, Convey Uncertainty." In other words, when researchers are merely displaying, describing, or summarizing their data—and not using the data they have collected to make inferences about the population that may have generated the data—supplying standard errors or confidence intervals is not necessary.[72] Because this suggestion has more bearing on the creation of graphical displays (which we do not want to unnecessarily clutter with, say, confidence intervals) we consider it in some detail in the next installment of this article. Suffice it to note for now that when Moore *described* a basic feature of her data—that patentees won 58% of the 1209 suits and alleged infringers won 42%—we see no purpose to presenting the associated uncertainty about those figures. If, on the other hand, Moore seeks to generalize from those figures of 58 and 42% to cases she has not collected (that is, to use her sample to make an inference), then some measure of error is in order.

C. Graph Data and Results

In a fascinating article in the *Harvard Law Review*, Richard L. Revesz attempts to demonstrate that different levels of preference for environmental protection in the states—and not public choice pathologies that cause systematic underrepresentation of environmental interests—help account for variation in states'

72. We take our cues here from Cleveland, though some scholars seem to disagree. CLEVELAND, ELEMENTS, *supra* note 26, at 213-15. *But see, e.g.*, Gelman et al., *supra* note 16, at 123 (transforming the following tahle, which originally appeared in an article by J.H. Ellenberg, into the graph directly below it:

Raters' characterization	Percent
Negative (1-1.9)	23.9
Neutral (2-2.9)	52.2
Positive (≥3)	23.9



If, from the table, the original author (Ellenberg) sought to make an inference about the ratings, then the display of uncertainty in the graph would be warranted. If, however, Ellenberg only sought to describe the data he collected, this step is probably overkill).

willingness to pass protective measures.[73] As part of that demonstration, Revesz supplies a table (reproduced here as Table 5) that summarizes ratings assigned by the League of Conservation Voters ("LCV") to congressional representatives in each state, as well as to the House delegation as a whole (the higher the score, the more pro-environment the representative).

To be sure, this table communicates interesting information; specifically, summary data on congressional voting (by state) over protective environmental measures. Were we to take a long look, we would see that the (lone) representative of Vermont always voted in support of environmental protection, thereby propelling his state into the number 1 ranking. The Oklahoma and Alabama delegations, in contrast, never voted in favor of protective legislation.

73. Richard L. Revesz, *Federalism and Environmental Regulation: A Public Choice Analysis*, 115 HARV. L. REV. 553 (2001).

Table 5

State	Median Democrat	Median Republican	Median Overall	Rank
Alabama	.00	.00	.00	48
Alaska	—	.06	.06	43
Arizona	.80	.07	.07	39
Arkansas	.53	.03	.17	31
California	.93	.07	.57	17
Colorado	1.00	.13	.13	33
Connecticut	.90	.83	.90	6
Delaware	—	.75	.75	10
Florida	.83	.13	.27	3
Georgia	.93	.12	.13	33
Hawaii	.93	—	.93	3
Idaho	—	.06	.06	43
Illinois	.93	.13	.67	15
Indiana	.73	.07	.17	31
Iowa	.47	.10	.20	27
Kansas	.80	.00	.07	39
Kentucky	.20	.07	.07	39
Louisiana	.33	.00	.00	48
Maine	.83	—	.83	8
Maryland	.83	.37	.73	11
Massachusetts	.97	—	.97	2
Michigan	.87	.13	.70	14
Minnesota	.80	.40	.73	11
Mississippi	.40	.03	.33	22
Missouri	.80	.10	.33	22
Montana	—	.06	.06	43
Nebraska	—	.13	.13	33
Nevada	.73	.07	.40	21
New Hampshire	—	.27	.27	24
New Jersey	1.00	.63	.93	3
New Mexico	.87	.03	.07	39
New York	.93	.37	.80	9
North Carolina	.80	.07	.20	27
North Dakota	.56	—	.56	18
Ohio	.87	.00	.20	27
Oklahoma	—	.00	.00	48
Oregon	.87	.07	.87	7
Pennsylvania	.80	.07	.47	19
Rhode Island	.93	—	.93	3
South Carolina	.77	.10	.27	24
South Dakota	—	.06	.06	43
Tennessee	.47	.13	.13	33
Texas	.60	.00	.10	38
Utah	—	.13	.13	33
Vermont	1.00	—	1.00	1
Virginia	.47	.07	.20	27
Washington	.87	.10	.60	16
West Virginia	.47	—	.47	19
Wisconsin	.93	.20	.73	11
Wyoming	—	.06	.06	43
Full House	.87	.07	.40	

Table 5: Results of Richard L. Revesz's analysis of the League of Conservation Voter records, by state delegations to the House of Representatives. Data are from 1999.

"—" means that no members of the state delegation affiliated with the political party in question. The higher the numbers (in the "Median" columns), the more "pro-environment" the state delegation.[74]

But is Table 5—or, rather, tables more generally—the best way to convey this information? If our purpose is to provide readers with the *exact* figures, then the answer is yes: tables always trump graphs. Figure 4, which is a graphical display of Revesz's data (specifically, dot plots of the state delegation medians), underscores this point. While we can, from the table, observe that the Pennsylvania delegation had a median score of exactly .47, we cannot make that observation with the same degree of precision from the figure.

Figure 4

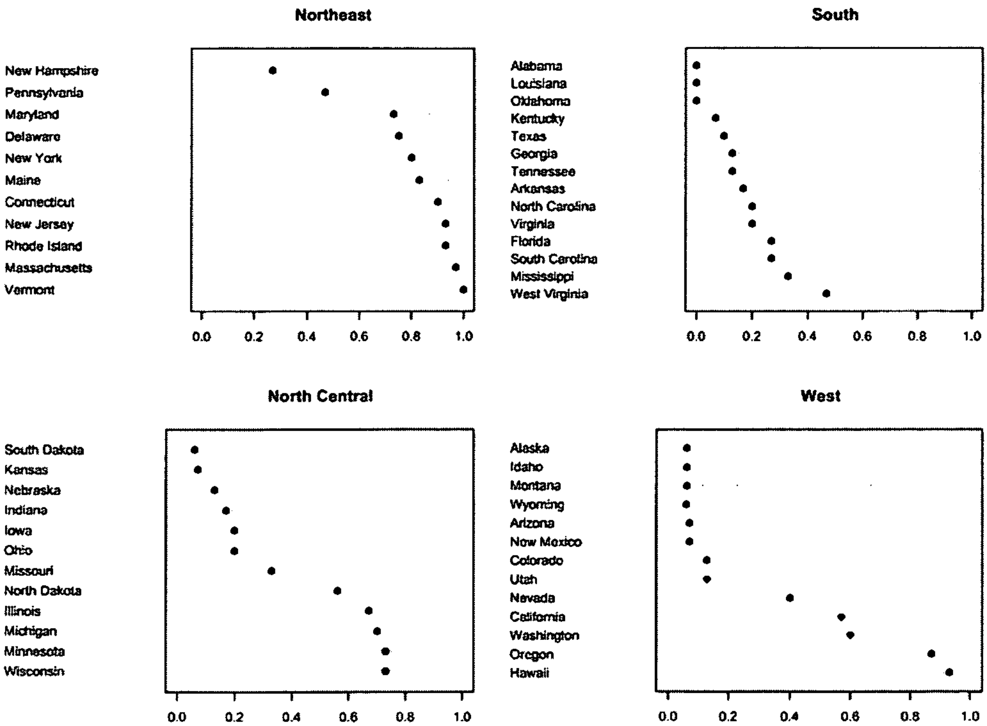


Figure 4: Dot plots, by region, of Revesz's analysis of the League of Conservation Voter records, by state delegations to the House of Representatives, by region (see Table 5). Data are from 1999. These figures depict the overall House median.[75]

74. *Id.* at 638.
75. *Id.*

More often that not, though, the degree of precision that only tables can convey is beside the point. Typically what we want to communicate to our audience (and to ourselves) is not exact values but comparisons, patterns, or trends. This certainly holds in the social sciences, and it is not easy to identify an exception in law. At the least, Revesz's article does not present an example of such an exception. What he wants us to take away from Table 5 is *not* the precise ranking of, say, Alaska's or North Dakota's delegation; he rather hopes to convey a sense of the relative voting records of the delegations (the highest support, the lowest, and so on) so that we can eventually draw comparisons between federal and state legislative support for environmental regulation. Surely, this purpose is better served by Figure 4 than Table 5.

Indeed, Figure 4 allows us to detect crucial patterns in the data that would be difficult, if not impossible, to discern from the tabular display. So, for example, we can observe that states in the Northeast are rather uniform supporters of the environment (with the exceptions of New Hampshire and Pennsylvania). In contrast, the South consists quite homogeneously of states that are not generally supportive of the environment. States in the North Central region cluster into essentially two groups, with North Dakota, Illinois, Michigan, Minnesota, and Wisconsin appearing to be solid environmental supporters, although not as strong as states in the Northeast. Distinct groups also emerge in the West, with one set on the left of the distribution and another to the right.[76]

More generally, if the point is to draw attention to trends or to make comparisons, as it almost always is, *graph* the data or results.[77] This advice reflects not only our own aesthetic preference but a growing consensus among scholars in the statistical and social sciences,[78] as well as journal editors.[79] Unless the author has a very

76. Another way to present geographic data is to shade states according to the intensity, here, of environmental support. Such plots provide information about geographic contiguity but, because they require readers to disentangle color gradients, they can obscure comparisons between and among states.

77. *See* Gary Klass, Constructing Good Charts and Graphs, http://lilt.ilstu.edu/gmklass/pos138/datadisplay/sections/goodcharts.htm (last visited Aug. 31, 2006).

78. *See, e.g.*, JACOBY, *supra* note 23, at 4-6 (reviewing the advantages of graphs); Gelman et. al., *supra* note 16, at 121 ("[W]e find well designed graphs to be superior to tables.") *See also supra* note 16 and accompanying text.

79. *See, e.g.*, Gelman et. al., *supra* note 16, at 121 ("[A] new editor of Memory and Cognition exhorted submitters to display data using well-designed graphs."). Also consider Suplee & Bradford's "Science and Engineering Visualization Challenge" encouraging researchers to rethink their data presentations for which, in asking for articles, they wrote "Data may be the gold standard of science, but they don't exactly glitter. A neat table of values cannot convey the significance, context, or excitement of research results to anyone besides other scientists in the

compelling reason to provide precise information to readers—a task better relegated to web sites than (precious) journal pages—"well designed graphs are superior to tables."[80]

This recommendation, we hasten to note, applies with equal force to the presentation of *results* (e.g., Figure 1) and *data* (or summaries of data) (e.g., Figure 4).[81] It also applies to small amounts of data. Despite Tufte's intuition that "tables outperform graphs when reporting on small data sets of *twenty numbers* or less" and that the "special power of graphics comes in the display of large data sets"[82] subsequent studies of perception have shown this questionable at best.[83]

Prove it to yourself by comparing the left and right panels of Figure 5, both of which depict the estimated ideology of the U.S. Courts of Appeals in 2000 (based on the ideology of the median judge in each). In the left plot we reproduce the precise figures; in the right, we graphically depict the percentages (both ordered from most liberal to most conservative). Surely if we stared at the numbers (eight short of Tufte's magic number of twenty) long enough, we could observe the patterns that emerge from the graph—e.g., the ideological closeness of the Second, Ninth, and Sixth Circuits on the left and the Fifth, Eleventh, and D.C. Circuits on the right, not to mention the gap between the Sixth and the Third Circuits or the large difference between the most liberal and most conservative courts. But it requires far more unnecessary cognitive work.

same subfield. No one else quite gets the picture—including the larger community that supports the global research enterprise". Curt Suplee & Monica Bradford, *Visualization and the Communication of Science*, 301 SCIENCE 1472, 1473 (2003).

80. Gelman et al., *supra* note 16, at 121.

81. CLEVELAND, ELEMENTS, *supra* note 26, at 215 (describing the only distinction the authors would draw as centering on the need to convey uncertainty). In keeping with our advice above, figures depicting results should also display confidence intervals. When the goal is simply to show the data, "then show the data," without conveying uncertainty. *Id.*

82. TUFTE, *supra* note 22, at 56 (emphasis added).

83. *See, e.g.*, Gelman at. al., *supra* note 16, at 121-22 (debating the common wisdom that tables are more effective for smaller sets of data); Douglas Gillian et al., *Guidelines for Presenting Quantitative Data in HFES Publications*, 40 HUM. FACTORS 28, 29 (1998) (positing that graphs may be superior to tables even for small sets of data); Spence & Lewandowsky, *supra* note 26, at 76 (finding that graphical displays had an advantage over tables, pie charts, and bar graphs).

Figure 5

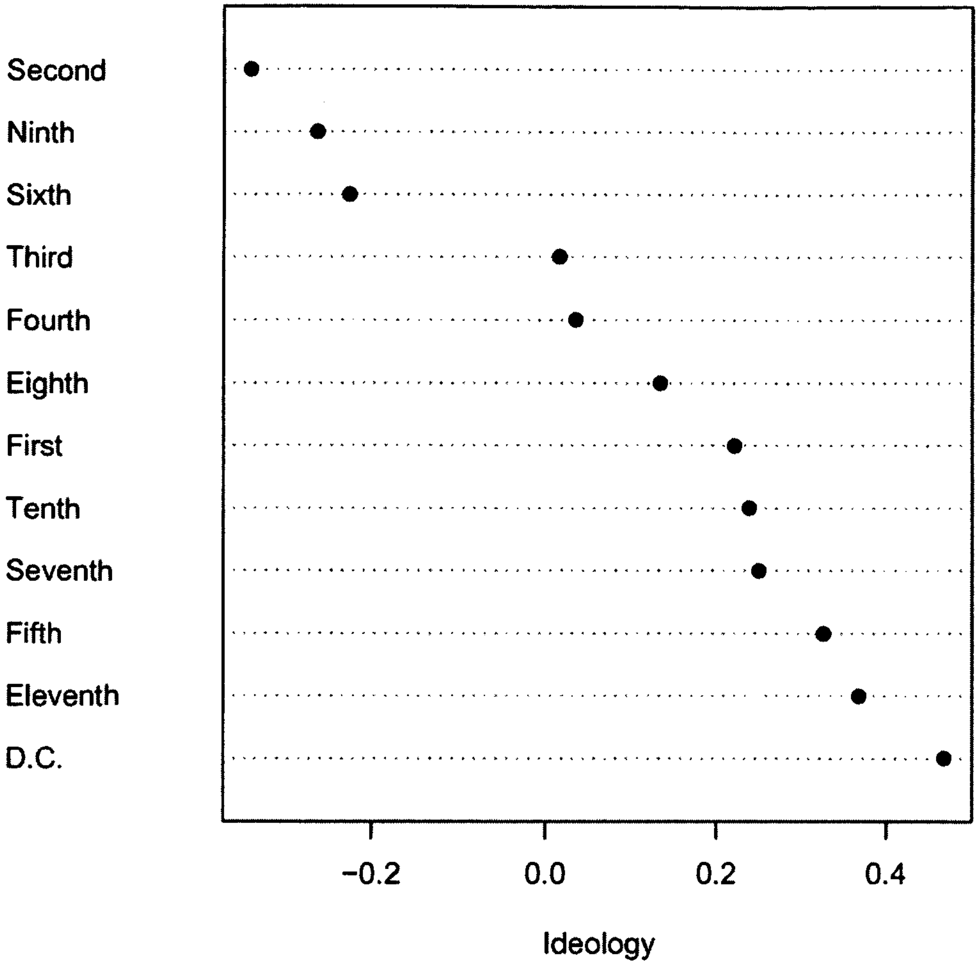


Figure 5: Estimated ideology of the U.S. Courts of Appeals, 2000. Both panels display the estimated ideology of the median member of the circuit court in 2000, where ideology is from most liberal (here, labeled as -0.2) to most conservative (labeled as 0.4). Though it is possible from the table to observe the patterns that jump out in the graph—e.g., the ideological closeness of the Second, Ninth, and Sixth Circuits on the left and the Fifth, Eleventh, and D.C. Circuits on the right—it requires far more (unnecessary) cognitive work.[84]

84. Lee Epstein et al., *The Judicial Common Space*, J. L. ECON. & ORG. (forthcoming in 2007.), *available at* http://epstein.law.northwestern.edu/research/JCS.html (identifying data regarding the appointment of judges); Michael W. Giles, Virginia A. Hettinger, & Todd Peppers, *Picking Federal Judges: A Note on Policy and Partisan Selection Agendas*, 54 POL. RES. Q. 623 (2001) (describing an analytical approach in which ideological estimates are based on the ideology of the median member of the circuit, where ideology is defined as follows: If a judge is appointed from a state where the President and at least one home-state Senator are of the same

IV. GENERAL PRINCIPLES FOR VISUALIZING DATA AND RESULTS

With these general principles now noted, we turn to the challenge of putting them into practice. In the next installment of this article, we outline specific strategies for presenting data and results. As we have suggested throughout, these are distinct tasks and so, to some degree, require distinct rules.

On the other hand, reflecting our view that graphs are generally superior to tables, we center both discussions on visualization via pictures. Hence we begin here with three basic rules that apply to *all* figures regardless of whether the task is to present data or results: (1) Aim for Clarity and Impact, (2) Iterate, *and* (3) Write Detailed Captions. These are critical to the enterprise of graphing data *and* results if only because, as Cleveland notes, "Visualization is surprisingly difficult. Even the most simple matters can easily go wrong."[85] Following these suggestions will, we hope, increase the odds of matters both simple and complex going right.

And *go right for both the audience and the researcher*. In other words, in the sections to follow, we do not differentiate between graphs for purposes of "prospecting" (that is, as part of the data analysis process) and "transferring" (that is, for communicating data and results)—even though some scholars do. Actually, more than a handful have argued that "different kinds of displays are needed" depending on whether the researcher is prospecting or transferring.[86]

We take the point: Just as the depiction of data and the depiction of results are distinct tasks, so too are the enterprises of analyzing a data set and presenting it. By way of example, consider "Anscombe's Quartet"—a famous demonstration set out by the statistician Francis Anscombe in 1973.[87] After listing four "fictitious" data sets—all of which consisted of eleven observations of one dependent and one independent variable—Anscombe presented the estimates, along with other summary statistics, of a linear regression analysis for each. Those estimates, as we depict in the caption of

party, the nominee is assigned the ideology (NOMINATE Common Space score) of the home-state Senator (or the average of the home-state Senators if both members of the delegation are from the President's party). If neither home-state Senator is of the President's party, the nominee receives the ideology (NOMINATE Common Space) score of the appointing President).

85. CLEVELAND, ELEMENTS, *supra* note 26, at 9.

86. Howard Wainer, *Graphical Visions from William Playfair to John Tukey*, 5 STAT. SCI. 340, 345 (1990); *see also* JACOBY, *supra* note 23, at 2 (differentiating between analytic graphs and presentation graphs, though noting that "carefully constructed analytic graphs also are quite effective for presentational purposes").

87. F.J. Anscombe, *Graphs in Statistical Analysis*, 27 AM. STATISTICIAN 17, 19-20 (1973).

Figure 6, are identical for the four data sets. But the underlying data in each are hardly identical, as the four plots in the figure make clear.

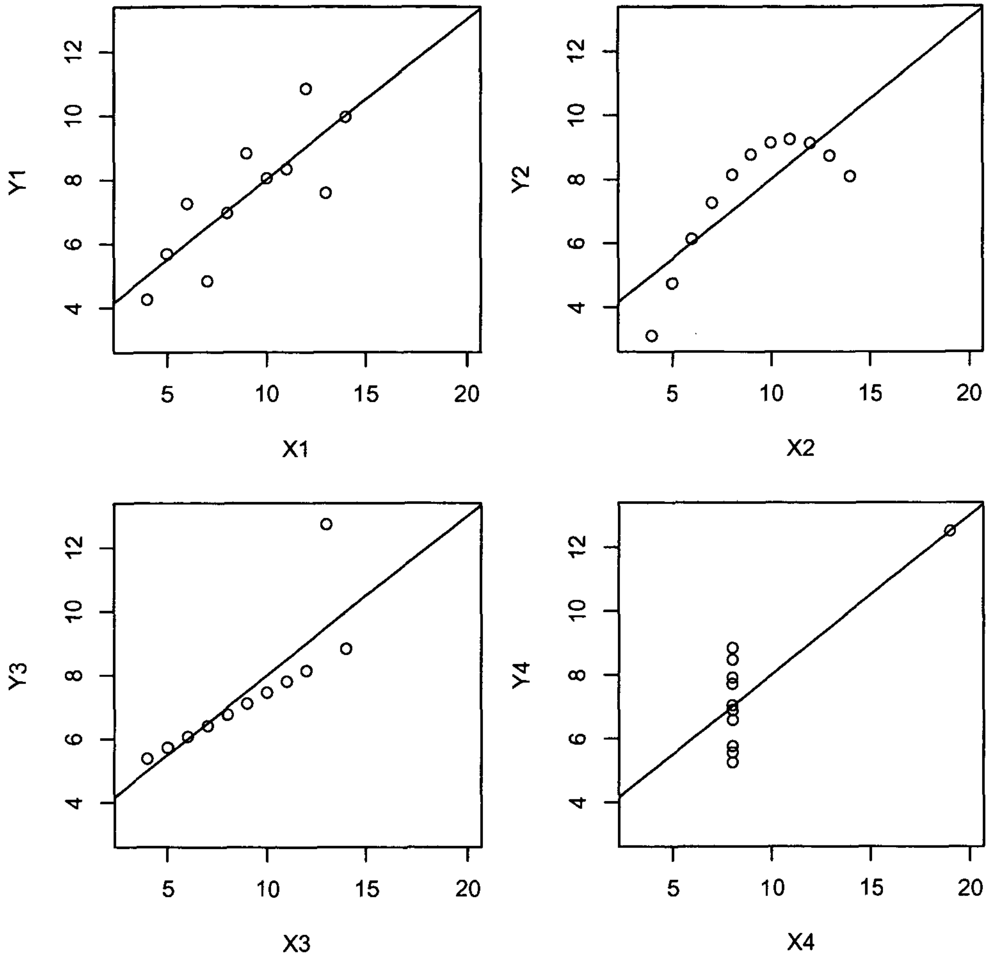


Figure 6

Figure 6: Taken collectively, the four panels form "Anscombe's Quartet"—a famous demonstration set out by the statistician Francis Anscombe in 1973. While the plots are quite distinct, the four data sets they represent each produce: (1) the same means on both the y and x variables (2) the same slope and intercept estimates in a regression of y on x, and (3) the same R2 and F values. Anscombe's point was to underscore the importance of graphing data before analyzing it. By creating the four plots, Anscombe was able to check the assumptions of his linear regression model, and he found them wanting for three of the four data sets (all but the top left).[88]

88. The data are from Anscombe, *id.*

Anscombe's point, of course, was to underscore the importance of graphing data before analyzing it. By creating the plots displayed in Figure 6, Anscombe was able to check the assumptions of his linear regression model, and found that they were violated in three of the four data sets (all but the top left panel). At the same time, though, his demonstration illustrates a point made by proponents of the "prospecting" versus "transferring" school: no doubt researchers in Anscombe's position would not "transfer" (i.e., present) all the plots they made during the data-analytic phase of their work. Many, if not most, would never see life beyond their designers' computer screens.

But to us this is the *only* major distinction between graphs designed for exploratory purposes and for presentation. We thus agree with Cleveland and others who suggest that while researchers may create more pictures when they are prospecting, the same general principles of graphic design apply regardless of the researcher's purpose.[89] Or, as Jacoby put it, "It is my experience that carefully constructed analytic graphs also are quite effective for presentational purposes."[90] And so it is to those principles for careful construction (and impact) that we now turn.

A. Aim for Clarity and Impact

In 1983 the graphic designer Edward Tufte issued his now (in)famous edict: when creating visual displays, "maximize the data-ink ratio," where "data-ink" is "the non-erasable core of a graphic, the non-redundant ink . . . ."[91] Surely, with these words Tufte was pushing researchers to strive for clarity in their graphs—no doubt an important goal. When we construct figures, we *encode* information. What we ask of our readers is to *visually decode* that information; if they cannot do that because our display lacks clarity, our graph fails. Pure and simple.[92]

89. CLEVELAND, ELEMENTS, *supra* note 26, at 1; *see also* John W. Tukey, *Data-Based Graphics: Visual Display in the Decades to Come*, 5 STAT. SCI. 327, 331 (1990) (seeming to agree with Cleveland, but vacillating). On the one hand, Tukey writes that "[t]here is no reason why a good strategy for prospecting will also be a good strategy for transfer." *Id.* On the other, Tukey claims that "[w]e all need to be clear that visual display can be very effective in serving two quite different functions [prospecting and transferring]." *Id.* He goes on to point out that the major difference between the two "is in prospecting's freedom to use multiple pictures." *Id.*

90. JACOBY, *supra* note 23, at 2.

91. TUFTE, *supra* note 22, at 93, 96.

92. CLEVELAND, ELEMENTS, *supra* note 26, at 64-67; *see also* TUFTE, *supra* note 22, at 55 ("A graphic does not distort if the visual representation of the data is consistent with the numerical representation.").

It turns out, though, that following Tufte's advice to the letter can have precisely the opposite effect: minimizing redundant ink may lead to inelegant, even silly, looking graphs that actually violate the principle of clarity. John W. Tukey, perhaps the most important contemporary figure in scientific graphing,[93] neatly made this point by comparing his now-famous box plot with Tufte's recommended revision (see Figure 7).[94] It is no therefore no surprise that, to our knowledge, no researcher has adopted the Tufte revision: By maximizing the data-ink ratio—thereby deemphasizing the "central clumping" of a distribution that the box is designed to highlight—it minimizes clarity.[95]

Figure 7

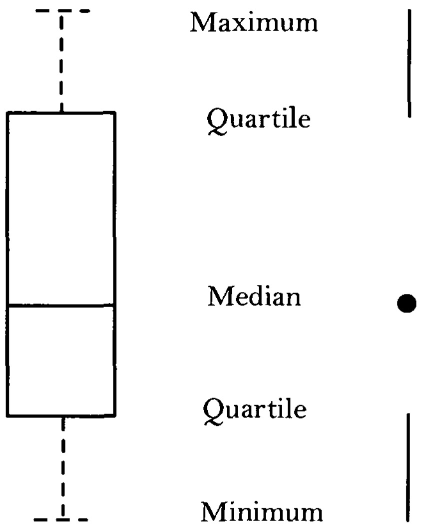


Figure 7: This figure shows two versions of a box plot, which Tukey created to display a visual summary of the distribution of a single variable. The left panel depicts Tukey's classic version (where the adjacent values are the minimum and maximum values in the data set); the right panel shows Tufte's revision, which he designed to maximize "data-ink." By eliminating the box, the Tufte plot may not sufficiently emphasize the middle half of a given distribution. It also lacks the visual impact of Tukey's original design.[96]

93. *See, e.g.*, TUKEY, *supra* note 35. However, it's possible that Tukey is best known for inventing the word "software." *See* David Leonhardt, *John Tukey, 85, Statistician; Coined the Word 'Software,'* N.Y. TIMES, July 28, 2000, at A19.

94. Tukey, *supra* note 89, at 329.

95. *See generally* Tukey, *supra* note 89, at 329 (exploring potential advances in computer-generated visual display).

96. Tukey's original box plot appears in TUKEY, *supra* note 35, at 48. Tufte's redesign is in TUFTE, *supra* note 22, at 125. The caption draws on Tukey's critique of Tufte. *See* Tukey, *supra* note 89, at 328-29.

The Tufte version also strips the box plot of its *impact*—hardly a trivial matter if we hope "to enforce the attention" of our readers.[97] As Tukey famously put it, "The greatest possibilities of visual display lie in the vividness and inescapability of the intended message."[98] He further wrote that Tufte's "less is more aesthetic" can interfere with the realization of those possibilities—and we agree.[99] While we certainly do not want to encourage researchers to create the "multi-colored, three-dimensional pie charts that clutter the pages of USA Today, Time, and Newsweek,"[100] neither do we want them to sacrifice impact. Indeed, had William Playfair (1759-1823), Charles Joseph Minard (1781-1870), and E.J. Marey (1830-1904)—three eminent developers of scientific graphs whose work Tufte seems to admire[101]—failed to focus on impact, there would be little to admire.[102]

Figure 8, in which we reproduce one of Playfair's most famous graphs in the top panel and a "less is more" version in the bottom panel, highlights this claim. No one could deny that the top is flawed in any number of ways,[103] nor could we say that the bottom fails to maximize the data-ink ratio. But the result in the bottom, though cleaner, is far less memorable.

---

97. Tukey, *supra* note 89, at 328.

98. *Id.* at 328. As several scholars point out, Playfair agreed. *E.g.*, Costigan-Eaves & Macdonald-Ross, *supra* note 16, at 319 ("Along with Playfair's desire to tell the story of history graphically was the desire to tell it dramatically."); *see also* Wainer, *supra* note 86, at 341-43.

99. To he fair, it is not even clear that Tufte would follow his own maxim under all circumstances. On the final page of his classic, THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION, TUFTE, *supra* note 22, at 191, he wrote: "Design is choice. The theory of the visual display of quantitative information consists of principles that generate design options and that guide choices among options. The principles should not be applied rigidly or in a peevish spirit; they are not logically or mathematically certain; and it is better to violate any principle than to place graceless or inelegant marks on paper."

100. This sentiment is adopted from Wainer, *supra* note 86, at 341 ("Austerity may serve certain purposes, but humans often prefer, even require, more. Although I shudder to consider it, perhaps there is something to be learned from the success enjoyed by the multi-colored, three-dimensional pie charts that clutter the pages of *USA Today*, *Time*, and *Newsweek*. I sure hope not much.").

101. TUFTE, *supra* note 22, at 32, 34 (calling Playfair one of "the two great inventors of modern graphical design" and deeming Marey's graphs "superb[ly] constructed"). Of Minard's famous depiction of the Napoleon army in Russia, Tufte said "it may well be the best statistical graph ever drawn." *Id.* at 40.

102. Wainer, *supra* note 86, at 341; *see also* Tukey, *supra* note 89, at 333 ("What would, for instance, an unremitting emphasis on 'data-ink ratio' leave of the famous [Minard] Napoleon-in-and-out-of-Russia chart?").

103. For a critique, see Michael Friendly & Howard Wainer, *Nobody's Perfect*, 17 CHANCE 51, 51-53 (2004).

Figure 8

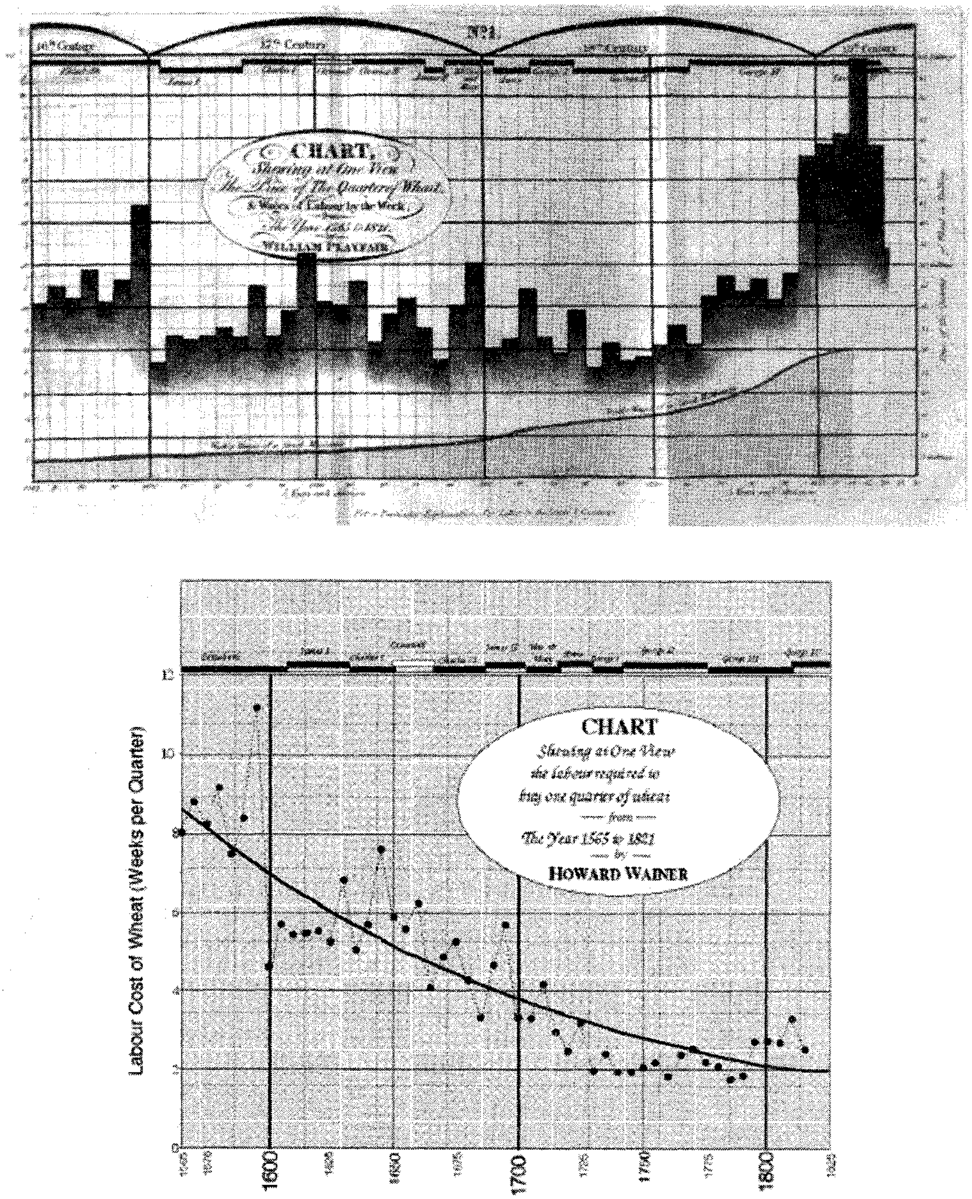


Figure 8: Both panels convey information about the labor required to buy wheat in England, 1565-1820. The top panel, the original produced by William Playfair in 1786, shows three time series: 1) the reigns of English monarchs, 2) the price of a quarter of

wheat (in the bars), and 3) the wages of a good mechanic (the line).[104] The bottom panel is Friendly and Wainer's redesign of Playfair's chart. The dots on the connected line show the number of weeks required to buy one quarter of wheat; the solid line is a fitted quadratic.[105] While the bottom panel has the virtue of clarity, some might contend that, relative to Playfair's plot, it lacks visual impact or the "ability to enforce attention."[106]

So how can researchers aim for clarity *and* impact? First, and foremost, they must eliminate what Tukey deems "busyness," what Tufte has famously labeled "chart junk," or what Cleveland calls "visual clutter"—in other words, irrelevant or distracting elements that stand in the way of decoding. Second, they should not underestimate their readers by dumbing down graphs and thus potentially muting the impact of their displays.

1. Eliminate Distracting Elements

If there is one principle of visualization on which graphic designers, statisticians, and social scientists agree, it is that researchers should eliminate irrelevant, distracting elements from their displays. And there is plenty to eliminate in the law reviews. We make this point in Figures 9 through 12, which depict some of the more common problems, as well as our correctives for eliminating superfluity—that is, for making the data stand out.[107]

Beginning with Figure 9, from Cloud et al.'s experimental work on whether the mentally disabled can understand Miranda warnings,[108] *we eliminate the depth cue* (i.e., we transform the graph from 3-D to 2-D). The added dimensionality is not only irrelevant to the data display; research (experimental research, ironically enough) has found that it can interfere with graph comprehension.[109] It is for these reasons that scholars in most other disciplines, and even graphic designers working for popular publications, are now eradicating superfluous dimensions—and we recommend that law professors follow suit.

---

104. The Playfair chart appears in all three editions of AGRICULTURAL DISTRESSES (1822). *See* Costigan-Eaves & Mcdonald-Ross, *supra* note 16, at 322; TUFTE, *supra* note 22, at 34 (reproducing the Playfair chart).

105. The chart is from Friendly & Wainer, *supra* note 103, at 53.

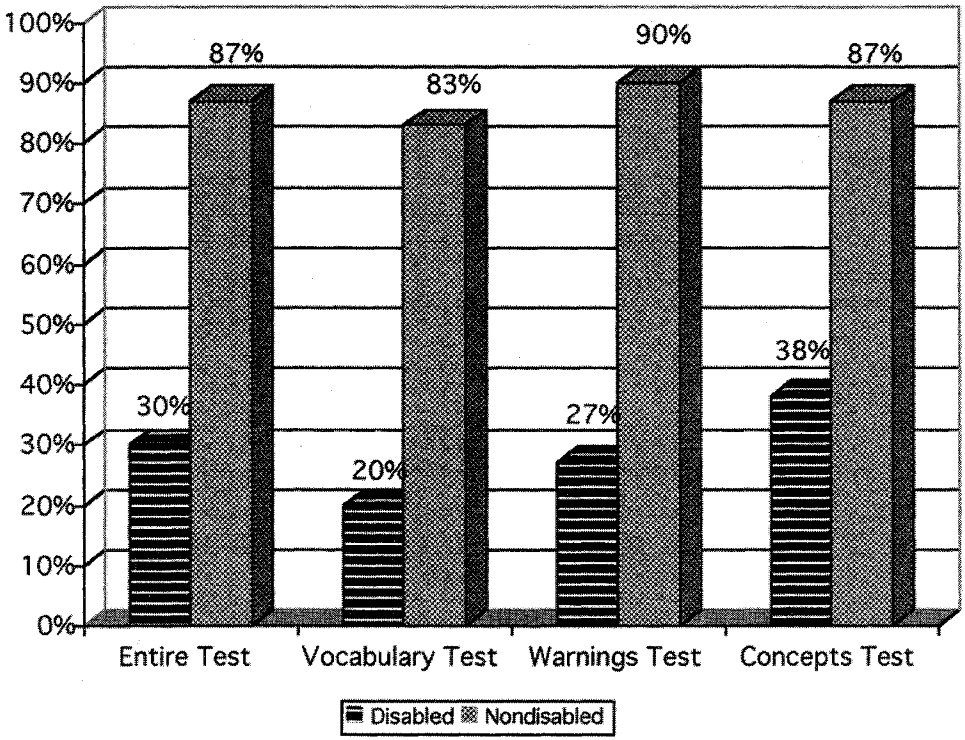
106. Tukey, *supra* note 89, at 328.

107. This is a general problem; others are more specific to particular types of graphs. Details will be discussed in the future article.

108. Morgan Cloud et al., *Words Without Meaning: The Constitution, Confessions, and Mentally Retarded Suspects*, 69 U. CHI. L. REV. 495, 539 (2002).

109. *See, e.g.*, Martin H. Fischer, *Do Irrelevant Depth Cues Affect the Comprehension of Bar Graphs?* 14 APPLIED COGNITIVE PSYCHOL. 151, 161 (2000).

Figure 9


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
87%
83%
90%
87%
30%
20%
27%
38%
Entire Test
Vocabulary Test
Warnings Test
Concepts Test
Disabled
Nondisabled

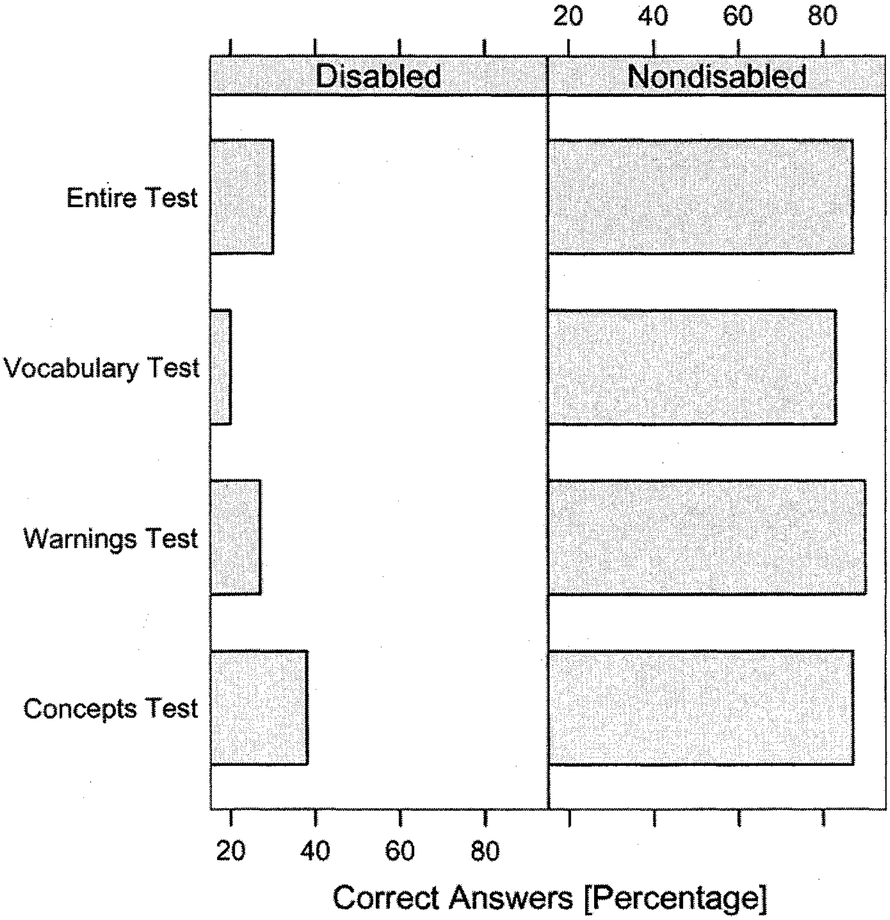


Figure 9: Both panels depict the percentage of correct results on three exams administered to the mentally disabled and nondisabled. The left panel is a reproduction of the original graph that appeared in Cloud, et al.; the panel on the right reflects our attempt to reduce clutter. Specifically, we eliminated (1) the depth cue, (2) the internal data labels, (3) tick marks on the horizontal axis, and (4) the legend. We also have reduced the number of tick marks on the vertical axis and supplanted the crosshatching with a solid color. The result is a graph far easier to decode but that still conveys the authors' message.[110]

Note that removing the depth cue is not the only change we made to Cloud's original figure. We further reduced clutter by, first, eliminating the internal data labels (e.g., 30%, 87%). As far as we can tell, the authors' primary purpose is to focus their audience on the *comparison* between the two groups of test takers, not on *precise values*, and so those values are unnecessary (if we are wrong, then Cloud et al. would have been better off with a table). Second, we

110. The original graph and data are from Cloud et al., *supra* note 108, at 539.

altered the tick marks on both axes, eliminating those on the horizontal (verbal descriptors need not be "ticked") and reducing by half the number on the vertical axis (typically, according to visualization studies, three to ten marks are sufficient).[111] Third, and again in line with extant work on graphic perception, not to mention good design practice, we filled the bars with a solid gray color rather than cross hatches, slanted lines, or other "pop-art" marks that can appear to vibrate.[112] Finally, we eliminated the legend. It too is unnecessary, and it can also interfere with decoding because of the tendency to look back and forth between the key and the data.

In Figure 9 we moved the legend to the heading. An alternative is to describe the data keys in the caption. This is an acceptable, even standard, practice (for more on the use of captions, *see infra* Part IV. C). Nonetheless, if it is possible to insert labels into the interior of the graph without interfering with visual assembly of the plotting symbols or lines, we recommend that step[113]—and have taken it in Figure 10. Note that the labels neither cause too much clutter nor obscure the data. Actually, because the reader need not consult a key (and, in this instance, a legend that actually interferes with decoding), they improve visualization—here, a comparison of criminal and civil cases filed per authorized judge.

111. *See, e.g.*, CLEVELAND, ELEMENTS, *supra* note 26, at 39.

112. *See, e.g.*, TUFTE *supra* note 22, at 107 ("Contemporary optical art relies on moiré effects, in which the design interacts with the physiological tremor of the eye to produce the distracting appearance of vibration and movement."). Tufte also argues that moiré vibration, caused by cross hatching and other nonsolid fill types, makes for "bad data graphics." *Id.* at 108-11. *See also* Tukey, *supra* note 89, at 332 (suggesting that one "avoid slanted lines within the elements").

113. See CLEVELAND, ELEMENTS, *supra* note 26, at 44-45 for more on this point.

Figure 10

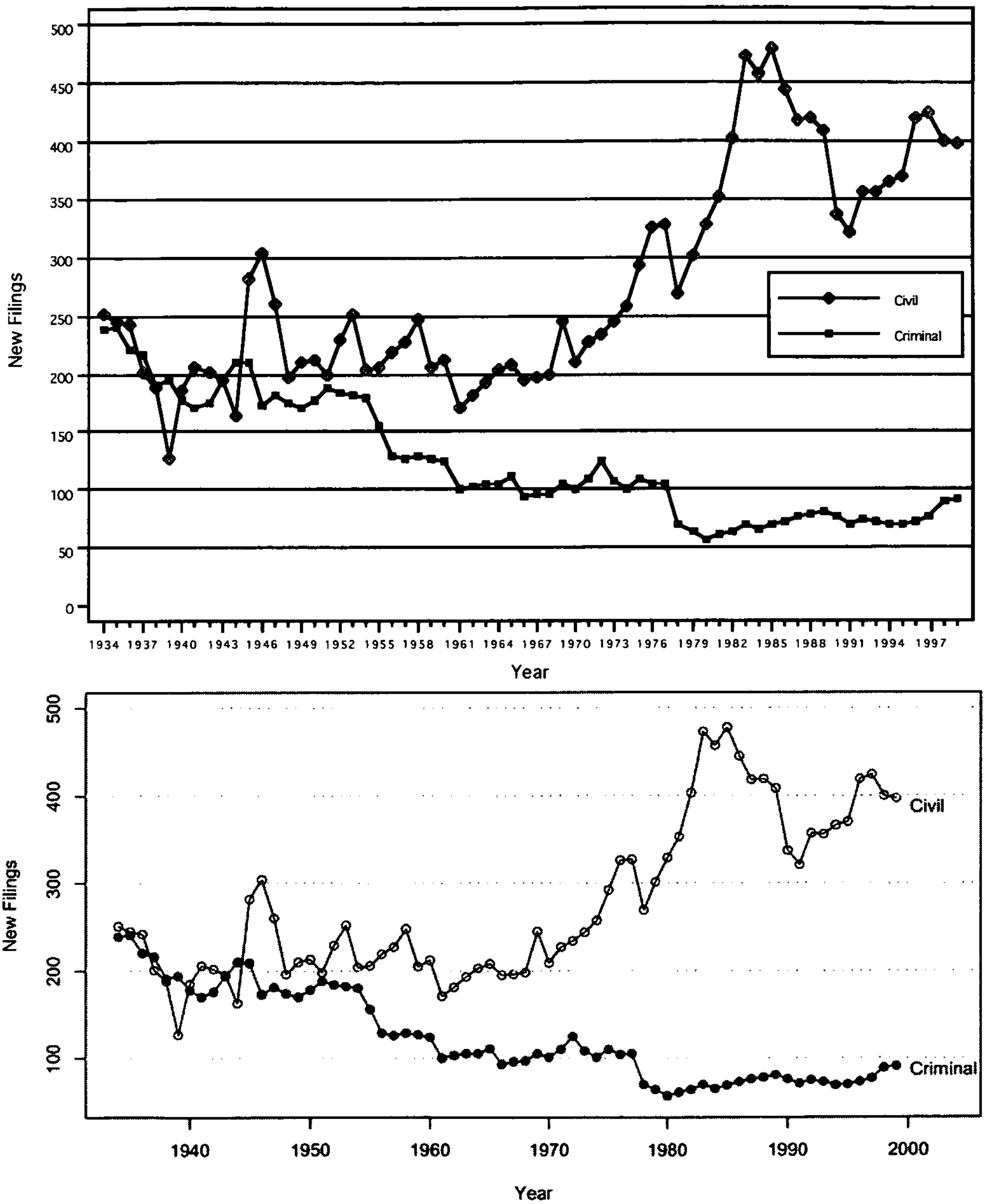


Figure 10: Both panels depict the number of criminal and civil cases filed per authorized judge, 1934-1999. The top panel is the original graph that appeared in a law review article by Michael Simons; the bottom panel reflects our attempt to improve visualization of the data. Specifically, we eliminated the legend (replacing it with internal data labels) and the dark grid, both of which obscure the data. We also reduced the number of tick marks and associated labels. Finally, we supplanted the non-circular

sub-elements with circular connectors, though if the author's purpose is merely to show trends in new filings, the circles may be unnecessary.[114]

Also observe that in altering the top panel of Figure 10, we changed the symbols connecting the lines from non-circular to circular elements. While many graphing packages offer users a dazzling array of plotting symbols, such as squares, triangles, diamonds, and so on, researchers should avoid almost all of them. More to the point, experimental results show that "unless there is a serious need for more distinctions,"[115] analysts should stick to circular forms. Typically, they can gain sufficient variation in the size and fill of circles to display prominently the data. If not, as we explain momentarily, we recommend a series of smaller plots within a single figure.

Now observe what we did not change about original Figure 10. First, we retained the horizontal axis as "New Filings" and the vertical, as "Year". This conforms to standard practice of placing explanatory variables on the x- (horizontal) axis and outcomes on the y- (vertical) axis.[116] Some exceptions to this convention do exist, however, most notably when a verbal descriptor accompanies each case. Figure 11, in which the author, Richard Lazarus, provides "environmental protection scores" for each justice, is an example.[117] Rather than running the descriptors (the names of the justices) vertically and within the bars (which interferes with visualization) we created a dot plot.

114. The original graph and data are from Michael A. Simons, *Prosecutorial Discretion and Prosecution Guidelines: A Case Study in Controlling Federalization,* 75 N.Y.U. L. Rev. 893, 914 fig.4, 964-65 (2000).

115. Tukey, *supra* note 89, at 332-33; *see also* CLEVELAND, ELEMENTS, *supra* note 26, at 154-64.

116. *See, e.g.,* Gelman et al., *supra* note 16, at 122 (noting that this is the standard convention).

117. Richard J. Lazarus, *Restoring What's Environmental About Environmental Law in the Supreme Court,* 47 UCLA L. REV. 703, 725, 812 (2000).

Figure 11

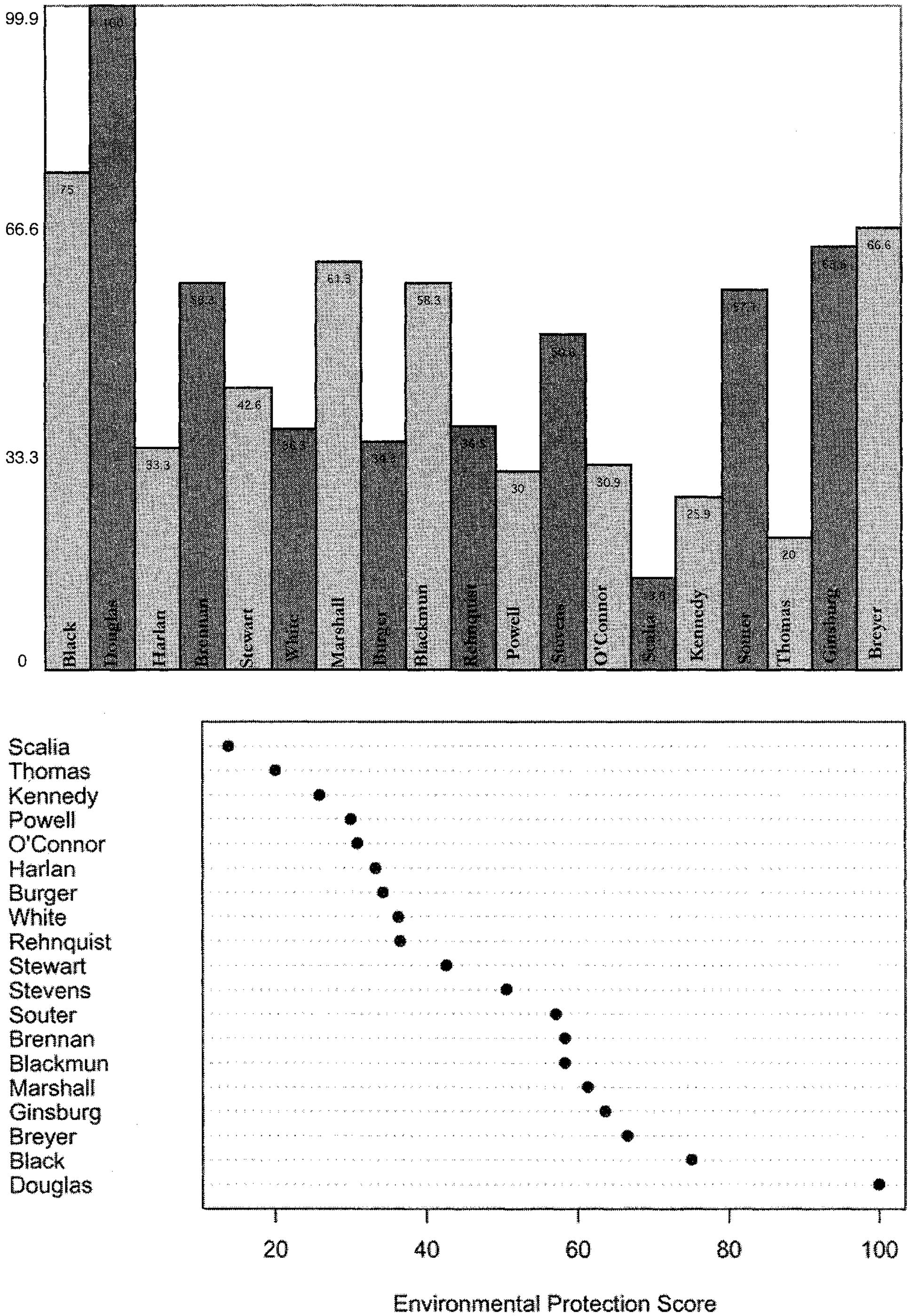

99.9
66.6
33.3
0
75
100
33.3
58.3
42.6
61.3
58.3
30
30.9
25.9
20
66.6
Black
Douglas
Harlan
Brennan
Stewart
White
Marshall
Burger
Blackmun
Rehnquist
Powell
Stevens
O'Connor
Scalia
Kennedy
Souter
Thomas
Ginsburg
Breyer
Scalia
Thomas
Kennedy
Powell
O'Connor
Harlan
Burger
White
Rehnquist
Stewart
Stevens
Souter
Brennan
Blackmun
Marshall
Ginsburg
Breyer
Black
Douglas
20
40
60
80
100
Environmental Protection Score

Figure 11: Both panels depict environmental protection scores developed by Richard J. Lazarus. The score is the number of pro-environmental votes cast by the justice over the total number of votes multiplied by one hundred. The top panel is a reproduction of the original graph that appeared in Lazarus' article; the bottom panel reflects our attempt to improve visualization of the data. Specifically, we transformed the chart into a dot plot so that the labels are easy to read and do not obscure the data. We also changed the tick marks, eliminated the value displays, reordered the justices (from lowest to highest level of support for environmental protection)—all with an eye toward facilitating comparisons.[118]

A second feature of Figure 10 (that is, the Simons figure on civil and criminal filings) that we did not change was the ordering of the labels on the horizontal axis, from the earliest (1934) to the latest (1999) year in the data set. Again, this is standard procedure for a time series plot of this sort but not for many other data displays. In fact, more typically researchers should order the labels by decreasing frequency (or another substantively motivated pattern), not, for example, alphabetically or even by time, as this can obscure interesting patterns.[119] Figure 4 (Revesz's figure) provides an example of an ordering that we changed to facilitate comparison, as does Figure 11. Lazarus placed the justices according to their date of appointment to the Court, but because his textual description focused on a comparison of the justices' scores across time, we ordered them by increasing support for the environment. Now, at the very least, the display draws attention to patterns of central concern to the author.

Finally, returning to Figure 10, note that we retained both the "Civil" and "Criminal" lines in a single picture. Because they do not clutter the display this is a reasonable decision here. But in other cases, graphing too much in one scale rectangle can obscure the data and should be avoided.[120] The easiest solution is to juxtapose smaller graphs within a single display.[121]

We followed this strategy in plotting Revesz's data (Figure 4), and we take it again in Figure 12, which we have drawn from Epstein et al.'s analysis on Senate voting over nominees to the Supreme Court.[122] The point of the original graph (in the top panel) is to show

---

118. *Id.*

119. *See* Gelman et al., *supra* note 16, at 122 (noting that display axes should be labeled such that interesting patterns are highlighted).

120. CLEVELAND, ELEMENTS, *supra* note 26, at 35-36.

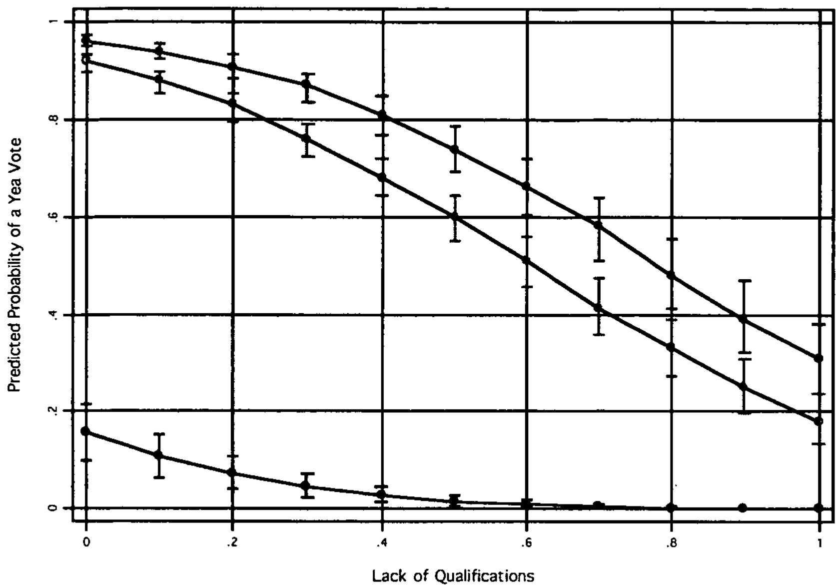
121. *See id.* at 38-39 (noting that juxtaposition allows the reader to clearly see each set of data); *see also* Gelman et al., *supra* note 16, at 122 ("[T]he most crucial tool is probably the juxtaposition of many small plots into a single figure . . . ."); EDWARD R. TUFTE, ENVISIONING INFORMATION 53 (1990) (noting that separating data can reduce noise and enrich the content of displays); Tukey, *supra* note 89, at 332 ("Be ready to avoid busyness by splitting one picture into two or more.").

122. Lee Epstein et al., *The Changing Dynamics of Senate Voting on Supreme Court Nominees*, 68 J. POLITICS 296, 301 (2006).

the predicted probability of a senator casting a yea vote over the range of a nominee's qualifications (0 indicates most qualified and 1 indicates least qualified), when the ideological distance between the senator and candidate is set at minimum, mean, and maximum levels. But the error bars (representing 95% confidence intervals) and the dark grid are so interfering that it is nearly impossible to get much of a feel for the range the bars represent, not to mention for the predicted probabilities.[123] There are so many lines that the message gets lost: the results do not sufficiently stand out. We cleared away the clutter by creating the four smaller plots depicted in the bottom panel and lightening the grid. Now the predicted probabilities (and error bars) are far easier to perceive, as are patterns in the results.

123. This example tracks Cleveland's approach in THE ELEMENTS OF GRAPHING DATA. *See* CLEVELAND, ELEMENTS, *supra* note 26, at 38-40 (illustrating this method by separating one cluttered graph into three juxtaposed panels of data).

Figure 12


Predicted Probability of a Yea Vote
0
.2
.4
.6
.8
1
Lack of Qualifications

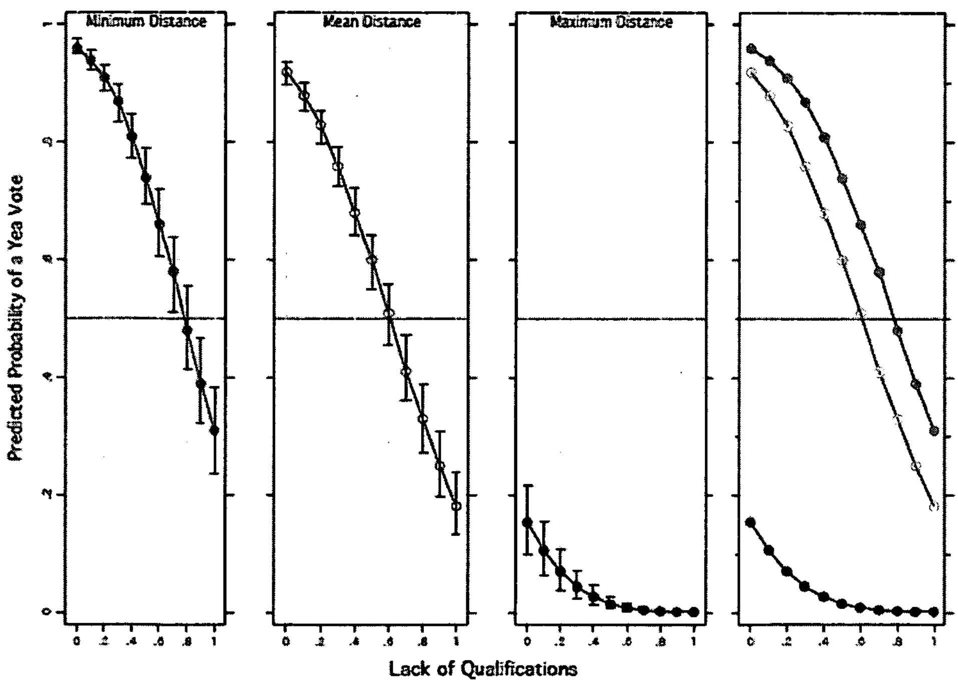

Minimum Distance
Mean Distance
Maximum Distance
Predicted Probability of a Yea Vote
Lack of Qualifications

Figure 12: Both panels depict the results of a multivariate model of Senate voting over Supreme Court nominees Specifically, they show the probability of a senator casting a yea vote over the range of a nominee's qualifications (0 indicates most qualified and 1 indicates least qualified), when the ideological distance between the senator and candidate is set at minimum, mean, and maximum levels. In both panels the vertical (capped) lines represent 95% confidence intervals. The top panel attempts to graph all three lines (minimum, mean, and maximum ideological distance, respectively) in the same picture; the bottom panel juxtaposes the lines in four panels. The smaller plots (along with a lightening of the dark grid) facilitate comparisons.[124]

Note, though, that in creating the smaller panels, we followed Cleveland's advice that "superposed data sets must be readily visually discriminated."[125] Accordingly, in each panel, the scales are identical, as are the gridlines and number of tick marks. Indeed, all that differs among the panels are the plotting symbols.

2. Trust Your Readers

The suggested revisions to Figures 9 through 12 have the benefit of reducing clutter and thus enhancing the reader's ability to decode the information they house. But the alterations do not—at least we hope not—have the effect of dumbing down the graphs. We have more faith in our readers than that and indeed, the principle we sought to follow in redesigning them is "Aim for *Clarity* and Impact, not Aim For *Simplicity* and Impact." To put it another way, to us, Strunk & White's advice for writers applies equally to graphic design: "no one can write decently who is distrustful of the reader's intelligence, or whose attitude is patronizing."[126]

Some scholars disagree, arguing that we ought judge graphs by their simplicity— by how many words they save or by how fast viewers can comprehend them. But these criteria are far too restrictive. As Tukey pithily writes, "A picture may be worth a thousand words, but it may take a hundred words to do it."[127] And Cleveland speaks to the issue of speedy comprehension: "While there is a place for rapidly-understood graphs, it is too limiting to make speed a requirement in science and technology, where the use of graphs ranges from detailed, in-depth data analysis to quick presentation."[128]

---

124. The underlying statistical model and the data used to construct the graphs is from Epstein et al., *supra* note 122, at 301.

125. CLEVELAND, ELEMENTS, *supra* note 26, at 50.

126. TUFTE, *supra* note 22, at 81 (quoting WILLIAM STRUNK, JR. & E.B. WHITE, THE ELEMENTS OF STYLE 70 (1st ed. 1959)).

127. Wainer, *supra* note 86, at 341.

128. CLEVELAND, ELEMENTS, *supra* note 26, at 94.

The same, we believe, applies to the social sciences and law. To argue otherwise would be to would be to eliminate classes of graphs (e.g., scatterplot matrices) that may require careful study but are otherwise extremely valuable, effective, and memorable. Likewise, elevating simplicity—a symptom, really, of lacking faith in our audience—can lead to graphs that lack clarity and elegance. So, for example, out of a belief that readers will not look at tick mark labels and will instead apply "the most trivial of quantitative reasoning"[129] researchers often feel compelled to start their scales with (an unnecessary) zero.[130] As we show in Figure 13, not only does this disrespect our audience and waste space (we should aim to fill the data rectangle) but it also may interfere with decoding. In this case, the author, Christopher Schroeder, wants to draw attention to the decline in Americans' trust of the government.[131] The trend he identifies seems real enough, but because he makes use of zero it is hardly discernible. Displaying the line more sensibly, as we do in the middle panel, facilitates a more effective judgment about the data.[132]

129. *Id.* at 78 (responding to DARRELL HUFF, HOW TO LIE WITH STATISTICS 64-65 (1954), which claims that excluding zero is downright dishonest).

130. We refer here to graphs that do not require zero. Of course, when zero is relevant, researchers must include it. *See, e.g.*, Gelman et al., *supra* note 16, at 122 (noting that zero is a relevant baseline in their research).

131. Christopher H. Schroeder, *Causes of the Recent Turn in Constitutional Interpretation*, 51 DUKE L.J. 307, 346 (2001).

132. *See* CLEVELAND, ELEMENTS, *supra* note 26, at 80-82 (illustrating a similar example of removing zero in order to "convey much more quantitative information").

Figure 13

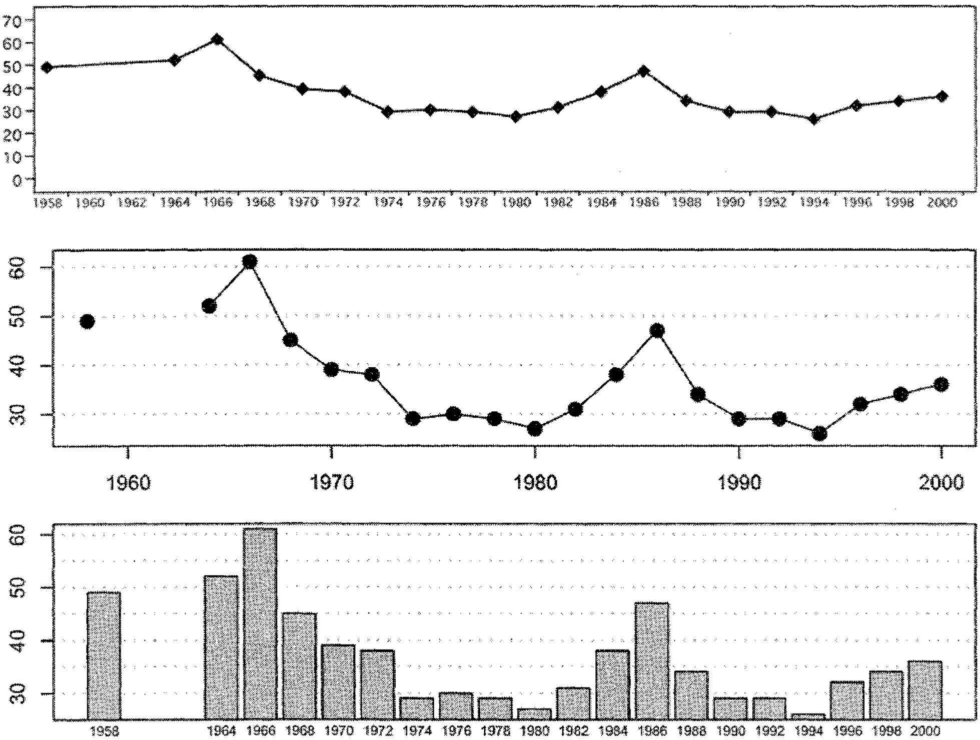


Figure 13: All three panels show the Trust in Government Index (developed from the American National Election Studies) on the vertical axis and year (from 1958-2000) on the horizontal axis. From the top panel, which appears in Schroeder's study, it is difficult to observe the decline in trust over time. We excluded zero in the middle pattern to facilitate a more effective judgment about the data, and did not connect 1958 to 1964 to accentuate data sparseness. A problem in both the top and middle panels, however, is that neither clearly delineates missing years. Accordingly, in the bottom panel we moved to a bar chart. Now readers can observe missing data, as well as the fact that the Index is available only for even-numbered years.[133]

This is just one example of how incorrect assumptions about the naïveté of our readers can interfere with principles of sound graphic design. Others are easy enough to summon, but the larger point is that we not only should but *must* assume that our audience will look closely at the graphs and understand them. Without this assumption, as Cleveland notes, "graphical communication would be far less useful."[134] Actually, we would go further and ask why we

133. The top panel appeared in Schroeder, *supra* note 131, at 347; the Trust in Government Index is available at http://www.umich.edu/~nes/nesguide/toptable/tab5a_5.htm.

134. CLEVELAND, ELEMENTS, *supra* note 26, at 79.

should bother with graphs or even tabular displays if we believe our readers will not bother to peruse them?

On the other hand, we are certainly not advocating that researchers fool their audience or themselves. And it is along these lines that line charts in the top and middle panels of Figure 13 are troubling. Because the author chose a connected line graph, it may well appear that data exist for each year between 1958 and 2000 (and the alternate tick marks in the original [top] panel do not help!). But this is not the case: The American National Election survey, from where the Trust in Government Index comes, is fielded only every other year (and the Index was not computed at all for 1960 and 1962). By moving to a bar chart, in the bottom panel of Figure 13, we are better able to alert readers to these "missing" years in the data set.

B. Iterate

To arrive at the bottom panel of Figure 13, we iterated, creating two (or more) depictions of the same data until we generated the clearest and most effective presentation. This is typical of graph making: In our experience, it is nearly impossible (whether for purposes of prospecting or transferring) to get it right on the first try. Practically, this means that it is necessary to have software that can easily reproduce graphs from scripts and provide fine control over all elements.

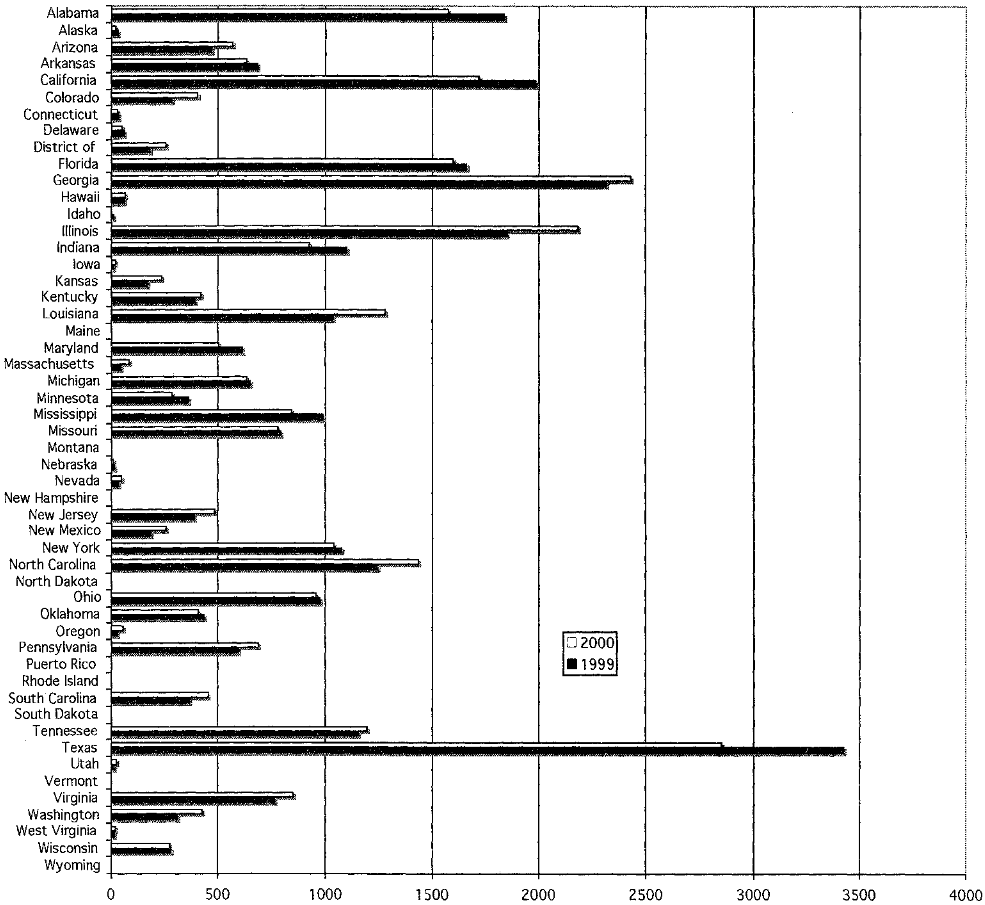
Consider another, perhaps more typical, example of the iterative process in action. In the top panel of Figure 14, we place a graph that appeared in Tanya Katerí Hernández's article on the role of "race ideology" in the enforcement of anti-discrimination laws in the United States and Latin America.[135] Though by no means a horrid visualization of the data, it contains a sufficient number of irrelevant or obscuring elements to make decoding difficult. Comparison and pattern detection are also no easy tasks because of the way the author ordered the data: alphabetically (which facilitates the look up of particular states, but not comparisons) rather than by magnitude. Finally, because of the distribution's range (from zero to 3424) some of the smaller data points (e.g., Nebraska) are difficult to see. A few alterations are thus in order.

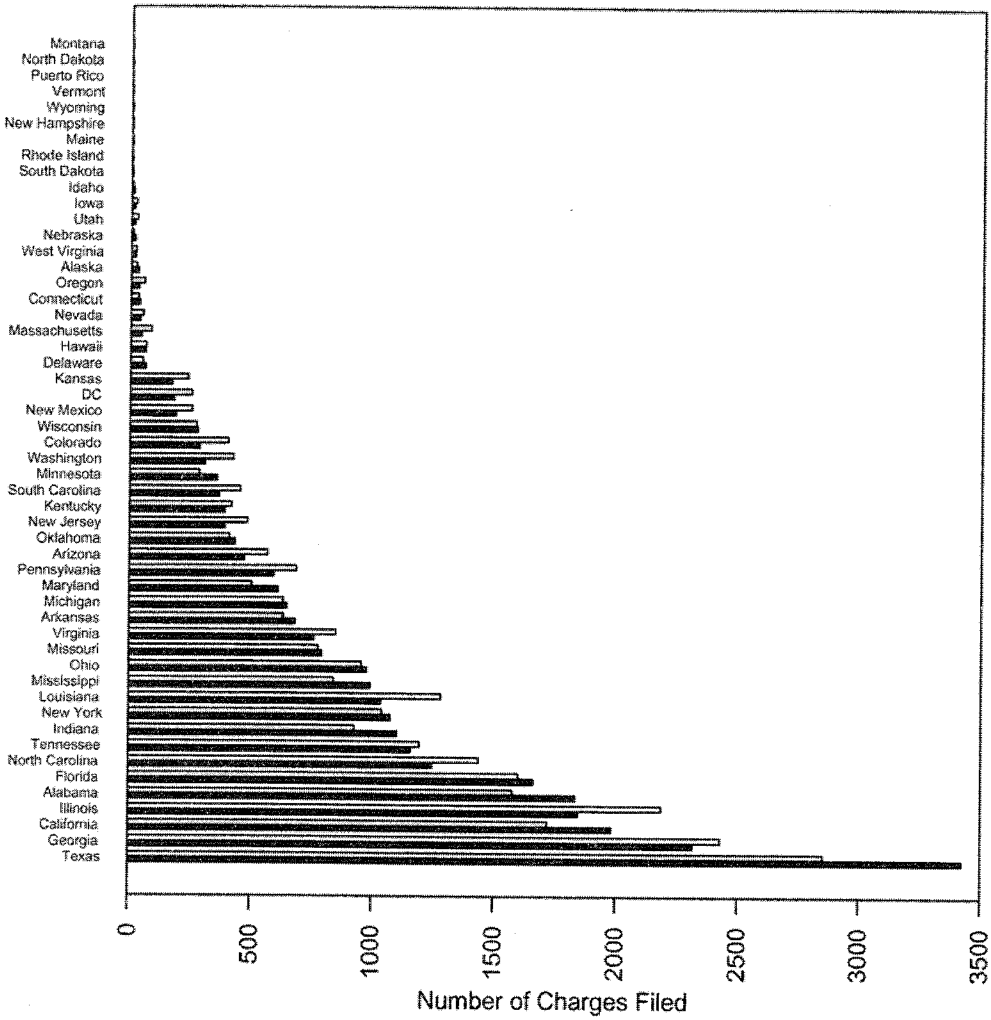
The bottom left panel represents our first attempt at improving visualization. We have removed some irrelevant elements and

135. Tanya Katerí Hernández, *Multiracial Matrix: The Role of Race Ideology in the Enforcement of Antidiscrimination Laws, A United States-Latin America Comparison*, 87 CORNELL L. REV. 1093, 1172 (2002). We took the liberty of using the states' full names rather than their postal abbreviations, which Hernández had displayed.

reordered the states, but the product remains less-than-satisfying; in particular, it is still difficult to discern within-state comparisons between 1999 and 2000. To correct the problem, we moved away from bar charts altogether and to a dot plot in the bottom right panel. Now, the result is a graph that is far easier to decode but took two iterations (actually many more) to create

Figure 14




Montana
North Dakota
Puerto Rico
Vermont
Wyoming
New Hampshire
Maine
Rhode Island
South Dakota
Idaho
Iowa
Utah
Nebraska
West Virginia
Alaska
Oregon
Connecticut
Nevada
Massachusetts
Hawaii
Delaware
Kansas
DC
New Mexico
Wisconsin
Colorado
Washington
Minnesota
South Carolina
Kentucky
New Jersey
Oklahoma
Arizona
Pennsylvania
Maryland
Michigan
Arkansas
Virginia
Missouri
Ohio
Mississippi
Louisiana
New York
Indiana
Tennessee
North Carolina
Florida
Alabama
Illinois
California
Georgia
Texas
0
500
1000
1500
2000
2500
3000
3500
Number of Charges Filed

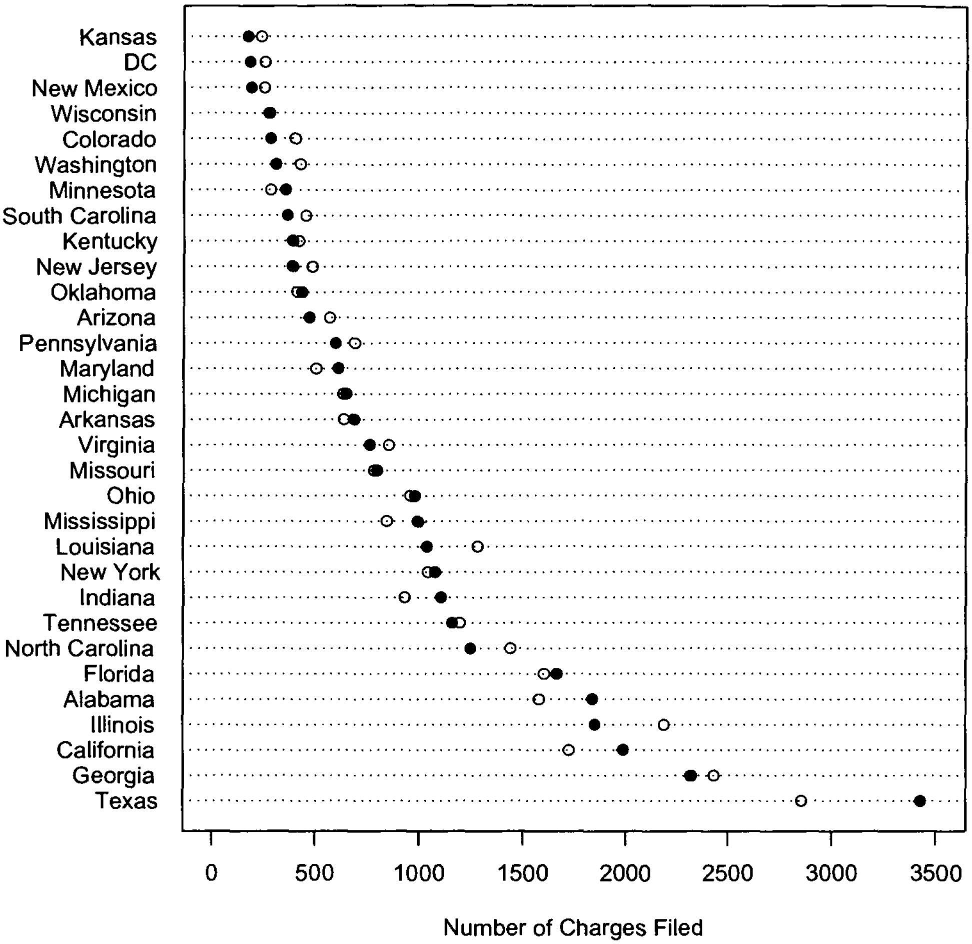


Figure 14: All three panels show Equal Employment Opportunity Commission ("EEOC") race charges filed, 1999-2000 by state. In all three lighter bars or dots are 1999; the darker bars or dots are 2000. The top panel appears in Tanya Katerí Hernández's article. Because the extraneous elements interfere with decoding, we eliminated them; we also reordered the states to enable readers to observe patterns in the data. The result, in the left bottom panel, is a more readable graph but one that still obstructs within-state comparisons for the years 1999 and 2000. To improve further comprehension and pattern detection, we moved to the dot plot depicted in the right bottom panel. Moreover, to enhance readability, we have presented states with more than one hundred EEOC race charges. If regional pattern were of interest, the data could be organized geographically (see Figure 4). The sequence of charts shores up Cleveland's advice: ". . . we should not hesitate to make two or more graphs of the same data."[136]

136. *Id.* at 1170-72; CLEVELAND, ELEMENTS, *supra* note 26, at 94.

C. Write Detailed Captions

For Figure 14 (and, in fact, for all the figures throughout this article), we wrote a detailed caption. This is standard operating procedure in many disciplines but not yet in law.[137] Indeed, our inspection of the legal publications turned up captions that were at best unhelpful guides to the figure and at worst, non-existent.

Illustrative is Figure 15, which we recreated from data in Wagner & Petherbridge's rigorous study of Federal Circuit's methodological approaches to claim construction.[138] While the data and methods throughout the article are entirely appropriate, this figure, *absent any guidance from the authors*, is difficult to interpret. We offer a corrective, in the form of a detailed caption that first explains what the graph displays and then draws "the reader's attention to salient features of the display."[139] Note that we do not clutter the graph with a legend; again, following standard practice, we provide a key in the caption.

More generally, we commend to you Cleveland's advice that captions should be "comprehensive and informative."[140] To wit, they should contain (some version of) the following information:

1. A description of everything in the display.

2. A sentence or two on particularly important features of the data or results.

---

137. Consider the following "Figure Captions" guidelines for manuscripts submitted to the JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION ("JASA"):

> Each figure must have a figure caption, including the figure number. Figures are numbered consecutively, using arabic numerals, as they are cited in text.
>
> Prepare the captions on a separate sheet and place them after the tables. They will be typeset and placed beneath the figures.
>
> Figures must be clearly described. The combined information of the figure caption and the text of the body of the paper should provide a clear and complete description of everything that is on the figure. Detailed captions can often be of great help to the reader. First, describe completely what is graphed in the display; then draw the reader's attention to salient features of the display and briefly state the importance of these features.
>
> Generally, it is a good idea to include the key to symbols in the caption to avoid cluttering the display. Abbreviations not already defined in text must be defined in the caption.
>
> Figures and their titles are editorially reviewed. The following examples illustrate these guidelines.

American Statistical Association Style Guide, http://www.amstat.org/publications/index.cfm?fuseaction=style-guide (example omitted).

138. R. Polk Wagner & Lee Petherbridge, *Is the Federal Circuit Succeeding? An Empirical Assessment of Judicial Performance*, 152 U. PA. L. REV. 1105, 1150 (2004).

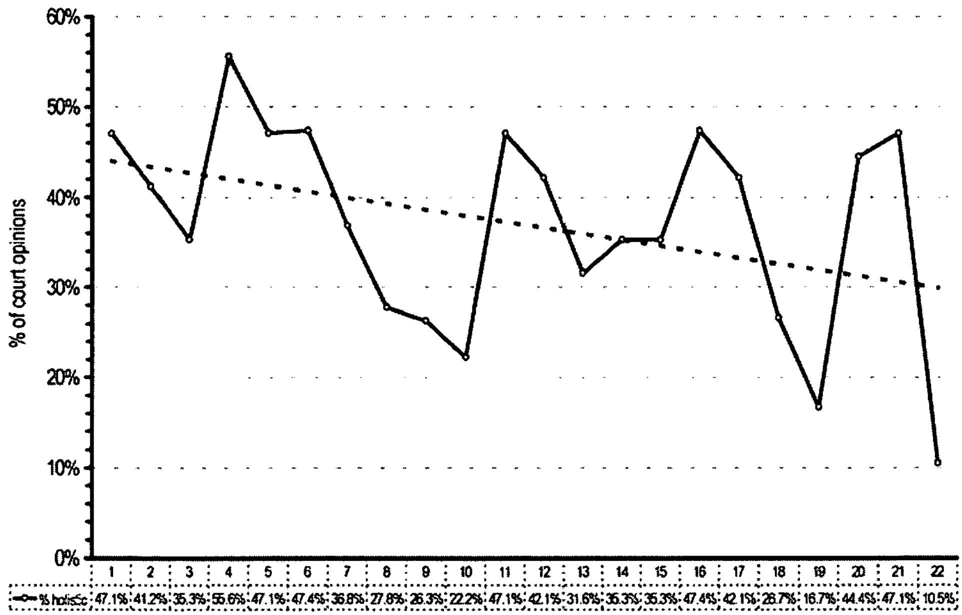
139. American Statistical Association Style Guide, *supra* note 137.

140. CLEVELAND, ELEMENTS, *supra* note 26, at 56.

3. A brief statement on why those features are important.[141]

On the other hand, we want to encourage flexibility: Captions will (and should) vary with the type of display and the data or results being displayed. Also, of course, it is possible to overdo a caption by providing too much (extraneous) information. But we cannot stress enough that, as a general principle, law professors will well serve their audience with a graph that is as self-contained as possible.[142]

Figure 15



141. We adapt this advice from the American Statistical Association Style Guide, *supra* note 137, and from CLEVELAND, ELEMENTS, *supra* note 26, at 57.

142. In its instructions to authors, the JOURNAL OF EMPIRICAL LEGAL STUDIES, *supra* note 20, makes this point as well: "Tables and figures should stand on their own. When appropriate, authors should include an explanatory note for a table or figure. The goal is to have the table or figure 'stand on its own' so that a busy reader can understand the table without reading the whole article."

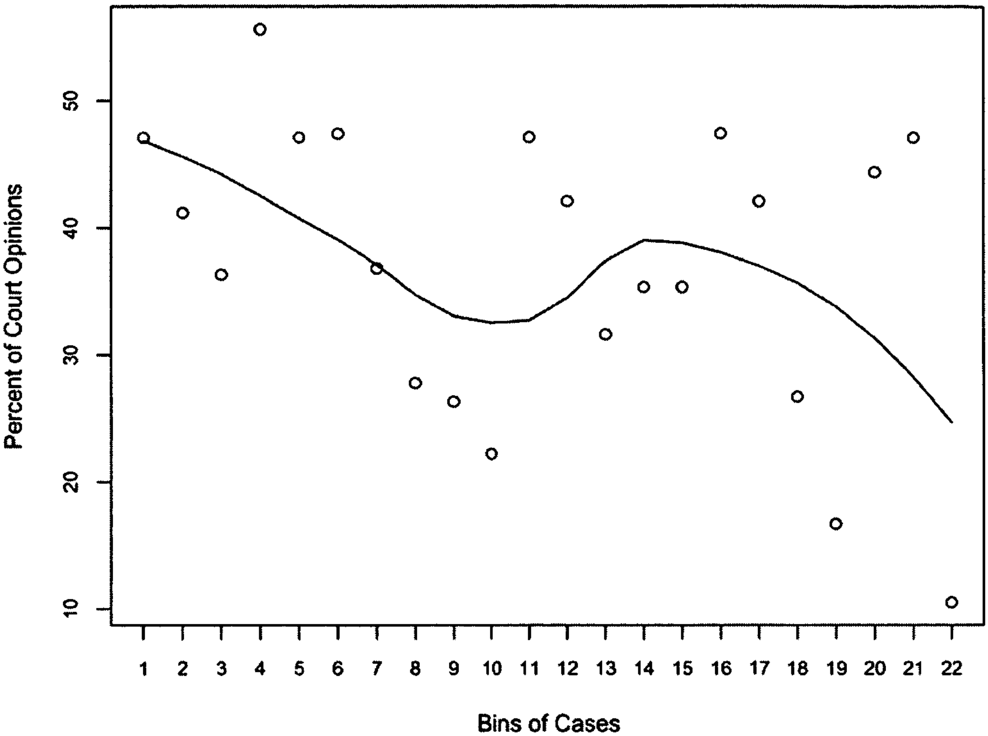


Figure 15: Both panels depict the percentage of opinions issued by the Federal Circuit that employed a holistic methodology in claim construction, by bin, 1996-2002. Bins are ordered chronologically, such that each bin represents two to four months of opinions during the period of study and each bin contains eighteen or nineteen opinions. Note that in the top panel, reproduced from R. Polk Wagner & Lee Petherbridge's article, the authors included a data table and used a connected line graph. In the bottom panel, we removed the data table; it is unnecessary and distracting. We also eliminated the lines since they convey a sense of connection that is not necessarily reflected in the data. In the bottom panel the circles represent the percentage of court opinions falling into each bin; tbe line is a local regression (loess) line, which shows a slight decline in the percentage of opinions making use of a holistic, as opposed to a procedural, methodology in the first ten bins and in the last eight.[143]

V. CONCLUSION

As quantitative empirical work is gaining traction in the legal academic community, and as members of that community are producing empirical work of increasingly high quality, the time has come to consider questions of communication.

Here we have attempted to do just that by supplying general suggestions for improving the presentation of empirical analyses. In

143. The top panel and data are from Wagner & Petherbridge, *supra* note 138, at 1150.

the next installment, we move to more specific strategies for communicating data and results, though we emphasize the latter. This reflects our belief, echoed throughout this Article, that an emphasis on sterile statistical results without an interrogation of their substantive importance disservices the research, the researchers, and their readers. On the other hand, analysts that assess the effect of their findings and are able to effectively and accessibly communicate that information will find the payoffs considerable.

Our goal in the Article to follow is to help scholars achieve that end, as well as to encourage editors of law reviews to ensure that they do. Hence, not only do we offer a set of strategies for researchers, but also a set of protocols centering on the presentation of statistical results for implementation by legal publications.

***