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

Lee Epstein,
Andrew D. Martin & Christina L. Boyd   *60 Vand. L. Rev. 801 (2007)*

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, they have been less effective at communicating the products of their labor. A strong devotion to tabular, rather than graphical, displays, and claims about "statistical significance" rather than substantive importance, are just two areas requiring improvement.

Here, as in Part I, we attempt to adapt a burgeoning literature in the social and statistical sciences to the unique interests of legal scholars. Our proposals are many in number, but none is particularly difficult to implement. More to the point, we believe that law professors should want to implement them. If other fields are any indication, moving toward more appropriate and accessible presentations of data will heighten the impact of empirical legal scholarship regardless of the audience—no doubt a desirable goal in a discipline that rightfully prides itself on its contributions to forming legal and public policy.

***

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

Lee Epstein, Andrew D. Martin,** & Christina L. Boyd****

801

I. INTRODUCTION

To assess the controversial claim that affirmative action in U.S. law schools causes blacks to fail the bar exam, Daniel E. Ho deploys an innovative approach.[1] Professor Ho "matches" students on all relevant observable variables,[2] except the key causal variable—the tier of their law school—and then compares bar passage rates. Figure 1, reprinted from Ho's article, displays the results.

Estimated causal effect on
probability of passing bar
of attending:

	White students	Black students
First tier versus second tier	(n = 3681)	(n = 48)
Second tier versus third tier	(n = 7825)	(n = 181)
Third tier versus fourth tier	(n = 11,177)	(n = 247)
Fourth tier versus fifth tier	(n = 6637)	(n = 46)
Fifth tier versus HBCU	(n = 546)	(n = 29)

-0.4   -0.2   0.0   0.2   -0.4   -0.2   0.0   0.2
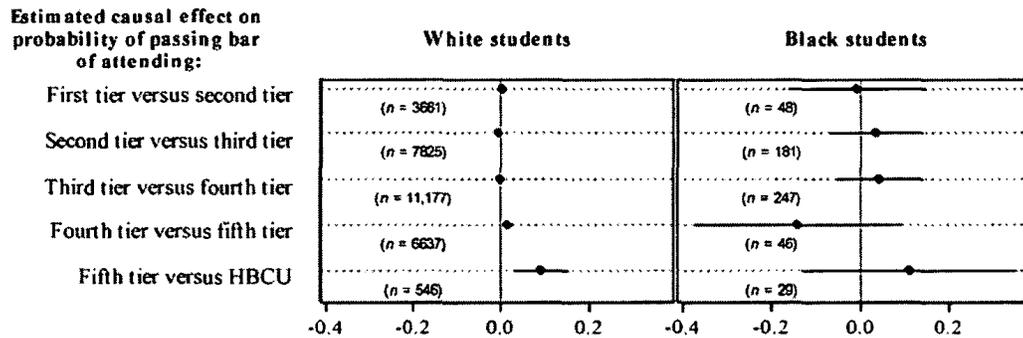
Figure 1: Daniel E. Ho's estimated causal effects on the probability of passing the bar for white and black students after attending different tiers of law schools. The horizontal lines represent 95% confidence intervals. All but one of these lines intersects with zero, indicating that the impact of school tier is not statistically significant.[3]

---

1.   Daniel E. Ho, Scholarship Comment, *Why Affirmative Action Does Not Cause Black Students to Fail the Bar*, 114 YALE L.J. 1997 (2005). Ho is responding to the controversial claims in Richard H. Sander, *A Systemic Analysis of Affirmative Action in American Law Schools*, 57 STAN. L. REV. 367 (2004).

2.   These observable variables include race, gender, LSAT score, and undergraduate GPA. Ho, *supra* note 1, at 1999.

3.   The figure appeared in Ho's work. *Id.* Ho utilizes data from Sander, *supra* note 1.

Ho's work has received no shortage of kudos,[4] but surely another is in order: the author knows how to communicate research results. From Figure 1, readers can easily grasp the study's key takeaway; namely, claims about the repercussions of affirmation action on bar passage rates are overblown. Similarly qualified black students, regardless of the tier of their law school, perform at the same (i.e., statistically indistinguishable) level.

Why Ho's graphic display is so powerful and, indeed, why it may help explain the impact of his article, is no mystery. First, while assessing the effect of school tier on bar performance required complex calculations, Ho deemphasizes them;[5] he instead focuses on communicating substance, not statistics. No one can look at Ho's figure and fail to see that all but one of the black circles and lines fall near zero (indicating no causal effect). Second, not only does the author well illustrate the substantive effect of school tier on bar passage rates, he also effectively conveys his uncertainty about that (non)effect. Because the dark horizontal lines (indicating 95% confidence intervals) intersect zero for all black students, we can safely conclude that the impact of tier is statistically indistinct from zero. A presentation depicting only the results and not Ho's uncertainty about them may well have led causal readers astray, especially about black students in the lowest tiers. Finally, we applaud Ho's use of a figure to convey his findings. Had he employed a tabular display, as do many scholars publishing in the law reviews, he would have missed an opportunity to present his results in the most accessible and powerful way possible.

In short, in conveying the findings of his important study, Ho followed the three key principles of effective communication:

1. Communicate Substance, not Statistics
2. When Performing Inference, Convey Uncertainty
3. Graph Data and Results

---

4. *See, e.g.*, Emily Bazelon, *Sanding Down Sander*, SLATE, Apr. 29, 2005, http://www.slate.com/id/2117745/ ("The forthcoming responses to Sander pounce on several of his moves (which they call causal inferences). To begin with, there is the problem of 'post-treatment bias,' which means that it's a bad idea to control for a factor that is itself a consequence of the cause you're studying. That no-no is explained by Daniel Ho . . ."); Vic Fleischer, *On Changing One's Mind*, A TAXING BLOG, May 9, 2005, http://vic.typepad.com/taxingblog/2005/05/on_changing _one.html ("Perhaps my initial agreement with Sander was in part out of an urge to defend him. In any event, I've changed my mind. Dan Ho's presentation changed my mind. Ask yourself— when was the last time an empirical paper changed your mind about an issue like affirmative action?").

5. Ho, *supra* note 1, at 2002 n.25.

Our earlier article, *On the Effective Communication of the Results of Empirical Studies, Part I* (hereinafter *Communication I*),[6] explores these principles in some detail. It also offers some general rules for creating visually effective displays of data.[7]

In other disciplines, adherence to these principles has generated benefits for the producers *and* consumers of empirical research, and we have no doubt that Law will see similarly salutary effects. Most crucially, as we explained in *Communication I*, moving towards more appropriate and accessible data presentations will enhance the impact of empirical legal scholarship—regardless of whether the intended audience consists of other scholars, students, policy makers, judges, or practicing attorneys.[8] At the same time, however, we realize that legal researchers require more than general guidelines; on-the-ground guidance may prove even more valuable for those who have carefully designed and executed their studies, and now must convey the fruits of their labor to their colleagues in the academy, to lay groups, or to both. Hence, in this second and final part in our series, we aim to get far more specific, offering analysts advice on how to translate their data (Part II) and results (Part III) into powerful visual presentations.

In setting out the various strategies to follow, we adhere to the general principles laid out in the earlier article,[9] but none more so than the very basic idea that researchers should almost always graph their data and results. Along these lines, we agree with Gelman and his colleagues: Unless the author has a very compelling reason to provide precise numbers to readers, a well-designed graph is a

---

6. Lee Epstein, Andrew D. Martin, & Matthew M. Schneider, *On the Effective Communication of the Results of Empirical Studies, Part I*, 59 VAND. L. REV. 1811 (2006) [hereinafter *Communication I*].

7. To wit:

Aim for Clarity and Impact

Iterate

Write Detailed Captions

Id. at 1845.

8. *Id.* at 1814; *see also* Gary King, Michael Tomz, & Jason Wittenburg, *Making the Most of Statistical Analyses: Improving Interpretation and Presentation*, 44 AM. J. POL. SCI. 347, 360 (2000) (arguing that such attention to interpretation and presentation "could help bridge the acrimonious and regrettable chasm that often separates quantitative and nonquantitative scholars, and make the fruits of statistical research accessible to all who have a substantive interest in the issue under study"). Much of the inspiration for our series (especially *infra* Part III) comes from this article.

9. We also assume that readers of this piece have at least skimmed *Communication I*, *supra* note 6. Accordingly, we do not reiterate, e.g., the basics of good graphic construction, among other topics, here.

superior choice to a table.[10] To put it another way, with only limited exceptions, we interpret the phrase "effective communication" in our title to mean "effective graphical presentations."

Just one final note of introduction: to be sure, our primary audience is the empirical legal scholar hoping to communicate her research to academics and the public, but it is not only the empiricist that we aim to reach. Because our goal here, as it was in *Communication I*, is nothing short of establishing a new norm in the presentation of empirical legal scholarship, we hope to enlist the entire legal community in our project. This should not be a difficult, as judges, policy makers, lawyers, academics, and students—the consumers of data work—have as much to gain as the producers from more insightful and accessible presentations. Nonetheless, to advance our goal, as well as to reinforce the basic lessons of our series, we supply, in Part IV, a set of guidelines for the communication and evaluation of data and results. We direct these suggestions primarily at those ideally situated to help elevate the quality of empirical work—journal editors. But we also hope that these proposals will prove valuable to others in the legal community who wish to become more informed evaluators of the data work now flooding the law reviews.

II. COMMUNICATING DATA

Scholars conducting empirical work generally seek to communicate two features of their research: the *data* they have collected and the *results* yielded by their analyses. If the researchers' sole goal is describing the information they have collected, then only the first, descriptions or summaries of the data, will come into play. More typically though, summarizing data is merely a prelude to drawing inferences, that is, to using observations the researcher has collected—her *sample*—to generalize about observations she has not collected—the *population* of interest.[11] While some studies stop at descriptive inference,[12] most studies aim to make claims that are causal in nature. For example, many studies deploy statistical procedures (e.g. regression analysis) to determine whether one or

---

10. *See generally* Andrew Gelman et al., *Let's Practice What We Preach: Turning Tables into Graphs*, 56 AM. STATISTICIAN 121 (2002). *See also Communication I, supra* note 6, at 1842-43 n.15.

11. For discussions of inference, see generally, e.g., *Communication I, supra* note 6; Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1 (2002).

12. *See* Epstein & King, *supra* note 11, at 29 ("[D]escriptive inferences are different than data summaries. We do not make them by summarizing facts; we make them by using facts we know to learn about facts we do not observe.")

more factors lead to (or cause) a particular outcome.[13] When conducting inferential analyses of these sorts, researchers will always communicate the results their methods yield; they will also frequently convey information about the data used in their procedures.

Daniel Schneider's analysis of the effect of appellate court judges' background characteristics on their decisions in tax cases is illustrative.[14] From social science theories of judging, Schneider develops several empirical implications about the relationship between background characteristics and outcomes; for example, he predicts that female judges and judges who are new to the bench will be more likely to rule in favor of taxpayers.

To assess these and other hypotheses, Professor Schneider drew a random sample of 416 federal tax decisions issued in the U.S. circuit courts between 1996-2000.[15] These 416 cases (and, more specifically, the 1295 judicial votes cast in them) were, in and of themselves, of little interest to Schneider. Rather, his ultimate objective was to use his sample to draw an inference about judging in all tax cases—an objective he intended to realize by evaluating the hypotheses of interest in multivariate statistical models. Nonetheless, prior to presenting the results of his statistical estimation, Schneider provided readers with information about the raw ingredients that went into the analysis—that is, about the data he had collected.[16] We learn, for example, that 82% of the 1295 votes were cast by male judges and 18% by females; that the number of years of service on the bench, on average, was twelve; and so on.[17]

Schneider's strategy of conveying information about the data he had collected, as well as the results of his statistical analysis is quite typical; it is also, we might add, quite appropriate. For readers to be able to evaluate the results of a statistical procedure, they require information about the data that went into producing those results.[18] What is less appropriate, however, and even problematic, is

---

13. To be clear, causal inference is the difference between two descriptive inferences. More specifically, a causal inference is the difference in the dependent variable between the situation where the treatment is applied and the situation where the control is applied. Different statistical models approach causal inference using varying modeling assumptions. *See, e.g.,* Epstein & King, *supra* note 11, at 36.

14. Daniel M. Schneider, *Using the Social Background Model to Explain Who Wins Federal Appellate Tax Decisions: Do Less Traditional Judges Favor the Taxpayer?,* 25 VA. TAX REV. 201 (2005).

15. *Id.* at 211, 221.

16. *Id.* at 221-22.

17. For more on Schneider's data, see *infra* Table 1.

18. For more on this point, see, e.g., *Communication I, supra* note 6, at 1819-21; EDWARD R. TUFTE, THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION 168 (2nd ed. 2001).

the typical manner in which such information is presented. If our tour through the law reviews, and even refereed legal journals, is any indication, authors more often than not communicate features of their *data* via tables, not figures; and when they do use figures, their choices are not optimal either for them or their audience. How scholars communicate the *results* of their analyses is even more troublesome. Unlike Ho's article,[19] the authors' (usually tabular) displays contain slews of estimated "coefficients" that are not only meaningless to virtually all of their readers but to themselves as well. Rarely do empirical legal researchers provide information about the substantive effects of their results (Ho's recent article is the exception, not the rule); and even more rarely do authors create a visual representation of those effects in a form that readers can easily grasp.

In the sections to follow, we offer some correctives. Specifically, in what directly follows in this Part we focus on communicating data; in Part III, we take up the presentation of results. We divide the material in this way because the presentation of data and of results are somewhat different tasks and are governed, to some extent, by distinct rules.[20] For example, as we discussed in *Communicating I*, when performing inference, authors have an obligation to convey the level of uncertainty about their results—as did Ho.[21] But when researchers are merely displaying or describing the data they have collected—and not using their sample to draw inferences about the population that may have generated the data—supplying measures of uncertainty, such as confidence intervals may be overkill.[22] On the other hand, reflecting our view that, for the purpose of communication, graphs are superior to tables, we generally focus both discussions on visualization via pictures—meaning that all the general principles we outlined in *Communication I* are operative here.[23]

With that cautionary note in mind, let us turn to the presentation of data, specifically to prescriptions for effectively visualizing (A) one variable and (B) the relationship between two or more variables.

---

19. Ho, *supra* note 1.

20. As we emphasize throughout this Article, there are different rules for describing data collected versus performing inference.

21. *See* Ho, *supra* note 1, at 2003 fig.1 (providing 95% confidence intervals for all estimated effects).

22. *See Communication I*, *supra* note 6, at 1838 n.72 (noting Gelman's apparent disagreement).

23. That is, whether presenting data or results, researchers must aim for clarity *and* impact, employ iterative efforts to improve visualization and craft detailed captions. *Id.* at 1811.

A. The One-Variable Case

The building blocks of most empirical analyses are variables—i.e., characteristics of some phenomenon that vary across instances of the phenomenon. In Ho's study, for example, bar passage is a variable that can take on one of two values: a student can pass or fail. In Schneider's data set, seniority on the bench varies, from less than one year to over forty. Gender, too, is among Schneider's variables: a judge is either a male or a female.[24] For purposes of designing their research projects, scholars tend to differentiate between *dependent* variables—the outcomes or responses the researcher is trying to explain—and *independent* variables—the factors that may help account for or explain the outcome. In Schneider's analysis, for example, seniority on the bench is an independent variable, which he expects to affect the outcome of tax cases, the dependent variable.

When researchers go about the twin tasks of analyzing and presenting data, another distinction between variables is equally important: quantitative (or numerical) versus qualitative (or categorical) variables. Schneider's study houses examples of both. Because it is numerical, his seniority variable—"years on the bench"—is quantitative.[25] To the extent that we can categorize judges with a descriptor—whether they are male or female—or differentiate them on the basis of this quality, gender is a qualitative and not quantitative variable. Indeed, while we could assign the number "1" to male judges and "2" to female judges, unless one believes that females are twice as good as males, those numbers associated with each category have no intrinsic meaning.[26]

Any scholar who has performed inference understands the distinction between quantitative and qualitative variables; it is fundamental to selecting the appropriate statistical model for analysis.[27] It is also crucial for selecting the appropriate tool for purposes of presentation—so much so that we divide the material to follow on this basis.

---

24. Schneider, *supra* note 14, at 213 n.36, 216 n.42.

25. Quantitative variables come in two varieties: those that can only take on a limited, or finite, number of values are discrete; and those that can be any possible number are continuous. *See* ALAN AGRESTI & BARBARA FINLAY, STATISTICAL METHODS FOR THE SOCIAL SCIENCES 16 (3d ed. 1997).

26. Categorical variables can be ranked (e.g. interval and ordinal variables) or unranked (e.g. nominal variables). *See id.*

27. As an illustration, if a dependent variable is quantitative, oftentimes a linear regression model is appropriate. If, however, a dependent variable is dichotomous, a logistic regression model would usually be appropriate. *See* J. SCOTT LONG, REGRESSION MODELS FOR CATEGORICAL AND LIMITED DEPENDENT VARIABLES (1997).

1. Quantitative Variables: Eliminate Tables of Summary Statistics

In an interesting study of judgments awarding attorneys' fees to the prevailing party, Michael Kao compares two continuous, quantitative variables—the hourly rate awarded in thirteen state and sixteen federal civil rights cases filed in California and terminated in 2000 or 2001.[28] Kao expects to find lower fees awarded in the federal cases but the data, he argues, reveal no meaningful differences between the two court systems. To shore up his claim, Kao deploys four different displays of the same data—two of which convey information about the individual observations: a table containing raw data and a univariate scatterplot (reproduced in Figure 2 below). The next two, a table of descriptive (summary) statistics and a three-dimensional bar chart, summarize the distribution of the data (see Figure 3).

We admire Kao's desire to be thorough, but we are troubled by his choices. Ironically enough, none of the four displays, taken individually or collectively, clearly conveys the researcher's primary message: that the structure of the two continuous variables (hourly rates awarded in the state and federal courts) is virtually indistinguishable.

Beginning with Kao's two attempts to convey each observation (case) in his data set, the first—the raw data table—is not simply unnecessary; it is distracting, even frustrating. While authors must make their data sets publicly available, and the law journals ought to ensure that they do,[29] analysts should avoid inserting them into the text of an article. As a general matter, raw data tables waste precious journal space and, worse still, they almost never serve the author's purpose: Even after careful study, most readers will be unable to discern patterns in Kao's state or federal cases, much less determine whether the patterns are similar or not. We simply cannot keep that many figures in our head, and the more observations in the study, the worse the problem grows.[30]

---

28.   Michael Kao, Comment, *Calculating Lawyers' Fees: Theory and Reality*, 51 UCLA L. REV. 825, 838 (2004).

29.   For recommendations on this point, see discussion *infra* Part IV.

30.   *See* WILLIAM G. JACOBY, STATISTICAL GRAPHICS FOR UNIVARIATE AND BIVARIATE DATA 47 (1997) ("[R]esearchers often have difficulty seeing the forest (i.e., a variable's distribution) because of the trees that it contains (i.e., the individual observations).").

STATE COURT			
Docket #	Fees Awarded	Hours	Hourly Rate
BC179938	$38,620.00	183.050	$210.98
BC181780	$80,000.00	235.720	$339.39
BC190354	$178,850.00	511.000	$350.00
BC1995383	$80,000.00	626.700	$127.65
BC196351	$800.00	3.070	$260.59
. . .			

FEDERAL COURT			
Docket #	Fees Awarded	Hours	Hourly Rate
98-10686	$119,206.19	308.57	$386.32
99-00720	$36,875.00	169.50	$217.55
98-05817	$485,822.27	2173.80	$223.49
99-08302	$21,335.00	39.30	$238.91
00-13363	$2,220.10	9.30	$238.91
. . .			
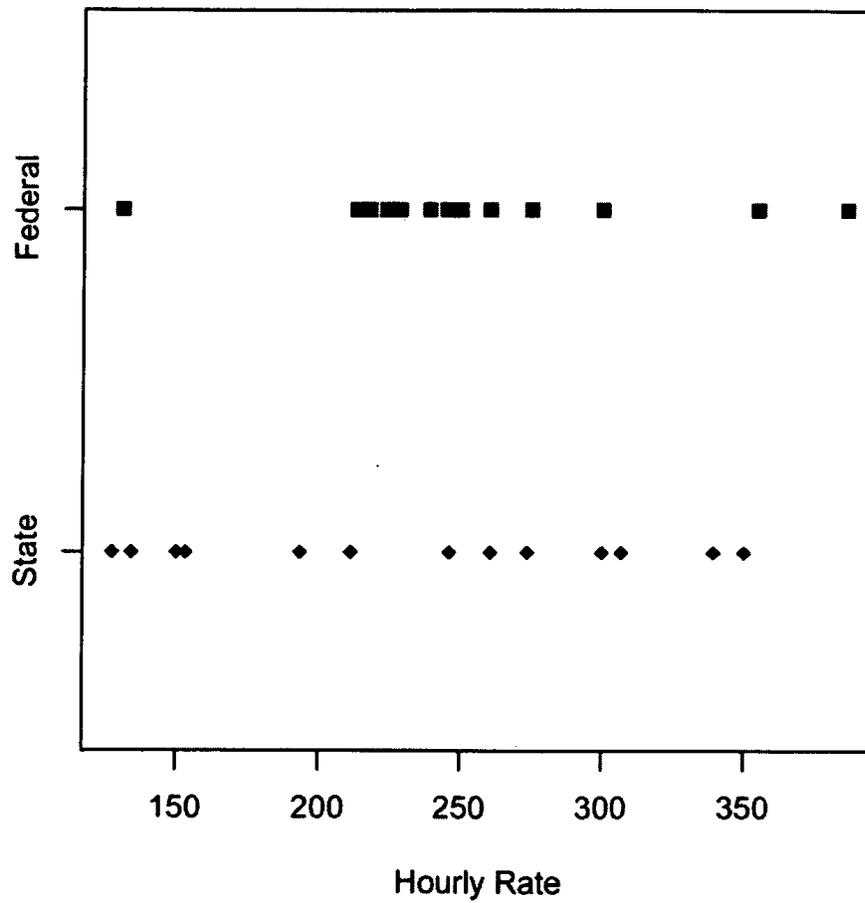
Figure 2: Michael Kao's data on attorney fees awarded in state and federal court cases. The panels on the top are partial reproductions of Kao's raw data on awarded fees. The panel on the bottom is Kao's univariate scatterplot representing the variance of awards in state and federal cases. The raw data tables make it difficult to decipher patterns, while the univariate scatterplot fails to capture the overall distribution of the state and federal awards.

With these words, we do not mean to pick on Kao. He is hardly the only law review author to violate the general principle of jettisoning raw data tables. But in only a very limited number of instances are those violations justifiable—chiefly, when the goal is to provide interesting substantive information to the readers or to facilitate the detection of the individual data points.[31] Kao's observations meet neither of these conditions.

---

31. Illustrative is Guhan Subramanian's study, *The Influence of Antitakeover Statutes on Incorporation Choice: Evidence on the "Race" Debate and Antitakeover Overreaching*, 150 U. PA. L. REV. 1795 (2002), which sought to join the "race to the top/bottom" debate by exploring whether managers migrate to states with anti-takeover statutes in place at the time of their decision to incorporate. As part of his demonstration that "bottom" proponents have the better argument, he presents and labels the measurements of a single continuous variable: the number of companies incorporating in a number of states. *Id.* at 1815 fig.2. In the left panel of the figure below we reproduce his display. *Id.*

Incorporation State [% of Number of Companies]
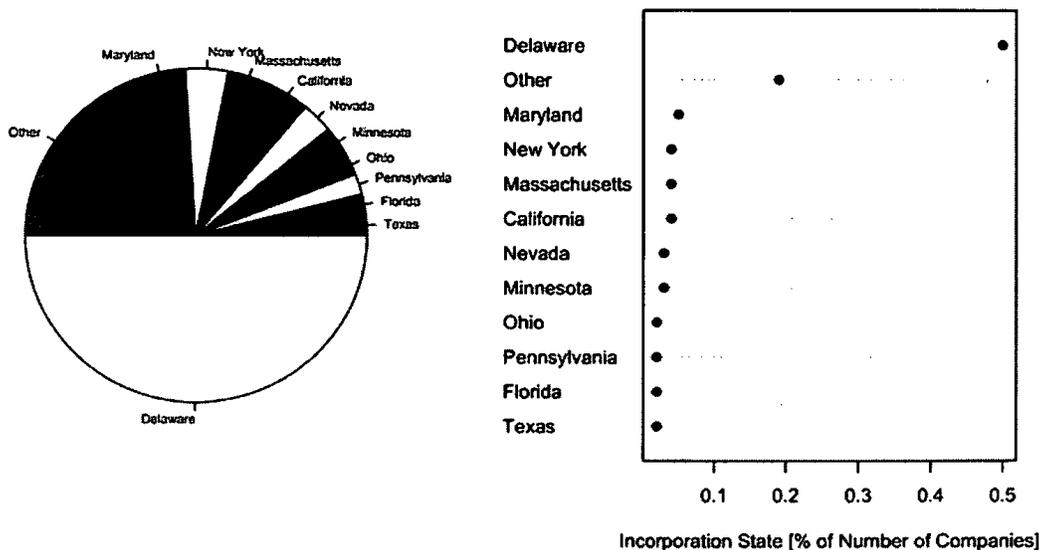
Unlike Kao's inclination to provide information on and label each case in his study, Subramanian's strikes us as entirely reasonable: the observations are small in number, familiar, and of clear substantive interest to participants in the debate he seeks to engage. Moreover, for all the reasons we discussed in *Communication I, supra* note 6, Subramanian shows good sense in graphing the data rather than presenting it in tabular form, as did Kao.

On the other hand, and again for the reasons we offered in the earlier paper, we would draw a line at his use of pie charts. These "pop" displays are never a good choice, and here the chart is particularly problematic. Subramanian's figure obscures the data, making visualization difficult—perhaps even more difficult than a tabular display. Subramanian, *supra*, at 1815 fig.2. Far better for purposes of decoding, as Cleveland demonstrates, is the dot chart, located in the right panel. *See* WILLIAM S. CLEVELAND, THE ELEMENTS OF GRAPHING DATA 262-63 (2d ed. 1994) ("[With a dot plot we] can effortlessly see a number of properties of the data that are either not apparent at all in the pie chart or that are just barely noticeable"); *see also* William S. Cleveland & Robert McGill, *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods*, 79 J. AM. STAT. ASS'N 531, 545 (1984) ("A pie chart can always be replaced by a bar chart .... [But] we prefer dot charts . . . ."). Indeed, we strongly recommend dot charts for those rather rare circumstances in which labeling the measurements

Better, but only marginally so, is Kao's other attempt to display his individual data points: the univariate scatter plot, which represents each observation as a point across the range of the variable(s) of interest. For Kao, as we can see in Figure 2, the observations are federal and state court cases, and the variables of interest are the hourly rates awarded in each.

Again, we applaud Kao's intuition here; namely, if the goal is to enable readers to *detect* information about each measurement, a univariate scatterplot can be an appropriate and valuable tool. The problems here are twofold. First, assuming Kao's goal is detection, the plot does not serve him well. Because of the relatively large number of cases (at least for a univariate scatter), and because the hourly wages for many are identical or nearly so, the individual data points are obscured. Overplotting of this sort can be reduced through a technique called jittering[32] or even by using different plotting symbols but it is difficult to eliminate entirely. Perhaps this explains why univariate scatterplots[33] are relatively rare.[34]

This brings us to a second problem with Kao's presentation: Like most researchers, Kao seems less interested in conveying the trees of his study (i.e., hourly rates in particular state and federal cases) than the forest (i.e. the distribution of hourly award rates by court type). If this is the goal, then univariate scatterplots, to continue the metaphor, can prevent readers from seeing the forest through the trees.[35] From Kao's scatterplot we get a far better feel for the relatively uninteresting individual observations than for the structure of the variables of interest, not to mention the comparison he wishes us to draw between them.

of a quantitative variable is desirable—that is, those circumstances presented by the Subramanian's study: a manageable number of substantively interesting cases. For more on these circumstances, see JACOBY, *supra* note 30, at 50 ("[D]ot plots are particularly suitable for *detection* or for the ability to discern individual data points in the graph.").

32. First developed in J.M. CHAMBERS ET AL., GRAPHICAL METHODS FOR DATA ANALYSIS (1983) 20-21, jittering helps to separate points in a univariate scatterplot by adding (or subtracting) a small amount to their value in order to set them off from other data points and thus aid in visual inspection. *See also* CLEVELAND, *supra* note 31, at 158 (defining jittering as "adding a small amount of random uniform noise to the data before graphing"); JACOBY, *supra* note 30, at 31 (describing jittering as the process of "displacing the points somewhat in the direction perpendicular to the variable's scale line"); Richard A. Becker & William S. Cleveland, *Brushing Scatterplots*, 29 TECHNOMETRICS 127, 134 (1987) (explaining that jittering is used to "alleviate overlap).

33. Bivariate scatterplots are a different matter altogether. See *infra* Part II.C for more on the use of these plots.

34. *See* JACOBY, *supra* note 30, at 32 ("When the number of observations is large, there will still be quite a bit of overplotting despite the jittering. Therefore, unidimensional scatterplots remain primarily useful for small data sets.").

35. This metaphor is borrowed from Jacoby. *Id.* at 47, 50.

Kao apparently appreciates the problem, and seeks to remedy it with two additional displays (see Figure 3), neither of which conveys much more information than the others. We need not say too much about the 3-D plot; we railed against this type of "pop chart" in *Communication I*, and here, the situation is compounded because the figure duplicates information listed in the table of descriptive statistics.

Court	Mean	Median	Standard Deviation	Variance
State	$234.33	$246.16	$78.34	$6,137.14
Federal	$253.44	$245.06	$58.11	$3,376.57
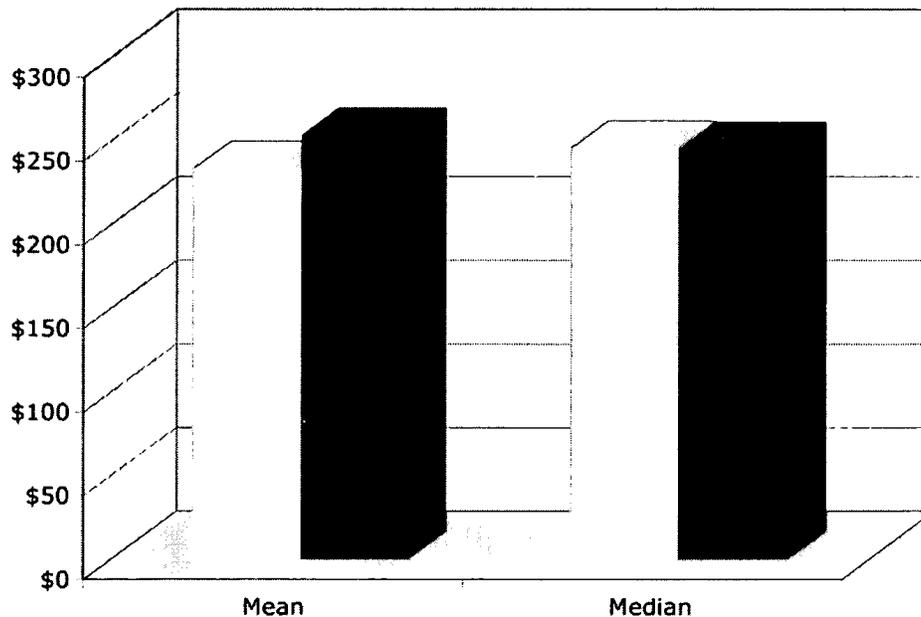
Figure 3: Descriptive statistics table and figure from Kao's study of awarded attorney fees. The top panel displays the descriptive statistics table from Kao's article while the bottom panel houses a reproduction of his 3-D plot. Kao aim is to provide summary information about the composition of the variables of interest through descriptive statistics, but raw numbers do not well serve his goals. The figure on the bottom is a better idea, but a 3-D plot that is distracting in its design and does not account for the distribution of the variables is not the best choice.

What of that table of descriptive statistics, a type of table that has become so standard in legal publications that virtually all articles with quantitative variables house one? Surely, by providing the precise value of the mean and median (measures of central tendency) and the standard deviation (a measure of dispersion), the author's purpose is to convey "useful" information about the structure of a

continuous variable(s).[36] This technique is an end, of course, that raw data tables or even graphical displays of each observation cannot reach, especially when the number of observations is large. *But because tables of descriptive statistics sacrifice visual clarity for the sake of artificial precision, they almost never meet that objective either.* Actually, if the goal is to convey information about a variable's structure—including its center, spread, and shape—as it almost always is, we strongly advise eradicating summary tables and replacing them with appropriate graphical displays.

Let us elaborate, beginning with means and medians. For continuous, quantitative variables, these measures of central tendency tell us about the "center" of the distribution (in Kao's case, there are two distributions, hourly rates awarded in federal civil rights cases and hourly rates awarded in state civil rights cases). This is important information, to be sure, but precise values are often unnecessary and, more problematically, can obscure the message the authors seek to convey. In most cases, researchers can make their point far more accessibly, powerfully, and nearly as easily with a figure. Several possibilities come to mind, though the *boxplot* is an excellent and time-tested option,[37] especially when analysts such as Kao hope to draw attention to a comparison between two or more continuous variables.

---

36. The mean is "the simple average," the median is "the middle of the distribution of cases," and the standard deviation is a measure of spread or dispersion of the data. Epstein & King, *supra* note 11, at 25-26; *see also* AGRESTI & FINLAY, *supra* note 25, at 45-58 (describing the mean and median).

37. The boxplot was developed decades ago by John W. Tukey, a giant in the field of scientific graphing. *See* JOHN W. TUKEY, EXPLORATORY DATA ANALYSIS 39-43 (1977).
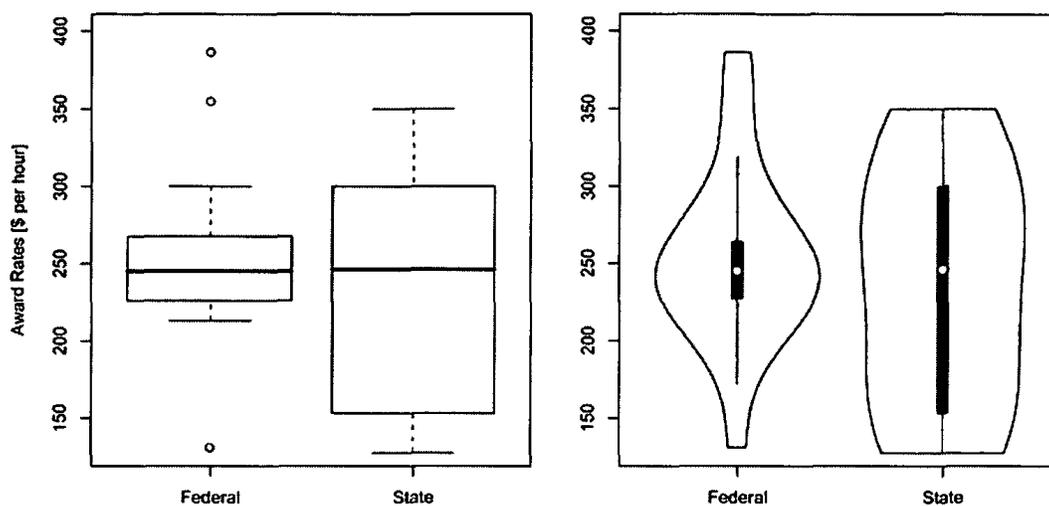
Figure 4: Kao's data (see Figures 2 and 3) presented in a more efficient and informative way, as box plots and violin plots. The box plots visually display the distribution of the award variable, with particular attention drawn to the median award, the interquartile range of the award, and any outliers. The violin plot provides similar information while conveying an even clearer picture of the shape of the variable's distribution. Here, it is easy to see that federal court awards are normally distributed while state court awards are far more uniformly distributed.

Why boxplots remain one of the most important and frequently used tools for data communication is no mystery: They are able to convey parsimoniously and clearly an enormous amount of information about the distribution of a single variable(s). In a simple plot, as we show in Figure 4, researchers can visually depict not only the median but also the interquartile range, the minimum and maximum values, and any observations that are unusually large or small (i.e., the outliers).[38] In short, these representations communicate precisely the right information without losing much, if any detail.

But the proof is in the pudding, and Figure 4 provides just that.[39] There we have designed a boxplot from Kao's data. Note that

---

38. As represented by the boxplot in Figure 4, the interquartile range covers the data points between the 25th and 75th percentiles. In other words, the box covers the middle 50% of the data. The minimum and maximum values are the first and last values of the data when the observations have been sorted from smallest to largest. Outliers, as represented by circles in the Kao boxplot, are data points that are located further than 1.5 interquartile range units from the upper or lower quartile; i.e., a great distance from center of the distribution. See WILLIAM S. CLEVELAND, VISUALIZING DATA 25-26 (1993) and TUKEY, *supra* note 37, at 39-43, for more information on the boxplot and its components.

39. See *Communication I, supra* note 6, where we make a series of suggestions for ensuring that the many things that can go wrong with graphing data go right. Because of the detail supplied there, we do not here dwell on design details (e.g., the appropriate plotting symbols, etc.) for this or any other plot.

the comparison Kao wishes to draw between hourly rates in state and federal courts now just pops. The median lines are so close that they are virtually indistinguishable, while equally as noticeable, the interquartile range is larger for the state cases.

Just as boxplots, relative to tables of means and medians, enhance visualization of the center and spread of a distribution, graphs perform far better than precise values in conveying information about the shape of the data. Think about it this way: Most of us understand that if a variable is normally distributed (i.e., looks bell-shaped), 95% of the observations fall within two standard deviations of the mean. What we sometimes forget is that this rule-of-thumb is useful *only* when we know the variable is symmetric and bell-shaped. Otherwise, knowing the precise value of the standard deviation is not terribly valuable.

And therein emerges an enormous drawback of tables of descriptive statistics: They do not reveal whether the data are normally distributed. Only by inspecting the shape of a distribution can researchers and their readers know whether this condition holds. And only via a plot of the data can they conduct this inspection.

Tools for plotting distributions abound but two excellent possibilities are *violin plots* and *kernel density plots*. Neither has received much attention in legal journals.[40] They should.

Now widely used in statistics and gaining traction in the social sciences,[41] the violin plot is a modern-day variant of the traditional box plot.[42] As we show in Figure 4, it too provides information on the

---

40. A Lexis search of U.S. law reviews and journals (conducted January 15, 2007) uncovers no relevant results for a search of "violin" within the same paragraph as "plot" or "graph." A Lexis search of the same journals (conducted contemporaneously) turns up 10 articles utilizing kernel density plots in some fashion.

41. *See* J. L. Hintze & Ray D. Nelson, *Violin Plots: A Box Plot-Density Trace Synergism*, 52 AM. STATISTICIAN 181 (1998) (developing the violin plot); *see also* Andrew G. Bunn & Scott J. Goetz, *Trends in Satellite-Observed Circumpolar Photosynthetic Activity from 1982 to 2003: The Influence of Seasonality, Cover Type, and Vegetation Density*, 10 EARTH INTERACTIONS 1, 10 (2006) (using violin plots to show the distribution of slopes for models of major forest types and categories of low growing vegetation); M. Jorgensen & Dag I.K. Sjoberg, *Impact of Experience on Maintenance Skills*, 14 J. SOFTWARE MAINTENANCE & EVOLUTION 123, 131 (2002) (noting that "the violin plot highlights the peaks and valleys of a variable's distribution"); Thomas R. Steinheimer & Kenwood D. Scoggin, *Fate and Movement of Atrazine, Cyanazine, Metolachlor, and Selected Degradation Products in Water Resources of the Deep Loess Hills of Southwestern Iowa, USA*, 3 J. ENVTL. MONITORING 126, 128 (2001) (using a violin plot to reveal "an observed distribution of values above the minimum detection limit"); Kenneth L. Weiss et al., *Clinical Brain MR Imaging Prescriptions in Talairach Space: Technologist and Computer-Driven Methods*, 24 AM. J. NEURORADIOLOGY 922, 926 fig.6 (2003) (featuring a violin plot of prescription errors).

42. *See* Hintze & Nelson, *supra* note 41, at 181 ("The violin plot . . . synergistically combines the box plot and the density trace (or smoothed histogram) into a single display that reveals structure found within the data.").

center of the variables (as indicated by the hollow white circles). But it also relays crucial information—and information we cannot obtain from tables of descriptive statistics—about the shape of the two variables: federal hourly rates appear normally distributed, while state rates are more uniform. The substantive impact of this result is that federal courts are extremely likely to award hourly rates very near the median value of $245.06. State courts, on the other hand, are just as likely to award hourly fees of $300.00 or $175.00 per hour as they are to award fees of $246.16, the median value. Whether this result speaks to inconsistencies among state courts or about the cases heard in those courts, we cannot say. Beyond speculation is that Kao's conclusion—"[t]he empirical evidence seems to suggest the counterintuitive conclusion that approximately the same amount of fees are awarded under the two competing methods"—now seems less powerful.[43]

The kernel density plot is a modern-day incarnation of a tool that has received some attention among legal academics—the histogram. Histograms are graphs of continuous (or nearly continuous) observations grouped into a series of vertical bars along the range of a variable's values. Although they can provide useful information, histograms have a number of disadvantages, including their arbitrarily designated "bins" and the relatively random assignment of observations to those bins.[44] By smoothing over the distribution with a continuous function, kernel density plots can ameliorate some of these problems. They work by essentially shrinking the bin-width of a histogram, and then using pieces of continuous functions to create a single, smooth curve that characterizes the distribution of the variable of interest.[45]

---

43. Kao, *supra* note 28, at 843.

44. *See* CLEVELAND, *supra* note 38, at 8 (arguing that although the histogram is over a century old and is widely used, "maturity and ubiquity do not guarantee the efficacy of a tool"); JACOBY, *supra* note 30, at 13-17 (noting that the arbitrary designation of bins impacts that shape of the histogram, the assignment of the number of observations in each bin impacts the bumpiness of the distribution, and the very nature of assigning bins means that data are forced to be assigned to one group or another).

45. *See, e.g.*, B. W. SILVERMAN, DENSITY ESTIMATION FOR STATISTICS AND DATA ANALYSIS (1986) (reviewing various approaches to kernel density estimation).

Histogram   **Kernel Density**

Seniority, Years [N = 1290]   Seniority, Years [N = 1290]
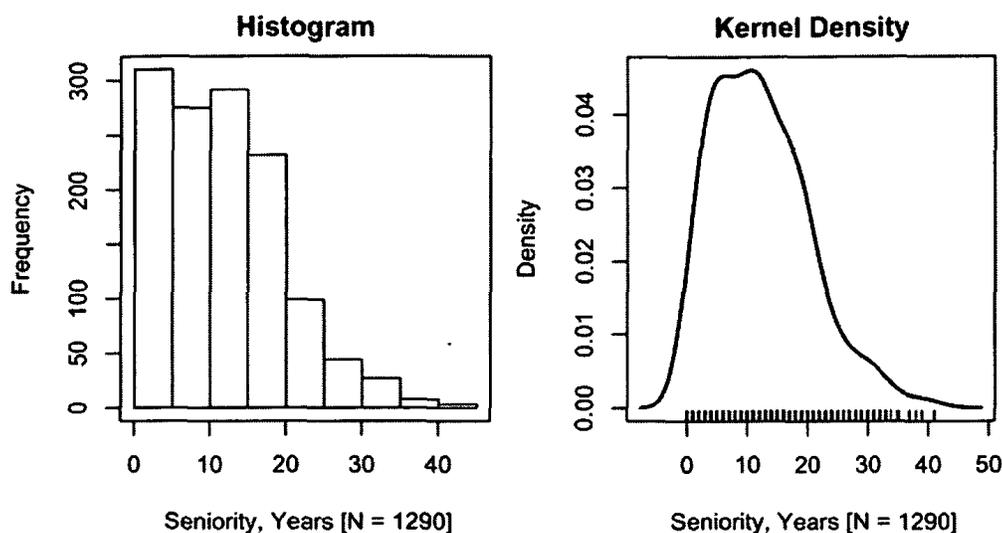
Figure 5: A histogram and a kernel density plot of Schneider's seniority variable. The histogram provides the reader with a richer understanding of the seniority variable and its distribution than does Schneider's descriptive statistics table. Arguably, the kernel density plot does an even better job because it makes the existence and location of the positive skew more apparent.

To provide an example, reconsider Schneider's article on judging in the tax context. Like Kao, Schneider provides a table of descriptive statistics housing the mean and standard deviation for his continuous variables, eliteness of undergraduate institution attended (mean=62; std. deviation=9) and seniority (mean=12; std dev.=8). And, as in the Kao study, we learn very little about the behavior of these variables from the precise figures in the table. By turning to visual displays that account for the distribution of continuous variables, we can remedy this deficit. In Figure 5 we take this step, creating a histogram and kernel density plot for one of Schneider's continuous variables: seniority. The histogram provides some help in understanding the distribution of these data. Even better, however, is the kernel density plot. As might be expected, seniority is positively skewed. The kernel density suggests that the mass of the distribution falls between five and fifteen years; it is more difficult to make that judgment from the histogram (and it is impossible to do so from the table of descriptive statistics).

After perusing the graphs in Figures 4 and 5 we hope readers can now understand why we so strongly recommend jettisoning tables of summary statistics: it seems to us nearly impossible to conclude that they are superior or even equal to visual depictions of a variable. Graphs have the advantage of communicating far more information— and far more useful information—parsimoniously, clearly, and powerfully. If the unusual circumstance arises and more precision is

needed, exact numbers are easy enough to present visually[46] or can appear in the caption.

2. Qualitative Variables: Jettison the Frequency Tables

Qualitative variables abound in the law literature. Race and gender occasionally figure into studies of criminal law.[47] Research on judging, regardless of the substantive context, often attends to the party affiliation of judicial appointees, or the political official appointing them.[48] And the method for disposing of a case, whether by settlement, non-trial adjudication, or trial, comes into play in many important studies of the litigation process.[49] While most researchers

---

46. *See, e.g., supra* fig.4, where we emphasize the median of the data.

47. *See, e.g.*, Marina Angel, *Criminal Law and Women: Giving the Abused Woman Who Kills a Jury of Her Peers Who Appreciate Trifles*, 33 AM. CRIM. L. REV. 229 (1996) (tracing historically the perspective of women in the political and judicial systems); Devon W. Carbado, *(E)Racing the Fourth Amendment*, 100 MICH. L. REV. 946 (2002) (arguing courts must recognize certain "racial realities" for minorities to receive full Fourth Amendment protection); R. A. Lenhardt, *Understanding the Mark: Race, Stigma, and Equality in Context*, 79 N.Y.U. L. REV. 803 (2004) (arguing courts should be sensitive to racial stigma); James S. Liebman, *Slow Dancing with Death: The Supreme Court and Capital Punishment, 1963-2006*, 107 COLUM. L. REV. 1 (2007) (discussing capital punishment's constitutionality); Erik Luna, *Race, Crime, and Institutional Design*, 66 LAW & CONTEMP. PROBS. 183 (2003) (surveying minority representation in the criminal process); Carolyn B. Ramsey, *Intimate Homicide: Gender and Crime Control, 1880-1920*, 77 U. COLO. L. REV. 101 (2006) (contrasting the degree of punishment between men and women convicted for killing their significant others); Laura E. Reece, *Women's Defenses to Criminal Homicide and the Right to Effective Assistance of Counsel: The Need for Relocation of Difference*, 1 UCLA WOMEN'S L.J. 53 (1991) (suggesting criminal defendants' differing perspectives be integrated into the substantive law); Victor L. Streib, *Gendering the Death Penalty: Countering Sex Bias in a Masculine Sanctuary*, 63 OHIO ST. L.J. 443 (2002) (examining capital punishment through sexual bias analysis).

48. *See, e.g.*, Stephen J. Choi & G. Mitu Gulati, *Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance*, 78 S. CAL. L. REV. 23 (2004) (creating a "tournament" judging the merits of potential Supreme Court nominees); Thomas J. Miles & Cass R. Sunstein, *Do Judges Make Regulatory Policy? An Empirical Investigation of Chevron*, 73 U. CHI. L. REV. 823 (2006) (examining policy judgments made by courts post-*Chevron*); Max Schanzenbach, *Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics*, 34 J. LEGAL STUD. 57 (2005) (analyzing the effect of jurisprudential characteristics on federal criminal sentencing); Emerson H. Tiller & Frank B. Cross, *A Modest Proposal for Improving American Justice*, 99 COLUM. L. REV. 215 (1999) (asserting that acknowledging the partisan component of judging would improve the federal judicial system).

49. *See, e.g.*, Orley Ashenfelter, Theodore Eisenberg & Stewart J. Schwab, *Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes*, 24 J. LEGAL STUD. 257 (1995) (examining jurisprudential predilection's effect on case outcome); James S. Kakalik et al., *Just, Speedy, and Inexpensive? An Evaluation of Judicial Case Management Under the Civil Justice Reform Act*, 49 ALA. L. REV. 17 (1997) (analyzing the CJRA's impact); Daniel Kessler, Thomas Meites & Geoffrey Miller, *Explaining Deviations from the Fifty-Percent Rule: A Multimodal Approach to the Selection of Cases for Litigation*, 25 J. LEGAL STUD. 233 (1996) (arguing that more reliable results are provided by looking beyond the party's simple divergent expectations in the selection of a case for litigation); Joel Waldfogel, *Reconciling Asymmetric Information and Divergent Expectations Theories of Litigation*, 41 J.L. & ECON. 451 (1998)

understand that conveying descriptive statistics for such variables (at least those with more than two categories) is uninformative, they have developed equally uninformative ways to convey the composition of those variables. Especially predominant in the law reviews are frequency, or one-way, tables that depict the number (and, typically, the percentage) of observations falling into each category of the variable. In Schneider's research on the background of federal judges, for example, he provides a table (part of which we reproduce in Table 1) showing the percentages and numbers of the judges in his dataset who are male and female; white, black, Latino, and Asian; attended elite law schools; and so on.

Discrete Variable Characteristic	Breakdown	Percentage
Gender (N=1295)	Male	82%
	Female	18%
Race (N=1290)	White	90%
	Black	5%
	Latino	4%
	Asian	1%
Eliteness of law school (N=1290)	Non-elite	52%
	Elite	48%
Prior work experience (N=1289)	Private practice	62%
	Judge	18%
	Government	10%
	Law School Professor	10%
Religious affiliation (N=966)	Protestant	57%
	Catholic	27%
	Jewish	15%
	Other	1%
Political affiliation of appointing President (N=1290)	Republican	58%
	Democrat	42%

Table 1: Descriptive statistics table from Daniel M. Schneider's study of the effect of judge background characteristics on tax case outcomes. Although the frequencies in the table provide details on the individual variables, it is unlikely that readers can quickly process the information.

To be sure, this table communicates useful information. But is Table 1—or, rather, frequency tables more generally—the best way to convey this information? If the purpose is to provide readers with the precise figures, then the answer is yes: frequency tables always trump graphs. Figure 1, which houses dot plots of Schneider's variables, underscores this point. While we can observe from the table that

---

(concluding that pretrial adjudication and settlement cause plaintiff win rates that tend toward central, rather than extreme, results).

exactly 48% of the judges attended elite law schools, we cannot make that observation with the same degree of precision from the figure.

More often than not, though, as we have stressed throughout, the degree of precision that frequency tables can convey is irrelevant. Typically what we want to communicate to our audience (and to ourselves) are comparisons, patterns, or trends, not exact values. Schneider's work is no exception. What he apparently wants us to take away from Table 1 is not the exact percentages of males and females, or Democrats and Republicans, or Protestants, Catholics and Jews in his data set but rather a sense of their relative proportions. Even for variables with fewer than three categories, Figure 6 better serves this purpose than Table 1.
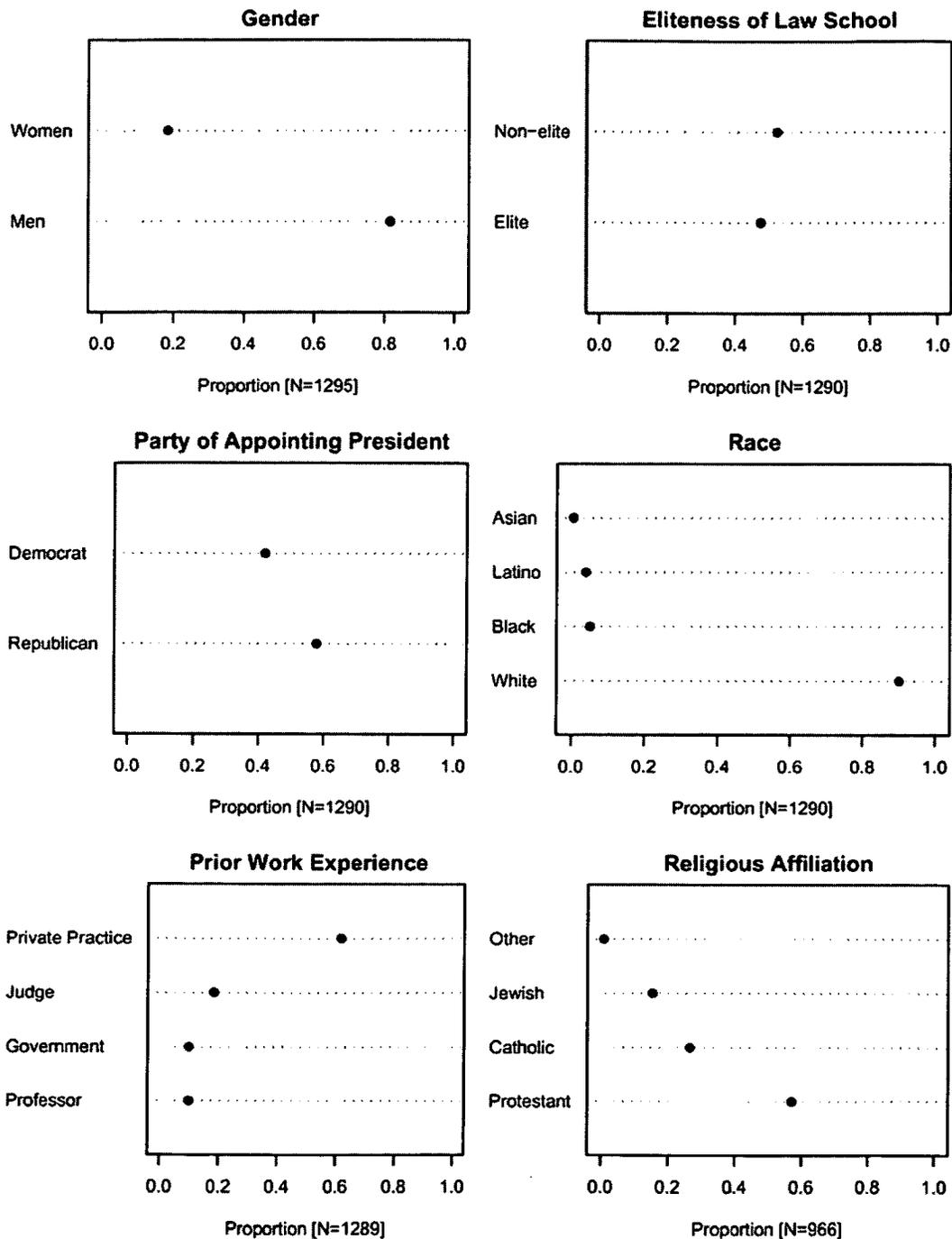
Figure 6: Juxtaposed against Schneider's table of descriptive statistics (see Table 1), the individual dot plots above provide a more visually and cognitively appealing solution to the problem of providing the reader with information about the composition of individual variables in a dataset.

A comparison of Schneider's data table with the dot plot of religious affiliation in Figure 1 clarifies this point. Surely if we stared

at the numbers long enough, we could observe the patterns that emerge from the graph—e.g., the comparative equivalence of Jewish and Catholic judges in his data base, not to mention the gap between the latter and Protestants. But it requires far more (unnecessary) cognitive work.

B. The Relationship Between Two or More Variables

In Schneider's study, conveying information about individual variables was a prelude to multivariate analyses designed to reach an inference about what causes judges to rule for or against taxpayers. This is not unusual. Unless the author's sole goal is to showcase particular variables in her sample, univariate displays are almost always just the first step toward the larger goal of inference.

Not so of analyses of the relationship between two or more variables. While some are certainly in the Schneider mold,[50] the authors' goals are more variegated. Take Gross and Barnes' paper on racial profiling in highway drug searches in Maryland between 1995-2000.[51] As part of their investigation, they present data, some of which is reproduced in the top panel of Figure 7, on the percentage of searches per year by the driver's race. They neither draw a statistical

---

50. Take Epstein and Segal's study of Senate confirmations of Supreme Court justices, in which they present a table of the relationship between a nominee's qualifications and ideology and the number of votes he received in the Senate. LEE EPSTEIN & JEFFREY A. SEGAL, ADVICE AND CONSENT: THE POLITICS OF JUDICIAL APPOINTMENTS 114 fig.4 (2005). The table, reproduced below, shows both the percentage of votes cast in favor of the nominee (the top number) and the total number of votes in that category (the lower number). Epstein and Segal use the table not to reach a causal inference about the effect of qualifications/ideology on votes—they realized that many other factors influence votes—but rather to communicate to readers the plausibility of such a relationship. With that demonstration in hand, they eventually moved toward a more sophisticated analysis containing a variable for qualifications, along with many others, designed to draw causal inferences.

Senate Voting Over Supreme Court Nominees Since 1953

Qualifications of The Nominee	Ideological Distance Between Nominee and Senator		
	Ideologically Very Close	Average	Ideologically Very Distant
Highly qualified	99.3 (602)	97.3 (299)	94.8 (231)
Qualified	97.6 (422)	83.0 (317)	44.9 (187)
Not Qualified	91.8 (182)	38.5 (96)	1.7 (115)

51. Samuel R. Gross & Katherine Y. Barnes, *Road Work: Racial Profiling and Drug Interdiction on the Highway*, 101 MICH. L. REV. 651 (2002).

inference from the table about the effect of race on highway stops, nor do they use the data it contains in a subsequent multivariate analysis. Rather, their primary purpose, it appears, is to convey trends in searches in Maryland.

		1995	1996	1997	1998	1999	2000 (Jan-June)
Number of Searches		564	309	116	374	607	352
Searches By Race	White	20.7%	22.0%	39.7%	47.3%	39.9%	39.2%
	Black	74.5%	65.0%	53.5%	45.5%	54.7%	53.4%
	Hispanic	3.6%	9.7%	6.9%	6.1%	5.8%	6.3%
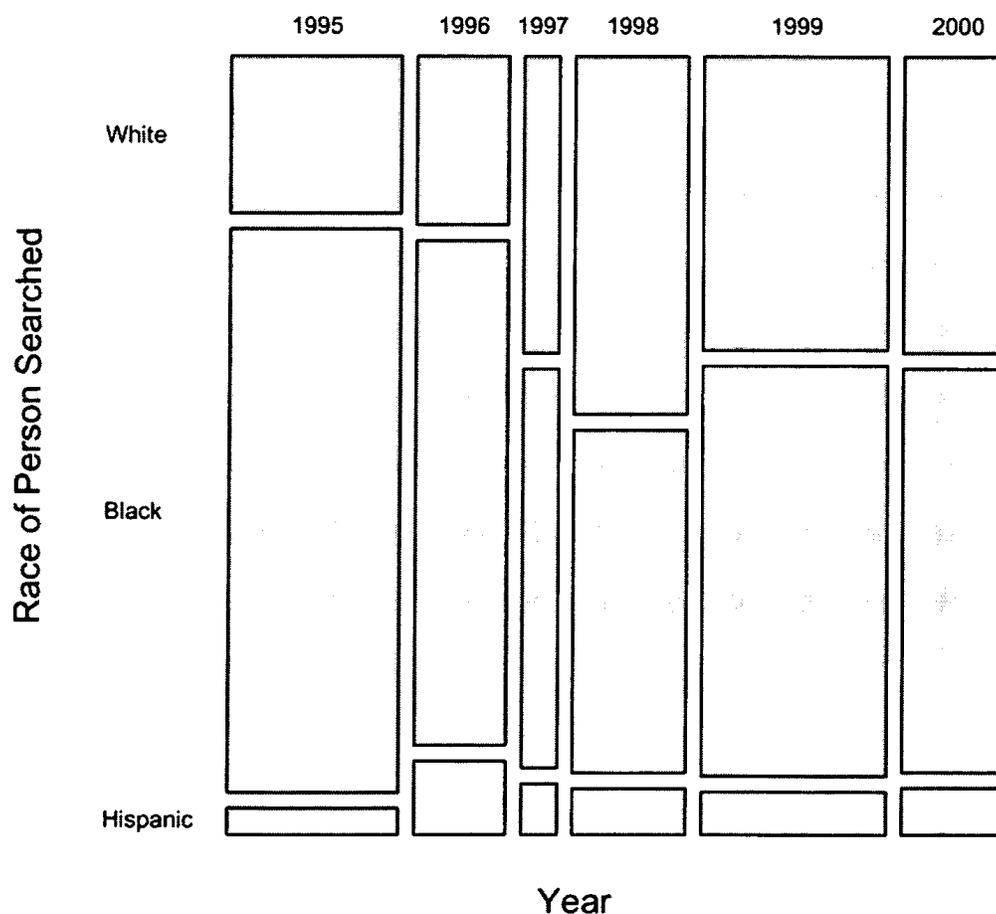
Figure 7: The table on the top is a partial replication of Table 23 in Samuel R. Gross and Katherine Y. Barnes' study on racial profiling on Maryland highways. The mosaic plot on the bottom presents the same data in a more concise and appealing fashion. The width of the bars depicts the number of searches per year while the height of each tile conveys the relative number of searches that are conducted on drivers of each race during each year. With this plot, it is much easier to see, for example, the large percentage of searches in 1996 that are of black drivers and how that percentage decreases sharply beginning in 1997.

We could say the same of Cumming and MacIntosh's study of how venture capitalists respond to the economic incentives in periods of boom and bust.[52] Among the researchers' arguments is that the existence of corruption during boom years (in their data set, 1999 and 2000) leads to considerable underpricing. To explore it, they present their raw data in tabular form (reprinted in the left of Figure 8). Again, the goal does not appear to be inference—the authors provide no statistics or measures of uncertainty—but rather to determine whether their argument and their data coincide.

Year	Millions of dollars of 2000 purchasing power	
	Market Price	Sales
1980	$183	$78
1981	$107	$54
1982	$118	$38
1983	$155	$86
1984	$85	$79
. . .		
1996	$365	$149
1997	$309	$167
1998	$600	$305
1999	$1,415	$343
2000	$1,528	$253
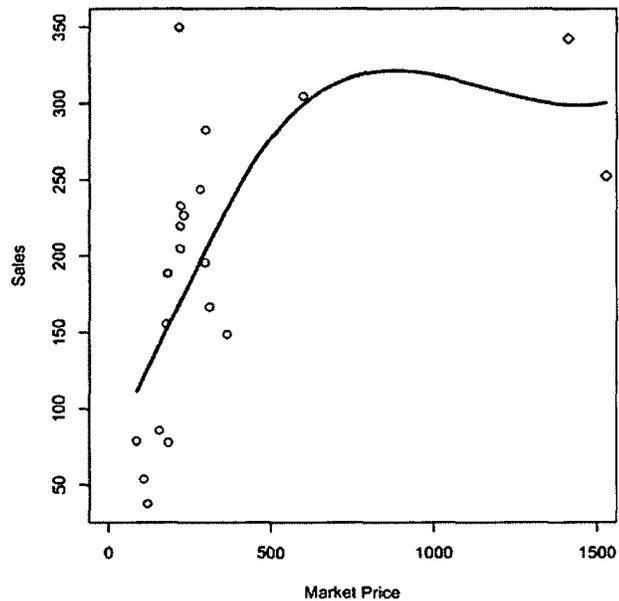
Figure 8: The market price and sales price for IPOs, by year. The left panel is a partial reproduction of Cumming and MacIntosh's raw data table. The right panel is a bivariate scatter plot of the full data set. The solid line is a smooth loess curve that summarizes the relationship between the market price and the sales price. Outlier points, indicated by the diamond shaped symbols, represent the data from 1999 and 2000, the two boom years in their study.[53]

---

52. Douglas Cumming & Jeffrey MacIntosh, *Boom, Bust, and Litigation in Venture Capital Finance,* 40 WILLAMETTE L. REV. 867 (2004).

53. These data appeared *id.* at 885 tbl.3. The table was reproduced by the authors from Tim Loughran & Jay R. Ritter, Why has IPO Underpricing Changed Over Time? (2003) (unpublished manuscript, on file with author).

Juliano and Schwab's study of federal sexual harassment cases is of a different order.[54] As even a mere glance at their table (reprinted in the top panel of Figure 9) would reveal (note, in particular, the chi-square statistic, along with a p-value), they are performing statistical inference. Here, the authors are using their data to learn about the association between their dependent variable, plaintiff success, and key independent variables: court type (district or appellate) and visibility (trial or not, published opinion or not).

Certainly the three studies depicted in Figures 7 through 9 investigate different actors—police, venture capitalists, and judges—and each deploys data for different reasons—to draw attention to trends and to make inferences. But they share two features: all seek to convey information about the relationship between two or more variables, and because they use tabular displays, all three fail to realize their potential to ensure successful decoding of the data by the reader.

| | DISTRICT COURT | | | | APPELLATE COURT | | | |
| | All Cases | | Cases w/ Trials | | All Cases | | Published Opinions | |
Circuit	% Pl Wins	# of Cases	% Pl Wins	# of Cases	% Pl Wins	# of Cases	% Pl Wins	# of Cases
1st	50.0	20	57.1	7	42.9	7	50.0	6
2d	50.0	62	38.1	21	80.0	5	80.0	5
3d	43.4	53	16.7	18	37.5	8	37.5	8
4th	38.5	39	33.3	12	21.1	19	30.0	10
5th	47.4	19	33.3	6	38.5	13	38.5	13
6th	58.3	24	55.6	9	28.0	25	44.4	9
7th	53.5	116	60.0	20	27.3	33	32.1	28
8th	50.0	30	50.0	18	72.7	11	72.7	11
9th	64.5	31	85.7	7	63.6	11	70.0	10
10th	50.0	64	47.1	17	33.3	24	31.6	19
11th	62.1	29	53.3	15	85.7	7	83.3	6
D.C.	53.3	15	28.6	7	0.0	1	0.0	1
Total	51.1	502	45.2	157	39.0	164	45.2	126
	chi2= 8.36	Pr.= 0.681	chi2= 15.94	Pr.= 0.143	chi2= 24.78	Pr.= 0.010	chi2= 17.41	Pr.= 0.096

Table 2: Plaintiff victory rates in sexual harassment cases across court type (district or appellate) and case type (trial or not, published or unpublished opinion) reproduced from Juliano and Schwab's Table 4. For a more effective way to present the data, see Figure 9 below.

---

54. Ann Juliano & Stewart J. Schwab, *The Sweep of Sexual Harassment Cases*, 86 CORNELL L. REV. 548 (2001).
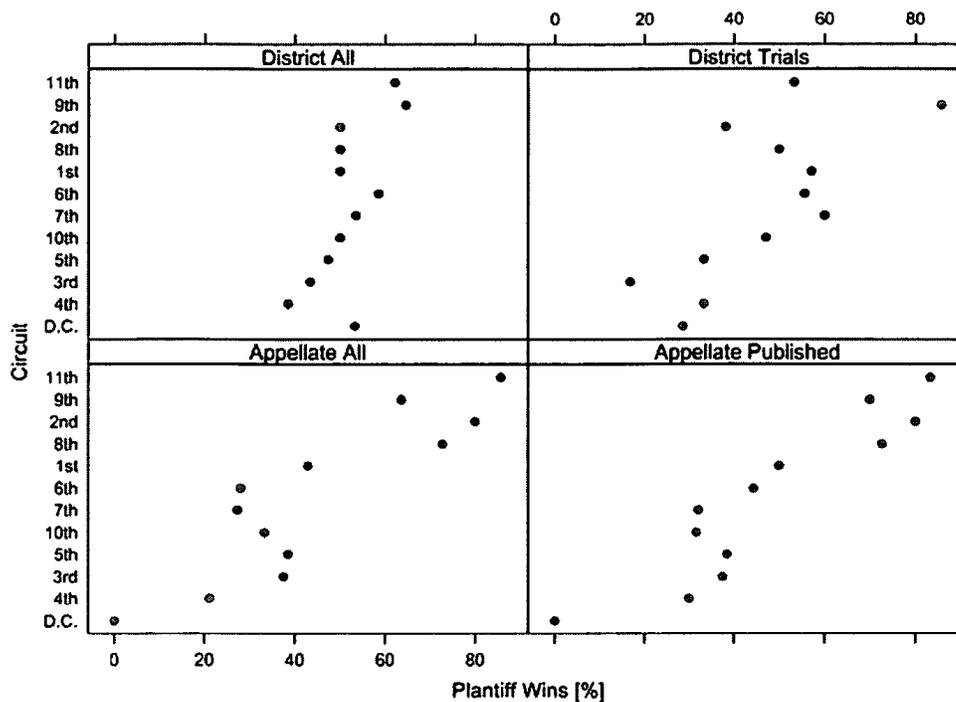
Figure 9: Plaintiff victory rates in sexual harassment cases across court type (district or appellate) and case type (trial or not, published or unpublished opinion). The raw data table, reproduced in Table 2, provides an overflow of information. These conditional dot plots provide an effective alternative, making it is easy to see at first glance that, e.g., D.C.'s circuit court has never ruled for the plaintiff in a sexual harassment case while the 11th Circuit often does.

To drive home this point, in Figures 7 through 9 we have converted the tables into plots. Note that the displays differ. This divergence, as we explain below, is completely appropriate given that the variables of interest are of different types: in the Gross and Barnes study, a *qualitative* variable; and in the case of Cummings and MacIntosh, *quantitative* variables. Juliano and Schwab's study mixes the two types but, because it analyzes the relationship between three variables, it presents something of a special case.

When an author seeks to present data over time, the data are typically numerical. Consider LoPucki and Kalin's study of bankruptcies, published in the pages of this journal.[55] As part of their analysis, the authors present the table depicted in the top panel of Figure 10, purporting to show that the rate of bankruptcy filings varies considerably from year to year. Because meaningful variation—or the lack thereof—is nearly impossible to discern from the precise

---

55. Lynn M. LoPucki & Sara D. Kalin, *The Failure of Public Company Bankruptcies in Delaware and New York: Empirical Evidence of a "Race to the Bottom"*, 54 VAND. L. REV. 231 (2001).

values in the table, we transformed the data into a time series plot depicting what seems to be the chief variable of interest to the authors—rate of filings.

Year	Public companies filing bankruptcy	Number of public companies	Rate of public company filing
1983	89	9,047	0.98%
1984	121	10,717	1.13%
1985	149	11,121	1.34%
1986	149	12,450	1.20%
1987	112	14,620	0.77%
1988	122	16,355	0.75%
1989	135	18,090	0.75%
1990	115	16,123	0.71%
1991	125	13,424	0.93%
1992	91	12,114	0.75%
1993	86	12,764	0.67%
1994	70	13,019	0.54%
1995	84	12,753	0.66%
1996	84	12,977	0.65%
1997	82	13,173	0.62%
1998	122	12,442	0.98%
1999	145	11,998	1.21%
All Years	1,881	223,187	0.84%
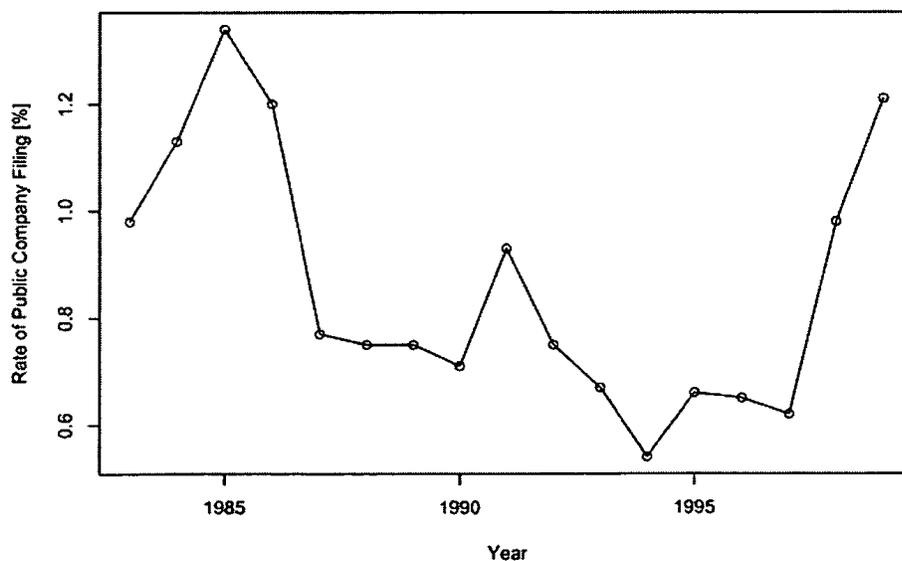
Figure 10: Rate of public company filings for bankruptcy, by year. Lopucki and Kalin's table, reproduced above, provides the raw data on bankruptcy filings, making it difficult to decipher trends in filings across time. Below their table, we provide a time series plot of the same data. Data points for each year are represented by a hollow circle. The time series plot draws attention to the high rate of filings in 1985 and 1999 and low rates in the intervening years.

We leave it to readers to determine whether they agree with LoPucki and Kalin's conclusion about the degree of variation, but at

least they are now equipped to form an opinion: decoding the information in the graph, as opposed to the table, is cognitively undemanding. More generally, simple time series plots are a handy solution when the variable of interest is continuous.

Studies seeking to depict qualitative variables—such as race in the Gross and Barnes study[56]—over time present more of a challenge. Because a time series plot would serve to hinder and not enhance decoding, scholars confronting this challenge tend to fall back on a rough-and-ready solution: the cross-tabulation ("cross-tab"), which displays the joint distribution of two (or more) variables in a (contingency) table.

This "solution" is ubiquitous in the law reviews but, like the table of descriptive statistics, the cross-tab should be banished, and banished for a similar reason: it often obscures, rather than clarifies, the very patterns the author wishes to highlight. Take Gross and Barnes' study.[57] The researchers hope to convey information about search trends, but with three categories of race dispersed over six time periods these trends are extremely difficult to detect (see Figure 7).

Enter the mosaic plot. These are created by using appropriately sized rectangles to illustrate the marginal and joint distribution of the variables. The width of each bar on the x-axis shows the marginal distribution of that variable. Within each bar, the plot shows the fraction corresponding to the variable on the y-axis.[58] Providing an example is Figure 7, in which we transformed Gross & Barnes' search data into a mosaic plot. Now we can visualize both the composition of the race variable in each year, as well as any trends over time. And, indeed, upon a quick glance at Figure 7, the reader simply cannot miss the decline in searches conducted of black drivers in 1995 and in 1998. It is also clear from the plot that the fewest searches were undertaken in 1997 and the most in 1999. Drawing the same conclusions via the authors' original cross tabulation would be possible but only with concerted effort.

C. *The Relationship Between Two Variables: Scatterplots*

Mosaic plots of the sort shown in Figure 7 work particularly well for the Gross and Barnes data, but they are not limited to

---

56. Gross & Barnes, *supra* note 51.

57. *Id.*

58. Mosaic plots were first developed in J.A. Hartigan & B. Kleiner, *Mosaics for Contingency Tables, in* COMPUTER SCIENCE AND STATISTICS: PROCEEDINGS OF THE 13TH SYMPOSIUM ON THE INTERFACE (W.F. Eddy ed., 1981) and were further refined in Michael Friendly, *Mosaic Displays for Multi-Way Contingency Tables*, 89 J. AM. STAT. ASS'N 190 (1994).

variables organized in a time series fashion. Indeed, unless researchers need to convey precise data values—which is almost never the case—we urge them to substitute mosaic plots for cross-tabs of two categorical variables.

That same advice does not hold for two continuous quantitative variables. The tiles on the plot would grow so small that it would make decoding impossible. In this situation, analysts ought to consider employing a graphic workhorse, the bivariate scatterplot.[59]

Earlier we encountered univariate scatterplots, which display all the observations of a single variable (see Figure 2). Bivariate scatterplots, as the name suggests, display the joint distribution of the observations of two quantitative variables. When constructed with sound graphing techniques in mind,[60] bivariate scatterplots can be quite useful for examining the relationship between two variables of interest.

Unfortunately, analysts all too often miss the opportunity to convey data using this important tool. Cumming and MacIntosh's study is a clear example. Recall the authors' basic claim: that underpricing will occur during economic booms (in their dataset, the period between 1999-2000). While they say this effect is recognizable in their data, from their tabular presentation (see Figure 8), it is hard to spot. A time series plot would provide an easy fix but one that does not follow from their basic claim: they are not suggesting a trend over time but rather an association between economic conditions and sales—a perfect application of the scatterplot.

We have taken advantage of this opportunity, and transformed Cumming and MacIntosh's raw data into a scatterplot (see Figure 8), with two embellishments: a loess fit and distinct symbols for the two boom years. A loess, or locally weighted regression, curve is a smooth plot[61] through the middle of the distribution of plotted observations. The smooth loess curve summarizes how the two plotted variables

---

59. In the physical, biological, and social sciences, the predominant graph is the scatterplot, appearing in its many variations; indeed, scholars have estimated that 75% of the graphs used in the sciences are scatterplots. *See* Ian Spence & Robert F. Garrison, *A Remarkable Scatterplot*, 47 AM. STATISTICIAN 12 (1993). Analysts often use simple scatterplots before analyzing their data, and the insights gained may stimulate the production of more complicated variations or may guide the choice of a model.

60. Like so many other visualization tools, bivariate scatterplots can go awry. To avoid unnecessary problems, Cleveland recommends the use of visually prominent plotting symbols, outward facing tick marks, and, where necessary, jittering, along with the avoidance of grid lines. CLEVELAND, *supra* note 31, at 158. See also JACOBY, *supra* note 30, at 52-56; *Communication I, supra* note 6, for a recommendation of similar approaches.

61. As the smoothness parameter α increases, so too increases the smoothness of the loess curve.

depend on one another.[62] In Figure 8, the loess curve shows a general and not unexpected trend in the data: as market price increases so to do sales. But more importantly, it draws attention to the two hypothesized outliers: the two boom years distinguished in the data as enlarged diamonds.

D. The Relationship Among More than Two Variables: Conditional Plots

If Figure 8 suggests anything, it is that Cumming and MacIntosh demanded too much from their readers. Relative to the scatterplot, the raw data table obscures their message.

This problem is even more acute in Juliano and Schwab's table (see Table 2). Because they are making claims about the relationship between one outcome variable (whether the plaintiff wins) and two explanatory variables (court and visibility), their tabular depiction is especially complex and the data it houses quite difficult to decode. Information overload is apparent. Breaking down the variables into smaller conditioning plots, as we recommended in *Communication I* and as we have now done in Figure 9, helps clarify the data enormously.[63] Note that many of the authors' key takeaway points—including that "the success rate of plaintiffs varies dramatically by circuit"[64]—are far easier to spot.

III. Presenting Results

By transforming their data into more attractive and informative displays, we certainly do not mean to portray Juliano and Schwab's—or any other authors'—work in a negative light. In all the articles we have discussed, the data work is quite sound. It is the use of data tables to which we object, even though we, of course, recognize that tables housing raw data or descriptive statistics have a long tradition in the law reviews. For decades now, to provide but one example, the *Harvard Law Review* has provided data summaries of the Supreme Court's term.[65]

---

62. *See e.g.*, CLEVELAND, *supra* note 31, at 170.

63. Also note our use of dot plots. As was the case for our earlier example regarding Subramanian's article, the number of measurements in Juliano and Schwab's study is small enough and the circuits well known enough that labeling each serves a substantive purpose.

64. Juliano & Schwab, *supra* note 54, at 574.

65. *Harvard Law Review's* practice of publishing Supreme Court statistics from the preceding term began in 1949. *The Supreme Court, 2004 Term: The Statistics*, 119 HARV. L. REV. 415, 415 (2005) (citing *The Supreme Court, 1948 Term—The Business of the Court*, 63 HARV. L. REV. 119 (1949)).

Beyond their use of tables, the Harvard project, Kao's analysis of hourly fees,[66] and Gross and Barnes' study of drug searches[67] have another common feature. They are largely descriptive efforts. Such may have dominated empirical legal scholarship for many years. But omnipresent now are attempts at performing causal inference, which, in law reviews, typically means invoking regression-based tools to assess the extent to which a variable(s) of interest causes an outcome or response. Examples we have discussed in our series on effective communication include Staudt's study of federal taxpayer standing, Epstein et al.'s work on the development of the norm of consensus on the U.S. Supreme Court, and Roe's analysis of the effect of national politics on the proportion of firms under diffuse ownership in rich countries.[68]

To be sure, these studies, along with the many others we could cite, vary in important ways. They raise different research questions and cover distinct substantive areas of the law; they even deploy different regression tools, including linear regression, probit, and logit. What does not vary much are the tools used to convey the research results: With only limited exceptions, they are uninformative both to laypersons and the statistically savvy alike. The ills are many, from (once again) a reliance on inaccessible tables, to displays and narrative that fail to convey substantive effects and uncertainly about those effects.

In what follows, we offer a step-by-step corrective. For the sake of clarity, we rely on a single running example throughout: the confirmation of nominees to the U.S. Supreme Court. Specifically, we consider how to convey the results of a statistical model that seeks to explain the individual votes cast by senators over Supreme Court nominees since 1937 (Black through Alito).[69] Following from work in the social sciences, the key causal variables of interest are (1) the degree to which a senator perceives the candidate as *qualified* for office[70] and (2) the *ideological distance* between the senator and the

---

66. Kao, *supra* note 28.

67. Gross & Barnes, *supra* note 51.

68. Lee Epstein, Jeffrey A. Segal, & Harold J. Spaeth, *The Norm of Consensus on the U.S. Supreme Court*, 45 AM. J. POL. SCI. 362 (2001); Mark J. Roe, *Political Preconditions to Separating Ownership from Corporate Control*, 53 STAN. L. REV. 539 (2000); Nancy C. Staudt, *Modeling Standing*, 79 VA. L. REV. 612 (2004).

69. This model follows from work by EPSTEIN & SEGAL, *supra* note 50; Charles M. Cameron, Albert D. Cover, & Jeffrey A. Segal, *Senate Voting on Supreme Court Nominees: A Neoinstitutional Model*, 84 AM. POL. SCI. REV. 525, 530 tbl.2 (1990); Lee Epstein et al., *The Changing Dynamics of Senate Voting on Supreme Court Nominees*, 68 J. POL. 296 (2006).

70. Previous work has assessed this by analyzing the content of newspaper editorials written from the time of the nomination until the vote by the Senate and then deriving a

candidate,[71] such that the more qualified the nominee and the closer the nominee is to the senator on the ideological spectrum, the more likely the senator is to cast a yea vote. Also following from the extant literature, we control for two other possible determinants of senators' votes: whether the president was "strong" in the sense that his party controlled the Senate and he was not in his fourth year of office; and whether a senator is of the same political party as the president.

To assess the extent to which these variables help us account for senators' votes, we employ probit regression, a common tool in legal scholarship, appropriate when the dependent variable is binary, as it is here.[72]

The top panel of Table 3 depicts the statistical estimates, and, crucially, depicts them in a way that seems to follow standard operating procedure in many law reviews.

---

qualifications score for each nominee. These Segal-Cover scores range from 0 (most qualified) to 1 (least qualified). *See* EPSTEIN & SEGAL, *supra* note 50, at 114 fig.4; Cameron, Cover & Segal, *supra* note 69, at 530 tbl.2..

71. Following Epstein et al., *supra* note 69, at 299, we measure the ideological distance between a senator and a nominee via the generation of Common Space scores for each nominee. More detail on the creation of this measure is available in Epstein et al., *supra* note 69, so suffice it to say here that these scores are generated by "bridging" candidates nominated by presidents whose party holds a majority of Senate seats. These "bridged" nominees receive the Segal-Cover scores, *see* discussion *supra* note 70, of their appointing president, and those scores, along with their Segal-Cover scores, permit a linear transformation. The result is that Common Space scores can be created for all nominees based on their Segal-Cover scores.

72. We could have also conducted our estimation with a logistic regression model. Like probit, logistic ("logit") regression is utilized when the dependent variable is dichotomous. The structural models of logit and probit are different, but they are related to each other in such a way that logit coefficients, when statistically significant, are approximately 1.7 times larger than probit coefficients, making the choice between the two models "largely one of convenience and convention." LONG, *supra* note 27, at 47-49, 83.

Variable	Coefficient	(Std. Err.)
Lackqual	-2.217**	(0.374)
Eucldist	-2.320**	(0.441)
Strngprs	0.589*	(0.286)
Sameprty	0.765**	(0.220)
Intercept	1.788**	(0.285)
N		3809
Log-likelihood		-928.282
$\chi^2_{(4)}$		80.406

Variable	Coefficient	(Std. Err.)
Nominee's Lack of Qualifications	-2.217*	(0.374)
Ideological Distance between the Nominee and Senator	-2.320*	(0.441)
President Controls Senate and is Not in His Last Year of Office	0.589*	(0.286)
Senator and President Share Party Affiliation	0.765*	(0.220)
Constant	1.788*	(0.285)
N		3809
Reduction of Error in Predicting Senators' Votes		26.5%

Table 3: Probit regression analysis of the effects on individual senators' votes on 41 Supreme Court nominees (Black through Alito). Cell entries are probit coefficients and robust standard errors clustered on the nominee (in parentheses). In the top table, * indicates $p \leq 0.05$ while ** indicates $p \leq 0.01$. In the bottom table * indicates $p \leq 0.05$. Although the two tables present the same statistical results, the bottom table, by, e.g., eliminating multiple stars for different levels of statistical significance and providing meaningful variable names, better enables readers to understand the results.

What we argue below is that this standard approach to presenting the results of a probit analysis—or, for that matter, any multivariate regression procedure—ought to be reconsidered. In particular, we suggest that authors (1) rework tables so that they not only stand alone from the text but are themselves informative; (2) convey the substantive effects of key variables of interest; and (3) communicate uncertainty.[73] Adhering to these rules will go some distance toward enhancing the impact of analysts' research projects if only because the audience will now better understand the results.

A. (How to) Produce Informative Tabular Displays of Statistical Results

Throughout our series we have counseled against tabular depictions of data. Frankly, and for the reasons we offer below, we feel no differently about tables displaying regression estimates (e.g., Table 3). Nonetheless, we understand that even in the social sciences readers have come to expect them; we also realize that, occasionally,

---

73. The recent work by Gelman et al. and King et al. has been particularly influential in establishing the need to convey uncertainty and providing the equipment to do so. Gelman, *supra* note 10; King et al., *supra* note 8. We follow their lead here.

they can convey valuable information if only to the statistically informed reader.[74]

As a result, though incorporating tables of estimates into presentations or papers is sometimes necessary, they surely need not be as utterly uninformative as the top panel of Table 3. One obvious problem is the variable names are not clear; e.g., from a mere glance at the table readers could not possibly know that "eucldist" means the "ideological distance between the nominee and the senator."

We understand how this problem comes about. When researchers enter data into a statistical package, they often shorten a variable's name; and when they estimate their regression model, they simply cut and paste the resulting table into their text file. This is good practice if and only if researchers are indifferent to their audience. Hoping that no reader of our series falls into this category, we suggest using descriptive names to label the variables as we have done in the bottom panel of Table 3. Note that we also clearly convey the dependent variable, a crucial piece of information, yet one surprisingly missing in many empirical studies. We use the caption for this purpose but other plausible locations include the table's title or column head.

Turning to the statistical estimates, both panels present the coefficient estimates and standard errors, as they should. These numbers convey information about the direction of the effect of the coefficient (indicated by the sign on the coefficient) and the presence of statistical significance (indicated by the relationship between the standard error and the coefficient),[75] even if the coefficients themselves are difficult to interpret substantively. Troubling, however, is the top panel's use of two different asterisks to denote statistical significance, * for $p \leq 0.05$ and ** for $p \leq 0.01$. Because this all-too-common practice may lead readers to inappropriately compare p-values,[76] we suggest omitting the asterisks altogether. Readers can

---

74. For example, for regression models, knowing the scale of the variable of interest allows the reader to contextualize the regression coefficient. It is necessary to know the scale of the dependent and independent variable for the reader to understand what a $\beta = -2.217$ actually means. Additionally, outside of the linear regression context, such substantive interpretations of coefficients are extremely difficult because they require the reader to make complex calculations that depend on the values of the independent variables. Thus, why ask the reader to do these calculations for a multivariate statistical analysis when conveying results in an easy-to-consume manner with figures is straight forward. We emphasize this point in the text to follow.

75. Without the presence of a statistically significant relationship between the coefficient and the dependent variable, there is no reason to test the substantive effect of a variable.

76. The arbitrary setting of a p-value simply means that a researcher wants to have a certain level of confidence (known as the α-level) in the accuracy of the estimated relationship of the variables. Providing differing levels of p-values merely amounts to having different levels of confidence in the statistical relationship. It would certainly be incorrect to say that, in Table 3,

discern whether a variable is "statistically significant" from the standard error. Alternatively, authors can choose a level (typically α = 0.05) and use just one * (see the bottom panel of Table 3).

Finally, to be complete, researchers should provide summary information about their model. The N (number of cases in the sample) is essential,[77] though it too is occasionally absent from many tabular displays. Ns can be placed in the caption or in a column, as we have done in Table 3. For models estimated by maximum likelihood, it is good practice to report the log-likelihood (or the deviance, which is -2 times the log-likelihood). This quantity is useful for a number of statistical tests. A measure of the predictive power of the model is also essential. In linear regression, R-squared is among the most common; for probit models of the sort we display in the table, a reduction of error assessment works well.[78] Conversely, we advise against reporting omnibus test statistics, such as the F statistic, (often used in linear regression) or likelihood ratio tests (e.g., the chi-square in the top panel of Table 3). Neither conveys information that is substantively useful. Indeed, if these tests are significant (and in practice, they always are), all one learns is that something in the model is related to the dependent variable.

the coefficient for the president controlling the Senate is less important of a finding than the others simply because it is only statistically significant at the .05 level as opposed to the .01 level. *See* AGRESTI & FINLAY, *supra* note 25, for more information on *p*-values, α–levels, and the proper use and interpretation of each.

77. Because some statistical techniques are inappropriate for studies with a small number of observations, reporting the N provides a check. For example, the properties of maximum likelihood estimation models, such as probit, do not hold when sample sizes are too small. *See* LONG, *supra* note 27, at 54 ("It is risky to use [maximum likelihood] with sample smaller than 100, while samples over 500 seem adequate. These values should be raised depending on characteristics of the model and the data.")

78. Reporting the proportional reduction in error from the newly estimated model provides information about the utility of the researcher's chosen model. With the assistance of statistical software, computing this reduction in error is also simple. For example, in Stata, after the model has been estimated, the user need simply install and use the "pre" command to yield this reduction in error. Within the software, this is computed by finding the errors when simple guessing is employed and then finding the number of errors after the model has been estimated. The final proportional reduction in error is computed by subtracting the number of errors in the model from the number of guessing errors and dividing that number by the number of guessing errors.
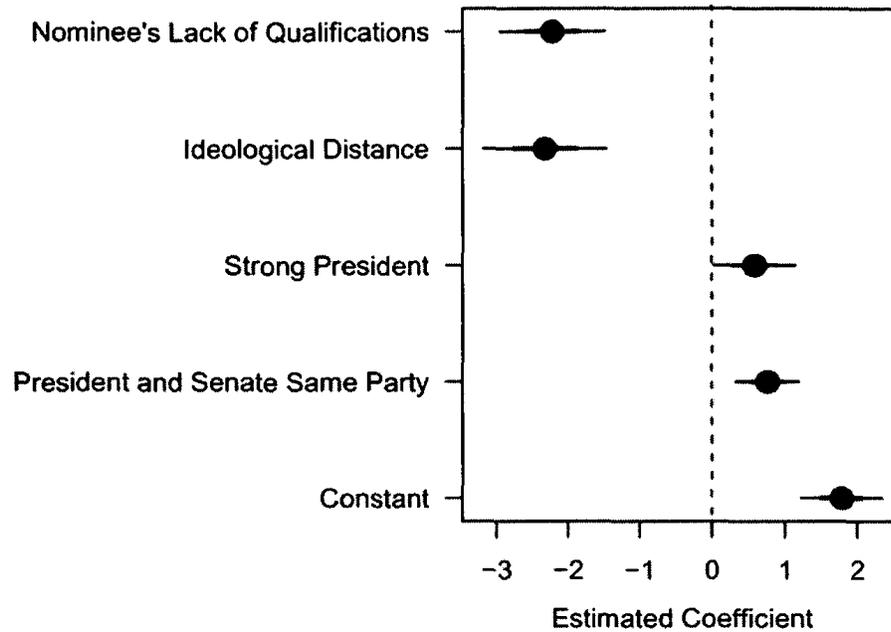
Figure 11: Nomogram depicting the results from Table 3's probit regression analysis of the effects on individual senators' votes on forty-one Supreme Court nominees (Black through Alito). Nomograms allow the reader to visually discern the estimated coefficients and the uncertainty around the estimates and to quickly discern whether the effects are statistically significant by looking at whether the confidence intervals cross zero.

These recommendations are designed to help authors create informative tables that do not require the reader to slog through the text in order to, for example, identify variables. For those analysts who need not present precise values but are more concerned with providing their audience with a feel for their estimates, a nomogram (see Figure 11) provides an ideal alternative to a table.[79] Nomograms are dotplots in which the estimated coefficient is represented by the dot, and the confidence interval is depicted by error bars. Visually, we can determine statistical significance by noting whether the error bars cross zero. It is also easy to compare the relative magnitude of coefficients, if that is meaningful for a given study.

B. Communicate Substantive Effects and Uncertainty About Those Effects

As researchers, we are inherently interested in whatever question we are investigating at the moment—to us, our work is exciting stuff. But conveying our results as we do in Table 3 and even in Figure 11 could not be less inviting in presentation. Displays of

---

79. For a similar graphical display of statistical results, see *supra* Figure 1.

estimated coefficients that fill the law reviews and even social science journals are not just ugly; they convey virtually no information of interest. Certainly, of our probit analysis we could say "The estimated coefficient on lackqual means that, holding all else constant, a one unit increase in lack of qualifications yields a -2.217 unit decrease on the cumulative Normal scale." Because (almost) no one would understand what this means, we usually just write, "the coefficient on lackqual is statistically significant at the .05 level." But even this is not a very informative statement to many readers—even those with knowledge of statistics.[80]

In short, the way we typically present regression-based results only works to dampen enthusiasm for our research. We can do better. More to the point, we should want to do better. How? In Figure 12, we provide a three-step process, moving from unacceptable to optimal communication.

---

80. See *Communication I, supra* note 6, at 1831-31, where we make a similar point. *See generally* King et al., *supra* note 8 (making similar arguments); Gelman, *supra* note 10 (same). Our inspiration for this section follows from their work, especially King et al.

The Ugly Table

Variable	Coefficient	(Std. Err.)
lackqual	-2.217**	(0.374)
eucldist	-2.320**	(0.441)
strngprs	0.589*	(0.286)
sameprty	0.765**	(0.220)
Intercept	1.788**	(0.285)

Good Communication: Estimate A Key Quantity of Interest

Other things being equal, when a nominee is perceived as highly unqualified the likelihood of a senator casting a yea vote is only 0.27. That probability increases to a near-sure bet yea vote (0.91) when the nominee is highly qualified.

Better Communication: Estimate A Key Quantity of Interest, Plus Uncertainty

Other things being equal, when a nominee is perceived as highly unqualified the likelihood of a senator casting a yea vote is only about 0.27 [.21, .33]. That probability increases to a near-sure bet yea vote 0.91 [.89, .93] when the nominee is highly qualified.

Best Communication: Graph Quantity of Interest Plus Uncertainty Across Values


Figure 12: An illustration of moving from unacceptable to optimal communication of research results. Adapting this schema to their own projects and needs should help researchers better relay their story. To generate the predicted probabilities and confidence intervals, we used Stata and the SPost package of post-estimation commands. Holding the ideological distance between a senator and a nominee at its mean value and the other two independent variables at 0 (meaning that the president is weak and the nominee and voting senator are of different political parties), we move from a highly unqualified nominee to a highly qualified nominee. Confidence intervals were computed using bootstrapping.

As a first step, the empiricist must ask him or herself, "What substantively interesting features of my results do I want to convey to my readers?" In our running example, several quantities of interest come to mind, but to keep it simple we begin with one: the probability that a senator will cast a vote for a highly unqualified (qualified) candidate. Communicating this piece of information is a start toward engaging on-going debates, such as whether a nominee's ideology is now so paramount that qualifications are irrelevant to senators.[81] It is also a rather straightforward way to begin the move away from a sole emphasis on statistical significance and toward a stress on substantive importance.

In Figure 12, under "Good Communication," we take this step by translating our inaccessible probit coefficients into substantively important quantities of interest: the odds of a yea vote when a candidate is highly qualified and when a candidate is highly unqualified for office, *other things being equal*. By that last phrase, we mean that the other variables in the model—ideological distance, strong president, and same party—are each set at fixed values. For example, for the statement in Figure 12, we set the ideological distance between a senator and nominee at its mean and the other two variables at 0 (weak president and senator of a different political party). But from our results we could have just as easily developed another counterfactual, such as the effect of qualifications when the ideology of the senator and nominee are very distant or when the president is "strong." Alternatively, we could have shifted focus entirely and considered the effect of ideological distance when we hold qualifications at its mean. For example,

> Other things being equal, when a nominee and senator are ideologically very distant the likelihood of a senator casting a yea vote is only 5%. That probability increases to a near-sure bet yea vote (90%) when the nominee and senator are ideologically very close.[82]

Estimating a quantity of interest is a good start. But even better communication, as we show in Figure 12, entails conveying error around that estimate. Because we covered uncertainty and its

---

81. *See, e.g.*, WERL, *On Tournaments for Appointing Great Justices to the U.S. Supreme Court*, 78 S. CAL. L. REV. 157, 157 (2005) (commenting on debate over "whether the current system [for appointing justices] does any better than [a] tournament in predicting skills that include the ability to compromise and negotiate, a talent for shaping national policy, and a gift for choosing among the thousands of petitions for certiorari filed with the Court").

82. This is the predicted probability of a senator casting a yea vote for a Supreme Court nomination when a nominee's (lack of) qualifications are held at their mean, when the president is weak (i.e., his party does not control the Senate), and when the voting senator and the president are of different political parties, changing the ideological distance between the nominee and the voting senator from distant to close.

importance in *Communicating I*, we need not go into detail here. Suffice it to note that most of us would be highly skeptical of a survey that failed to provide readers with the margin of error or a table of regression estimates that omitted standard errors or confidence intervals. We should be equally skeptical of claims about substantive effects that fail to do so (via, e.g., confidence intervals). To see why, consider two hypothetical versions of the claim above:

> 1. *Other things being equal, when a nominee and senator are ideologically very distant the likelihood of a senator casting a yea vote is 30%, though it could be as low as 25% or as high as 35%.*

> *versus*

> 2. *Other things being equal, when a nominee and senator are ideologically very distant the likelihood of a senator casting a yea vote is 30%, though it could be as low as 1% or as high as 60%.*

In both examples, the (point) estimate of the impact of ideology is identical (30%) but our certainty about that estimate differs dramatically. So dramatically, in fact, that we should be highly skeptical of the second claim: because the confidence interval goes beyond 50% we cannot eliminate the real possibility of a yea vote even when the senator and candidate are ideologically very dissimilar.

More generally, the examples above and in Figure 12 go some distance toward bridging the gap between researchers and their audience. Unlike the terms "statistical significance," "coefficient" or "0.01 level," statements containing quantities of interest and error are easy to understand and, crucially *to evaluate*, even by the most statistically challenged among us.[83]

We would thus be delighted if every article published in the law reviews supplanted the sterile "statistically significant at the .05 level" with the substantively informative, "other things being equal . . . ." We would be even more delighted, as would all readers of empirical work, if researchers took it to the next level and graphed their results. Actually, the advantages of generating visual displays are so great that analysts should need very little encouragement to move in this direction.

First, while substantive claims of the form ". . . when a nominee is perceived as highly unqualified the likelihood of a senator casting a yea vote is only about 0.27 [.21, .33]" may be informative, these claims exclude a lot of information, such as the values in

---

83. King et al., *supra* note 8, at 359-60, makes this point, and we adopt it here.

between "highly unqualified" and "highly qualified." To provide these quantities, we could generate a long series of statements—e.g., when a nominee is perceived as minimally qualified, on average qualified, and so on. But graphing the results is a far more parsimonious, pleasing, and, for our readers, cognitively less demanding approach.

Underscoring these points is the bottom display in Figure 12. Here we juxtapose Lack of Qualifications against Ideological Distance. Specifically, in the first three panels we show the probability of a senator casting a yea vote across the range of Lack of Qualifications and when we set Ideological Distance at its minimum, mean, and maximum levels. In that triptych we depict our uncertainty, in the form of 95% confidence intervals, with vertical lines. To avoid cluttering the fourth panel, we eliminate the confidence intervals and simply show the three sets of probabilities.

This display, we believe, is a good example of what we mean by parsimony. It conveys a great deal of information—actually it encodes 132 pieces of information—quite efficiently. Or at least more efficiently than the 132 sentences it would have taken to describe each and every result depicted in the four panels.

A second and perhaps even more important virtue of graphing results centers on pattern detection. From the display in Figure 12, several results are immediately apparent. Chiefly, we observe the conditional nature of the relationship between qualifications and ideology: the effect of a nominee's qualifications, in other words, depends at least in part on the nominee's ideology vis-à-vis the senator. So, for example, professional merit has far less of an impact on nominees who are extremely ideologically distant from senators than on those who are more proximate; the former, even those who are highly qualified (0 in Figure 12), confront low odds (about .15) in their quests for confirmation. Surely, this is crucial information for both researchers and their readers, but it is virtually undetectable from the "ugly table" of coefficients.

C. How to Communicate Substantive Effects and Uncertainty

We believe the advantages of the sorts of narratives and graphs we depict in Figure 12 are obvious. For readers, they need not struggle to make sense of regression estimates that even the analyst may have trouble understanding. Nor are critical questions left dangling, such as "are the results substantively important?" and "how sure is the researcher about the findings?" As we have stressed throughout this series of articles, the advantages for analysts are equally obvious,

ranging from the detection of patterns in their own work to the ability to impart their own excitement to their audience.

If there is a downside for authors, it is that communicating quantities of interest and uncertainty requires more time and more thought. To see why, think about current practices (at least as they appear to us): The legal empiricist estimates a regression model, hopefully performs some diagnostic checks, and then cuts and pastes the resulting table into a Word file. End of story. As Figure 12 suggests, much more is needed—and the "more" mandates that researchers learn about procedures enabling them to estimate substantive effects (and confidence intervals). We remain catholic as to the precise tools one should use to perform this work, but we can say that researchers can implement nearly all our suggestions using common statistical packages, such at Stata or SPSS. Three add-on packages that make computing substantive effects easy are CLARIFY[84] and SPost[85] for Stata, and Zelig[86] for the R language.

Whatever statistical package researchers decide to employ, it is the more general message that we hope they do not miss: collecting and compiling data, and then estimating a model, should not complete the task. Similar thought and care should be used to effectively communicate the results of a study.

IV. IMPLEMENTING CHANGES IN THE COMMUNICATION OF DATA AND RESULTS

As we draw to the close of our series, we cannot help but hope we have provided legal researchers with some guidance on how to more effectively communicate their data and results. The benefits of following the standards that we have articulated, we believe, well outweigh the costs. Sure, empiricists must now familiarize themselves with a new set of tools for presenting their work, but once they do numerous advantages will accrue. Because they will be better situated to detect patterns in their own data, their own work will improve; because their audience will be better able to understand their work, its impact will be greater.

These should be sufficient incentives for change. But recognizing that additional prodding may be necessary, we want to

---

84. CLARIFY can be found at Michael Tomz et al., *CLARIFY: Software for Interpreting and Presenting Statistical Results*, June 1, 2001, http://gking.harvard.edu/clarify/docs/clarify.html.

85. SPost can be found at J. Scott Long, *SPost: Post-estimation with Stata*, Mar. 1, 2007, http://www.indiana.edu/~jslsoc/spost.htm.

86. Zelig can be found at Kosuke Imai, *Zelig: Everyone's Statistical Software*, http://gking.harvard.edu/zelig/ (last visited Mar. 14, 2007).

encourage readers of empirical work and editors of legal journals to play a role as well. Demanding that authors adhere to basic standards for communicating their data and results will go some distance toward pushing empirical legal scholarship to new heights.

Along these lines we provide a set of guidelines that build upon the advice we have offered to authors in this article and in *Communicating I*. First and foremost, scholars, editors, policy makers, judges, and practicing lawyers—that is, all consumers of empirical work—should *press authors to move beyond sterile claims about statistical significance*. When researchers present tables full of coefficients and asterisks, and make claims about statistical significance, audience members should push them to provide quantities of interest. Journal editors should do the same. It is simply not enough to report statistical significance without conveying substantive meaning.

Second—and relatedly—editors and consumers should be highly skeptical of "point estimates" that do not supply sufficient information on how they were calculated or on the author's uncertainty about them. We have dwelled enough on the latter point. But the first is equally important. To see why, reconsider an estimate we offered earlier, in Figure 12:

> *Other things being equal, when a nominee is perceived as highly unqualified the likelihood of a senator casting a yea vote is only about 0.27.*

Now consider a second claim developed using the same data set and the same statistical model:

> *Other things being equal, when a nominee is perceived as highly unqualified the likelihood of a senator casting a yea vote is 0.56.*

How can it be that the odds of a yea vote for a highly unqualified candidate shift from unlikely (.27) to likely (.56)? The answer lies with the phrase "other things being equal": in the first example, the senator and president are of different parties; in the second, they are both Democrats or both Republicans. Clearly, senators who share a party affiliation with the president are more inclined to support his candidates—a fact that the researcher ought communicate. At the least, readers and editors should compel the author to explain how he or she developed the given counterfactual so that they can evaluate its plausibility.

These two suggestions apply to both consumers and editors. For the latter we have an additional set of recommendations, all centering on the conditions they should impose on potential contributors. While editors of legal journals often follow or create style

guidelines, those we have consulted say next to nothing about the communication of data and results.[87] They should. At a minimum, editors ought to establish policies governing three areas of the data-communication process.

First, we, like others before us,[88] implore editors to develop standards regarding replication.[89] Indeed, it strikes us as just plain odd that the law reviews, in particular, are so concerned with ensuring the availability and providing the exact location of unpublished papers,[90] which are typically tangential to an article, but not with data sets, which may be at the article's core. Even if law review editors are unwilling to adopt a full-blown replication policy, they should, at the very least, require authors to submit their databases and the code used to generate the research results. This will allow someone at or hired by the law journal to review the study and ensure its replicability. The salutary effects of taking even this small step would be many, not the least of which would be to help reduce (though not completely eliminate) concerns about the lack of peer review in legal scholarship.

---

87. *Communication I, supra* note 6, at

88. *See* Epstein & King, *supra* note 11, at 132 (recommending law reviews require documentation and archiving of empirical data that would enable replication).

89. Professor Gary King at Harvard offers the following, easy-to-implement replication policy for journals:

> Authors submitting quantitative papers to this journal for review must address the issue of data availability and replication in their first footnote. Authors are ordinarily expected to indicate in this footnote in which public archive they will deposit the information necessary to replicate their numerical results, and the date when it will be submitted. The information deposited should include items such as original data, specialized computer programs, lists of computer program recodes, extracts of existing data files, and an explanatory file that describes what is included and explains how to reproduce the exact numerical results in the published work. Authors may find the Publication-Related Archive of the Inter-university Consortium for Political and Social Research (ICPSR) a convenient place to deposit their data. Statements explaining the inappropriateness of sharing data for a specific work (or of the necessity for periods of embargo past the publication date) may fulfill the requirement .... Authors of works relying upon qualitative data should submit a comparable footnote that would facilitate replication where feasible. As always, authors are advised to remove information from their datasets that must remain confidential, such as the names of survey respondents.

Gary King, *An Example Replication Policy for Journals*, http://gking.harvard.edu/repl.shtml (last visited Feb. 18, 2007). We should also note that a replication standard such as this is in the best interest of authors, as it requires them to give a little extra effort at the time of submission to centrally organize the data and statistical analysis (something that we often wish we would have done better for our own projects).

90. The ever-present "paper on file with the journal" or "with the author."

Second, we strongly recommend that editors write out a set of instructions for the preparation of tables and graphs. Although these will vary to some extent from journal to journal, many of the rules we have covered here and in our earlier article are universal. Journals should adopt guidelines for what constitutes an acceptable table or figure, and require that authors meet those guidelines. This is common practice in many social science journals, and one we hope catches on in the law reviews.

Finally, editors must work with authors to ensure the integrity of their data presentations. After spending time and energy to produce high-quality graphical displays, we have been, on occasion, surprised (read: disappointed) by the published results. It seems that the editors merely cut and paste graphics files into word processing programs. Their readers—not to mention authors!—deserve better. Editors should request graphs in scalable forms, such as Adobe PDF files or encapsulated postscript files. Then modern software tools should be used to typeset the graphics in such a way as to maximize their readability. The advantage of scalable files is that graphics specialists can resize images without degrading the quality of the presentation. The days of india ink, photo-ready illustrations may be gone, and the technology replacing it may be terrific, but it gets us nowhere if we do not use it.

V. CONCLUSION

We end where we began our series: 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, these authors have been less effective at communicating the products of their labor. A strong devotion to tabular, rather than graphical, displays, and claims about "statistical significance" rather than substantive importance, are just two areas requiring improvement.

What we have attempted to do here and in *Communication I* is adapt a burgeoning literature in the social and statistical sciences to the unique interests of legal scholars. Our proposals have been many in number, but none are particularly difficult to implement. More to the point, we believe that law professors should want to implement these suggestions. If other fields are any indication, moving toward more appropriate and accessible presentations of data will heighten the impact of empirical legal scholarship regardless of the audience— no doubt a desirable goal in a discipline that rightfully prides itself on its contributions to the formation of legal and public policy.