

Stat 5810, Section 004
Statistical Visualization I
Fall 2017

Dr. Jürgen Symanzik

Utah State University

Department of Mathematics and Statistics

3900 Old Main Hill

Logan, UT 84322–3900

Tel.: (435) 797–0696

FAX: (435) 797–1822

e-mail: symanzik@math.usu.edu

Web: <http://www.math.usu.edu/~symanzik/>



Figure 1: <http://www.phdcomics.com/comics/archive.php?comicid=1541>, Cartoon.

Contents

Acknowledgements	1
1 Examples of Good and Bad Graphics	1
1.1 How to Display Data Badly	1
1.1.1 Don't show much data	2
1.1.2 Show the data inaccurately	7
1.1.3 Obfuscate the data	14
1.2 Bad Graphics are Everywhere — In Space and in Time	30
1.3 Rules for Good Data Displays	40
1.4 Further Reading	44
References	1

Acknowledgements

This course uses some of the course materials provided by Dr. Mike Minnotte (formerly USU, now with the University of North Dakota) as held in the Fall 2006 semester. Additional materials have been taken from other Statistical Graphics courses, such as the ones offered by Dr. Di Cook (formerly Iowa State University; now Monash University: <http://dicook.org/>) and Dr. Dan Carr (George Mason University: <http://mason.gmu.edu/~dcarr/>). Other examples and R code originate from Heike Hofmann, Carson Sievert, Martin Theus, Antony Unwin, Simon Urbanek, Hadley Wickham, Lee Wilkinson, and others. We are likely to include parts from additional authors and sources that will be specified later during the semester.

Thanks are also due to around 30 students and guests who took the former “Stat 6560: Graphical Methods” with me since the Spring 2009 semester for their valuable comments that helped to improve, correct, and extend these lecture notes.

Jürgen Symanzik, October 2, 2017.

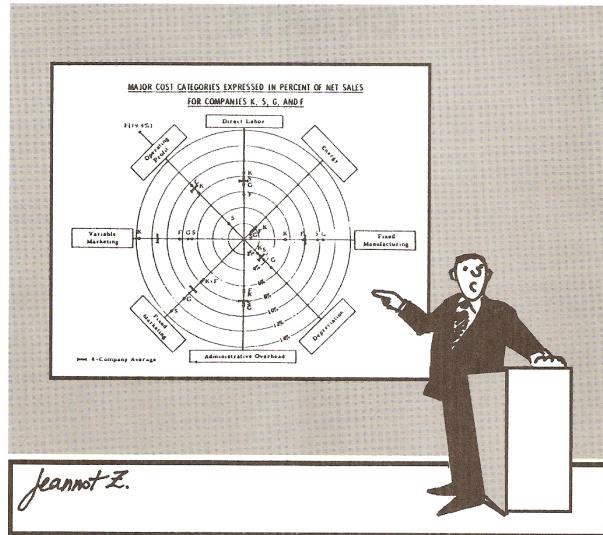


Figure 2: Zelazny (2001), p. x, Cartoon.

1 Examples of Good and Bad Graphics

(Based on Wainer (1997), Chapter 1 & Tufte (1983), Chapter 2)

1.1 How to Display Data Badly

Wainer (1997), p. 12, states:

“The aim of good data graphics is to display data accurately and clearly.

[...]

Thus, if we wish to display data badly, we have three avenues to follow.

- A. Don’t show much data.
- B. Show the data inaccurately.
- C. Obfuscate the data.”[†]

Let us follow these strategies:

[†]Show the data unclearly.

1.1.1 Don't show much data

Rule 1: Show as little data as possible (minimize the data density).

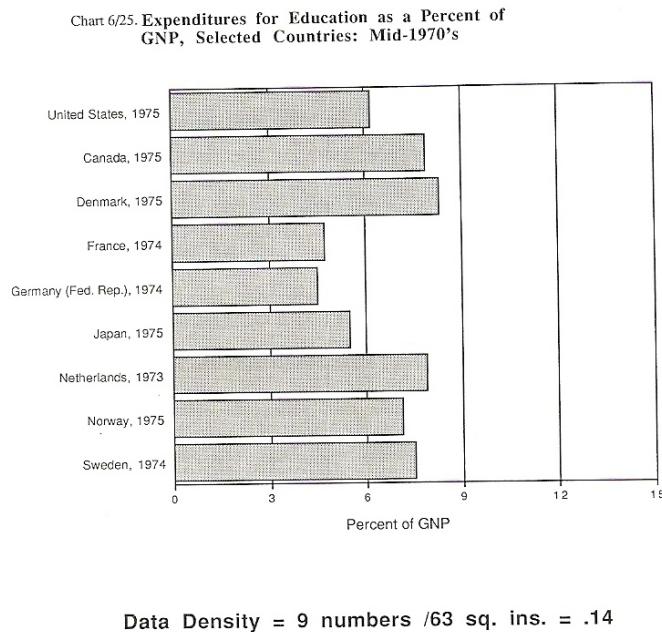


FIGURE 2. Chart 6/25 from *Social Indicators III* showing expenditures for education for nine countries as a function of GNP.

Figure 3: Wainer (1997), p. 13, Figure 2.

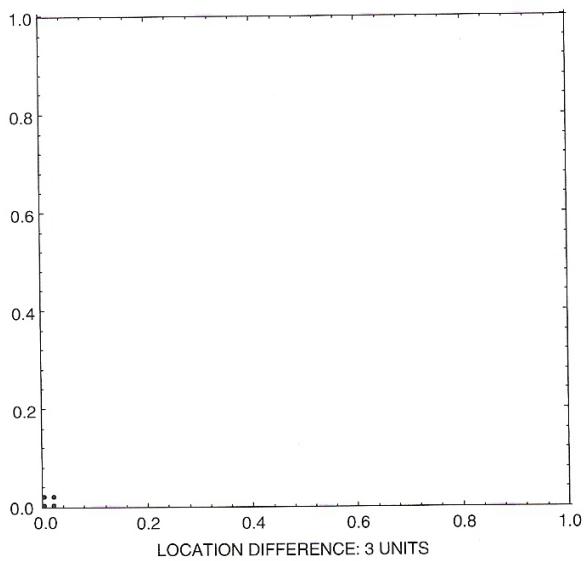


FIGURE 3. A graph of obviously low data density.

Figure 4: Wainer (1997), p. 13, Figure 3.

Labor Productivity: U.S. vs Japan

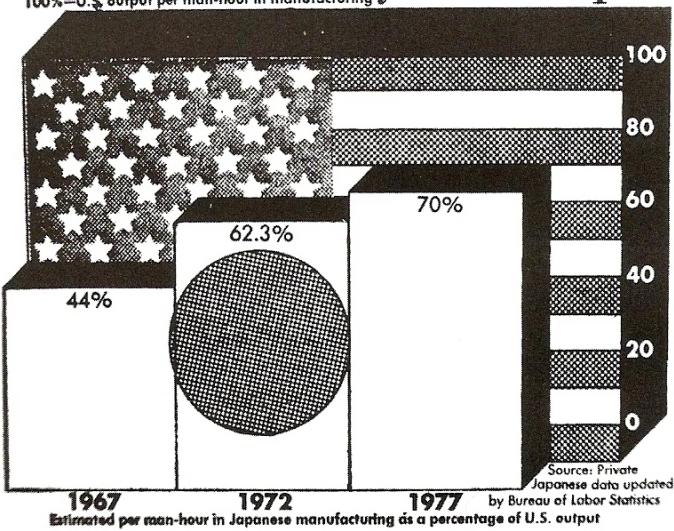


FIGURE 6. A graph with low data density filled in with chartjunk from the *Washington Post*, 1978.

Figure 5: Wainer (1997), p. 16, Figure 6.

Rule 2: Hide what data you do show (minimize the data/ink ratio).

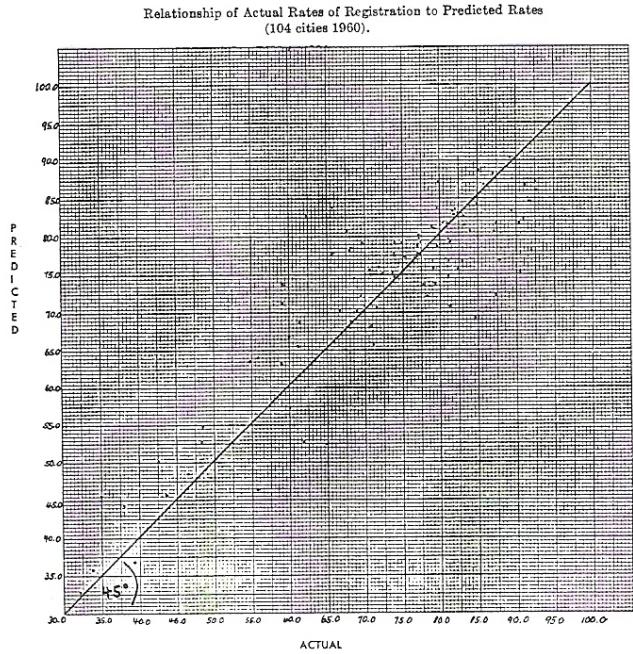


FIGURE 8. Hiding the data in the grid.

Figure 6: Wainer (1997), p. 17, Figure 8: Hiding the data in the grid.

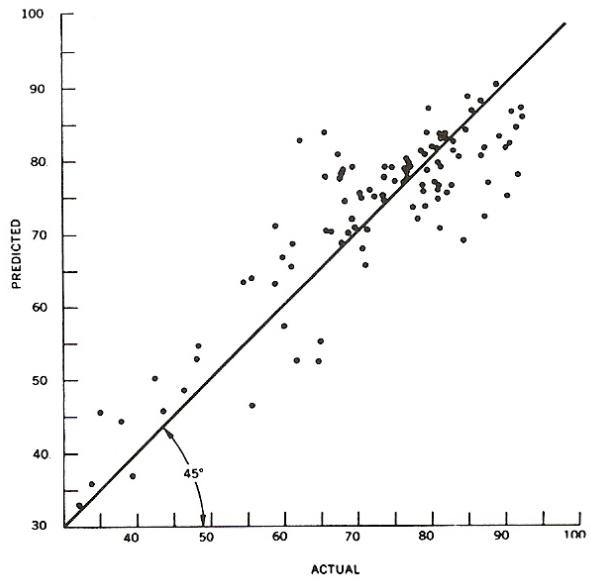


FIGURE 10. A redone example of the data from figure 8.

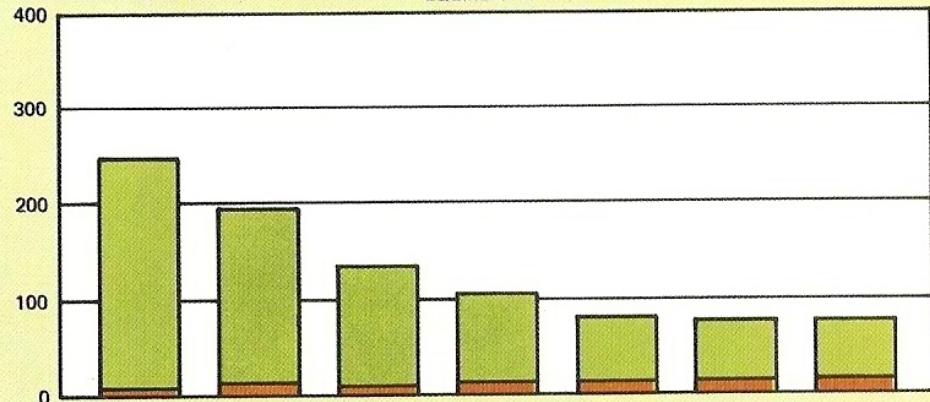
Figure 7: Wainer (1997), p. 18, Figure 10: Wainer (1997), p. 17, Figure 8, improved.

Chart 6/4. **Public and Private Schools, by Level, Selected Years: 1929-1977**

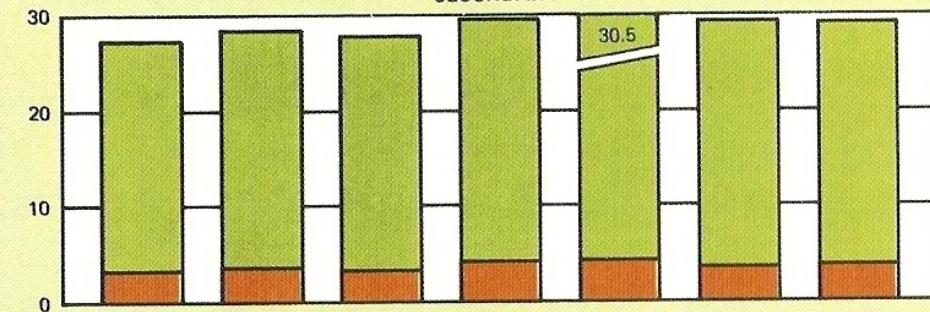
Public
Private

Thousands of schools

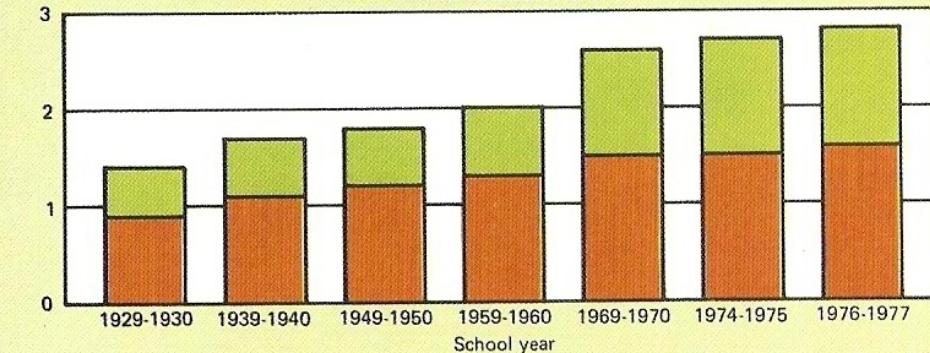
ELEMENTARY



SECONDARY



HIGHER EDUCATION



CHAPTER 1, FIGURE 11. Hiding the data in the scale.

Figure 8: Wainer (1997), p. 20A, Figure 11: Hiding the data in the scale.

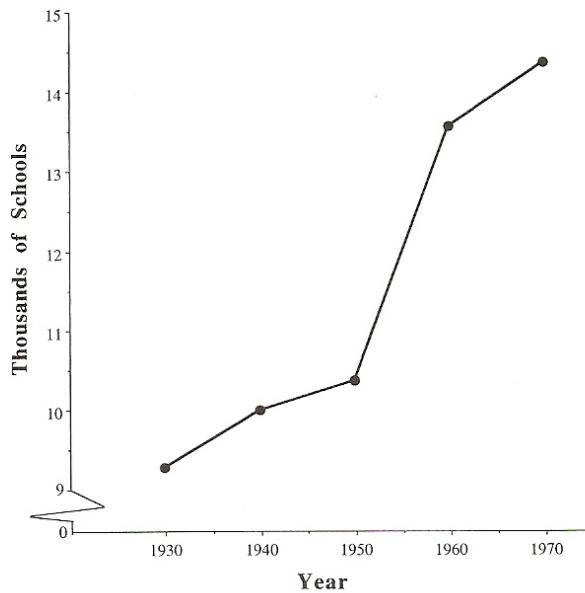


FIGURE 12. Expanding the scale and showing the data for the number of private elementary schools from figure 11.

Figure 9: Wainer (1997), p. 20, Figure 12: Wainer (1997), p. 20A, Figure 11, improved.

1.1.2 Show the data inaccurately

Rule 3: Ignore the visual metaphor altogether.

FIGURE 13. Ignoring the visual metaphor by letting a longer bar segment represent a smaller amount of coal (from the *New York Times*, 1978).

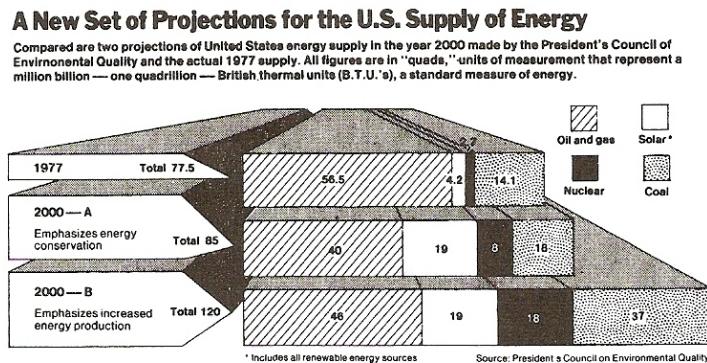


Figure 10: Wainer (1997), p. 20, Figure 13.

U.S. trade with China and Taiwan

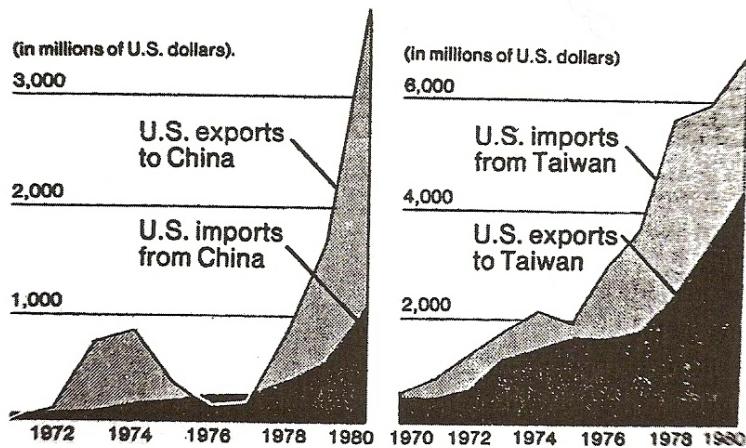


FIGURE 14. Reversing the metaphor in mid-graph while changing scales on both axes (from the *New York Times*, June 14, 1981).

Figure 11: Wainer (1997), p. 21, Figure 14.

FIGURE 15. Figure 14 redone with a consistent scale and visual metaphor.

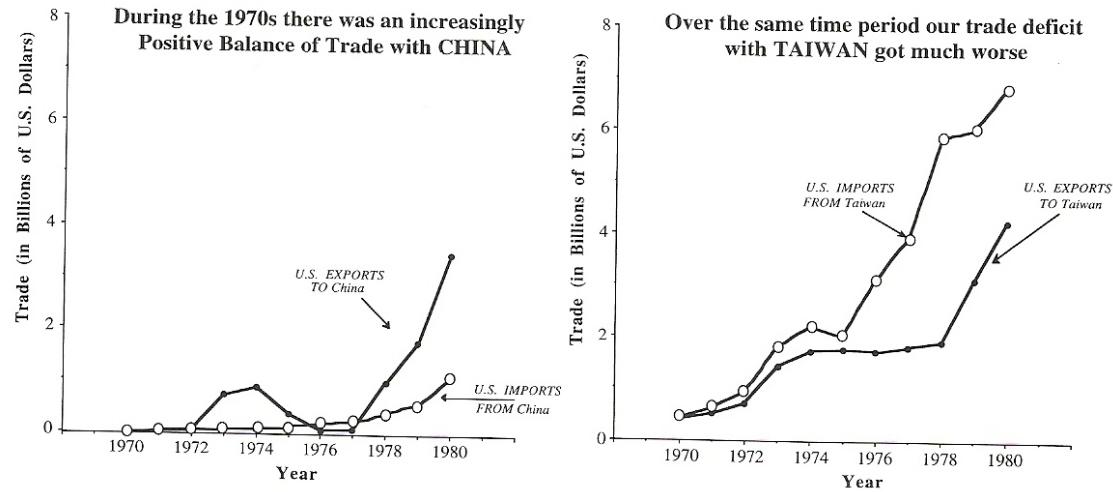


Figure 12: Wainer (1997), p. 21, Figure 15: Wainer (1997), p. 21, Figure 14, improved.

Rule 4: Only order matters.



FIGURE 17. An example of how to goose up the effect by squaring the eyeball.

Figure 13: Wainer (1997), p. 23, Figure 17.

Additional details will be provided after class.

Figure 14: Wainer (1997), p. 24, Figure 18: Wainer (1997), p. 23, Figure 17, improved.

FIGURE 19. Cubing the visual effect and choosing the origin to yield a near record lie factor of over 131,000% (from the *Washington Post*).

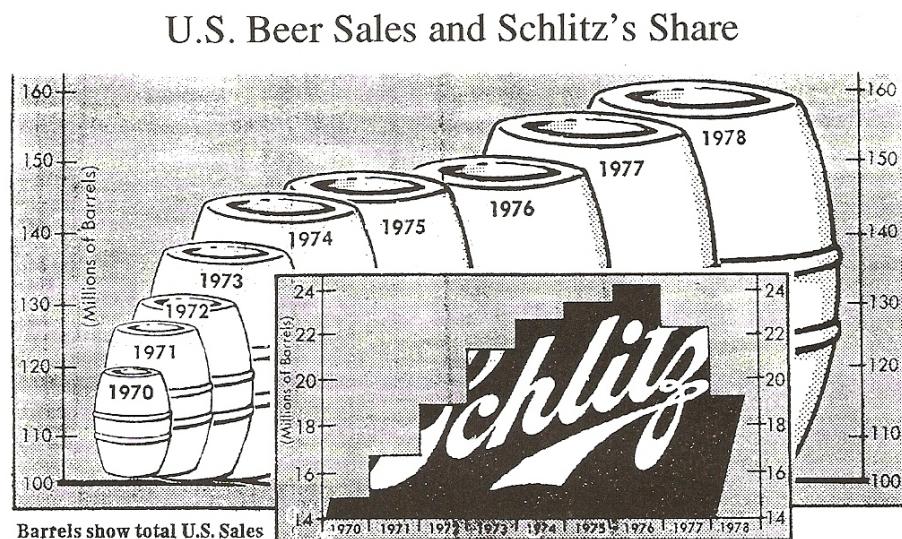


Figure 15: Wainer (1997), p. 24, Figure 19.

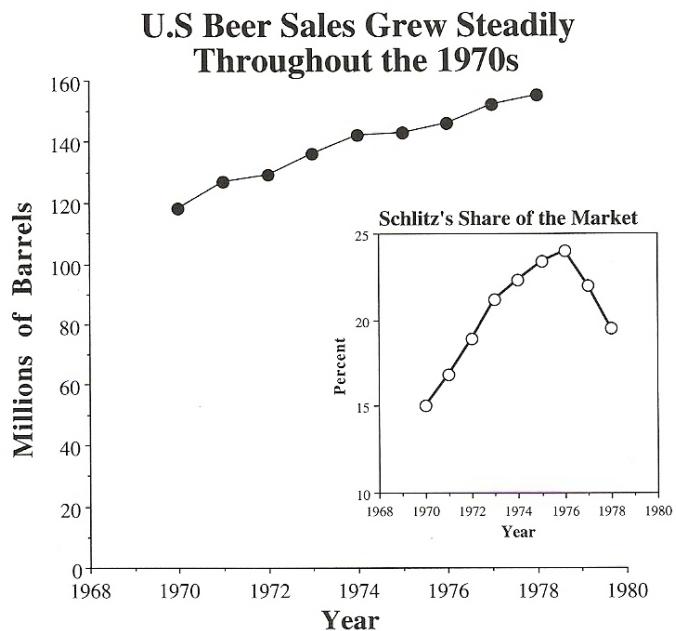


FIGURE 20. Data from figure 19 redone without tricks (from Wainer, 1980).

Figure 16: Wainer (1997), p. 25, Figure 20: Wainer (1997), p. 24, Figure 19, improved.

Rule 5: Graph data out of context.

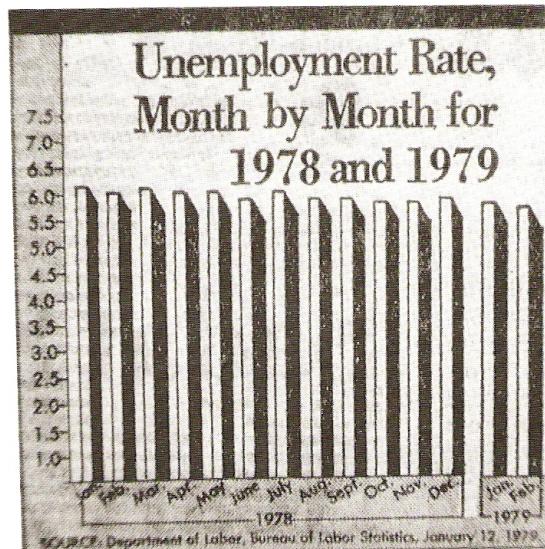


FIGURE 21. Hiding the effect by the careful choice of scale and origin (from the *Washington Post*).

Figure 17: Wainer (1997), p. 26, Figure 21.



FIGURE 22. Regraph of data from figure 21 with expanded scale, different starting point, and previous year's average added for context (from Wainer, 1980).

Figure 18: Wainer (1997), p. 26, Figure 22: Wainer (1997), p. 26, Figure 21, improved.

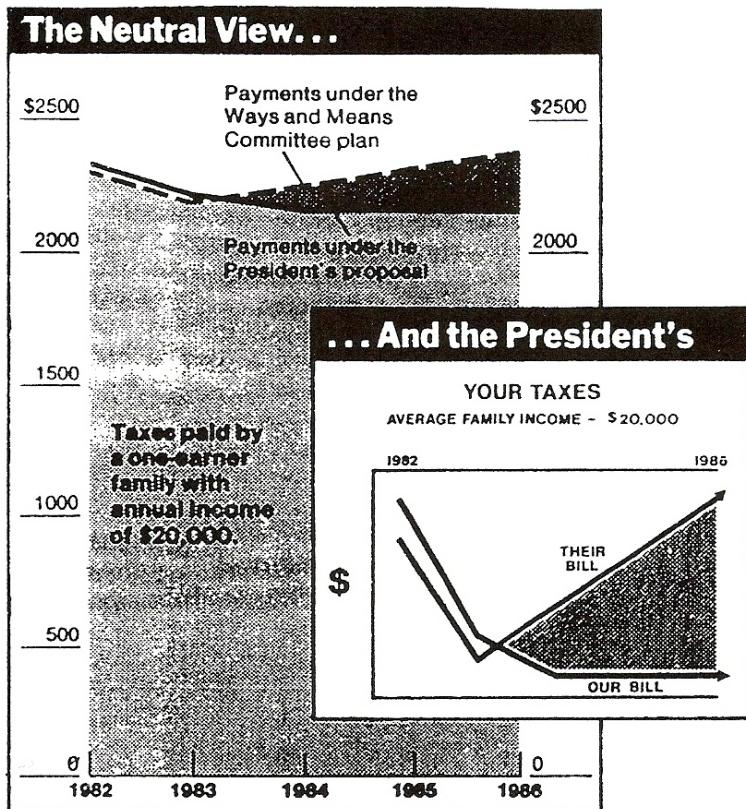


FIGURE 23. *New York Times* graphs showing how lack of context changes our perceptions about alternative tax bills.

Figure 19: Wainer (1997), p. 27, Figure 23.

1.1.3 Obfuscate the data

Rule 6: Change scales in mid-axis.

FIGURE 24. Changing the scale in mid-axis to make large differences seem small (from the *New York Post*, May 12, 1981).

The soaraway Post — the daily paper New Yorkers trust

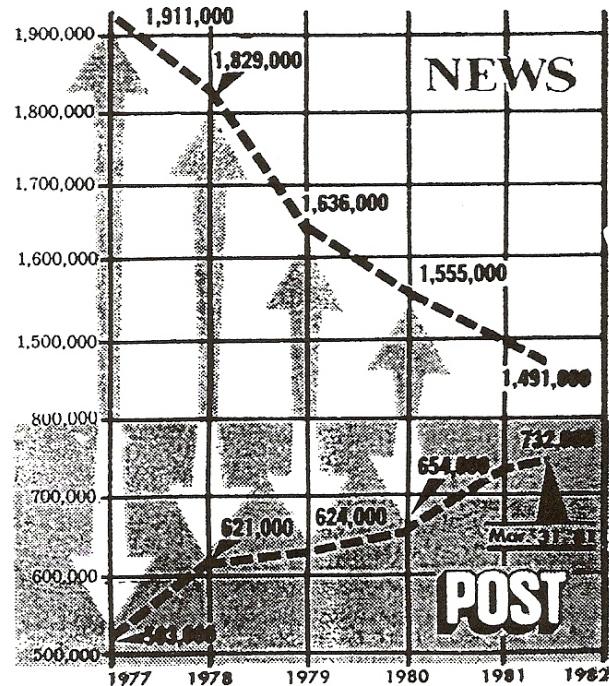


Figure 20: Wainer (1997), p. 28, Figure 24.

FIGURE 25. Changing scale in mid-axis to make exponential growth linear (from the *Washington Post*, Jan. 11, 1979, in an article titled "Pay, Practices of Doctors on Examining Table" by Victor Cohn and Peter Milius).

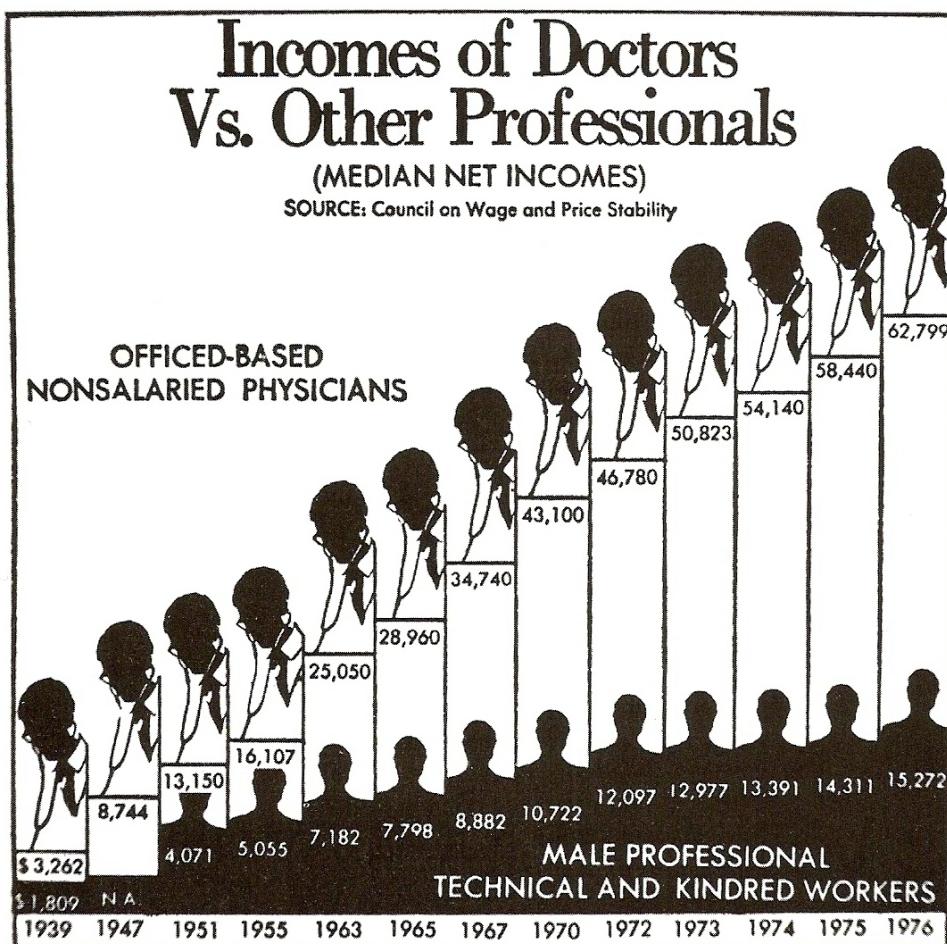


Figure 21: Wainer (1997), p. 29, Figure 25.

FIGURE 26. Data from figure 25 redone with a linear scale (from Wainer, 1980).

Physicians' income has grown exponentially since 1939
Whereas other professionals' income has gone up linearly

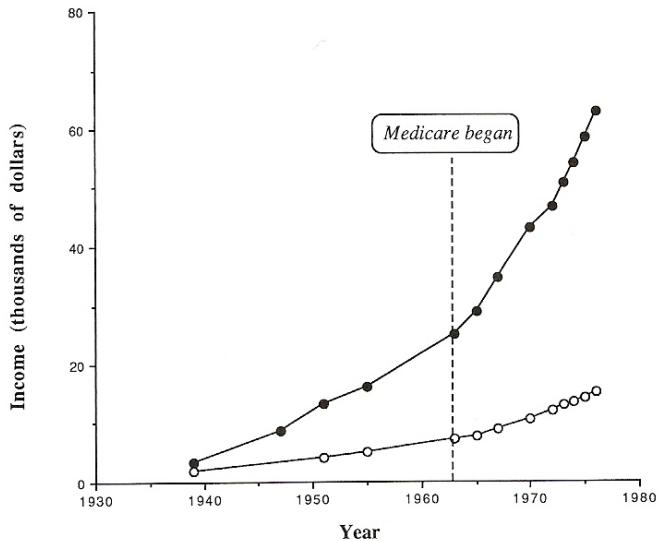
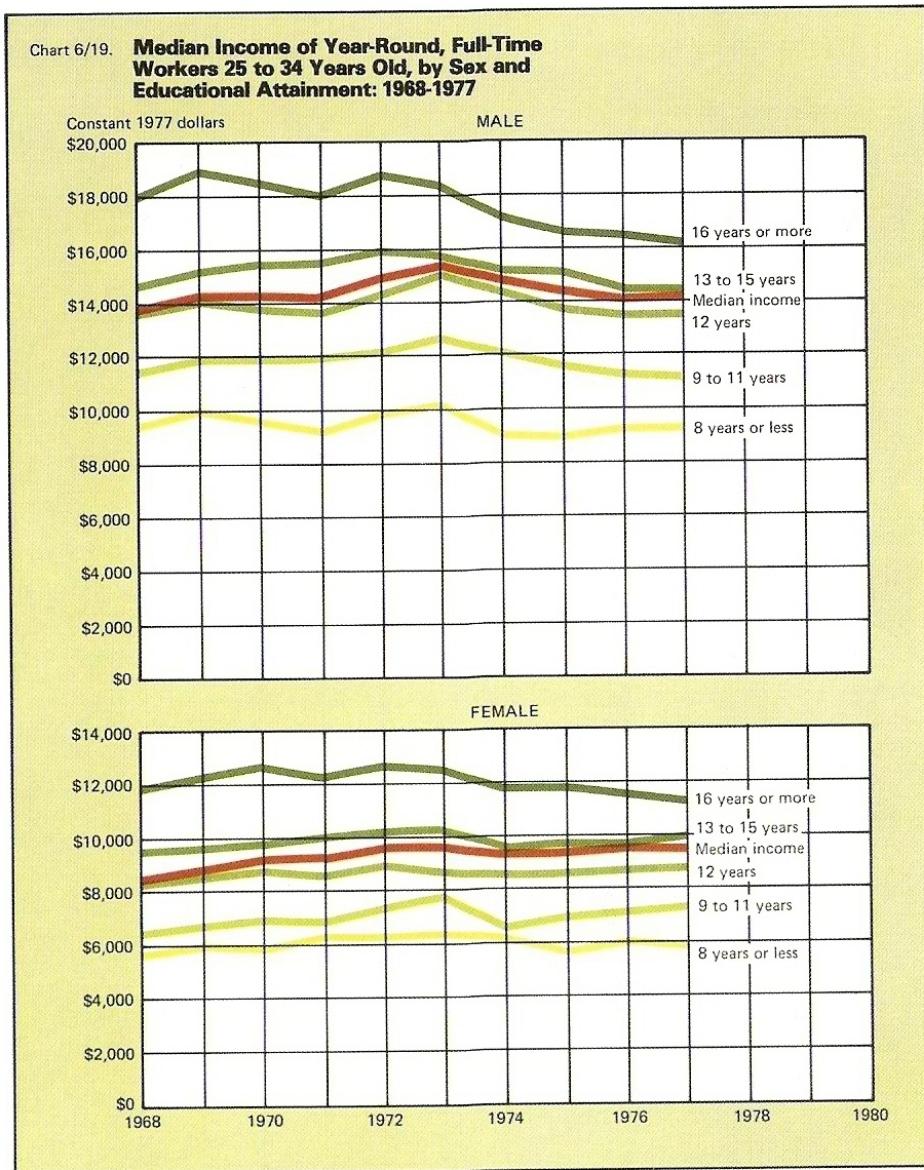


Figure 22: Wainer (1997), p. 30, Figure 26: Wainer (1997), p. 29, Figure 25, improved.

Rule 7: Emphasize the trivial (ignore the important).



CHAPTER 1, FIGURE 27. Emphasizing the trivial: Hiding the main effect of sex differences in income through the vertical placement of plots.

Figure 23: Wainer (1997), p. 20A, Figure 27.

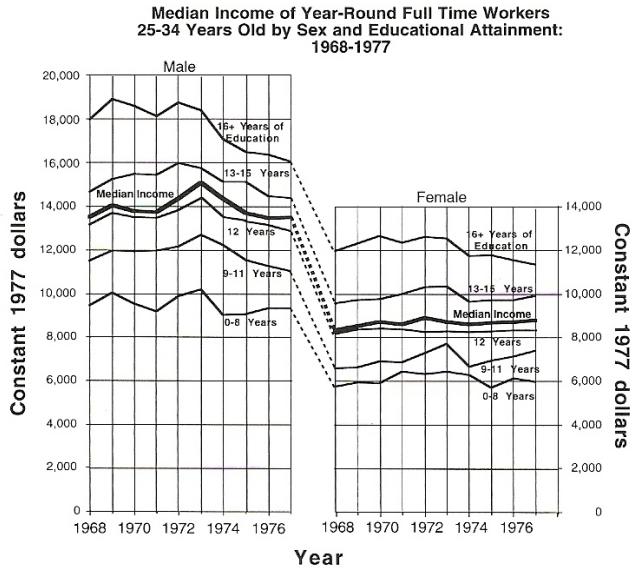


FIGURE 28. Figure 27 redone with the two plots horizontally opposed, showing the size of sex differences more clearly.

Figure 24: Wainer (1997), p. 31, Figure 28: Wainer (1997), p. 20A, Figure 27, improved.

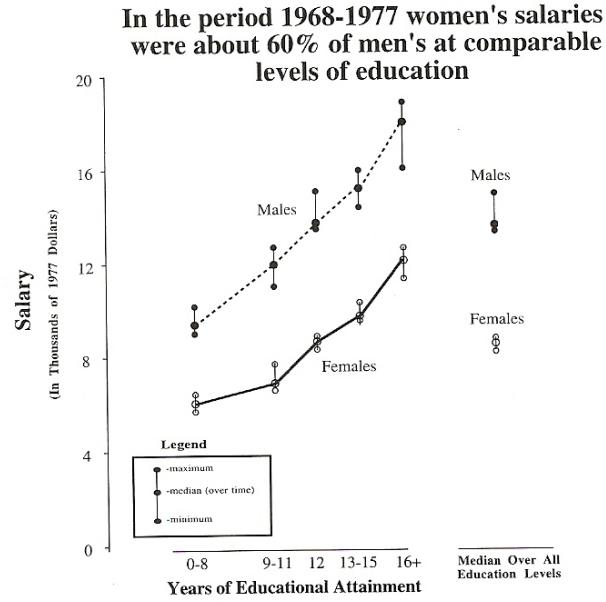
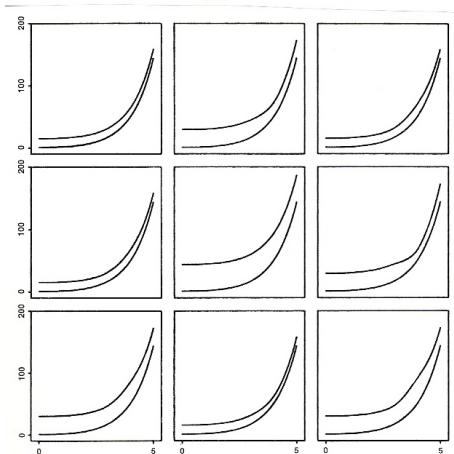


FIGURE 29. Figure 28 redone with the large effects of sex and education emphasized and the small-time trend suppressed.

Figure 25: Wainer (1997), p. 32, Figure 29: Wainer (1997), p. 31, Figure 28, further improved.

Rule 8: Jiggle the baseline.

FIGURE 30. A graphical experiment (from Cleveland and McGill, 1984). Without looking at the corresponding right panel, try to determine the difference between the two curves in the left panel.



*Sorry, these plots
got scrambled ...*

Figure 26: Wainer (1997), p. 33, Figure 30.

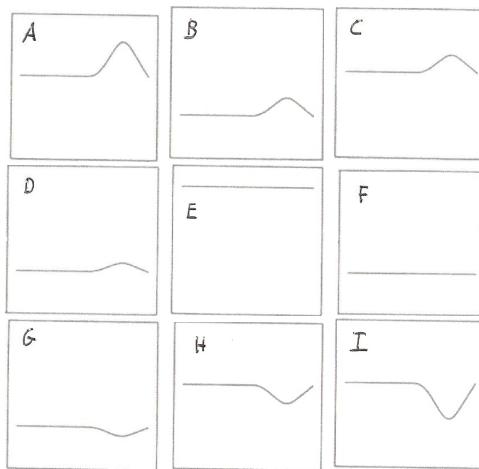
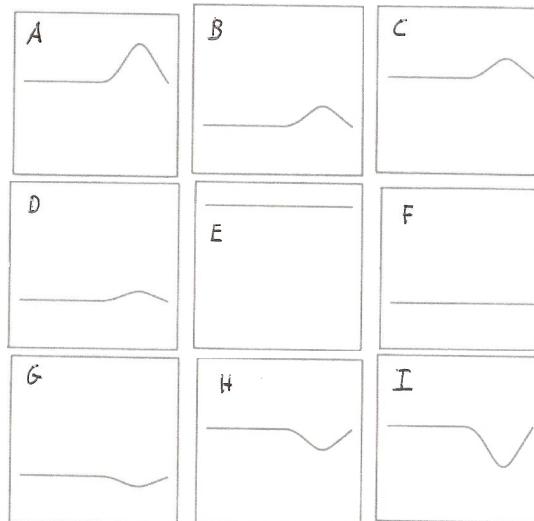
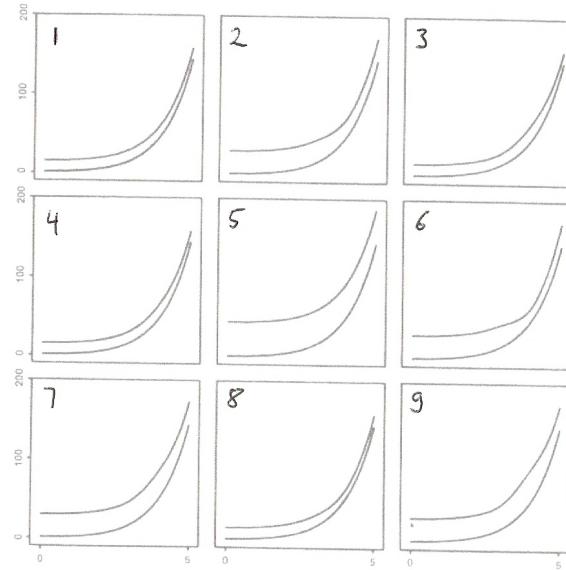


Figure 27: Wainer (1997), p. 33, Figure 30: Scrambled differences. The horizontal axis covers the interval 0 to 5, the vertical axis covers the interval 0 to 50.

Worksheet

Your Name: _____



Task: Match each original (labeled 1 to 9) with the plot (labeled A to I) that shows the difference between upper and lower line in the original plot.

Answer:

- 1: _____ 2: _____ 3: _____
4: _____ 5: _____ 6: _____
7: _____ 8: _____ 9: _____

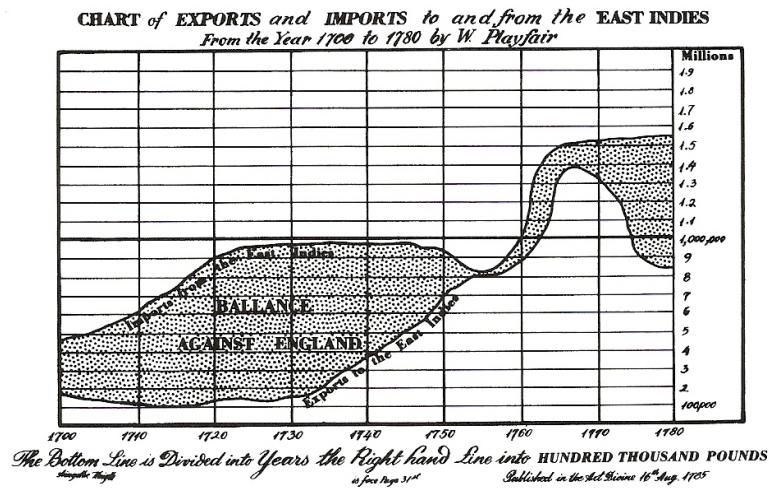


FIGURE 31. William Playfair's eighteenth-century graph of England's imports and exports with the East Indies (from Cleveland and McGill, 1984).

Figure 28: Wainer (1997), p. 34, Figure 31: One of William Playfair's few mistakes.

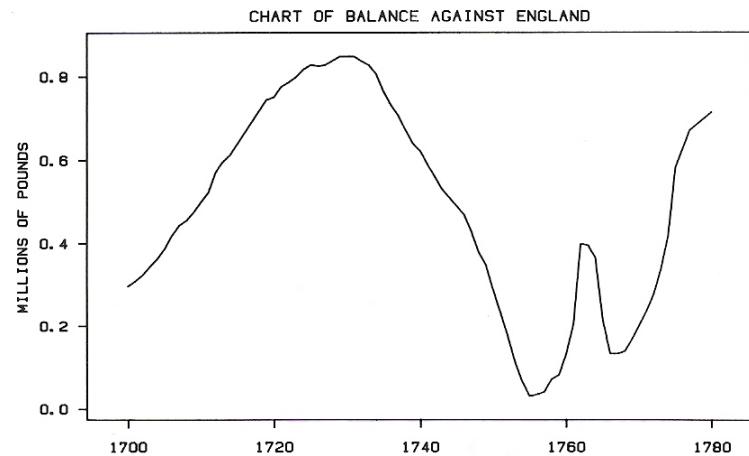


FIGURE 32. A graph of the difference between West Indies imports and exports showing explicitly the previously invisible jump in the 1760s (from Cleveland and McGill, 1984).

Figure 29: Wainer (1997), p. 34, Figure 32: Wainer (1997), p. 34, Figure 31, improved.

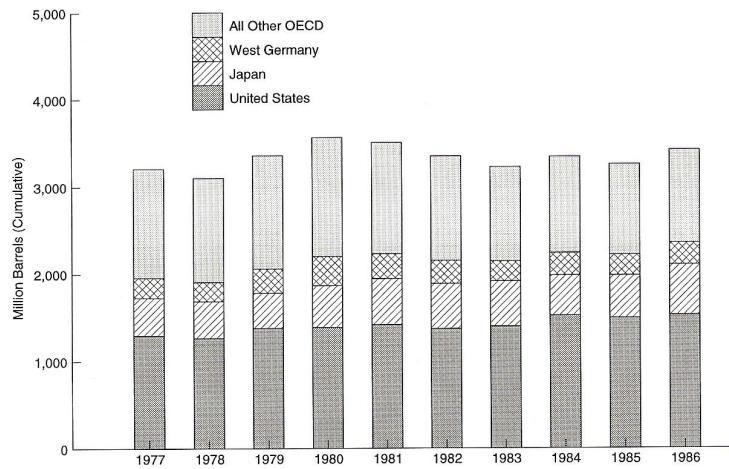


FIGURE 33. From the U.S. Department of Energy's *Annual Energy Review, 1986*, showing the changes in primary stocks of petroleum in OECD countries.

Figure 30: Wainer (1997), p. 35, Figure 33.

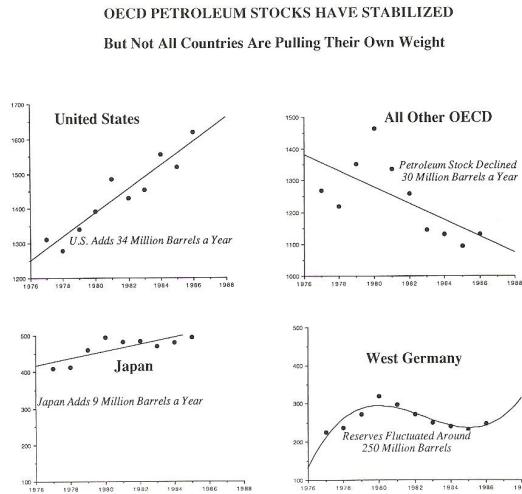
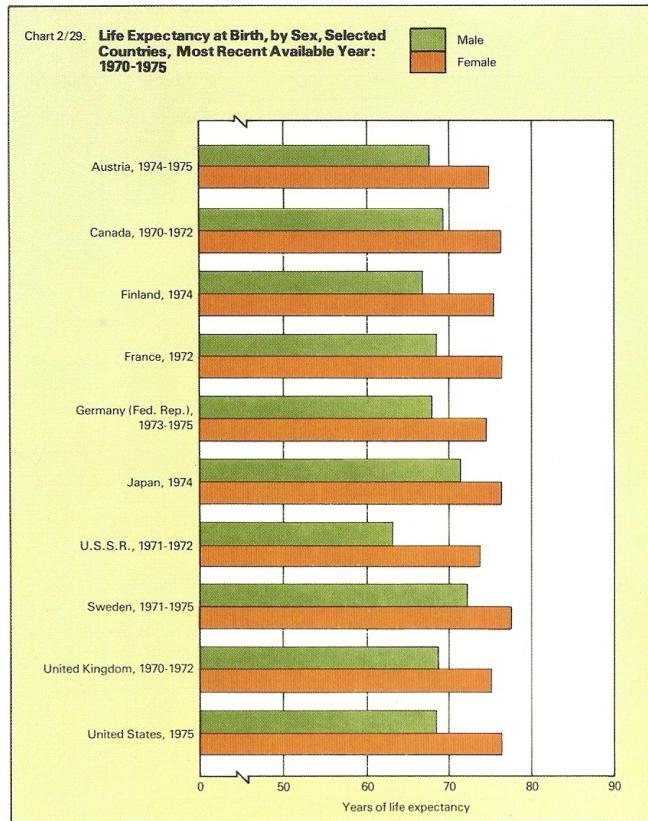


FIGURE 34. Regraphing of the data from figure 33 in which each country's data are shown relative to a straight line.

Figure 31: Wainer (1997), p. 36, Figure 34: Wainer (1997), p. 35, Figure 33, improved.

Rule 9: Alabama first!



CHAPTER 1, FIGURE 35. Austria first! Obscuring the data structure in some life expectancy data by alphabetizing the plot.

Figure 32: Wainer (1997), p. 20B, Figure 35.

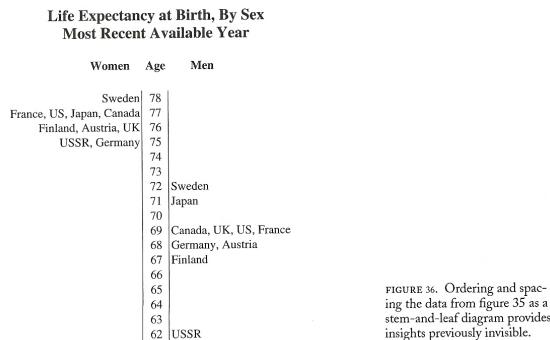
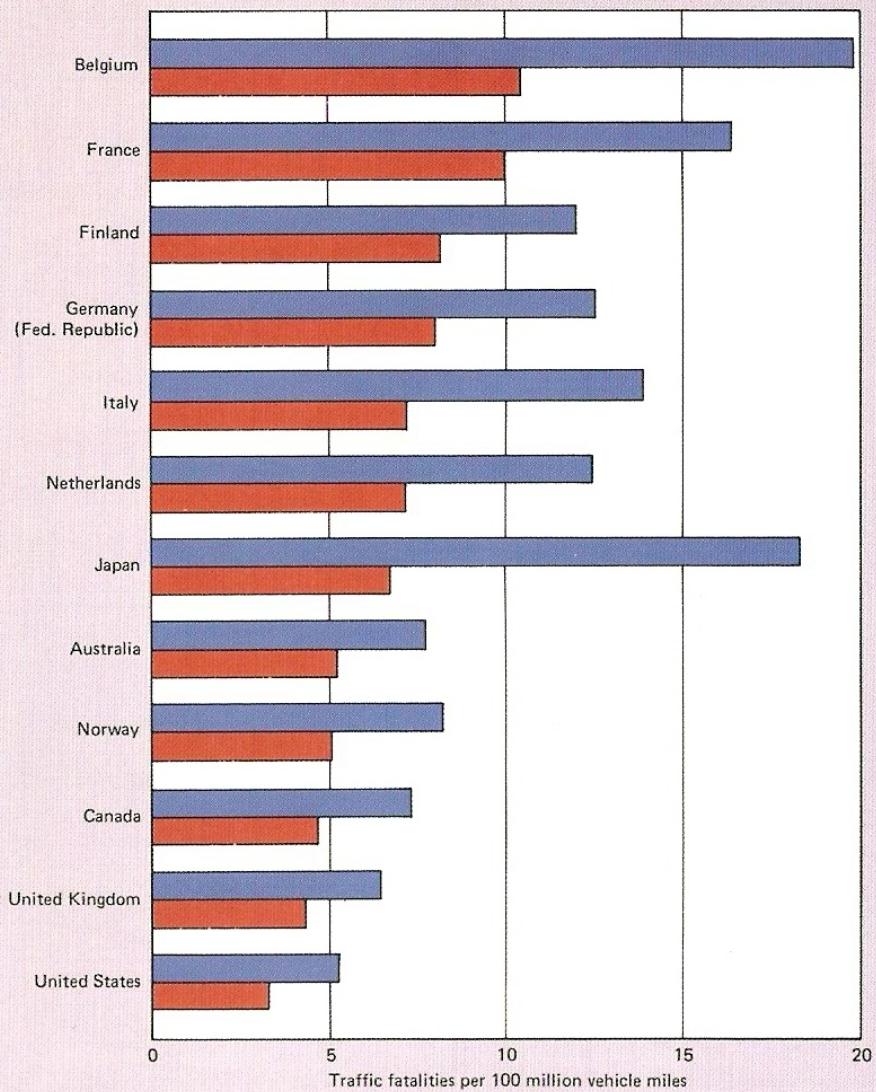


Figure 33: Wainer (1997), p. 37, Figure 36: Wainer (1997), p. 20B, Figure 35, improved.

Chart 4/25.

**Motor Vehicle Fatality Rates,
Selected Countries: 1969 and 1976**

1969
1976


CHAPTER 1, FIGURE 37. Ordering the bar chart by the data tells the tale a bit more clearly.

Figure 34: Wainer (1997), p. 20B, Figure 37: Layout similar to Wainer (1997), p. 20B, Figure 35, but improved due to ordering.

Rule 10: Label: (a) illegibly, (b) incompletely, (c) incorrectly, and (d) ambiguously.

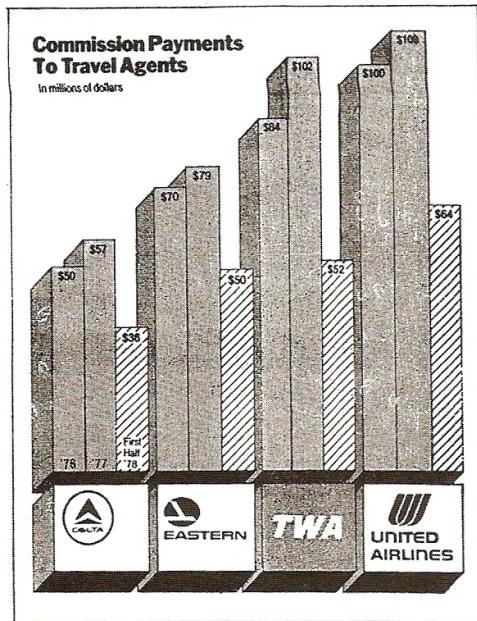


FIGURE 38. Mixing a changed metaphor with a tiny label reverses the meaning of the data.

Figure 35: Wainer (1997), p. 39, Figure 38.

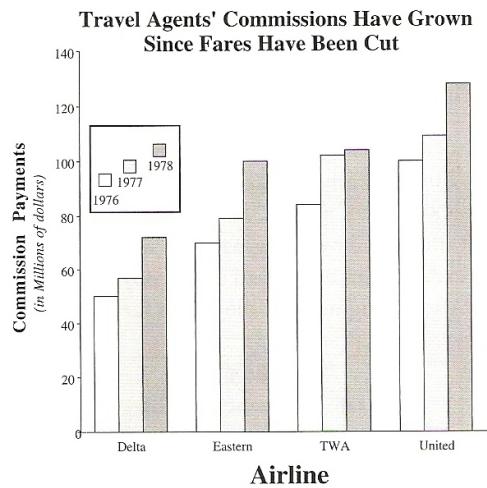


FIGURE 39. Figure 38 redrawn with 1978 data placed on a comparable basis shows that the fare cuts have been a boon to travel agents.

Figure 36: Wainer (1997), p. 39, Figure 39: Wainer (1997), p. 39, Figure 38, improved.

Rule 11: More is murkier: (a) more decimal places and (b) more dimensions.

TABLE 1		
Life Expectancy at Birth		
Country	Male	Female
Argentina	56.90	61.40
Brazil	39.30	45.50
Canada	67.61	72.92
Iceland	66.10	70.30
Japan	65.37	70.26
Mexico	37.92	39.79
Netherlands	71.40	74.80
New Zealand	68.20	73.00
Norway	71.11	74.70
Spain	58.76	63.50

Figure 37: Wainer (1997), p. 40, Table 1.

TABLE 2		
Life Expectancy at Birth		
Country	Male	Female
Netherlands	71	75
Norway	71	75
New Zealand	68	73
Canada	68	73
Iceland	66	70
Japan	65	70
Spain	59	64
Argentina	57	61
Brazil	39	46
Mexico	38	40

Figure 38: Wainer (1997), p. 40, Table 2: Wainer (1997), p. 40, Table 1, improved.

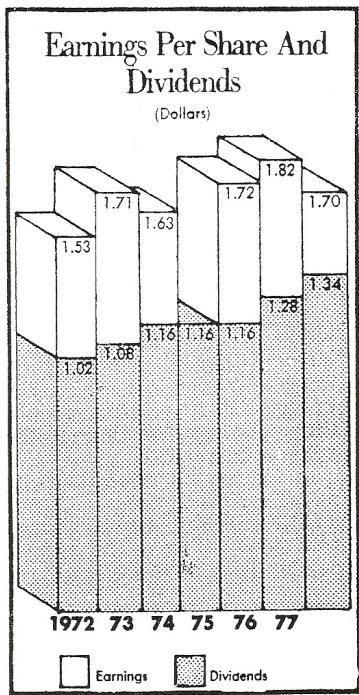


FIGURE 41. An extra dimension on earnings and dividends confuses even the grapher (from the *Washington Post*, 1979).

Figure 39: Wainer (1997), p. 41, Figure 41.

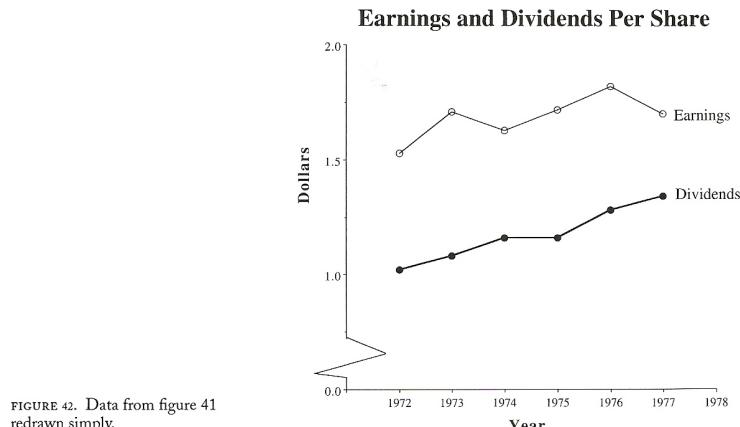
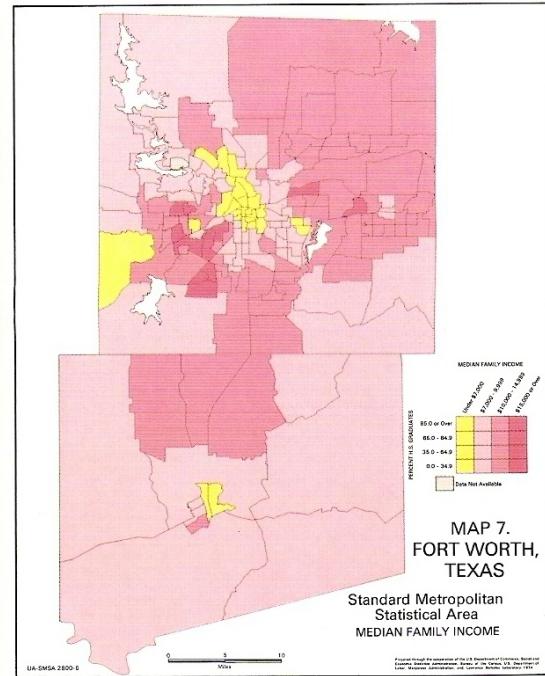


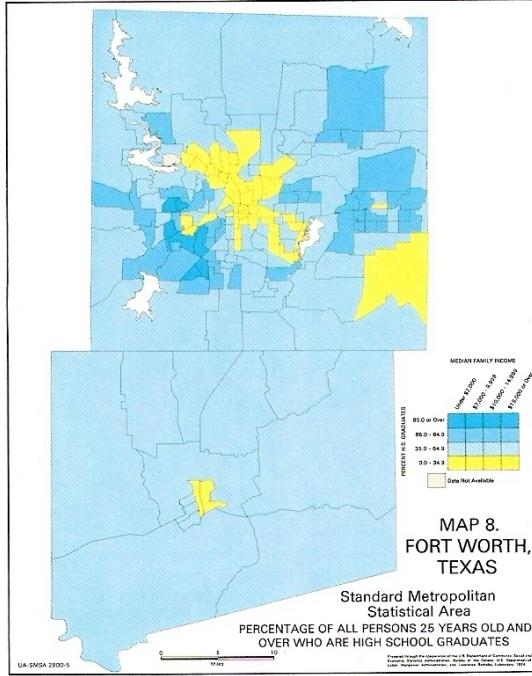
FIGURE 42. Data from figure 41 redrawn simply.

Figure 40: Wainer (1997), p. 42, Figure 42: Wainer (1997), p. 41, Figure 41, improved.

Rule 12: If it has been done well in the past, think of a new way to do it.

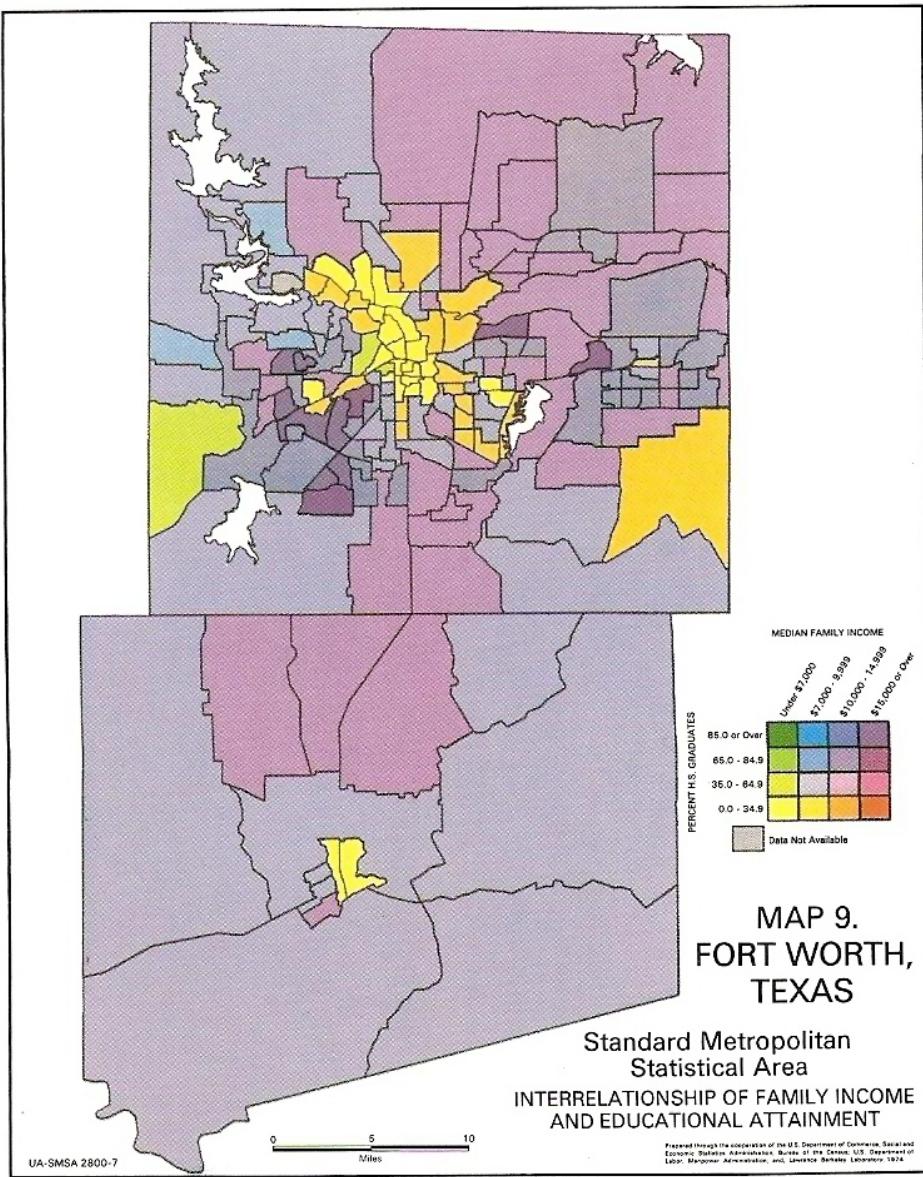


CHAPTER 1, FIGURE 44. The geographic distribution of median family income in Fort Worth, Texas, in 1974.



CHAPTER 1, FIGURE 45. The geographic distribution of percentage of high-school graduates in Fort Worth, Texas, in 1974.

Figure 41: Wainer (1997), p. 20C, Figures 44 & 45: Traditionl maps.



CHAPTER 1, FIGURE 46. The geographic distribution of both median family income and percentage of high-school graduates in Fort Worth, Texas, in 1974, shown as a two-variable color map.

Figure 42: Wainer (1997), p. 20C, Figure 46: Wainer (1997), p. 20C, Figures 44 & 45, modified but **not** improved.

1.2 Bad Graphics are Everywhere — In Space and in Time

Example 1: Zion National Park, UT, Shuttle Parking Lot, December 28, 2002

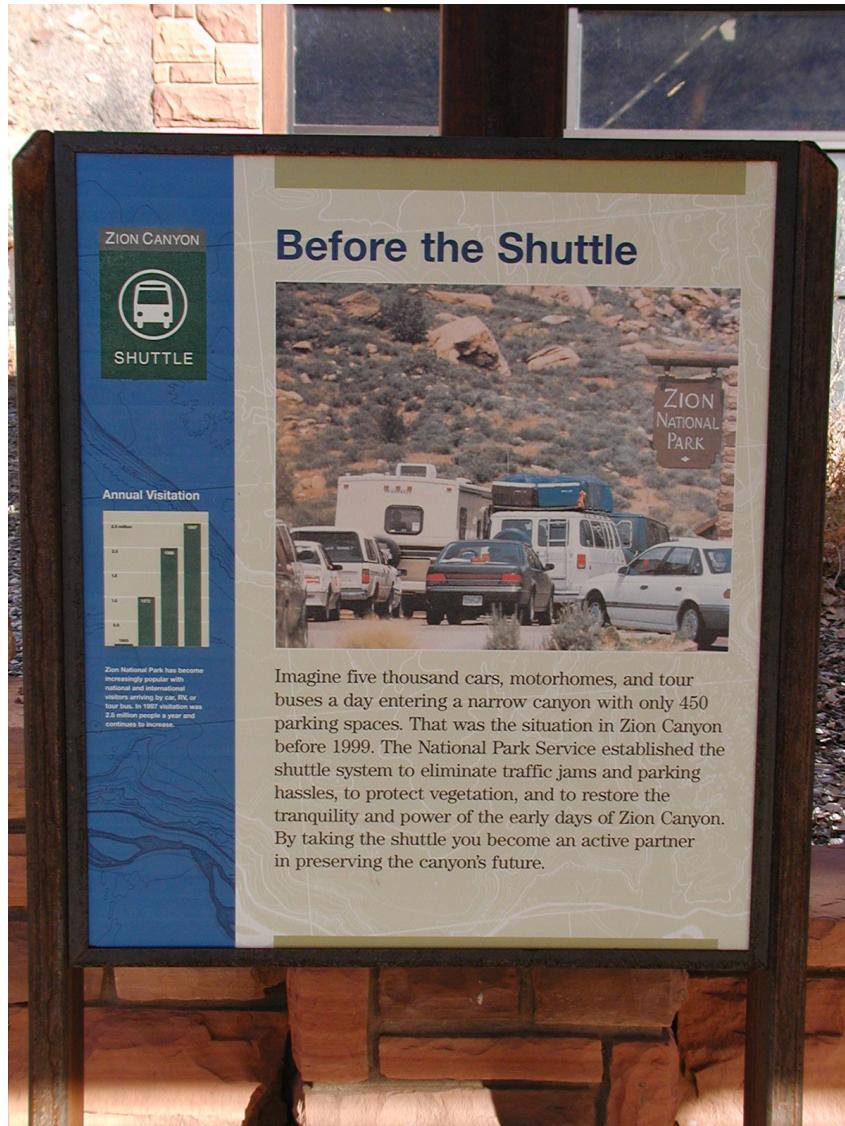


Figure 43: Personal Photograph: From the distance, the annual visitation appears to increase linearly,

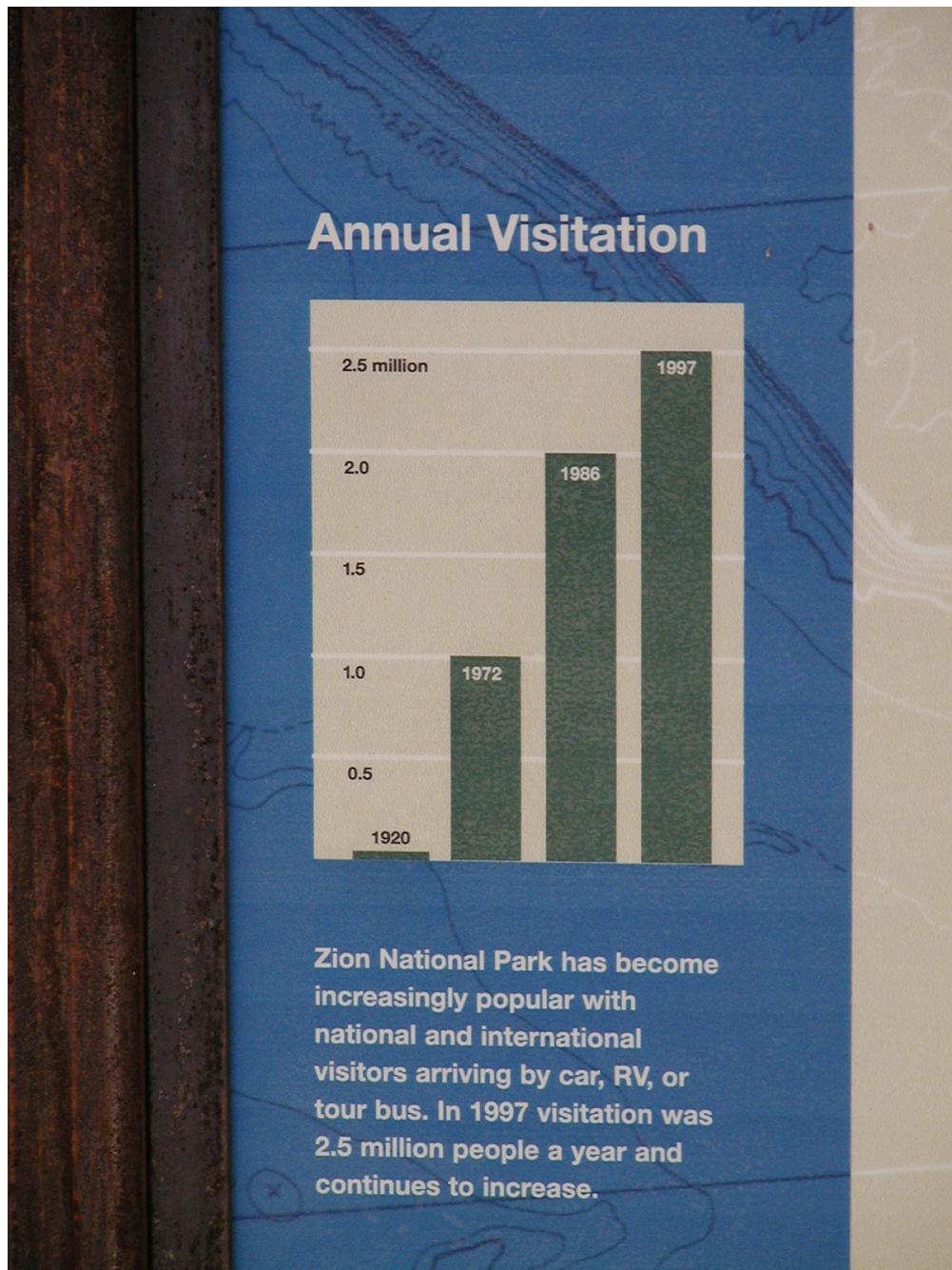


Figure 44: Personal Photograph: ... but at a closer view, this is certainly not the case.

Rules followed (to make this a bad graphic):

Improved Version:

Example 2: Berlin, Germany, August 20, 2006



Figure 45: Personal Photograph: Exhibit at the 1936 Berlin Olympic Site, related to the history of the Olympic area from 1909 to 1936 to 2006.



Figure 46: Personal Photograph: Historical graphic (from the late 1920ies), dedicated to the development of women's gymnastics as part of the *Deutsche Turnerschaft* (the governing body of German gymnastics).

Rules followed (to make this a bad graphic):

Improved Version:

Example 3: Wikipedia, 2009

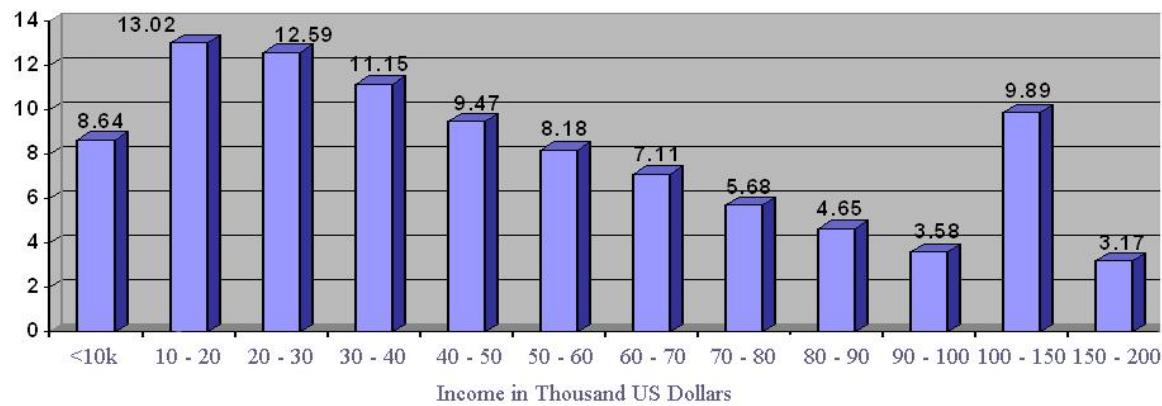


Figure 47: Figure taken from http://en.wikipedia.org/wiki/Household_income_in_the_United_States on 1/13/2009.

Rules followed (to make this a bad graphic):

Improved Version:

Example 4: Computational Statistics, 2002

430

age, surface and contour graphs. Various types of graphs created by KyPlot are shown in Figures 3 and 4.

Almost every component of each graph can be customized through dialog boxes. Double-clicking an axis of a graph brings up a dialog box through which one can change various settings for the axis interactively. The scales of the x- and y-axes of graphs can be individually set as either linear or logarithmic. Error bars can be attached to either x- or y-values, or both, and the attributes of individual error bars can also be customized. (For example, in Figure 3A, the error bars for two data points of a line graph have been partially suppressed to avoid overlapping.) A break along an axis can be set, over a specific range and at a specific location, to indicate that a range of values has been omitted (Figure 3B).

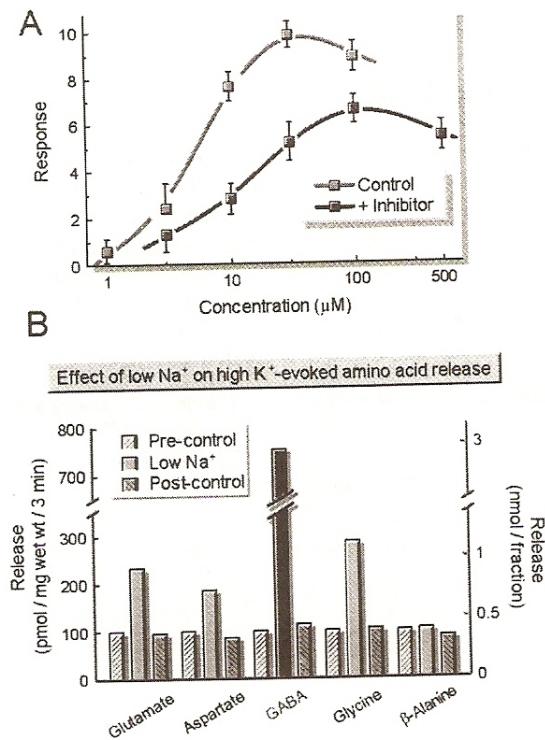


Figure 3: Line and bar graphs created with KyPlot

Figure 48: Yoshioka (2002), p. 430, Figure 3: Intended (!) features of the KyPlot software package for statistical data analysis and visualization.

Rules followed (to make this a bad graphic):

Improved Version:

1.3 Rules for Good Data Displays

Wainer (1997), p. 46, suggests:

- “1. Examine the data carefully enough to know what they have to say, and then let them say it with a minimum of adornment.
2. In depicting scale, follow practices of “reasonable regularity.”
3. Label clearly and fully.”

Tufte (1983), p. 77, suggests:

“Graphical integrity is more likely to result if these six principles are followed:

- The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.
- Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.
- Show data variation, not design variation.
- In time-series displays of money, deflated and standardized units of monetary measurements are nearly always better than nominal units.
- The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.
- Graphics must not quote data out of context.”

Robbins (2005), pp. 375–377, provides a “*Checklist of Possible Graph Defects*” in her Appendix A:

“Can the reader clearly see the graphical elements?

- Do the data stand out? Are there superfluous elements?
- Are all graphical elements visually prominent?
- Are overlapping plotting symbols visually distinguishable?
- Can superposed data sets be readily visually assembled?
- Is the interior of the scale-line rectangle cluttered?
- Do data labels interfere with the quantitative data or clutter the graph?
- Is the data rectangle within the scale-line rectangle?
- Do tick marks interfere with the data?
- Do tick mark labels interfere with the data?
- Are axis labels legible?
- Are there too many tick marks?

Fig. 6.1 Terminology

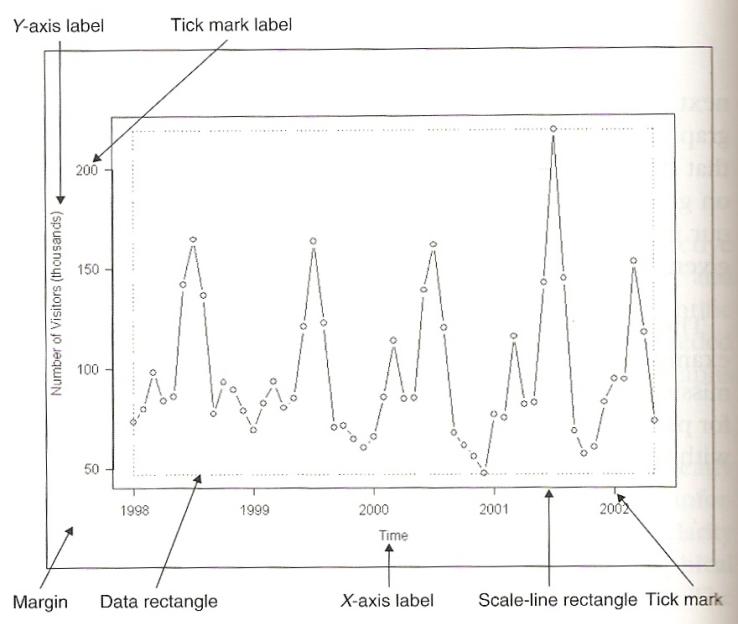


Figure 49: Robbins (2005), p. 156, Figure 6.1.

- Are there too many tick mark labels?
- Do the grid lines interfere with the data?
- Are there notes or keys inside the scale-line rectangle?
- Will visual clarity be preserved under reduction and reproduction?

Can the reader clearly understand the graph?

- Are the data drawn to scale?
- Is there an informative title?
- Is area or volume used to show changes in one dimension?
- Are there too many dimensions in the graph (more than in the data)?
- Are common baselines used wherever possible?
- Are all labels associated with the correct graphical elements?
- Is the reader required to make calculations?
- Are groups of charts drawn consistently?

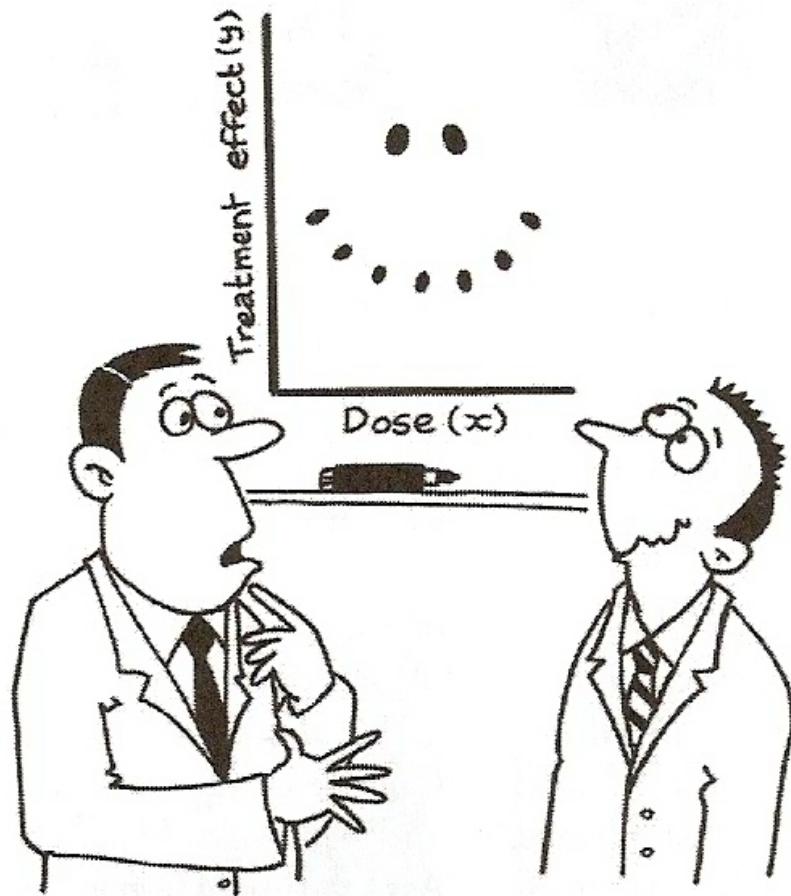
Are the scales well chosen and labeled?

- Is zero included for all bar graphs?
- Are there any unnecessary scale breaks?
- Is there a forceful indication of a scale break?
- Are there numerical values on two sides of a scale break that are connected?
- Does the aspect ratio allow the reader to see variations in the data?
- Are scales included for all axes?
- Are the scales labeled?
- Are tick marks at sensible values?
- Do the axes increase in the conventional direction?
- Does the data rectangle fill as much of the scale-line rectangle as possible?
- Are uneven time intervals handled correctly?
- Are the scales appropriate when different panels are compared?"

1.4 Further Reading

In addition to Wainer (1997), Tufte (1983), and Robbins (2005), cited so far in this chapter, many other sources exist that compare bad graphics with good graphics. Some of these additional sources are:

- Bertin (1977) and Bertin (2005) (first published in 1967)
- Henry (1995)
- Holmes (1991): check the author credentials and then decide whether this book is a source for good or bad graphics
- Huff & Geis (1954)
- Jones (2000)
- Kosslyn (1994) and Kosslyn (2006)
- Krämer (1991)
- Wainer (2005)
- Wainer (2007)
- Wallgren et al. (1996)
- Zelazny (2001)



"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

A CAUSE-commissioned cartoon that is part of the CAUSEweb collection and available for free noncommercial use by statistics teachers. Cartoon by John Landers ©. Provided by permission.

Figure 50: Amstat News, January 2009, p. 25, Cartoon.

References

- Bertin, J. (1977), *La Graphique et le Traitement Graphique de l'Information*, Flammarion, Paris, France.
- Bertin, J. (2005), *Sémiologie Graphique: Les Diagrammes — Les Réseaux Les Cartes (4e édition)*, Les ré-impressions des Éditions de l'École des Hautes Études en Sciences Sociales, Paris, France.
- Henry, G. T. (1995), *Graphing Data: Techniques for Display and Analysis*, Sage Publications, Thousand Oaks, CA.
- Holmes, N. (1991), *Designer's Guide to Creating Charts & Diagrams (Paperback Edition)*, Watson-Guptill Publications, New York, NY.
- Huff, D. & Geis, I. (1954), *How to Lie with Statistics*, W. W. Norton & Company, New York, NY.
- Jones, G. E. (2000), *How to Lie with Charts*, toExcel Press, Lincoln, NE.
- Kosslyn, S. M. (1994), *Elements of Graph Design*, W. H. Freeman and Company, New York, NY.
- Kosslyn, S. M. (2006), *Graph Design for the Eye and Mind*, Oxford University Press, New York, NY.
- Krämer, W. (1991), *So lügt man mit Statistik (3. Auflage)*, Campus Verlag, Frankfurt/Main, Germany.
- Robbins, N. B. (2005), *Creating More Effective Graphs*, Wiley, Hoboken, NJ.
- Tufte, E. R. (1983), *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT.
- Wainer, H. (1997), *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot*, Copernicus/Springer, New York, NY.
- Wainer, H. (2005), *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*, Princeton University Press, Princeton, NJ.
- Wainer, H. (2007), 'Improving Data Displays: Ours and the Media's', *Chance* **20**(3), 8–15.

Wallgren, A., Wallgren, B., Persson, R., Jorner, U. & Haaland, J.-A. (1996), *Graphing Statistics & Data: Creating Better Charts*, Sage Publications, Thousand Oaks, CA.

Yoshioka, K. (2002), 'KyPlot — A User–Oriented Tool for Statistical Data Analysis and Visualization', *Computational Statistics* 17(3), 425–437.

Zelazny, G. (2001), *Say it with Charts: The Executive's Guide to Visual Communication (Fourth Edition)*, McGraw–Hill, New York, NY.

— THE END —

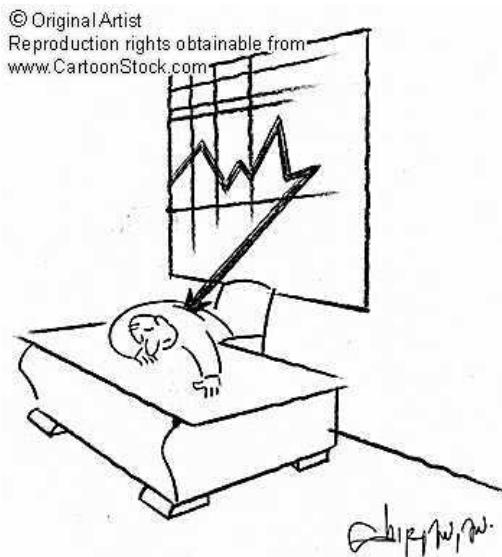


Figure 51: http://www.cartoonstock.com/blowup_stock.asp?imageref=vsh0184&artist=Shirvanian,+Vahan&topic=statistics+, Cartoon.