

# Statistical Computing and Graphics

---

## Let's Practice What We Preach: Turning Tables into Graphs

Andrew GELMAN, Cristian PASARICA, and Rahul DODHIA

---

Statisticians recommend graphical displays but often use tables to present their own research results. Could graphs do better? We study the question by going through the tables in a recent issue of the *Journal of the American Statistical Association*. We show how it is possible to improve the presentations using graphs that actually take up less space than the original tables. We find a particularly effective tool to be multiple repeated line plots, with comparisons of interest connected by lines and separate comparisons isolated on different plots.

**KEY WORDS:** Data reduction; Graph; Table; Visual display.

---

### 1. INTRODUCTION

Statisticians recommend graphical displays but often do not follow this recommendation in presenting their own research. We decided to figure out what's going on by examining a recent issue of the *Journal of the American Statistical Association* (JASA), the leading journal of applied and theoretical statistics. The articles in the issue contained 72 figures and 60 tables presenting numerical information. We considered the motivation for each of the tables and considered how they could be made into graphs, following the principles we have learned and taught in our statistics courses.

In this article, we go through seven interesting and representative examples of tables from the March 2000 issue of JASA. We conclude, after studying these examples in detail, that the main motivation for displaying numerical information in statistics articles is to make comparisons, and for this purpose, we find well-designed graphs to be superior to tables. In the examples we consider, multiple repeated line plots are particularly effective, with the comparisons of interest performed within the individual plots, and separation into small plots to isolate the separate comparisons.

In addition to general recommendations about the kinds of plots that work well for presenting the results of statistical research, we supply several specific suggestions about the placement of graphs, the use of labels and baselines, and other details that are crucial in constructing small yet readable graphical dis-

plays. Our advice follows well-known principles of data display (see, e.g., Tufte 1983; Cleveland 1985) but applied to the presentation of research results as well as raw data.

### 2. DISPLAYING NUMERICAL RESULTS

Statistical research requires the display of many different kinds of numerical results, including raw numbers, data reductions, inferences, and—for research in theory and methodology—summaries of probability distributions, most notably frequency properties of statistical procedures. Unfortunately, it is still standard to display these latter summaries as tables rather than graphs. In this section of our article, we briefly review the research on data display and perception, and then we discuss the goals of numerical displays for statistical purposes.

#### 2.1 Literature on Graphs and Tables

The pages of the *Journal of the American Statistical Association* and *The American Scientist* have occasionally carried reports on ways to display data (examples from different eras include Eels (1926); Croxton and Stein 1932; Cleveland 1984; and Cook and Weisberg 1999). Many other authors have also reported studies trying to figure out the best methods of displaying data (e.g., Wilkinson 1999; Day 1994). Many such articles have been described in reviews of the literature; Cleveland (1987) gave a bibliography of graphical statistics, Dibble (1997) provided a number of sources that have contrasted tables and graphs, and Lewandowsky and Spence (1989) focused on the perception of graphs.

The clearest conclusion from most of these studies is that tables are best suited for looking up specific information, and graphs are better for perceiving trends and making comparisons and predictions (Meyer, Shamo, and Gopher 1999; Dibble 1997; Jarvenpaa and Dickson 1988; Carter 1947; Washburne 1927). Therefore, Loftus (1993), as a new editor of *Memory and Cognition*, exhorted submitters to display data using well-designed graphs. Gillan, Wickens, Hollands, and Carswell (1998), in an effort to increase the standards of graphs in articles submitted to the journal *Human Factors*, suggested that graphs may present data better than tables even for very small datasets, a point we illustrate in Figure 1 of this article. If precise values are needed along with a visualization of the relations among the data, these values may be displayed as quantitative labels on a graph. A similar point was made by Gelman and Price (1999) when discussing maps of estimated rates and proportions: the typical (and appropriate) use of such a map is not as a look-up table but rather for making comparisons, and the statistical properties of the map should be evaluated keeping that in mind. Although tables are sometimes perceived as less intimidating than graphs (Lusk and

---

Andrew Gelman is Professor, and Cristian Pasarica is Ph.D. Student, Department of Statistics, Columbia University, New York, NY 10027 (E-mail: gelman@stat.columbia.edu). Rahul Dodhia is Postdoctoral Researcher, NASA Ames Research Center, Moffett Field, CA 94035. We thank Seth Roberts and several reviewers for helpful suggestions and the National Science Foundation for financial support.

Kersnick 1979) they do not lead to better performance (Vernon 1952). Subjects who indicated a preference for tables were generally novices: laypersons unused to statistical information or beginning college students. Studies using more sophisticated subjects showed a preference for graphs (Zmud 1978).

Much of the debate between tables and graphs is fueled by intuitive ideas and personal preferences, but a body of empirical studies provides more objective results. In general, these studies provide subjects with information in various display formats and ask a variety of questions based on this information: reading off specific information, comparing data, and so forth. Spence and Lewandowsky (1991) compared pie charts, bar graphs, and tables and found a definite advantage for graphical displays. Feliciano, Powers, and Kearn (1963) found that subjects were faster and more accurate in making comparisons when given information in horizontal bar graphs rather than in tables. (Ironically, the authors presented the results of their study in a table.) Performance improved when the data displays complemented textual descriptions (which we attempt to do in this article via extensive captions, as is the style for American Statistical Association journals). Stock and Watson (1984) gave subjects information in tables or faces and found that subjects were more accurate at detecting temporal changes using faces rather than tables.

Meyer, Shinar, and Leiser (1997) found that tables had an advantage over graphs in their study. A few other studies also had similar results (see, e.g., Remus 1984; Coll, Coll, and Thakur 1994; and Lohse 1993). Meyer et al. (1999) argued that these results came about because the tasks used in these studies ignored the inherent organizing ability of graphs. For example, Meyer et al. (1997) considered a very small dataset that could be summarized in a 3 by 5 table, and its values ended in “0” or “5”. The questions asked of the subjects did not exploit the visual arrangement of the graphs.

## 2.2 The Goals and Principles of Statistical Data Display

Constructing the graphs shown in this article actually took a lot of time and several iterations, with each step provoking further thought on the motivation behind various techniques of graphical display. Here, we attempt to identify some of the key ideas, with the hope of making the task easier for future researchers.

Tables are more effective if the goal is to read off exact numbers. However, the interest in a statistical paper typically lies in comparisons, not absolute numbers. For example, there is no reason for the reader to care that the standard error for a particular parameter estimate is 0.029 as computed under a certain method and 0.054 under another method. Rather, the point is that the second number is almost twice as high. When dozens of such potential comparisons are possible, they can be seen much more clearly in a well-chosen graphical display than in a table (see Figure 3). The idea of comparisons provides a theoretical framework for statistical graphics (as is implicit in Tukey 1972, 1977).

Within each plot, we tend to follow the standard convention, with descriptions or explanatory variables on the (horizontal)  $x$ -axis and the outcome of interest on the (vertical)  $y$ -axis, but with two exceptions to this rule. First, if the outcome is binary (or discrete with few categories), it can be effective to use the  $x$ -

and  $y$ -axes for two continuous predictor variables and indicate the outcome using symbols (see Figure 7). Second, if each case has a verbal descriptor, it can be more convenient to lay out the outcome horizontally so that the display is more readable (see Figures 1 and 8); this advice was also given by Cleveland (1985) for line plots. For example, if Figure 8 were displayed as a histogram with vertical bars, it would be much harder to place the labels so as to make the figure readable all at once.

Display axes are precious resources to be used wisely. We try not to order by ID number or alphabetically, since this can obscure interesting patterns. Sometimes a graphical display simply eliminates the need for ordering (as in Figures 2 and 7); other times, there may be a natural order (see Figure 1) or we can order by decreasing frequency (see Figure 8) or thematically (see the  $x$ -axes in Figure 6).

We spent a fair amount of effort considering the location and size of labels (see, e.g., Figures 2 and 5). It is hard to state general principles here, except that we have found it useful to create multiple plots rather than cramming too much information in a single plot. We have also been careful to make plots square if and only if the  $x$ - and  $y$ -axes are on the same scale (as in Figure 2).

Baselines are an important tool in focusing comparisons. For most of our plots, zero is a relevant baseline—this could correspond to parameter values of zero, standard errors of zero, probabilities of zero, or whatever. In some examples, additional baselines are relevant, and they can be shown by dotted lines; for example, Figure 6 displays estimated odds ratios, for which 0 means zero risk and 1 means no effect.

Within any plot, we align key comparisons side by side. Where possible, we connect these by lines (see Figures 3–5) except where it seems to us that connecting lines would confuse an otherwise-clear picture. We still rely on subjective aesthetic judgments at this point.

We also find useful many standard methods from data display. In general, we keep lines thin, avoid unnecessary shading and distractions, keep symbols distinct, jitter discrete values on scatterplots, and make axis labels readable. When labels are necessary, we put them directly on the graph rather than using a code that requires the reader to go back and forth between the legend and the graph (see, e.g., Tufte 1983; Cleveland 1985). As noted, the most crucial tool is probably the juxtaposition of many small plots into a single figure, as recommended by Tufte (1990).

Having written this article and made these graphs, we are in a much better position to apply these ideas in other settings. In fact, while this article was in preparation, one of us was inspired to remake into graphs all three of the tables in an unrelated research manuscript (Berkhof, Van Mechelen, and Gelman 2001). The collaborators on this other project were very pleased with the new displays and judged them to present the material much more clearly and in less space.

## 2.3 Using a Graphics Package

In preparing this article, we have found the statistical package S (or its freeware version, R, or its enhanced version, S-Plus) to be a particularly flexible and powerful tool for statistical graphics (Becker, Chambers, and Wilks 1988; R project 2000; MathSoft 2000). S (or R or S-Plus) works well in exploratory analysis—it is easy to make scatterplots, add material to existing plots, use

Raters' characterization of sentences (absolute score range)		Percent
Negative (1–1.9)		23.9
Neutral (2–2.9)		52.2
Positive ( $\geq 3$ )		23.9
Total		100.0

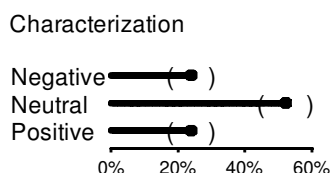


Figure 1. Top panel: Table from Ellenberg (2000) shows the relative frequencies of three categories in a set of 46 ratings of sentences. Bottom panel: Graphical display allows direct comparison without distractions of irrelevant decimal places. Parentheses show  $\pm 1$  standard error bounds based on the implicit binomial distribution with  $n = 46$ .

color, and display multiple plots on a single screen—and also in the presentation stage, where it is easy, after a little bit of practice, to alter axis labels, add text to plots, and do the other little bits that make a graph readable. Other programs could also work well if they have this level of flexibility. Microsoft Excel is an example of a package that does *not* allow enough user control to consistently make readable and concise graphs from tables.

### 3. EXAMPLES

In this section, we show how we would take seven of the tables in the March 2000 issue of *JASA* and change them into figures. We consider the motivations for the original tables and the details of how to display the information graphically, where in many cases we apply general principles of graphical display as discussed in books such as Tufte (1983), Cleveland (1985), and Wainer (1997). We also briefly discuss the other tables in the journal issue, since we do not have the space to consider them all individually here.

#### 3.1 Comparing Three Percentages

The first article in the *JASA* issue is an overview of statistics (Ellenberg 2000), and its first table (reproduced as the top panel of Figure 1 here) displays the relative frequencies of three categories in a set of 46 ratings of sentences in newspaper articles. The first thing we notice is the over-precision of the stated numbers, for example, 23.9%. Based on a simple calculation of binomial variation, we would expect this number to have an associated uncertainty of  $\sqrt{0.24(1 - 0.24)/46} = 6\%$ , so any precision beyond single percentage points is clearly overkill. This problem will automatically be corrected in a graphical display, since the eye will not detect fractions of percentages in any case.

Our simple graph appears in the bottom panel of Figure 1: the lengths of the lines indicate the percentages, and they are aligned vertically for direct visual comparison. We display as a line plot, with zero as a baseline, because the magnitude of

the percentages are relevant (e.g., the proportion of “neutral” responses is approximately twice that of “negative” or “positive”). The response categories are ordered, and we keep that ordering in our display. The parentheses indicate  $\pm 1$  standard error bounds based on the binomial distribution, which makes it clear which comparisons are statistically significant. We have suppressed the  $y$ -axis and the box that would, by default, enclose the graph, because we felt those lines would be distractions. We have also shrunk the tick marks on the  $x$ -axis so that they would not interfere with the visual comparison of the three percentages. This all took a bit of work but we believe the graph shows the comparisons of interest more clearly (and in no more space) than the table.

Once we move from a table to a graph, we can think of ways to display more information in this example. The measurements appear to be on a continuous scale, and so they could be presented as a histogram with more than just three categories. There must certainly be more information on each case (e.g., the length

Profession	Frequency of recent citations	1996 total employed (1,000)	Relative frequency
Lawyers	8101	880	9.2
Economists	1201	148	8.1
Architects	1097	160	6.9
Physicians	3989	667	6.0
Statisticians	34	14	2.4
Psychologists	479	245	2.0
Dentists	165	137	1.2
Teachers (not university)	3938	4724	0.8
Engineers	934	1960	0.5
Accountants	628	1538	0.4
Computer programmers	91	561	0.2
Total	20,657	11,034	1.9

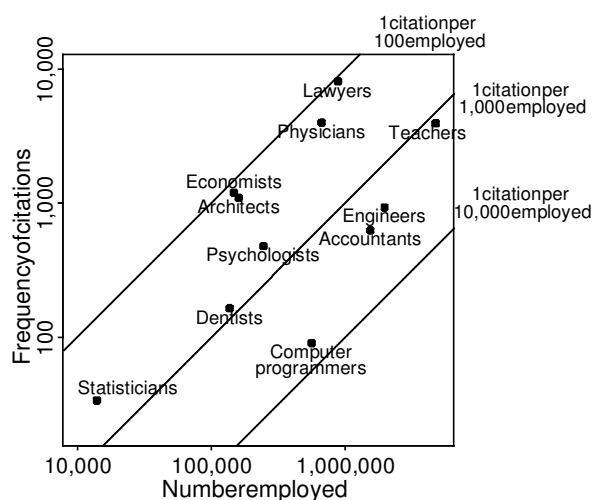


Figure 2. Top panel: Table from Ellenberg (2000) displays counts and rates of citations of various professions from the New York Times database. Bottom panel: Display as a figure shows the relative positions of the different professions much more clearly. The log-log display allows comparison across several orders of magnitude. The  $x$  and  $y$  axes are on the same scale, so that any  $45^\circ$  line indicates a constant relative frequency.

Parameter	1991					1993				
	Estimate	ASE	RSE	SSE	BSE	Estimate	ASE	RSE	SSE	BSE
Intercept	−8.921	2.125	2.055	2.194	4.591	−8.551	1.260	1.064	1.539	1.441
Day	−.230	.030	.030	.029	.054	−.269	.023	.019	.024	.026
ln(day+1)	5.668	.853	.866	.867	1.836	5.693	.555	.475	.649	.652
Rain	.890	.602	.374	.590	.514	.943	.283	.150	.293	.154
Temp	.051	.019	.017	.017	.017	.041	.012	.012	.012	.011
Wind	.030	.030	.034	.030	.048	.088	.021	.030	.022	.031
sqrt( $\hat{\phi}$ )	5.353					3.318				

Figure 3a. Table from Brumback et al. (2000) shows parameter estimates and results from comparing four different methods of estimating standard errors for a regression model.

of each sentence, its position in the newspaper, the date of its appearance), and so the 46 cases could be displayed as a scatter-plot (as we shall illustrate in Figure 7 for a different example) with the rating on the  $y$ -axis, one of the other measurements on the  $x$ -axis, and a dichotomized version of yet another measurement indicated by open and full circles. We do not display such a plot here, however, because the point of this article is not to increase the information displayed (this is the concern of the authors of the original articles) but rather to display the existing information more accessibly in a graph.

### 3.2 Comparing Ratios With Numerators and Denominators Varying Over Several Orders of Magnitude

The next table in the Ellenberg (2000) article displays the frequencies at which various professions are mentioned in the *New York Times* database in comparison with their employment in the United States. We reproduce the table as the top panel of Figure 2; it clearly shows the ranking (with lawyers at the top and computer programmers at the bottom), but it is hard to make much of the numerical frequencies except that some are below 100 and others are in the thousands.

The bottom panel of Figure 2 shows the same information compactly in a graph, using the log-log scale to capture the wide range in the data. We made the graph square and carefully put the two axes on a common scale (they do not have the same range, but a factor of 10 takes up the same distance on either axis). As a result, 45° lines correspond to constant ratios. We aid this interpretation by including reference lines corresponding to various relative frequencies. One side benefit of displaying as a graph is that we can clearly show employment and relative frequency in absolute numbers, rather than in thousands or per thousand as is necessary for the table. The graph of course also makes it much more clear where the different professions stand in reference to each other.

### 3.3 Comparing Methods for Computing Standard Errors

Ellenberg (2000) has two more tables of percentages; these data could be displayed as in Figure 1 and we do not discuss them further here. The article also includes a table of nonnumerical data—a list of names—which we would not attempt to show as a graph. The next article in the journal, Ramsay (2000), has several figures but no tables at all, so we continue to the

following article, Brumback et al. (2000), which contains three tables summarizing inferences from a set of nonlinear regression models. Here we consider the first of their tables, which reports estimates and four different estimated standard errors for each of several parameters in two instances of a generalized linear model fit.

We reproduce the table as Figure 3a. The main problem with

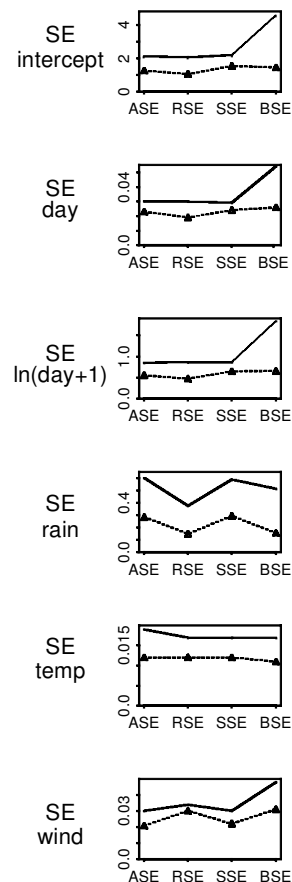


Figure 3b. Graphical display of Figure 3a from Brumback et al. (2000). The figure allows immediate comparisons of the four methods. Solid lines show standard error estimates for the 1991 regression; dotted lines show 1993. The  $y$ -axes on the plots are on different scales, but they all use zero as a reference point. Separate plots for the different parameters allow the comparisons of interest to be made without distractions.

	Penetration assumption		
	Full	3/4	1/2
Effective carrier sample size	222	182	127
prbrca	-.3359	-.4038	-.6471
SE(prbrca)	.2377	.2849	.3988
Radiation	.2073	.2059	.2032
SE(Rad)	.0600	.0599	.0599
prbrca*Radiation	.3729	.4685	.7228
SE(prbrca*Rad)	.4362	.5326	.7857

Figure 4a. Table from Iversen et al. (2000) presents the results of a Cox proportional hazards regression for three different assumptions about "penetration."

the table can be seen with a quick glance: just about the only easily noticeable pattern is that all the numbers in the first row, and some of the numbers in the third row, are greater than 1 in absolute value, and the others are less than 1. This fact is essentially irrelevant since the purpose of the table is to compare the numbers in different columns corresponding to different standard error estimates. This display violates the precepts of Ehrenberg (1975), who recommended that numbers to be compared be stacked vertically, not horizontally, in a table. We also note that unnecessary significant figures are being displayed; for example, an estimate of  $-8.921 \pm 2.125$  could just as well be summarized as  $-8.9 \pm 2.1$  (or even, arguably, as  $-9 \pm 2$ ). Rather than proceed further along these lines, however, we move to a graphical display.

The major challenge in presenting this information cleanly as a table or a graph is the variation in scales: the estimated standard errors for the different parameters vary by more than two orders of magnitude. We sidestep this problem by displaying the information for each parameter on a separate, individually scaled plot, as shown in Figure 3b. Within each plot, two lines show the different standard error estimates for the 1991 and 1993 regressions. On each plot, the  $y$ -axis extends all the way down to zero because when comparing standard errors, the absolute magnitude is relevant. For example, if several standard errors vary between 4.3 and 4.4, then a 2% change would appear large on a graph in that range, but it would still be irrelevant for practical purposes, and that would be shown by a plot that includes zero. (This argument often, but does not necessarily, hold for parameter estimates; in some applications, small relative differences can be important in practice.)

The only numbers in the table (Figure 3a) not included in Figure 3b are the parameter estimates. These could be included, either by extending the  $y$ -axes of the plots to include the parameter values, or by displaying the results as estimates  $\pm$  standard errors (as in the example shown in Figure 6). We did not do so here because we judged the key issue in the data display to be the relative values of the standard error estimates, not the parameter estimates or their statistical significance.

In this example, we suspect the best way to display the parameter estimates would be on separate plots, showing the estimates for each parameter for 1991 and 1993 under the several different models fit to these data, as described elsewhere in the Brumback et al. (2000) article.

### 3.4 Comparing Regression Coefficients Under Different Model Assumptions

Continuing through the March 2000 issue of *JASA*, the next article, Iversen, Parmagiani, Berry, and Schildkraut (2000), includes three tables presenting parameter estimates, standard errors, and other inferential summaries. We reproduce the first of their tables as Figure 4a here; it displays effective sample size and parameter estimates for a Cox regression model under three assumptions about an unknown population parameter called "penetration."

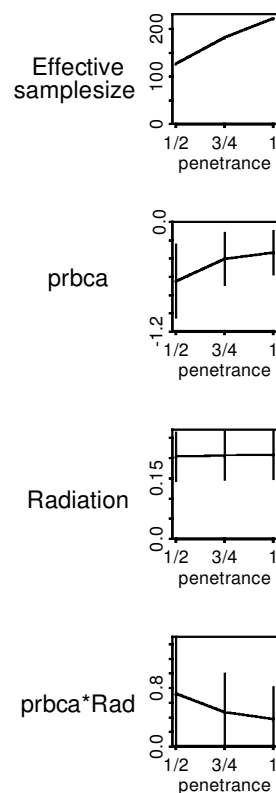


Figure 4b. Graphical display of Figure 4a (Iversen et al. 2000), showing comparisons and uncertainties directly. The  $y$ -axes on the plots are on different scales, but they all use zero as a reference point. The plots allow immediate understanding of the sizes of the estimates, their statistical significance, and how they vary as a function of the penetration assumption.

	Estimate under $H_0$	Type 1 error	Estimate under $H_A$	Standard error	$MSE^{1/2}$	Coverage
<i>Distribution A</i>						
$DIF^A$	.000	.05	.012	.004	.004	.95
$DIF^B$	-.004	.12	.01	.005	.005	.93
$LRR^A$	.00	.05	.23	.070	.07	.95
$LRR^B$	-.05	.12	.15	.070	.110	.80
<i>Distribution C</i>						
$DIF^C$	.000	.05	.015	.005	.005	.95
$DIF^B$	.000	.05	.012	.004	.005	.90
$LRR^C$	.00	.05	.29	.10	.10	.95
$LRR^B$	.00	.05	.22	.080	.100	.87
<i>Distribution AC</i>						
$DIF^{AC}$	.000	.05	.015	.005	.005	.95
$DIF^B$	-.004	.11	.012	.005	.006	.90
$LRR^{AC}$	.00	.05	.28	.09	.09	.95
$LRR^B$	-.05	.11	.16	.07	.15	.55

Figure 5a. Table from Baker (2000) compares the asymptotic performances of estimates from model B to those from models A, C, and AC, according to distributions based on each of the models.

To display this numerical information in a graph, we use a similar strategy as in our previous example:

- Isolate the individual comparisons of interest in separate plots with individually scaled  $y$ -axes.
- Use a common  $x$ -axis on all the plots to index different values of the key parameter.
- Align the plots vertically (because they have different  $y$ -axes but a common  $x$ -axis).
- Within each plot, indicate each comparison with a line graph.
- Within each plot, include zero as a  $y$ -axis boundary. (For effective sample size, zero is a natural baseline; for the parameter estimates with standard errors, separation from zero indicates statistical significance.)
- In the plots of parameter estimates, display  $\pm 1$  standard error bounds.

### 3.5 Summarizing Asymptotic Properties of a Statistical Method

The next article in the journal, Baker (2000), has a single table, reproduced here as Figure 5a, summarizing the asymptotic performances of a default model (model B) compared to the appropriate model (A, C, or AC) under three different assumptions. The models are fit to a problem in cancer screening, and the table displays results for estimated differences (DIF) and relative risks (LRR).

In designing a graphical display, we must identify the information that the table is intended to convey. The key comparison is of the statistical procedure based on the true (assumed) model, compared to the procedure based on model B. This comparison is evaluated under three different assumptions (A, C, and AC), for two outcomes (DIF and LRR), and six measures (corresponding to the six columns of the table) are being examined.

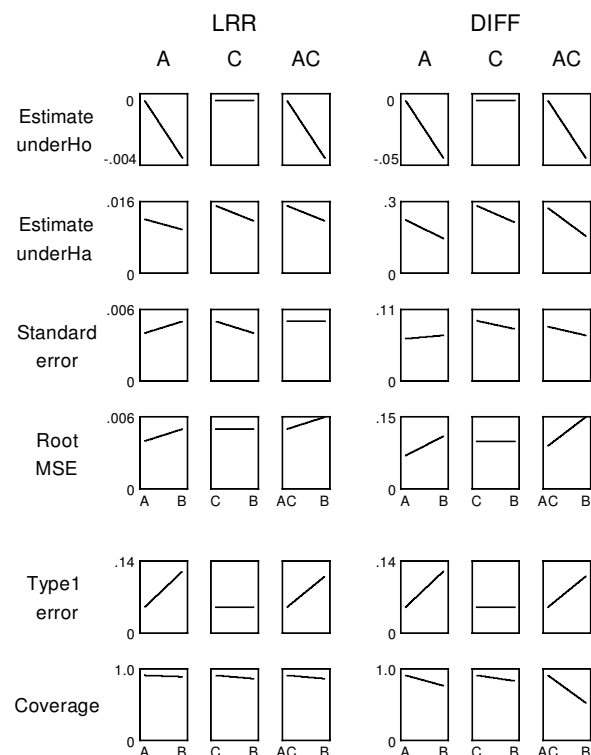


Figure 5b. Graphical display of Figure 5a with each plot showing a direct comparison between the estimate under the assumed model and under model B. Each cluster of three plots shows results for three different assumed distributions, and the two outcomes (DIF and LRR) are displayed separately. Each of the six measures are shown on separate rows, which makes sense since there is no reason for them to be directly compared. In addition, the six measures are divided into two groups: first, the four measures (estimates and standard errors) that are specific to the outcome being studied, then, the two dimensionless measures (error probability and coverage) that are on the same scale for the two outcomes.

Method	Saturated fat ( $\leq 30$ g/day; $\geq 30$ g/day)	Total energy intake (1,000 kcal/day)	Alcohol (20 g/day; $\geq 20$ g/day)	Age (10 years)
UC	.94 (.75–1.17)	.95 (.78–1.16)	1.35 (1.05–1.74)	1.60 (1.43–1.87)
RC	.78 (.33–1.80)	.80 (.27–2.38)	1.57 (1.09–2.28)	1.54 (1.42–1.67)
ML-WICI	.65 (.18–2.44)	.89 (.26–3.05)	1.55 (1.06–2.27)	1.53 (1.33–1.76)
ML-WGCI	.65 (.22–1.99)	.89 (.26–3.01)	1.55 (1.05–2.27)	1.53 (1.33–1.77)
ML-PLCI	.65 (0–2.13)	.89 (.19–3.61)	1.55 (1.07–2.60)	1.53 (1.25–1.75)
ML-RbCI	.65 (.13–3.22)	.89 (.24–3.30)	1.55 (1.05–2.28)	1.53 (1.34–1.75)
GEE a*-RbCI	.71 (.21–2.38)	.88 (.25–3.15)	1.53 (1.05–2.25)	1.53 (1.38–1.75)
GEE b-RbCI	.69 (.15–3.10)	.93 (.20–4.42)	1.52 (1.01–2.30)	1.55 (1.37–1.76)
GEE c-RbCI	.69 (.19–2.44)	.94 (.24–3.71)	1.52 (1.03–2.23)	1.55 (1.38–1.75)

Figure 6a. Table from Spiegelman et al. (1990) compares several methods of estimating odds ratios and 95% intervals in a logistic regression.

In the graphs, we thus want to align the measures for the two statistical procedures (“True” and “B”), but the six separate measures should be considered separately (there is no reason, e.g., for an estimate and a Type 1 error to be on the same scale). Similarly, the values for DIF and LRR are not necessarily directly comparable, so it makes sense to put these in separate plots. Finally, we want to see how the difference between “True” and “B” depends on the model assumption (A, C, or AC), so these should be either put into common plots or in neighboring plots with the same scale.

Figure 5b (the array of 36 plots) shows our solution: for each of the six measures and each of the two outcomes, we display the comparisons of interest in separate plots, with adjoining plots for each of the three models. We experimented with graphs with multiple lines per plot but found the separate display to be clearest. As with many other of the plots in this article, we use zero as a baseline, with a common scale for each set of three plots. The six plots in each row are divided into two sets of three to emphasize their structure, following the “small multiples” idea of Tufte (1990).

Finally, we have ordered the six measures into two clusters. The first four measures (two parameter estimates and two measures of error) are separately scaled for each outcome, with different  $y$ -axes for LRR and DIFF. The last two measures (Type 1 error and coverage) are dimensionless, and so it makes sense, for each, to use the same  $y$ -axis for LRR and DIFF.

### 3.6 Comparing Several Methods of Estimating Odds Ratios

The next article, Spiegelman, Rosner, and Logan (2000), contains two tables of inferences followed by six tables summarizing asymptotic properties. Here, we just consider their first table, shown in Figure 6a, which displays estimated odds ratios and standard errors for a model of diet and cancer. The rows of the table correspond to nine different methods for obtaining point estimates and 95% confidence intervals. It is hard to see much in this table, beyond that in the first two columns the estimated odds ratios are less than 1, with confidence intervals bracketing 1, and in the last two columns the estimated odds ratios and their confidence intervals all exceed 1. But the table is clearly intended to convey more than that.

We display this information graphically in Figure 6b. We start

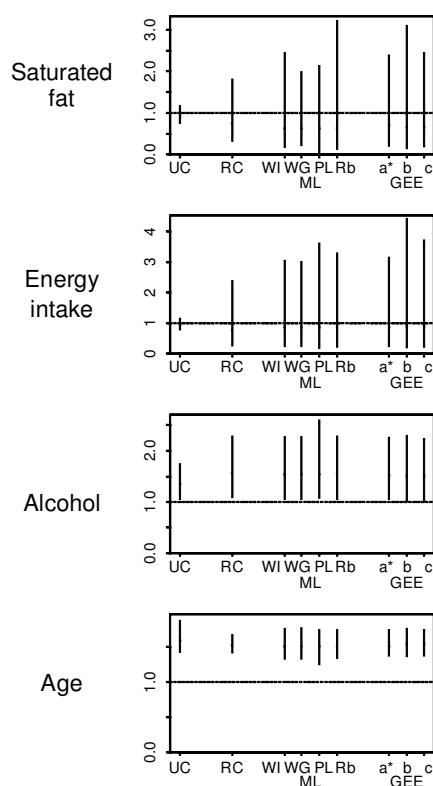


Figure 6b. Graphical display of Figure 6a separates the comparisons for the four odds ratios. For each, the estimates and standard errors under the different methods are displayed, with the different forms of ML and GEE estimates grouped. The  $y$ -axes of the graphs are on different scales, but they all include zero at the bottom and have dotted lines showing the reference odds ratio of 1.

by lining up the key comparisons for each parameter, which are the nine estimates and confidence intervals. For each graph, we include zero as a baseline (since zero risk is an important comparison) and use a dotted line to indicate an odds ratio of 1 (i.e., a null effect). As in Figure 4, we use separate plots with different scales for the different parameters in the model. Finally, we group the four ML comparisons and the three GEE comparisons into identifiable groups, which allows us to better read each plot. This layout was not trivial, even in S, which is generally a very flexible graphics package: we had to lay out the  $x$ -axis labels

ID	RSP	DOSE	DAYS
1	0	24.4	4
2	0	1.1	3
3	0	4.6	2
4	0	0.5	1
5	0	1.3	2
6	1	51.5	3
7	0	1.1	1
8	0	14.7	4
9	0	0.3	1
10	1	39.9	8
11	1	27.2	4
12	1	83.8	8
13	0	4.5	3
14	0	15.8	3
15	1	4.8	3
16	0	5.0	3
17	1	4.3	5
18	1	32.4	4
19	0	1.1	3
20	0	1.3	2
21	0	0.3	1
22	1	0.7	1
23	1	99.2	13
24	0	4.3	3
25	1	33.2	8
26	1	29.3	4
27	0	8.9	4
28	0	33.3	11
29	0	10.7	3
30	1	0.1	1
31	0	18.5	4
32	0	10.0	6
33	0	2.3	2
34	0	0.9	1
35	0	3.1	2
36	0	7.3	3
37	0	24.4	5

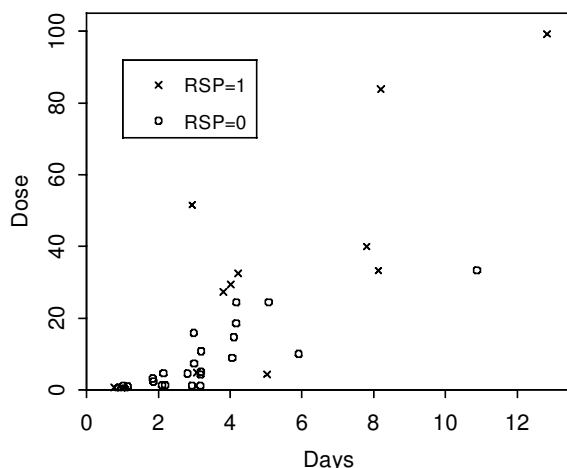


Figure 7. Top panel: Table from Mehta et al. (2000) shows data with a binary outcome and two numerical predictors. Bottom panel: Display of all the information (except the ID ordering) as a scatterplot. In addition to using less space than the table, the plot clearly shows the positions of the 0 and 1 outcomes given the two predictors. The values for Days are jittered slightly so that points will not overlap. Both x- and y-axes extend to zero, since this is a natural reference point for both Dose and Days.

so that they would be readable while still fitting in the available space.

As always, we find that compressing the graphs into a space smaller than the original table in fact increases readability because it allows the eye to see all the relevant comparisons at once.

### 3.7 A Three-Variable Data Display

We now skip through several articles that include tables of large-sample simulation results, numerical data summaries, and comparisons of inferences under different assumptions. All these tables can be made into graphs following the examples shown above. We stop at Mehta, Patel, and Senchaudhuri (2000), which includes, among other tables, a dataset reproduced here as the top panel of Figure 7. The table shows, for each of 37 cases, an ID number, a binary response, and two numerical predictors.

The bottom panel of Figure 7 reproduces these data compactly as a graph, with the 0 and 1 cases indicated by open and solid circles, respectively. The two axes extend to zero as is appropriate since this is a reasonable baseline for both “Dose” and “Days,” and the graph is deliberately *not* square since the two variables are on different scales. Since “Days” is integer-valued, we have *jittered* it slightly—that is, added a small amount of randomness to this variable (Chambers, Cleveland, Kleiner, and Tukey 1983)—to separate the points clearer without confusing their interpretation. The only data from the table that have been discarded in the graph are the ID numbers, but these do not appear to be relevant in the data or the subsequent analysis.

For the graph, it might make more sense to plot Dose on the log scale, given that it is always positive and ranges over almost three orders of magnitude; we do not do so here because the model fit by Mehta et al. (2000) includes this predictor on the original untransformed scale. We would certainly not argue against this transformation, however.

### 3.8 The Other Tables in the Journal

What of the other tables in the March 2000 issue of *JASA*? We examined each of the 64 tables in that issue and categorized them as indicated in Figure 8. As we have already indicated above, most of these tables have similar structures to those reproduced in Figures 1–7 of this article.

One of the most common kinds of table is the summary of parameter estimates and uncertainty statements (confidence intervals, standard errors, and so forth), with comparisons being made between models, datasets, or different methods of inference. We would recommend reconfiguring these as graphs, using separate plots for each parameter, lining up the comparisons of interest within each plot, with reference lines as appropriate. Examples appear in Figures 3, 4, and 6. We also see some tables of other inferential summaries (goodness-of-fit measures, *p* values), which can be similarly displayed.

The other very common category in Figure 8 is the table of frequency properties of statistical methods, either computed analytically or using simulation. We recommend displaying these sorts of results, too, in graphs (see, e.g., Figure 5), using the same divide-and-conquer approach as used in the display of inferences.

The *JASA* issue also contained some tables of raw data and data summaries. We have shown in Figures 1, 2, and 7 how



we would display some of these as graphs. Displaying more complicated datasets is more of a challenge, but we would argue that in these cases it is more important than ever to present the information graphically (see, e.g., Tukey 1977; Cleveland 1985, 1993; Wainer 1997).

The journal also includes a few tables with nonnumerical information (e.g., a list of models and formulas) which we do not attempt to display graphically.

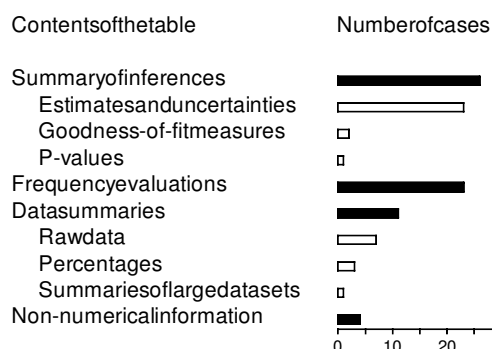


Figure 8. Categorization of the 64 tables appearing in the March, 2000, issue of the *Journal of the American Statistical Association*. The categories are listed by group, with groups and categories within each group listed in decreasing order of frequency. We think that all these tables (except the 4 tables displaying non-numerical information) could be more clearly displayed as graphs, as exemplified by Figures 1–7 of this paper.

Before writing this article, we would have presented the information in Figure 8 in a table. In constructing the figure, we have followed the principles described in this article. The first principle is to identify the key comparisons: here, they are between frequencies of the major categories (as indicated by the solid bars), and of the subcategories (the open bars). We have aligned the bars vertically for direct comparisons (as in Figure 1) and have left enough space for the category labels to be readable.

## 4. CONCLUSIONS

### 4.1 Tabular and Graphical Displays in General

We have looked at a single issue of *JASA*; how representative is this of tabular display in general? As Figure 8 shows, more than one-third of the tables in the journal were summaries of frequency evaluations, which is a task specific to statistical research. In applied work, we would expect to see many more summaries of raw data. In fact, there is already a vast and useful literature on the graphical display of data. Our contribution in this article is to show how *all* numerical displays can be shown graphically, including inferences such as estimated regression coefficients and statistical research results such as frequency evaluations that are typically displayed in tabular form.

For this article, we have constructed graphs that are comparable in size to the original tables—this is to address the possible objection that graphs require too much space to be practical. However, we are aware that a graph can require more explanation than a table, especially for unfamiliar users. For both the creator and the reader of the display, we find that, in general, graphs require a larger initial investment of effort than tables.

However, once the graph is understood, it can be used to make comparisons in a way that is nearly impossible from a table. For example, the numbers in the tabular form of Figure 3 have essentially no practical meaning until they are compared with each other. For another example, the complexity of the matrix of graphs in Figure 5 is inherently related to the multidimensionality of the summaries. The apparent simplicity of the original table masks the comparisons that need to be made for its results to be relevant to the reader.

### 4.2 Summary

This article is certainly not intended as a criticism of the authors of the articles in the May 2000 issue of *JASA*. All of us, as statisticians, occasionally take the easy way out and include unwieldy and essentially unreadable tables in our articles. Certainly the current authors are no exception (for a particularly ugly example, see Table 1 of Gelman, Bois, and Jiang 1996). One thing we have learned in this research is that there is a good reason to be lazy—it takes a lot of work to make nice graphs! But the other, more important, thing we’ve learned is that nice graphs are possible, especially when we think hard about why we want to display these numbers in the first place.

A common first reaction to rants about data display is: Yeah, yeah, I know that already. But, as we have seen, we—the statistical profession—don’t know it yet! If the world’s leading statistical journal doesn’t do it right, there is obviously still room for progress. And the first steps are recognizing why the problem exists and forming a step-by-step plan to fix it. The steps we recommend, for most data displays in statistical research, are: (1) identify the key comparisons of interest; (2) display these on small individual plots with comparison lines or axes where appropriate; and (3) establish enough control over the graphical display so that small legible plots can be juxtaposed as necessary.

The graphs in this article are extremely simple, and serious exploratory data analysis benefits from many elaborations, including color, multiple linked (trellis) plots, and dynamic displays (see, for a start, Chambers et al. 1983; Cleveland 1985, 1993). We have included only the simple plots here because these appear to be the most appropriate replacements for the tables we saw in the *JASA* articles. We expect that as statistical researchers begin to think more seriously about displaying their results as comparisons, they will begin to display more complex summaries for which more sophisticated graphical displays. This article provides “proof of concept” that statistical papers can benefit by changing tables into graphs.

[Received March 2001. Revised January 2002.]

## REFERENCES

- Baker, S. G. (2000), “Analyzing a Randomized Cancer Prevention Trial With a Missing Binary Outcome, An Auxiliary Variable, and All-or-Nothing Compliance,” *Journal of the American Statistical Association*, 95, 43–50.
- Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988), *The New S Language: A Programming Environment for Data Analysis and Graphics*, Pacific Grove, CA: Wadsworth.
- Berkhof, J., Van Mechelen, I., and Gelman, A. (2001), “A Bayesian Approach to the Selection and Testing of Mixture Models,” Technical report, Department of Psychology, Katholieke Universiteit Leuven, Belgium.
- Brumback, B. A., Ryan, L. M., Schwartz, J. D., Neas, L. M., Stark, P. C., and Burge, H. A. (2000), “Transitional Regression Models, With Application to

- Environmental Time Series," *Journal of the American Statistical Association*, 95, 16–27.
- Carter, L. F. (1947), "An Experiment on the Design of Tables and Graphs Used for Presenting Numerical Data," *Journal of Applied Psychology*, 31, 640–650.
- Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983), *Graphical Methods for Data Analysis*, Pacific Grove, CA: Wadsworth.
- Cleveland, W. S. (1984), "Graphs in Scientific Publications," *The American Statistician*, 38, 261–269.
- (1985), *The Elements of Graphing Data*, Monterey, CA: Wadsworth.
- (1987), "Research in Statistical Graphics," *Journal of American Statistical Association*, 82, 419–423.
- (1993), *Visualizing Data*, Summit, NJ: Hobart Press.
- Coll, R. A., Coll, J. H., and Thakur, G. (1994), "Graphs and Tables: A Four-Factor Experiment," *Communications of the ACM*, 37, 76–86.
- Cook, R. D., and Weisberg, S. (1999), "Graphs in Statistical Analysis: Is the Medium the Message?" *The American Statistician*, 53, 29–37.
- Croxtan, F. E., and Stein, H. (1932), "Graphic Comparisons by Bars, Squares, Circles, and Cubes," *Journal of the American Statistical Association*, 27, 54–60.
- Day, R. A. (1994), *How to Write and Publish a Scientific Paper*, Phoenix, AZ: Oryx Press.
- Dibble, E. (1997), *The Interpretation of Tables and Graphs*, Seattle, WA: University of Washington.
- Eels, W. C. (1926), "The Relative Merits of Circles and Bars for Representing Component Parts," *Journal of the American Statistical Association*, 21, 119–132.
- Ehrenberg, A. S. C. (1975), *Data Reduction: Analysing and Interpreting Statistical Data*, New York: Wiley.
- Ellenberg, J. H. (2000), "Statisticians' Significance," *Journal of the American Statistical Association*, 95, 1–8.
- Feliciano, G. D., Powers, R. D., and Kearn, B. E. (1963), "The Presentation of Statistical Information," *Audio Visual Communication Review*, 11, 32–39.
- Gelman, A., Bois, F. Y., and Jiang, J. (1996), "Physiological Pharmacokinetic Analysis Using Population Modeling and Informative Prior Distributions," *Journal of the American Statistical Association*, 91, 1400–1412.
- Gelman, A., and Price, P. N. (1999), "All Maps of Parameter Estimates are Misleading," *Statistics in Medicine*, 18, 3221–3234.
- Gillan, D., Wickens, C., Hollands, J. G., and Carswell, C. M. (1998), "Guidelines for Presenting Quantitative Data in HFES Publications," *Human Factors*, 40, 28–41.
- Iversen, E. S., Parmigiani, G., Berry, D. A., and Schildkraut, J. M. (2000), "Genetic Susceptibility and Survival: Application to Breast Cancer," *Journal of the American Statistical Association*, 95, 28–42.
- Jarvenpaa, S. L., and Dickson, G. W. (1988), "Graphics and Managerial Decision Making: Research-Based Guidelines," *Communications of the ACM*, 40, 764–774.
- Lewandowsky, S., and Spence, I. (1989), "The Perception of Statistical Graphs," *Sociological Methods and Research*, 18, 200–242.
- Loftus, G. R. (1993), "Editorial comment," *Memory and Cognition*, 21, 1–3.
- Lohse, G. L. (1993), "A Cognitive Model for Understanding Graphical perception," *Human-Computer Interaction*, 8, 353–358.
- Lusk, E. J., and Kersnick, M. (1979), "Effects of Cognitive Style and Report Format on Task Performance: The MIS Design Consequences," *Management Science*, 25, 787–798.
- MathSoft (2000), S-Plus, <http://www.splus.mathsoft.com/>
- Mehta, C. R., Patel, N. R., and Senchaudhuri, P. (2000), "Efficient Monte Carlo Methods for Conditional Logistic Regression," *Journal of the American Statistical Association*, 95, 99–108.
- Meyer, D. E., Kuskin, S. M., and Gopher, D. (1999), "Information Structure and the Relative Efficacy of Tables and Graphs," *Human Factors*, 41, 570–587.
- Meyer, J., Shinar, D., and Leiser, D. (1997), "Multiple Factors that Determine Performance with Tables and Graphs," *Human Factors*, 39, 268–286.
- R Project (2000), "The R Project for Statistical Computing," <http://www.r-project.org/>.
- Ramsay, J. O. (2000), "Functional Components of Variation in Handwriting," *Journal of the American Statistical Association*, 95, 9–15.
- Remus, W. (1984), "An Empirical Investigation of the Impact of Graphical and Tabular Data Presentations on Decision Making," *Management Science*, 30, 533–542.
- Spence, I., and Lewandowsky, S. (1991), "Displaying Proportions and Percentages," *Applied Cognitive Psychology*, 5, 61–77.
- Spiegelman, D., Rosner, B., and Logan, R. (2000), "Estimation and Inference for Logistic Regression With Covariate Misclassification and Measurement Error in Main Study/Validation Study Designs," *Journal of the American Statistical Association*, 95, 51–61.
- Stock, D., and Watson, C. J. (1984), "Human Judgment Accuracy, Multidimensional Graphics, and Humans Versus Models," *Journal of Accounting Research*, 22, 192–206.
- Tufte, E. R. (1983), *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press.
- (1990), *Envisioning Information*, Cheshire, CT: Graphics Press.
- Tukey, J. W. (1972), "Some Graphic and Semigraphic Displays," in *Statistical Papers in Honor of George W. Snedecor*, ed. T. A. Bancroft, Ames, IA: Iowa State University Press, pp. 293–316.
- (1977), *Exploratory Data Analysis*, New York: Addison-Wesley.
- Vernon, M. D. (1952), "The Use and Value of Graphical Material in Presenting Quantitative Data," *Occupational Psychology*, 26, 22–34.
- Wainer, H. (1997), *Visual Revelations*, New York: Springer-Verlag.
- Washburne, J. N. (1927), "An Experimental Study of Various Graphic, Tabular and Textual Methods of Presenting Quantitative Material," *Journal of Educational Psychology*, 18, 361–376.
- Wilkinson, L. (1999), *The Grammar of Graphics*, New York: Springer-Verlag.
- Zmud, R. W. (1978), "An Empirical Investigation of the Dimensionality of the Concept of Information," *Decision Sciences*, 9, 187–195.