

## Improving Graphic Displays by Controlling Creativity

“Those who cannot remember the past are condemned to repeat it.”

— George Santayana (1863–1952)

I could not help but think of George Santayana’s observation about the importance of knowing the past while I was reading a 15-page letter that John Tukey wrote to Linda Pickle in May of 1998 (see [www.amstat.org/publications/chance](http://www.amstat.org/publications/chance) and click on “Supplemental Material”). The letter was his response to the *Atlas of United States Mortality* that Pickle and her colleagues at the National Center for Health Statistics published in 1996. He said, “It is by far the best job of this sort that I have seen. It probably deserves a grade of between 94 and 98 out of 100.” He then spent the rest of his note with suggestions for improvement. But, before he turned to those suggestions, he added the following:

Things like the *Atlas* evolve over a substantial period of time, and I have no reason to believe that a document responsive to the emendations that follow (be they good or bad) would represent the end of the evolutionary process. Whatever the next step of advance may be, once taken, will, I trust with near certainty, open up our thinking to new possibilities beyond those we have so far imagined.

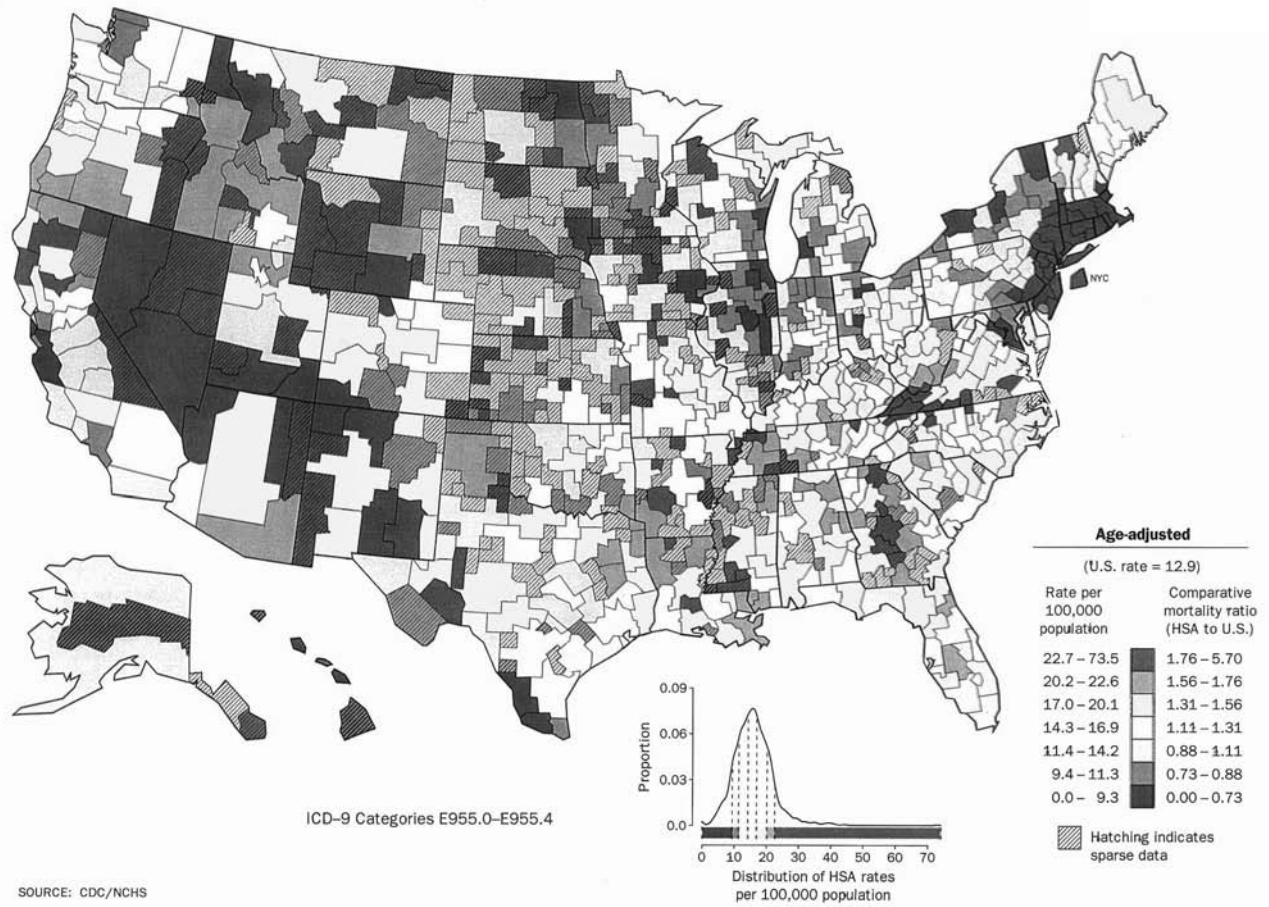
These are wise words, indeed, and suggest a pathway of evolution for statistical reports in which the direction of improvement is likely to be monotonic, with only small local variants.

I will not attempt to describe all the design decisions that went into the *Atlas* that Tukey gave such a high grade, but instead recommend to all who have not yet had the pleasure to immediately run out and get a copy of their own. I will, however, focus on one characteristic of that fine report, germane to my topic today. Specifically, improving the quality of a report through the tight control of creativity.

### Example 1. *The Atlas of U.S. Mortality*

The *Atlas* has a very regular structure. Its body is made up of 18 sections, each concerned with mortality from a specific cause (e.g., cancer, stroke, motor vehicle injuries, diabetes, firearms, etc.). Then, each section, in its turn, is subdivided into four sections corresponding to white females, black females, white males, and black males. Each of those is comprised of a large colored map whose HSAs (Health Service Areas) are shaded and colored to represent the age-adjusted death rates for that cause and demographic group. On the facing page are three smaller maps showing the variations for ages 40 and 70 and a comparison of the death rates compared with the overall U.S. rate. There is also a small plot showing regional variation. The first chapter of the *Atlas* goes over this format, also carefully explaining the various statistical adjustments and smoothings that yielded the figures presented.

Obviously, an enormous amount of attention went into the basic design and, I assume, what we see is the result of endless discussion and compromises. Once the design was set, it was then repeated exactly from chapter to chapter. This makes it easier on the reader who has only to master the design once and can then read the *Atlas* through with no further education. The *Atlas* accomplishes this so gracefully that the reader can be blissfully unaware of how hard it must have been to control rampant creativity. I am sure someone must have argued desperately to use a histogram made up of miniature Colt 45s to indicate firearms deaths. But, happily, editorial wisdom has shaded our eyes from such creative brilliance.

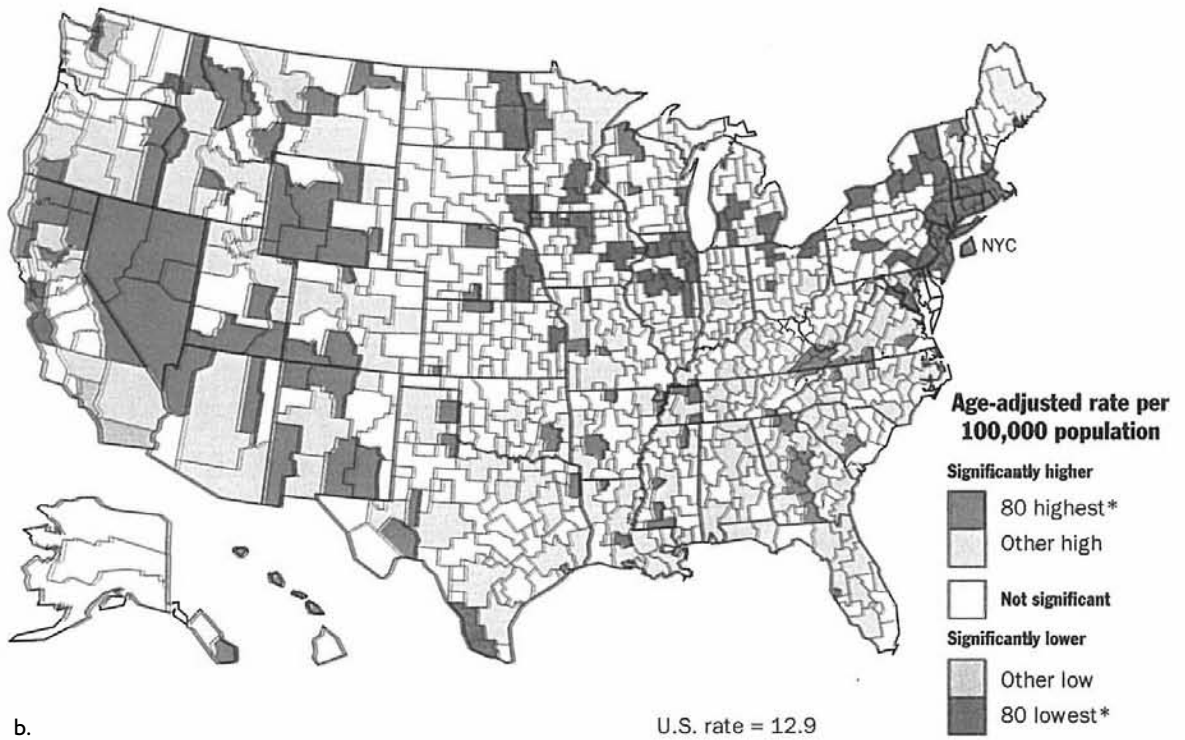


a.

Figure 1a-e. Age-adjusted death rates for firearm suicides for white males, 1988-1992. These displays were on facing pages and constituted the standard format for the entire atlas. (To view the color version of the *Atlas*, go to [www.amstat.org/publications/chance](http://www.amstat.org/publications/chance) and click on "Supplemental Material.")

DEATH RATES OF EACH HSA  
 COMPARED WITH U.S. RATE

FIREARM SUICIDE  
 WHITE MALE



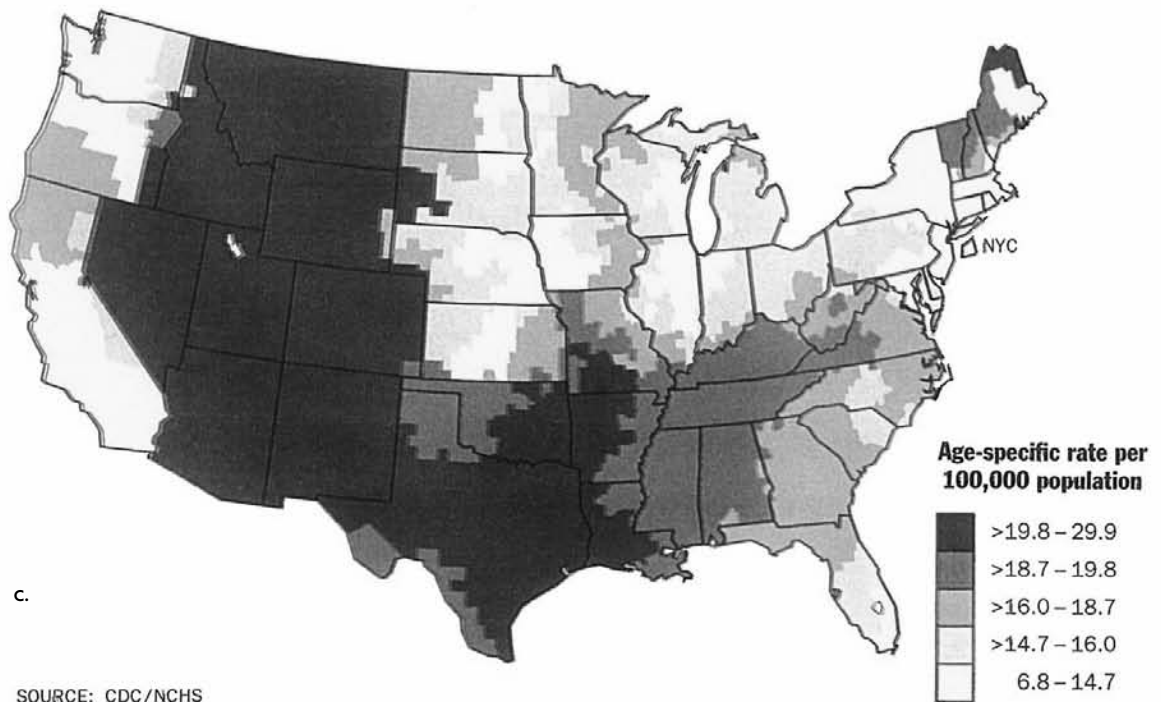
b.

SOURCE: CDC/NCHS

\* See text

SMOOTHED DEATH RATES  
 FOR AGE 20

FIREARM SUICIDE  
 WHITE MALE

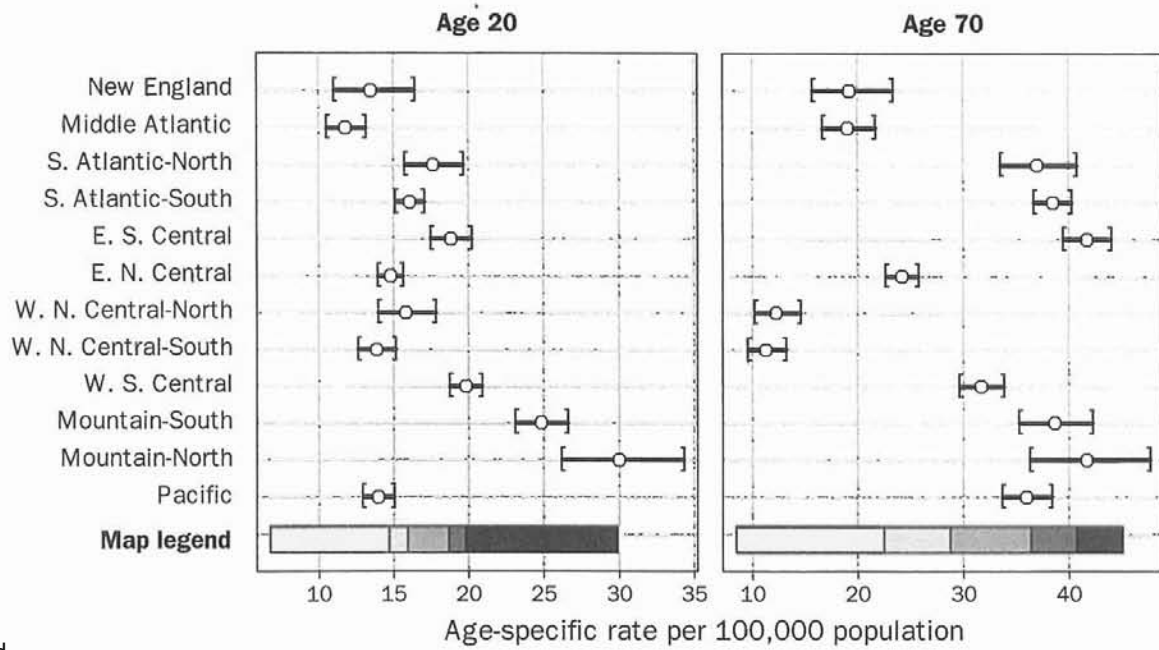


c.

SOURCE: CDC/NCHS

PREDICTED REGIONAL RATES  
FOR SMOOTHED RATE MAPS

FIREARM SUICIDE  
WHITE MALE

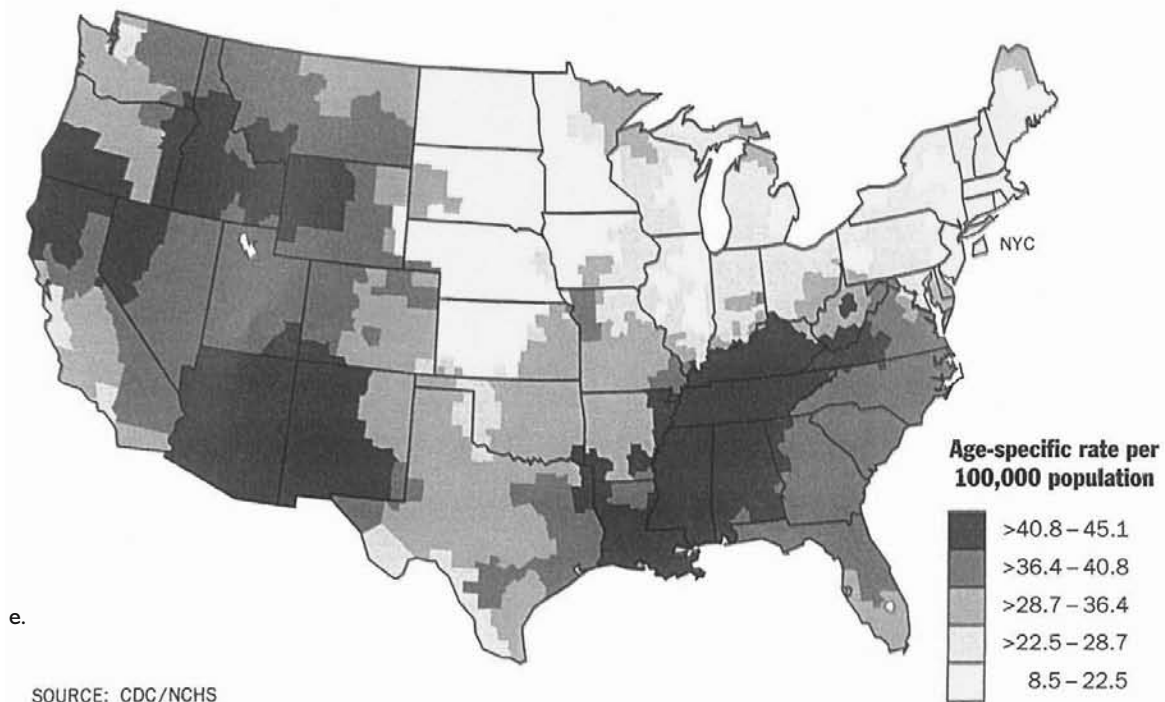


d.

NOTE: Brackets indicate 95% confidence limits.  
SOURCE: CDC/NCHS

SMOOTHED DEATH RATES  
FOR AGE 70

FIREARM SUICIDE  
WHITE MALE

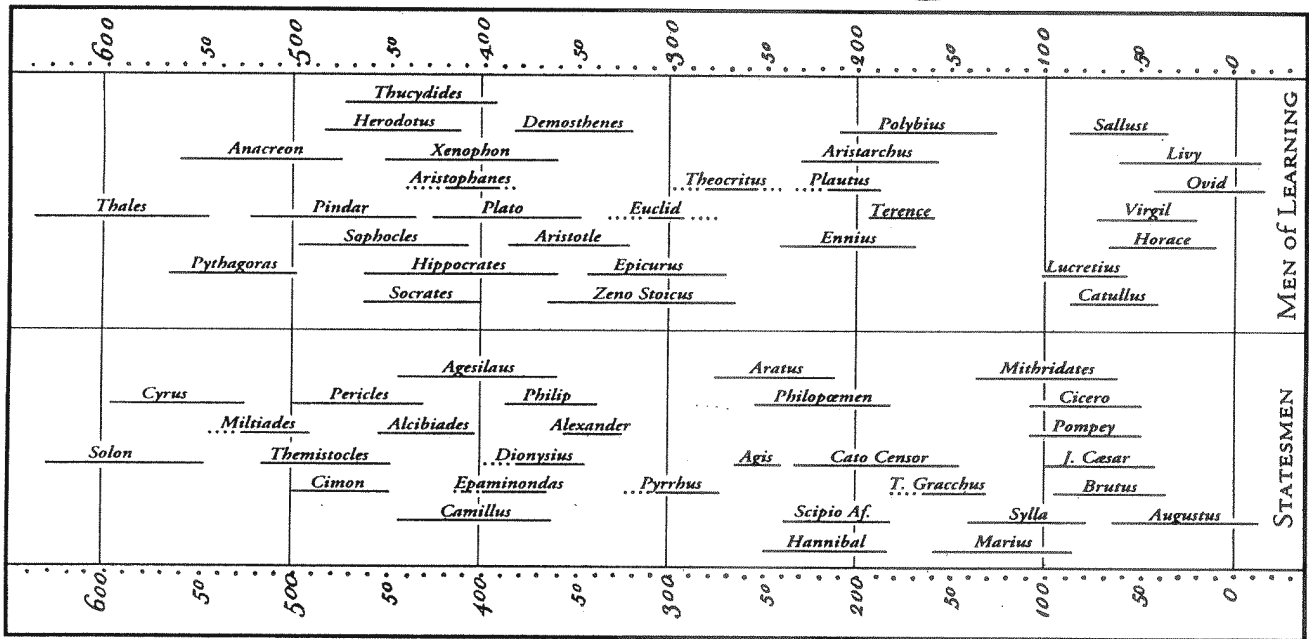


e.

SOURCE: CDC/NCHS

# A Specimen of a Chart of Biography.

6 June p. 210



J. Priestly L.L.D. F.R.S. Inventor del.

Figure 2. Joseph Priestly's chart of biography from his 1765 publication. See Chapter 5 of *Graphic Discovery: A Trout in the Milk and Other Visual Adventures* by Howard Wainer for a full description.

When we choose a display format, there can be competing forces. On the one hand, we may invent a specific format that conforms exactly to the data and demands associated with communicating the message contained within those data, yet such a format would be foreign to the audience. On the other hand, there may be a standard format familiar to the readers that does the job almost as well. Which do we choose? Convention is powerful, and, unless the gains from defying convention are monstrous, it is usually a mistake to opt for the innovative. The odds change, however, if we are designing an extensive statistical report, in which we often have the opportunity to reuse the unconventional display. In this situation, it may be worth the reader's time to learn the new format. The *Atlas* uses a moderately unusual display format, but it is only new the first time.

The earliest example I can recall of how quickly people can learn is an early bar chart: Joseph Priestly's 1765 plot (see Figure 2) of the lives of famous men in history. When it first appeared, it was accompanied by an extensive textual description, ostensibly to help the reader who had surely never seen anything like it before. Yet, in his 1769 elaboration, Priestly included essentially no further explanation.

## Example 2. Understanding USA

Stifling creative urges has obviously been too difficult for some authors, even at the cost of reducing the effectiveness of communication. For example, in *Understanding USA*—a chart book put out by TED Conference LLC—every page uses a different

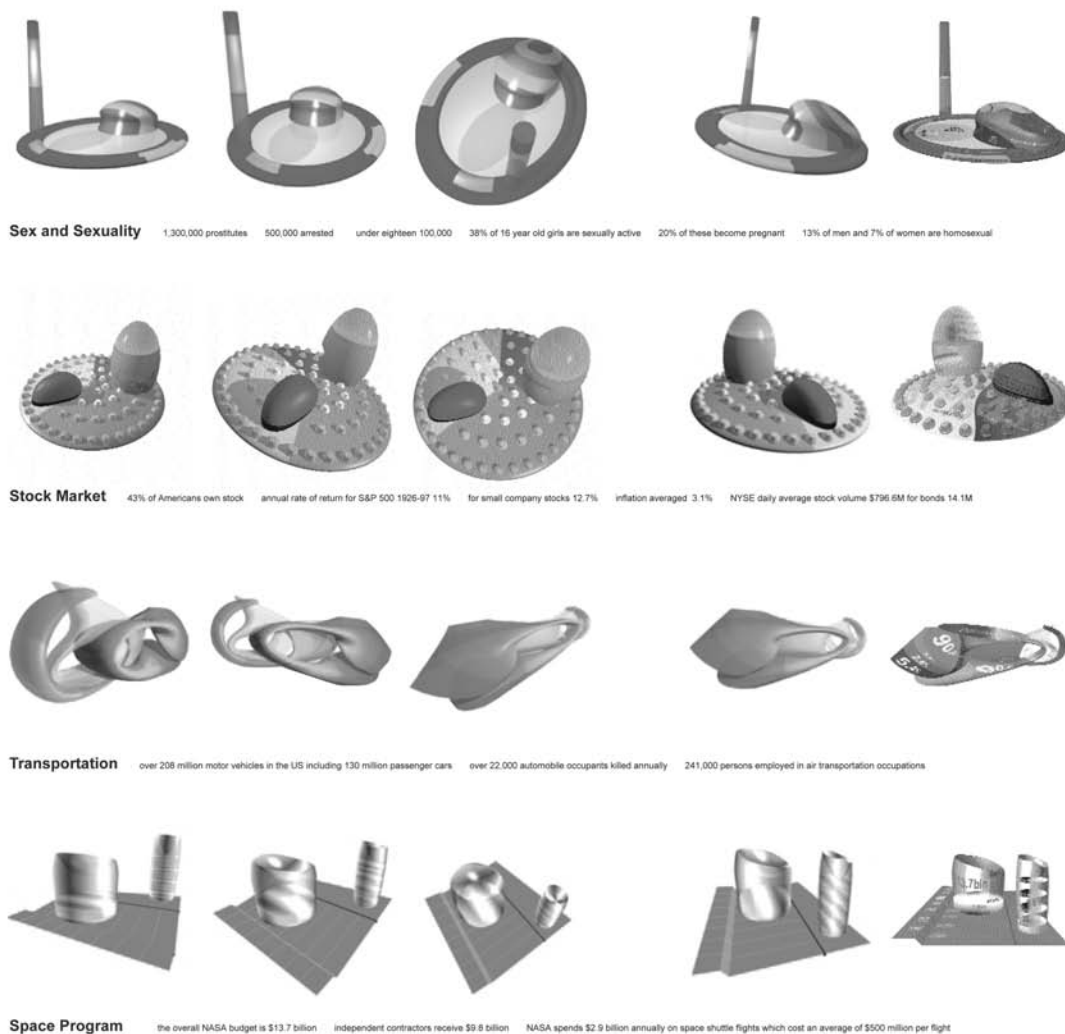
display format. The only aspect they seem to have in common is that they are all mostly indecipherable. I believe that if they had followed the path provided by Pickle and her colleagues in the *Atlas* and agreed to a common graphical format, two problems would have been solved. Obviously, the problem of deciphering a new format on each page would disappear, but also, if each proponent of a particular format had to convince the other authors/designers of its efficacy, the weaknesses of that format would be exposed and corrected. Moreover, by finding a general format suitable for a broad range of data, simplicity would surely have trumped chart junk.

I include, as Figure 3, an example from a chapter by Hani Rashid and Lise Ann Couture. As hard as it may be to believe, this display is not notably worse than many of the others contained in this remarkable volume.

## Example 3. Cancer Trends Report

On December 12, 2007, the National Cancer Institute provided their annual Cancer Trends Progress Report (<http://progressreport.cancer.gov>). In it, they followed the model provided by Pickle and her colleagues a decade earlier. Each chapter frames questions about cancer, its detection, and its risk factors, and then caps the questions with various sound-bite-suitable responses. Each section then culminates with a graph.

The graphical format is simple and clear and always the same (see Figure 4). The figures are apparently produced in some automatic way and so some unfortunate choices of color and label placement are made, perhaps by accident (or perhaps by an imperfect algorithm). However, even though the figure



## Datascaping USA

How can we better understand America's characteristics?



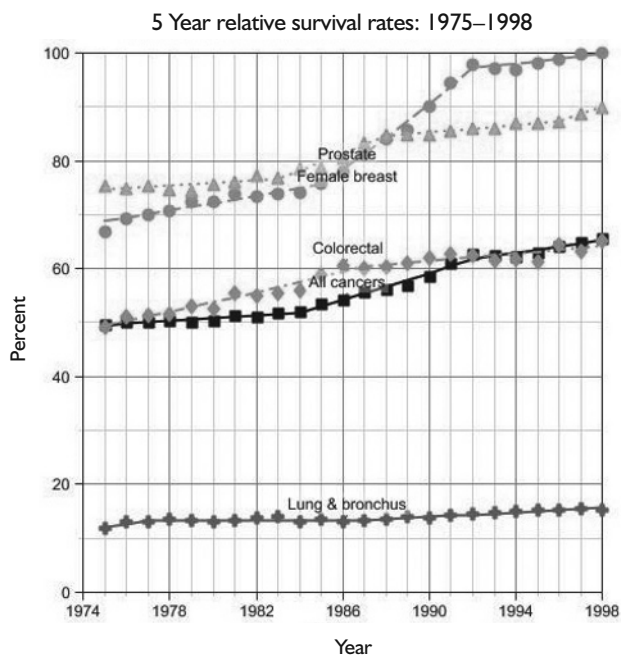
The eight models contained on these pages created by Asynquore, exist as Virtual Reality Markup Language (VRML) models that are accessible and operable on the Internet. The models can be found at the TEDX website [www.understandingsa.com](http://www.understandingsa.com). Each three dimensional model can be manipulated and adjusted according to certain criteria allowing for different readings and means of understandings. The adjustability and transformative aspects of these data entities allow for each subject to be analyzed and scrutinized from different viewpoints. These are different "vantage" points one can assume in order to reference information in a variety of ways. *Timelining*, for example allows for certain adjustments of the data to be made where one can reverse or fast forward a timeline to see how the datascapes alters and transforms. By choosing one element in the data model and adjusting it according to some criteria, other items that are related can also be re-configured and adjusted accordingly. Here various correlations can be made on data that might otherwise be missed or misconstrued. This new method of reading, recording and gathering information creates a much richer and deeper understanding of information in contrast to conventional graphing and charting systems.

The datascapes consist of three main components that allow them to be interactive entities for eventual porting over the web. They were initially modeled as rudimentary wireframes in Maya software and kept "lightweight"

Figure 3. An incomprehensible plot. Courtesy of Richard Saul Wurman

is reasonably clear, there are still improvements that can be made. So, in the spirit of Tukey's suggestions a decade ago, let me offer 10 suggestions here (implemented in Figure 5) so the 2008 version will be still better and possibly suggest further avenues for progress.

1. Obviously, light colors (such as yellow) should be avoided, as their visibility is easily compromised. Moreover, not all users of a report will have easy access to color printing, and so it is important, when possible, for all the colors used be completely readable if the plot was ever to be reproduced in black and white.
2. When lines cross, ambiguity is reduced if both ends are labeled.
3. Axes should be spaced logically. In this instance, why should the x-axis be spaced in four-year intervals? Such a convention makes sense if the phenomenon being plotted happens at four-year intervals (e.g., U.S. presidential elections). Otherwise, it is sensible to stay with the convention of five- or 10-year intervals that are derivative of our
- base-10 society. It is especially suitable for these data to emphasize the five-year survival criterion.
4. Labels must be large enough to be easily read and positioned so as to not have their referent be confused.
5. The category "ALL" is special. It should be made darker and bigger to differentiate it from its components.
6. The x-axis label should be made both complete and explicit; the partial label "year" is ambiguous. It could be the year of diagnosis (my guess) or the year the survey noted they were still alive.
7. Too many extra grid lines add little but visual noise. They should be elided if their loss yields no loss of information. I sketched in just four major horizontal lines to aid orientation (e.g., lung cancer five-year survival rates are less than 20%) and to add extra horizontal references that emphasize the gentle positive slopes for all the curves (even lung cancer) that constitute some of the good news contained in the report.



Source: SEER Program, National Cancer Institute. Rates are from the SEER 9 areas (<http://seer.cancer.gov/registries/terms.html>). Data are not age-adjusted.

Figure 4. Five-year survival rates from various kinds of cancer, showing the improvements over the past two decades (from the National Cancer Institute)

8. There should be space between the axes and the first and last data points so that no points are obscured by sitting on an axis.
9. Plotting points can be deleted once they serve their purpose of showing where the connecting function needs to go. Leaving them in is like leaving up the scaffolding after a building is complete.
10. A friendlier font than Helvetica may be found less off-putting to readers. Helvetica is a clean, austere, serious-looking font; it is frequently a good choice. But, a document focusing on cancer does not need anything extra to impress its seriousness on the reader. A little visual gentleness may serve us well.

### Lessons Learned

The path to improved display is endless, but mostly monotonic, if we learn from the past and continue to innovate, standing on the shoulders of our predecessors.

Innovation should be controlled; too much may increase the load on the viewers beyond their capacity. Also the graphical inventors of the past were not idiots, and the inventions that have survived time have done so because of their usefulness over a broad range of areas of application. It is possible that we can invent something entirely new and superior to all that has come before, but the odds are against it. Charles Joseph Minard did, but such ideas don't come along all that often, which is why we still celebrate his flow maps more than a century later. Control hubris.

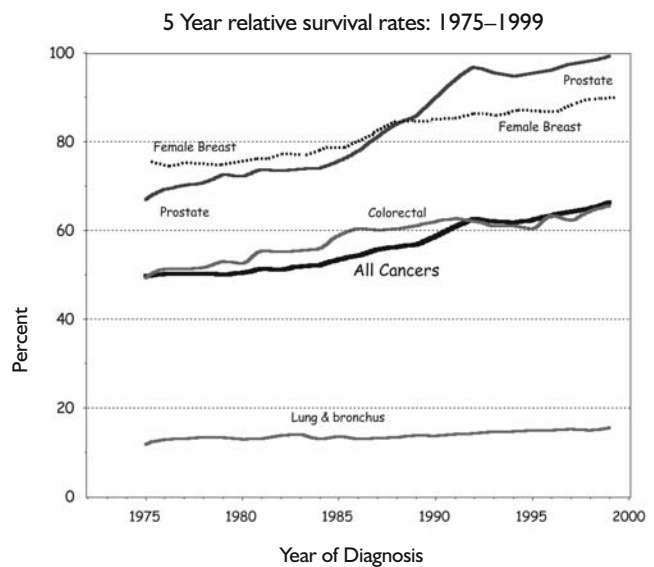



Figure 5. Figure 4 redrafted with 10 changes

When trying to prepare a coherent report on a single, possibly broad, topic, the displays should also be coherent. Repeating the same format with different data eases the decoding task of the viewer. It is usually a mistake to think such repetitiveness will bore the readers—quite the opposite. It will allow them to focus on the content of the displays and not their format. In the end, they will be grateful.

A complex statistical report often has many authors, each preparing a separate section. If only a single presentation format is to be used throughout, there must be considerable cooperation among the authors and strong leadership from the editor. The multiple eyes and minds looking at each section this approach requires are almost sure to lead to improved quality. It is an important benefit of cooperation. 

### Further Reading

- Pickle, L.W., Mungiole, M., Jones, G.K., and White, A.A. (1996). *Atlas of United States Mortality*. Hyattsville, Maryland: National Center for Health Statistics.
- Priestley, J. (1765). *A Chart of Biography*. London: William Eyres.
- Priestley, J. (1769). *A New Chart of History*. Reprinted: 1792, New Haven: Amos Doolittle.
- Santayana, G. (1905). *Life of Reason*, Vol. 1, Chapter 12. New York: Charles Scribner & Sons.
- Wurman, S. (ed.) (2000). *Understanding USA*. New York: TED Conference LLC.
- Wainer, H. (2005). *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*. Princeton, New Jersey: Princeton University Press.

**Column Editor:** Howard Wainer, Distinguished Research Scientist, National Board of Medical Examiners, 3750 Market Street, Philadelphia, PA 19104; [hwainer@nbme.org](mailto:hwainer@nbme.org)