Understanding Graphs and Tables

HOWARD WAINER

Quantitative phenomena can be displayed effectively in a variety of ways, but to do so requires an understanding of both the structure of the phenomena and the limitations of candidate display formats. This article (a) recounts three historic instances of the vital role data displays played in important discoveries, (b) provides three levels of information that form the basis of a theory of display to help us better measure both display quality and human graphicacy, and (c) describes three steps to improve the quality of tabular presentation.

Educational Researcher, Vol. 21, No. 1, pp. 14–23.

Although there have been many contributors to the development of graphical methods for the depiction of data, William Playfair (1759–1823) was the most influential of innovators. He was a popularizer and propagandist whose inventions found immediate acceptance because they worked so well. In his own, somewhat immodest, words,

> I found the first rough draft gave me a better comprehension of the subject, than all that I had learnt from occasional reading, for half my lifetime. (Advertisement on prelims, *An Inquiry*, 1805)

The unrelenting forcefulness inherent in the character of a good graphic presentation is its greatest virtue. We can be forced to discover things from a graph without knowing in advance what we were looking for.

How Graphics Have Given Rise to Discoveries

There are many examples of important discoveries in which graphics have played a vital role. From these I have selected three to present here. I chose a strategy that may appear to be overkill in order to counteract the common misunderstanding of the role of graphs in theory development. Tilling (1975), in a history of experimental graphs, restates this misconception:

> Clearly an ability to plot an experimental graph necessarily precedes an ability to analyze it. However, although any map may be considered as a graph, and carefully constructed maps had been in use long before the eighteenth century, we do not expect the shape of a coastline to follow a mathematical law. Further, although there are a great many physical phenomena that we do expect to follow mathematical laws, they are in general so complex in nature that direct plotting will reveal little about the nature of those laws. . . . (p. 193)

Example 1

Attitudes like these have hindered appropriately serious regard for such theories as that of continental drift (see Figure 1), whose initial evidence (noticed by every school child) is solely graphical.

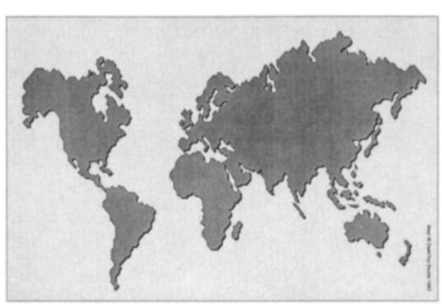


***FIGURE 1.* A familiar map projection that fairly screams "continental drift."**

The Source of a Cholera Epidemic

Dr. John Snow plotted the locations of deaths from cholera in central London in September of 1854 (see Figure 2). Deaths were marked by dots, and, in addition, the area's 11 water pumps were located by crosses. Snow observed that nearly all of the cholera deaths were among those who lived near the Broad Street pump. But before he could be sure that he had discovered a possible causal connection, he had to understand the deaths that had occurred nearer some other pump. He visited the families of 10 of the deceased. Five of these, because they preferred its taste, regularly sent for water from the Broad Street pump. Three others were children who attended a school near the Broad Street pump. On September 7, Snow described his findings before the vestry of St. James Parish. The graphic evidence was sufficiently convincing for them to allow him to have the handle of the contaminated pump removed. Within days the neighborhood epidemic that had taken more than 500 lives ended.[1]

At the time Snow did his investigation, very little was known about the vectors of contagion of disease. Theories of "foul vapors" and "divine retribution" were still considered viable. The map that resulted from Snow's methodical work did not uncover the bacterium *Vibrio cholerae*, which current theories consider cholera's cause, but it drew the causal connection between the transmission of cholera and drinking

HOWARD WAINER *is Principal Research Scientist, Educational Testing Service (21-T), Princeton, NJ 08541. His areas of specialization include statistical graphics and psychometrics.*

***FIGURE 2.* The map of a section of London that was drawn by John Snow in 1854 showing cholera deaths and water pumps. It is often used as a landmark in epidemiology.**

***Note.* From *The Visual Display of Quantitative Information* (p. 24), by Edward R. Tufte, 1983, Copyright 1983 by Edward R. Tufte. Reprinted by permission.**

from the Broad Street pump. His work is often cited as an early example of what has grown into modern epidemiology.

Armoring Airplanes

Abraham Wald, in some work he did during World War II that has only recently become available (Mangel & Samaniego, 1984; Wald, 1980), was trying to determine, on the basis of the pattern of bullet holes in returning aircraft, where to add extra armor to planes. His conclusion was to carefully determine where returning planes had been shot and *to put extra armor everyplace else!*

Wald made his discovery by drawing an outline of a plane (crudely shown in Figure 3) and then putting a mark on it where a returning aircraft had been shot. Soon the entire plane had been covered with marks *except* for a few key areas. He concluded that since planes had probably been hit more or less uniformly, those aircraft hit in the unmarked places had been unable to return, and so those were the areas that required more armor.

Taking Graphics for Granted

Graphs are so basic to our understanding that we cannot easily imagine the world without them. This was brought home to me some years ago (Wainer, 1980a) when I was reading a technical report that examined the *London Bills of Mortality*[2] and their analysis by three early statisticians (Arbuthnot, 1710; Brakenridge, 1755, Graunt, 1662). The aim of the paper, according to Zabell (1976), was

> to see how much these writers were able to extract from the *Bills* that we might reasonably expect them to—for example, how sensitive they were to questions of data quality, data consistency and data aggregation—we deliberately avoid the use of modern statistical methods. . . and limit ourselves to what is, in effect, a simple form of data analysis. (p. 2)

The result of these simple analyses was the discovery of a variety of errors that should have been seen by these early investigators but were not. Zabell concluded:

> Although we have deliberately avoided all but the simplest of statistical tools, a remarkable amount of information can be extracted from the *Bills of Mortality*, much of it unappreciated at the time of their publication. (p. 27)

The "simple" methods of data analysis he used were graphical. Such data characteristics as clerical errors in the *Bills* literally stuck out like sore thumbs. Yet, Zabell's carefully researched work was flawed. The graphical method, on which his analysis leans so heavily, was developed after the scholars he discussed did their work. Thus despite his desire to play 18th-century scholar and to use only techniques of analysis available at the time, Zabell fell into an anachronism. This incorrect assumption is but one indication of how ubiquitous the notion of graphical depiction has become; it is hard to imagine the world without it.

Measuring Graphicacy

Graphs work well because humans are very good at seeing things.[3] A child can tell that one-third of a pie is larger than a fourth long before being able to judge that the fraction 1/3 is greater than 1/4. I used to think that this was evidence supporting the power of pie charts. I was wrong. It is because the ability to understand spatial information is so powerful that humans can do it well even with flawed graphs.

Thus you can understand my dismay when a recent headline blared, **"Only 50% of American 17-year-olds can identify information in a graph of energy sources."** If the ability to read graphs is pretty much hard-wired in, how do we explain this headline?

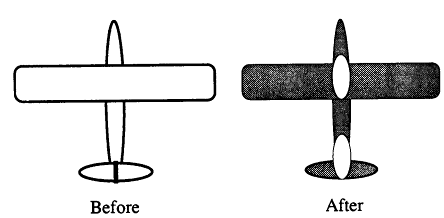


***FIGURE 3.* Abraham Wald drew his incredible conclusion about armoring airplanes only after he drew "maps" of bullet holes on returning aircraft.**

***Note.* From Wainer 1989. Copyright 1989 by the American Educational Research Association and the American Statistical Association. Reprinted by permission.**

The graphical item referred to is shown in Figure 4. The graphical item and the results associated with it were reported at the beginning of June 1990 in *From School to Work* and were taken *in toto* from one form of the *National Assessment of Educational Progress (NAEP)* that, in turn, had taken the graph from the *Annual Energy Review*. It is a flawed item in a variety of ways. If the graph were redrawn (see Figure 5), the answer to the question asked would be obvious.

Characterizing an examinee's ability to understand graphical displays on the basis of a question paired with a flawed display is akin to characterizing someone's ability to read by asking questions about a passage full of spelling and grammatical errors. What are we really testing?

One might say that we are examining whether or not someone can understand what is *de facto* "out there." I have some sympathy with this view, but what is the relationship between the ability to understand illiterate versus proper prose? If we measure the former, do we know anything more about the latter? Yet how often do we encounter well-drawn graphs in the everyday world? Should we be testing what is? Or what should be?

A more practical problem is that if a graph is properly drawn, most commonly asked questions are easily answered. That is the nature of graphics and human information-processing ability. A well-drawn graph invites deeper questions. Figure 5, for example, suggests questions about the accuracy of the obviously pre-Chernobyl predictions of the growth of nuclear power.

How can we measure someone's proficiency in understanding quantitative phenomena that are presented in a graphical way (an individual's graphicacy)? There are test items written that purport to do exactly this; the item in Figure 4 is an all too typical example. We can do better with the guidance of a formal theory of graphic communication. What follows is an expansion of a theory proposed more than a decade ago (Wainer, 1980b).

Rudiments of a Theory of Graphicacy

Fundamental to the measurement of graphicacy is the broader issue of what kinds of questions graphs can be used to answer. These are my revisions of Bertin's (1973) three levels of questions:

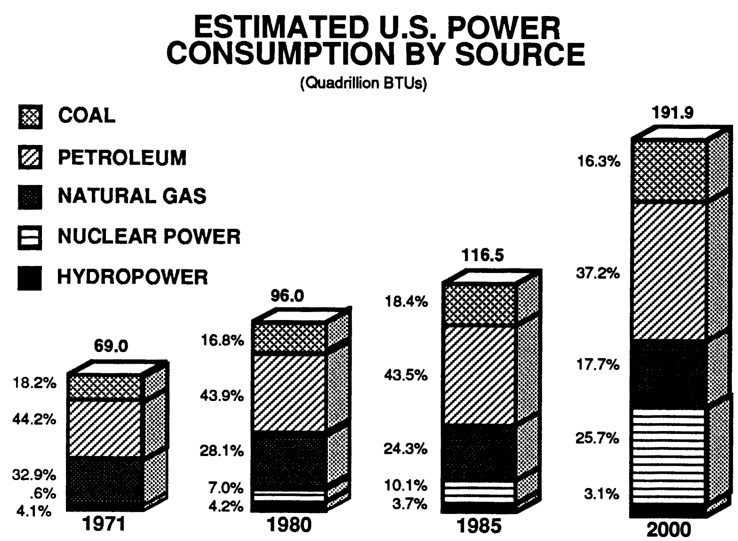
- Elementary level questions involve data extraction, for example, "What was petroleum use in 1980?"
- Intermediate level questions involve trends seen in parts of the data, for example, "Between 1970 and 1985 how has the use of petroleum changed?"
- Overall level questions involve an understanding of the deep structure of the data being presented in their totality, usually comparing trends and seeing groupings, for example, "Which fuel is predicted to show the most dramatic increase in use?" or "Which fuels show the same pattern of growth?"

The three levels are often used in combination; for example, Zabell referred to their use in the detection of outliers—unusual data points. To accomplish this objective, we need a sense of what is usual (e.g., a trend = level 2), and then we look for points that do not conform to this trend (level 1).

Note that although these levels of questions involve an increasingly broad understanding of the data, they do not necessarily imply an increase in the empirical difficulty of the questions.[4]

The epistemological basis of this formulation was clearly stated by the Harvard mathematician and philosopher Charles Sanders Peirce (1891). He felt that all things could be ordered into monads, dyads, and triads, which he often characterized as firstness, secondness, and thirdness.

Firstness considers a thing all by itself—for example, redness. Secondness considers one thing in relation to another—for example, a red apple. Thirdness concerns two things "mediated" by a third—for example, an apple falling from a tree. The tree and the apple are linked by the relation "falling from." Peirce applied firstness, secondness, and thirdness to every branch of philosophy. There is no need, he argued, to go on to fourthness or fifthness, and so on, because in almost every case these higher relations can be reduced to combinations of firstness,



In the year 2000, which energy source is predicted to supply less power than coal?

A Petroleum
B Natural Gas
C Nuclear Power
D Hydropower
E I don't know

Source: U.S. Department of Interior United States Energy Through the Year 2000
BTU: Quantity of heat required to raise temperature of one pound of water one degree Fahrenheit

COPYRIGHT 1973 CONGRESSIONAL QUARTERLY, INC.

FIGURE 4. **A graphical item as it appeared as part of the National Assessment of Educational Progress.**

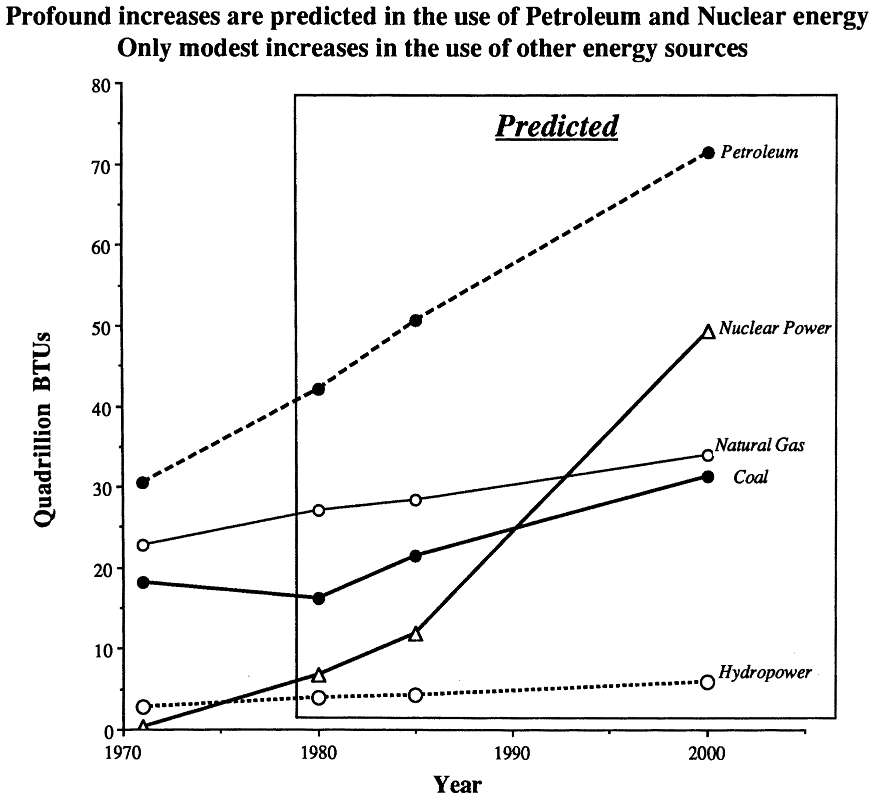


FIGURE 5. **Redrawing Figure 4 makes the correct answer to the item obvious.**

secondness, and thirdness. On the other hand, genuine thirdness can no more be reduced to secondness than can genuine secondness to firstness.[5]

Peirce traces the origins of this architecture of theory to Kant's *Critique of Pure Reason*, but enough is uniquely Peirce's to credit him as its progenitor. One can think about it linguistically as firstness being like a noun, secondness like adjective-noun combinations, and thirdness as including a verb. Once again we can see that each level cannot be constructed from a lower one and that we have no need for a concept of fourthness or more. How does this apply to the measurement of graphicacy?

Reading a graph at the intermediate level is clearly different from doing so at the elementary level; a concept of trend requires the notion of connectivity. If the horizontal axis in Figure 5 were not four years but instead four countries ordered alphabetically, the idea of an increasing trend would be meaningless. Comparing trends among different fuels likewise requires an additional notion of connectivity, but this time across the dependent variable (BTUs). This connectedness is characterized by a common vertical axis.

This theory makes explicit the limitations of double y-axis graphs. Consider the plot shown in Figure 6 from the May 14, 1990, issue of *Forbes* magazine. It purports to show that while per pupil expenditures for education have gone up precipitously over the last decade, student performance (as measured by mean SAT scores) has not responded. The conclusion, of course, is that we ought not waste our money on education. The author is asking us to make an observation of the third kind (a comparison of trends) when the lack of a common y-scale does not support it. By manipulating the two y-axes separately, we can make the graph tell exactly the opposite story (Figure 7).

I hope that this brief introduction conveys a sense of how this formal structure can make it easier to construct tests of graphicacy and to understand better which characteristic of graphicacy we are measuring. Of course, to ask questions at higher levels requires data of

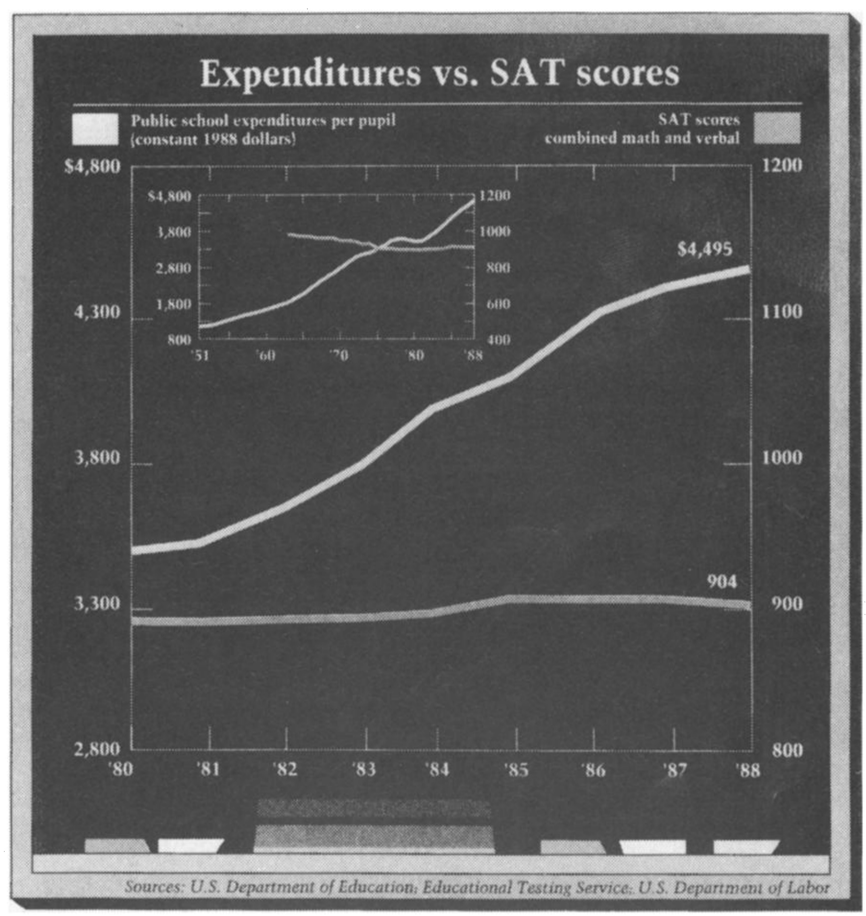


FIGURE 6. **A "double *y*-axis" graph drawn by the artists at *Forbes* magazine is but one example of why this misleading format should be expunged from use.**

Note. **Reprinted by permission of FORBES magazine, May 14, 1990. © Forbes Inc., 1990. (Vol. 145, No. 10, p. 82.)**

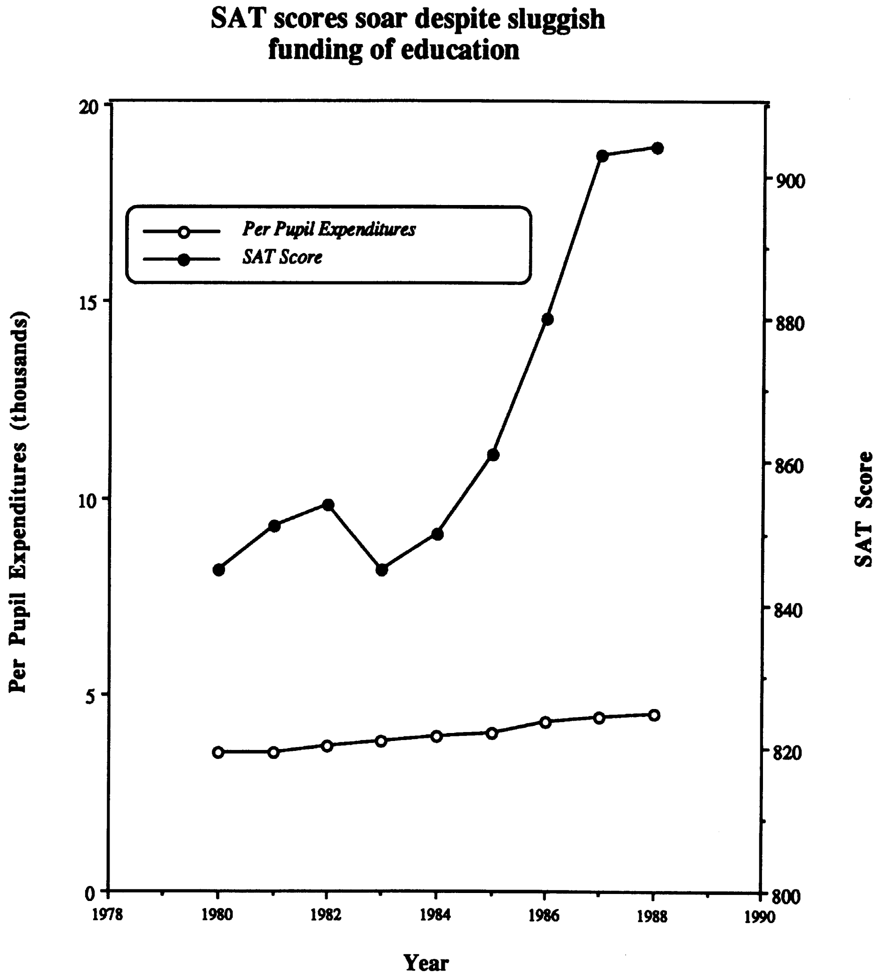


FIGURE 7. **Redrawing Figure 6 shows exactly the opposite effect. Neither inference can be properly drawn from these data.**

sufficient richness to support them, as well as graphs clear enough for the quantitative phenomena to show through.[6] I suspect that it would be much more difficult to answer second- or third-level questions from Figure 4 than from Figure 5.

My experience is that test items asociated with graphics tend to be questions of the first kind, although often they are compounded through the use of nongraphical complexity. This is not an isolated practice confined to the measurement of graphicacy. In the testing of verbal reasoning, it is common practice to make a reasoning question more difficult simply by using more arcane vocabulary. This practice stems from the unalterable fact that it is almost impossible to write questions that are more difficult than the questioner is able. When we try to test the upper reaches of reasoning ability, we must find item writers who are more clever still.

Of course, when we record a certain level of performance by an examinee on a graph-based item, we can only infer a lower bound on someone's graphicacy,[7] a better graph of the same data ought to make the item easier. Similarly, a more graphicate audience makes a graph appear more efficacious.

It is beyond my immediate purpose here to describe any specific ways of improving graphic presentation, although my suggestions for improving tables in the next section do generalize. Those interested in good graphical display are referred to Bertin (1973), Cleveland (1985), Tufte (1983), Tukey (1990), and some of my more recent works (Wainer, 1984, 1990a, 1990b, 1991a, 1991b; Wainer & Thissen, 1981, 1988). The careful reading of these works will be rewarded with increased ability to draw graphics properly, for even though there is ample evidence that the ability to understand graphically presented material is hard-wired in, there is even more evidence that the ability to draw graphs well is not. It requires instruction; remember Margerison's observation in the Prologue!

Tabular Presentation

> Getting information from a table is like extracting sunlight from a cucumber. (Farquhar & Farquhar, 1891)

The disdain shown by the two 19th-century economists quoted above reflected a minority opinion at that time. Since then the use of graphs for data analysis and communication has increased, but since Playfair's death, their quality has, in general, deteriorated. Tables, spoken of so disparagingly by the Farquhars, remain, to a large extent, worthy of contempt.

Test items involving tables are almost exclusively concerned with questions of the first kind. A typical usage[8] contains a poorly constructed table with four or five questions about specific entries. Increased difficulty is often obtained by first requiring multiple values to be extracted and then asking for algebraic manipulations of those values; thus, difficulty is not obtained by moving to a deeper level of inference but rather by requiring multiple steps at the same level. The same theoretical structure described in the section *Measuring Graphicacy* generalizes quite directly to the measurement of numeracy with tabular presentations; we extract single bits of information (firstness); we look for trends and groupings (secondness); and we make comparisons among groups (thirdness). My primary focus in this section is the improvement of tabular presentation. Toward this end I will discuss and illustrate three simple rules for the preparation of useful tables.

Driving these rules is the orientating attitude that a table is for communication, not data storage. Modern data storage is accomplished well on magnetic disks or tapes, optical disks, or some other mechanical device. Paper and print are meant for human eyes and human minds.

We begin with Table 1, which is Table 5/19 in the Bureau of the Census' well-known book *Social Indicators III* (1980). Any redesign task must first try to develop an understanding of purpose. The presentation of this data set must have been intended to help the reader answer such questions as:

1. What is the general level (per 100,000 population) of accidental death in the countries chosen?
2. How do the countries differ with respect to their respective rates of accidental death?

3. What are the principal causes of accidental death? Which are the most frequent? The least frequent?

4. Are there any unusual interactions between country and cause of accidental death?

These are obviously parallel to the questions that are ordinarily addressed in the analysis of any multifactorial table—overall level, row, column, and interaction effects.

Before going further, I invite you to read Table 1 carefully and see to what extent you can answer these four questions. But don't peek ahead!

The first rule of table construction is to:

1. *Order the rows and columns in a way that makes sense.* We are almost never interested in "Austria First." Two useful ways to order the data are:

a. Size places—Put the largest first. Often we look most carefully at what is on top and less carefully further down. Put the biggest thing first! Also, ordering by some aspect of the data often reflects ordering by some hidden variable that can be inferred.

b. Naturally—Time is ordered from the past to the future. Showing data in that order melds well with what the viewer might expect. This is always a good idea.

Table 2 is a redone version of Table 1. A few typos have been corrected, some uninformative columns removed, and the rows ordered by the total death rate. The columns were already ordered in a reasonable way and so were left unaltered. Now we can begin to answer Questions 1 and 2 above. We see that France is the most dangerous place, having an accidental death rate of about 78 per 100,000; that is more than twice that of Japan (about 30 per 100,000), which, at least by this measure, appears to be the safest country. Now that the rows are ordered, the overall death rate (taken as an unweighted median) can be easily calculated—count down eight countries—and is around 50 per 100,000.

Note that when I referred to the actual rates, I rounded. This is very important. The second rule of table construction is to:

2. *Round—a lot!* This is so for three reasons:

a. Humans cannot understand more than two digits very easily.

b. We can almost never justify more than two digits of accuracy statistically.

c. We almost never care about accuracy of more than two digits.

Let us take each of these reasons separately.

Understanding. Consider the statement that "This year's school budget is $27,329,681." Who can comprehend or remember that? If we remember anything, it is almost surely the translation, "This year's school budget is about $27 million."

Table 1

Deaths Due to Unexpected Events, by Type of Event, Selected Countries: Mid-1970's

(Rate per 100,000 population)

Country	Year[1]	Deaths due to all causes	Deaths due to unexpected events					
			Total	Transport accidents	Natural factors[2]	Accidents occurring mainly in industry[3]	Homicides and injuries caused intentionally[4]	Other causes[5]
Austria	1975	1,277.2	75.2	34.8	29.7	4.3	1.6	4.8
Belgium	1975	1,218.5	62.6	25.0	25.8	1.5	9	9.4
Canada	1974	742.0	62.1	30.9	18.0	3.9	2.5	6.8
Denmark	1976	1,059.5	41.1	18.3	15.6	1.0	7	5.5
Finland	1974	952.5	62.3	23.7	26.0	2.9	2.6	7.1
France	1974	1,049.5	77.8	23.8	31.0	1.0	9	21.1
Germany (Fed. Rep.)	1975	1,211.8	66.4	24.8	31.6	1.8	1.2	7.0
Ireland	1975	1,060.7	48.6	19.8	20.1	1.9	1.0	5.8
Italy	1974	957.8	47.2	22.8	19.2	1.9	1.1	2.2
Japan	1976	625.6	30.5	13.2	9.7	2.1	1.3	4.2
Netherlands	1975	832.2	40.3	17.8	18.2	1.0	7	2.6
Norway	1976	998.9	48.4	17.3	25.1	1.9	7	3.4
Sweden	1975	1,076.6	55.8	17.2	27.9	1.3	1.1	8.3
Switzerland	1976	904.1	48.4	20.6	20.4	2.1	9	4.4
United Kingdom	1976	1,217.9	34.8	13.0	13.9	1.3	1.1	5.5
United States	1975	888.5	60.6	23.4	15.8	2.6	10.0	8.8

[1]Most current year data available.
[2]Includes fatal accidents due to poisoning, falls, fire, and drowning.
[3]For some countries data relate to accidents caused by machines only.
[4]By another person, including police.
[5]Includes accidents caused by firearms, war injuries, injuries of undetermined causes, and all other accidental causes.

Source: United Nations, World Health Organization, *World Health Statistics Annual*, 1978, vol. I, *Vital Statistics and Cause of Death*. Copyright; used by permission.

Note. From the U.S. Bureau of the Census publication *Social Indicators III, December 1980, p. 252.*

Table 2

Table 1 With Rows Ordered by Overall Death Rate, Typographical Errors Corrected, and Uninformative Columns Removed

(Rate per 100,000 population)

Country	Total unexpected deaths	Transport accidents	Natural factors	Industrial accidents	Homicides	Other causes
France	77.8	23.8	31.0	1.0	0.9	21.1
Austria	75.2	34.8	29.7	4.3	1.6	4.8
Germany	66.4	24.8	31.6	1.8	1.2	7.0
Belgium	62.6	25.0	25.8	1.5	0.9	9.4
Finland	62.3	23.7	26.0	2.9	2.6	7.1
Canada	62.1	30.9	18.0	3.9	2.5	6.8
United States	60.6	23.4	15.8	2.6	10.0	8.8
Sweden	55.8	17.2	27.9	1.3	1.1	8.3
Ireland	48.6	19.8	20.1	1.9	1.0	5.8
Norway	48.4	17.3	25.1	1.9	0.7	3.4
Switzerland	48.4	20.6	20.4	2.1	0.9	4.4
Italy	47.2	22.8	19.2	1.9	1.1	2.2
Denmark	41.1	18.3	15.6	1.0	0.7	5.5
Netherlands	40.3	17.8	18.2	1.0	0.7	2.6
United Kingdom	34.8	13.0	13.9	1.3	1.1	5.5
Japan	30.5	13.2	9.7	2.1	1.3	4.2

Statistical justification. The standard error of any statistic is proportional to one over the square root of the sample size. God did this, and there is nothing we can do to change it. Thus, suppose we would like to report a correlation as .25. If we don't want to report something that is inaccurate, we must be sure that the second digit is reasonably likely to be 5 and not 6 or 4. To accomplish this, we need the standard error to be less than .005. But since the standard error is proportional to $1/\sqrt{n}$, the obvious algebra ($1/\sqrt{n} \sim .005 \Rightarrow \sqrt{n} \sim 1/.005 = 200$) yields the inexorable conclusion that a sample size of the order of 200^2, or 40,000, is required to justify the presentation of more than a two-digit correlation. A similar argument can be made for all other statistics.

Who cares? I recently saw a table of average life expectancies.[9] It proudly reported that the mean life expectancy of a male at birth in Australia was 67.14 years. What does the 4 mean? Each unit in the hundredth's digit of this overzealous reportage represents 4 days. What purpose is served in knowing a life expectancy to this accuracy? For most communicative (not archival) purposes 67 would have been enough.

Table 3 contains a revision of Table 2 in which each entry is rounded to the nearest integer. Because the original entries had only one extra digit, the clarifying effect of rounding is modest. In this version of the table, the unusual homicide rate of the United States jumps out at us. At a glance, we can see that it is an order of magnitude greater than the rate found in any civilized nation. We also see an unusual entry for France under "other causes," which raises questions about definitions.

The effect of too many decimal places is sufficiently pernicious that I would like to emphasize the importance of rounding with another short example. Equation 1 is taken from *State Court Caseload Statistics: 1976:*

$$\mathrm{Ln(DIAC)} = -.10729131 + 1.00716993 \times \mathrm{Ln(FIAC)}, \quad (1)$$

where DIAC is the annual number of case dispostions, and FIAC is the annual number of case filings. This is obviously the result of a regression analysis with an overgenerous output format. Using the standard error justification for rounding, we see that to justify the eight digits shown we would need a standard error that is of the order of .000000005, or a sample size of the order of 4×10^{16}. This is a very large number of cases—the population of China doesn't put a dent in it. The actual n is the number of states, which allows one digit of accuracy at most. If we round to one digit and transform out of the log metric, we arrive at the more statistically defensible equation

$$\mathrm{DIAC} = .9\ \mathrm{FIAC}. \quad (2)$$

This can be translated into English as

> "There are about 90% as many dispositions as filings."

Obviously, the equation that is more defensible statistically is also much easier to understand. A colleague, who knows more about courts than I do, suggested that I needed to round further, to the nearest integer (DIAC = FIAC), and so a more correct statement would be

> "There are about as many dispositions as filings."

Table 3

Table 2 With Entries Rounded to Integers

(Rate per 100,000 population)

Country	Total unexpected deaths	Transport accidents	Natural factors	Industrial accidents	Homicides	Other causes
France	78	24	31	1	1	21
Austria	75	35	30	4	2	5
Germany	66	25	32	2	1	7
Belgium	63	25	26	2	1	9
Finland	62	24	26	3	3	7
Canada	62	31	18	4	3	7
United States	61	23	16	3	10	9
Sweden	56	17	28	1	1	8
Ireland	49	20	20	2	1	6
Norway	48	17	25	2	1	3
Switzerland	48	21	20	2	1	4
Italy	47	23	19	2	1	2
Denmark	41	18	16	1	1	6
Netherlands	40	18	18	1	1	3
United Kingdom	35	13	14	1	1	6
Japan	31	13	10	2	1	4

A minute's thought about the court process reminds one that it is a pipeline with filings at one end and dispositions at the other. They must equal one another, and any variation in annual statistics reflects only the vagaries of the calendar. The sort of numerical sophistry demonstrated in Equation 1 can give statisticians a bad name.[10]

The final rule of table construction is:

3. *ALL is different and important.* Summaries of rows and columns are important as a standard for comparison—they provide a measure of usualness. What summary we use to characterize *ALL* depends on the purpose. Sometimes a sum is suitable, more often a median. But whatever is chosen, it should be visually different from the individual entries and set apart spatially.

Table 4 makes it clearer how unusual the United States' homicide rate is. The column medians allow us to compare the relative danger of the various factors. We note that although "transport accidents" are the worst threat, they are closely followed by "natural factors." Looking at the entries for the United States, we can see that "natural factors" are under somewhat better control than in most other countries.

Can we go further? Sure. To see how requires that we consider what distinguishes a table from a graph. A graph uses space to convey information. A table uses a specific iconic representation. We have made tables more understandable by using space—making a table more like a graph. We can improve tables further by making them more graphical still. A semigraphical display like the stem-and-leaf diagram (Tukey, 1977) is merely a table in which the entries are not only ordered but are also spaced according to their size. To put this notion into practice, consider the last version of Table 1 shown as Table 5.

The rows have been spaced according to what appear to be significant gaps (Wainer & Schacht, 1978) in the total death rate, thus dividing the countries into five groups. Further investigation is required to understand why they seem to group that way, but the table has provided the impetus.

The highlighting of single entries points out the unusually high rate of transport accidents in Canada and Austria, as well as the unusually low rates of death due to natural factors in the United States and Canada. The determination that these values are in-

Table 5

Table 4 With Rows Spaced by Total Death Rate and Unusual Values Highlighted

(Rate per 100,000 population)

Country	**Total unexpected deaths**	Transport accidents	Natural factors	Industrial accidents	Homicides	Other causes
France	**78**	24	31	1	1	21
Austria	**75**	35	30	4	2	5
Germany	**66**	25	32	2	1	7
Belgium	**63**	25	26	2	1	9
Finland	**62**	24	26	3	3	7
Canada	**62**	31	18	4	3	7
United States	**61**	23	16	3	10	9
Sweden	**56**	17	28	1	1	8
Ireland	**49**	20	20	2	1	6
Norway	**48**	17	25	2	1	3
Switzerland	**48**	21	20	2	1	4
Italy	**47**	23	19	2	1	2
Denmark	**41**	18	16	1	1	6
Netherlands	**40**	18	18	1	1	3
United Kingdom	**35**	13	14	1	1	6
Japan	**31**	13	10	2	1	4
Median	**53**	**22**	**20**	**2**	**1**	**6**

= an unusual data value

Table 4

Table 3 With Column Medians Calculated and Total Highlighted

(Rate per 100,000 population)

Country	**Total unexpected deaths**	Transport accidents	Natural factors	Industrial accidents	Homicides	Other causes
France	**78**	24	31	1	1	21
Austria	**75**	35	30	4	2	5
Germany	**66**	25	32	2	1	7
Belgium	**63**	25	26	2	1	9
Finland	**62**	24	26	3	3	7
Canada	**62**	31	18	4	3	7
United States	**61**	23	16	3	10	9
Sweden	**56**	17	28	1	1	8
Ireland	**49**	20	20	2	1	6
Norway	**48**	17	25	2	1	3
Switzerland	**48**	21	20	2	1	4
Italy	**47**	23	19	2	1	2
Denmark	**41**	18	16	1	1	6
Netherlands	**40**	18	18	1	1	3
United Kingdom	**35**	13	14	1	1	6
Japan	**31**	13	10	2	1	4
Median	**53**	**22**	**20**	**2**	**1**	**6**

deed unusual was done by additional calculations in support of the display (subtract out row and column effects and look at what sticks out). But the viewer can appreciate the result without being aware of the calculations. Spacing tables commensurate with the values of their entries and highlighting unusual values are often useful techniques but are not as universally important as the three rules mentioned previously.

The version of Table 1 shown as Table 5 is about as far as we can go. It may be that for special purposes other modifications might help, but Table 5 does allow us to answer readily the four questions about these data phrased earlier. Some aspects are memorable. Who can forget the discovery of the gigantic disparity between the homicide rate in the United States and that of the other 15 nations reported.[11]

Conclusions

In this account I have tried to further the effective display of quantitative phenomena by accomplishing three things.

- To illustrate how effective display can help us, indeed sometimes force us, to discover what we were not expecting. I chose only three examples; there are many more.
- To aid the understanding of displays by adapting Peirce's "architecture of theory" to this context. The formalism of this theory helps to show why some sorts of common displays are unacceptable for the most plausible purposes. This same theory provides a framework for the development of measures of human graphicacy (the extent to which people understand a particular figuration) and thus helps us to avoid the erroneous conclusion fostered by such tests as that represented by the NAEP item shown in Figure 4.
- To explicitly show how the much-maligned table can be used to effectively display even rather complex phenomena. The display rules that I report owe much to Andrew Ehrenberg's (1977) advice, although he should certainly not be held responsible for where I have taken them. Note that these rules differ from what are often held to be the standards in scientific publications. APA standards, for example, frown upon the profligate use of extra spaces. I hope that those sorts of standards[12] can be modified to reflect the changing role of a table—modern electronic storage provides a far better way for archiving data—as well as to reflect what we have learned about effective display.

Of good data displays, it may be said what Mark Van Doren observed about brilliant conversationalists: "In their presence others speak well." I hope that the theory and practice illustrated here can improve the quality of our displays and thus allow our data to speak more clearly.

Notes

This article is drawn from an invited address titled "Graphical Visions" at the Annual Meeting of the American Educational Research Association given on April 4, 1991, in Chicago, Illinois. Its preparation was partially supported by funds provided to me by the Trustees of the Educational Testing Service in the 1990–1992 "Senior Scientist Award." I am pleased to be able to acknowledge their generosity. In addition, this presentation has benefited enormously from the advice and criticism of a number of my colleagues. First is my friend, colleague, sometime coauthor and, best of all, my wife, Linda S. Steinberg. In addition my thanks to David Berliner, Michael Friendly, Drew Gitomer, Samuel Livingston, Keith Reid-Green, David Thissen, Neal Thomas, Edward Tufte, and John Tukey.

[1]The Broad Street pump is now gone. In its place is the John Snow Pub. See Gilbert (1958) and Jaret (1991) for more details.

[2]In London of the 1530s parish clerks were requested to submit weekly reports on the number of plague deaths. These bills of mortality were meant to tell authorities when measures should be taken against the epidemic. In 1604 publication of the *London Bills of Mortality* by the Company of Parish Clerks began.

[3]This statement may seem vacuous, but Yogi Berra is famed for observations indistinguishable from this one. This idea was put forward in a more scholarly way by cartographers Arthur H. Robinson and Barbara Bartz Petchenik (1976), who said, "There is fairly widespread philosophical agreement, which certainly accords with common sense, that the spatial aspects of all existence are fundamental. Before an awareness of time, there is an awareness of relations in space." They conclude their book with the observation that "the concept of spatial relatedness. . . is a quality without which it is difficult or impossible for the human mind to apprehend anything."

[4]Nevertheless, one small empirical study among 3rd-, 4th-, and 5th-grade children (Wainer, 1980b) showed that, on average, item difficulty increased with level and graphicacy increased with age.

[5]This paragraph is a rather close paraphrasing of a description by Martin Gardner (1978, p. 23).

[6]*Purgamentum init, exit purgamentum.*

[7]It is like trying to decide on Mozart's worth as a composer on the basis of a performance of his works by Spike Jones on the washboard.

[8]See, for example, page 132 (items 22–25) of Form GR85-3 of the Graduate Record Exam in *Practicing to take the GRE—General Test-No. 3*, Princeton, NJ: Educational Testing Service, 1985.

[9]*UN Demographic Yearbook*, 1962.

[10]I sometimes hear from colleagues that my ideas about rounding are too radical, that such extreme rounding would be "OK if we knew that a particular result was final. But our final results may be used by someone else as intermediate in further calculations. Too early rounding would result in unnecessary propagation of error." Keep in mind that tables are for communication, not archiving. Round the numbers and, if you must, insert a footnote proclaiming that the unrounded details are available from the author. Then sit back and wait for the deluge of requests.

[11]These data are more than 15 years old, but their message certainly stayed with me enough so that when a newspaper article in the *New York Times* on August 13, 1989, reported that Detroit and Washington, DC had annual homicide rates of about 60 per 100,000, I knew enough to be horrified. Tables with memorable content can be memorable.

[12]These date back at least to 1914 and the standards published by the American Society of Mechanical Engineers. A recent update (American National Standards Institute, 1979) replaces the 1914 recommendations for pen-nib size with a specification for number of pixels, but otherwise remains remarkably the same.

References

American National Standards Institute. (1979). *American national standard illustrations for publication and projection* (ANSI Y15.1M). New York: American Society of Mechanical Engineers.

Arbuthnot, J. (1710). An argument for Divine Providence taken from the constant regularity in the births of both sexes. *Philosophical Transactions of the Royal Society*, London, *27*, 186–190.

Barton, P., & Coley, R. (1991). *Performance at the top: From elementary through graduate school.* ETS Policy Information Center, Princeton, NJ: Educational Testing Service.

Bertin, J. (1973). *Semiologie Graphique* (2nd ed.). The Hague: Mouton-Gautier. English translation by William Berg & Howard Wainer (1983) and published as *Semiology of graphics*, Madison, WI: University of Wisconsin Press.

Brakenridge, W. (1755). A letter from the Reverend William Brakenridge, D.D. & F. R. S. to George Lewis Scot, Esq., F. R. S., concerning the *London Bills of Mortality. Philosophical Transactions of the Royal Society*, London, *48*, 788–800.

Cleveland, W. S. (1985). *The elements of graphing data.* Boston: Duxbury.

Costigan-Eaves, P. (1984). *Data graphics in the 20th century: A comparison and analytic survey.* Unpublished doctoral dissertation, Graduate School of Education, Rutgers University, New Brunswick, NJ.

Educational Testing Service. *From school to work.* (1990, June). *Policy information report of the policy information center.* Princeton, NJ: Author.

Ehrenberg, A. S. C. (1977). Rudiments of

numeracy. *Journal of the Royal Statistical Society, Series A, 140*, 277-297.
Farquhar, A. B., & Farquhar, H. (1891). *Economic and industrial delusions: A discourse of the case for protection.* New York: Putnam.
Gardner, M. (1978). On Charles Sanders Peirce: Philosopher and gamesman, *Scientific American, 240*, 18-26.
Gilbert, E. W. (1958). Pioneer maps of health and disease in England. *Geographical Journal, 124*, 172-183.
Graunt, J. (1662). *Natural and political observations on the Bills of Mortality*, London: Martyn.
Jaret, P. (1991). The disease detectives. *National Geographic, 179*, 114-140.
Mangel, M., & Samaniego, F. J. (1984). Abraham Wald's work on aircraft survivability. *Journal of the American Statistical Association, 79*, 259-267.
Margerison, T. (1965, January 3). Review of *Writing Technical Reports* by Bruce M. Copper. *Sunday Times.* In R. L. Weber (Compiler) & E. Mendoza (Ed.), *A random walk in science* (p. 49). New York: Crane, Russak.
Peirce, C. S. (1891, January). The architecture of theories. *The Monist*, 161-176.
Playfair, W. (1786). *The commercial and political atlas.* London: Corry.
Playfair, W. (1805). *An inquiry into the permanent causes of the decline and fall of powerful and wealthy nations.* London: Greenland & Norris.
Priestly, J. B. (1765). *A chart of biography.* London.
Priestly, J. B. (1769). *A new chart of history.* London.
Robinson, A. H., & Petchenik, B. B. (1976). *The nature of maps: Essays toward understanding maps and mapping.* Chicago: University of Chicago Press.
Social indicators III. (1980). Washington, DC: U.S. Bureau of the Census.
State court caseload statistics: 1976. (1976). Williamsburg, VA: National Center for State Courts.
Tilling, L. (1975). Early experimental graphs. *British Journal of Historical Science, 8*, 193-213.
Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.
Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.
Tukey, J. W. (1990). Data based graphics: Visual display in the decades to come. *Statistical Science, 5*, 327-329.
UN Demographic Yearbook. (1962). New York: United Nations Press.
Wainer, H. (1980a). A timely error. *Royal Statistical Society News and Notes. 7*, 6.
Wainer, H. (1980b). A test of graphicacy in children. *Applied Psychological Measurement, 4*, 331-340.
Wainer, H. (1984). How to display data badly. *The American Statistician, 38*, 137-147.
Wainer, H. (1989). Eelworms, bullet holes, and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. *Journal of Educational Statistics*, 14, 121-140.
Wainer, H. (1990a). Graphical visions from William Playfair to John Tukey. *Statistical Science, 5*, 340-346.
Wainer, H. (1990b). Measuring graphicacy. *Chance, 3*, 52 & 58
Wainer, H. (1991a). Double Y-axis graphs. *Chance, 4*, 50-51.
Wainer, H. (1991b). Humble pie. *Chance*, 4, 52-53.
Wainer, H., & Schacht, S. (1978). Gapping. *Psychometrika*, 43, 203-212.
Wainer, H., & Thissen, D. (1981). Graphical data analysis. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual Review of Psychology* (pp. 191-241). Palo Alto, CA: Annual Reviews. A revision reprinted in G. Keren & C. Lewis (Eds.), *Methodological and quantitative issues in the analysis of psychological data.* Hillsdale, NJ: Erlbaum, 1992.
Wainer, H., & Thissen, D. (1988). Plotting in the modern world: Statistics packages and good graphics. *Chance, 1*, 10-20.
Wald, A. (1980, July). A method of estimating plane vulnerability based on damage of survivors (CRC 432). (These are reprints of work done by Wald while a member of Columbia's Statistics Research Group during the period 1942-1945. Copies can be obtained from the Document Center, Center for Naval Analyses, 2000 N. Beauregard St., Alexandria, VA 22311.)
Zabell, S. (1976). Arbuthnot, Heberden and the *Bills of Mortality* (Tech. Rep. No. 40). Chicago: Department of Statistics, The University of Chicago.