Chapter 9

Plots, Graphs, and Pictures

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Thought Question 1:

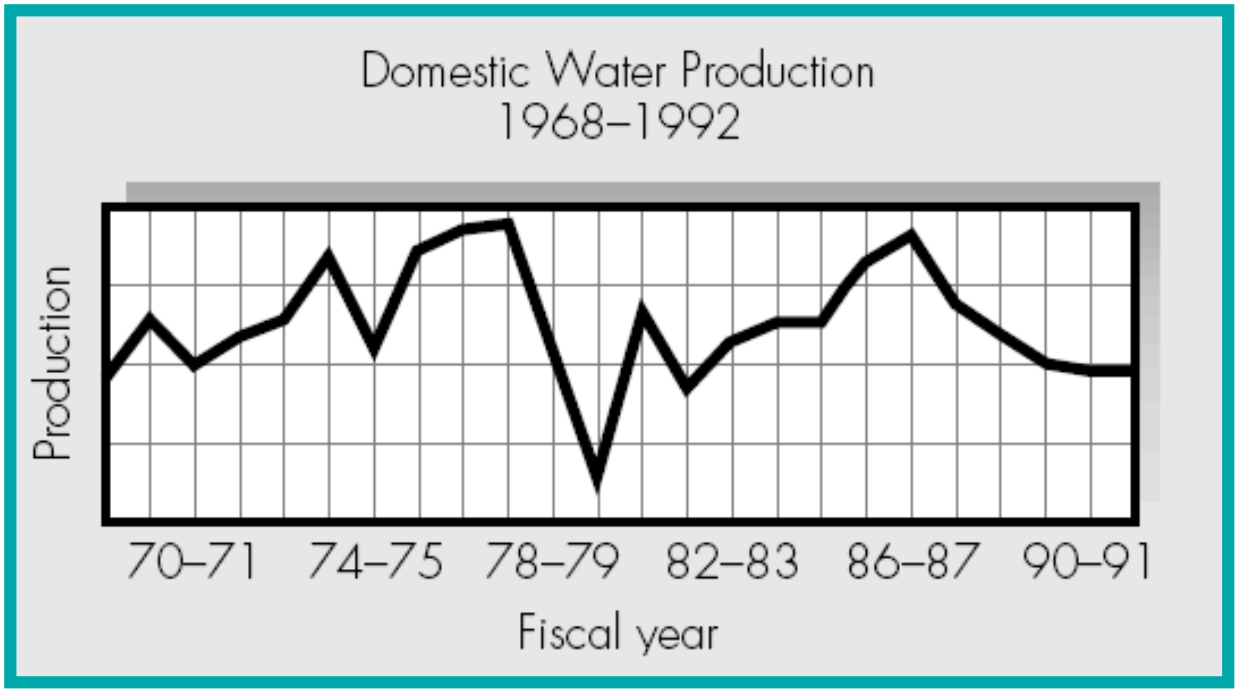
Suppose you have been keeping track of your living expenses and find that you spend 50% of your money on rent, 25% on food, and 25% on other expenses.

Draw a pie chart and a bar graph to depict this information. Discuss which is more visually appealing and useful.

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Thought Question 2:

Here is a plot that has some problems.
Give two reasons why this is not a good plot.



Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Thought Question 3:

Suppose you had a set of data representing two measurement variables—namely, height and weight—for each of 100 people.

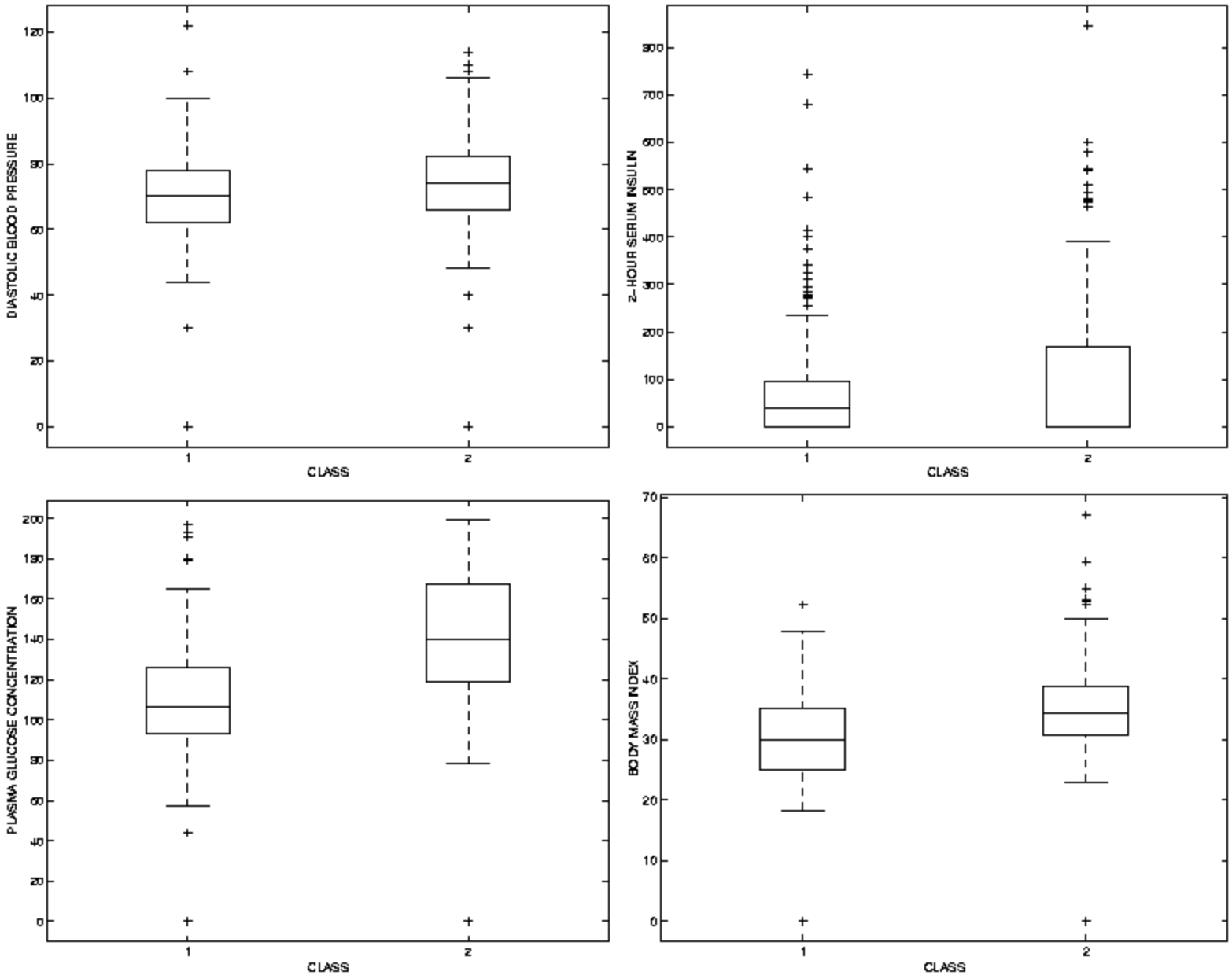
How could you put that information into a plot, graph, or picture that illustrated the relationship between the two measurements for each person?

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

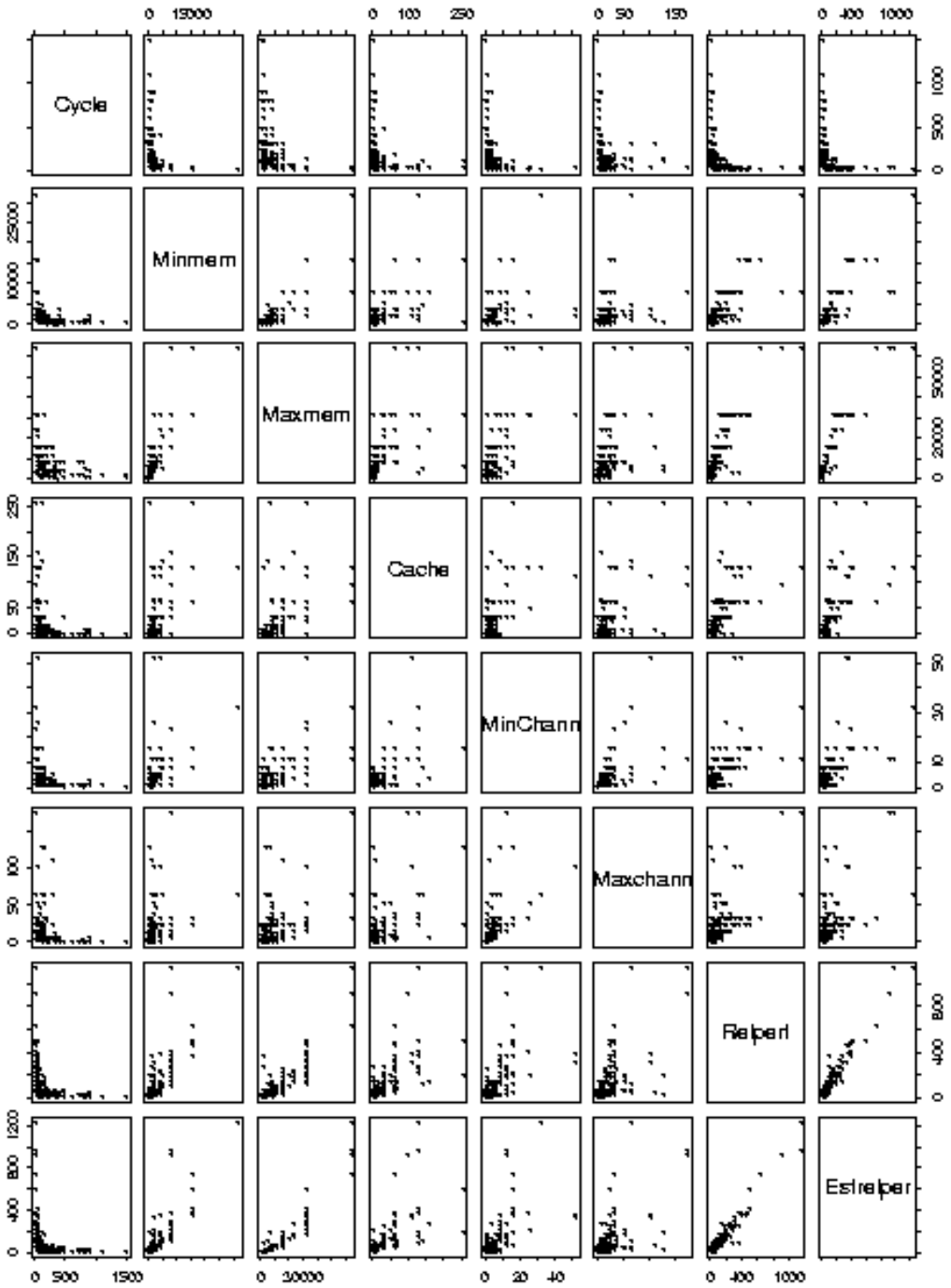
Thought Question 4:

Suppose you own a company that produces candy bars and you want to display two graphs. One graph is for customers and shows the price of a candy bar for each of the past 10 years. The other graph is for stockholders and shows the amount the company was worth for each of the past 10 years. You decide to adjust the dollar amounts in one graph for inflation but to use the actual dollar amounts in the other graph. If you were trying to present the most favorable story in each case, which graph would be adjusted for inflation? Explain.

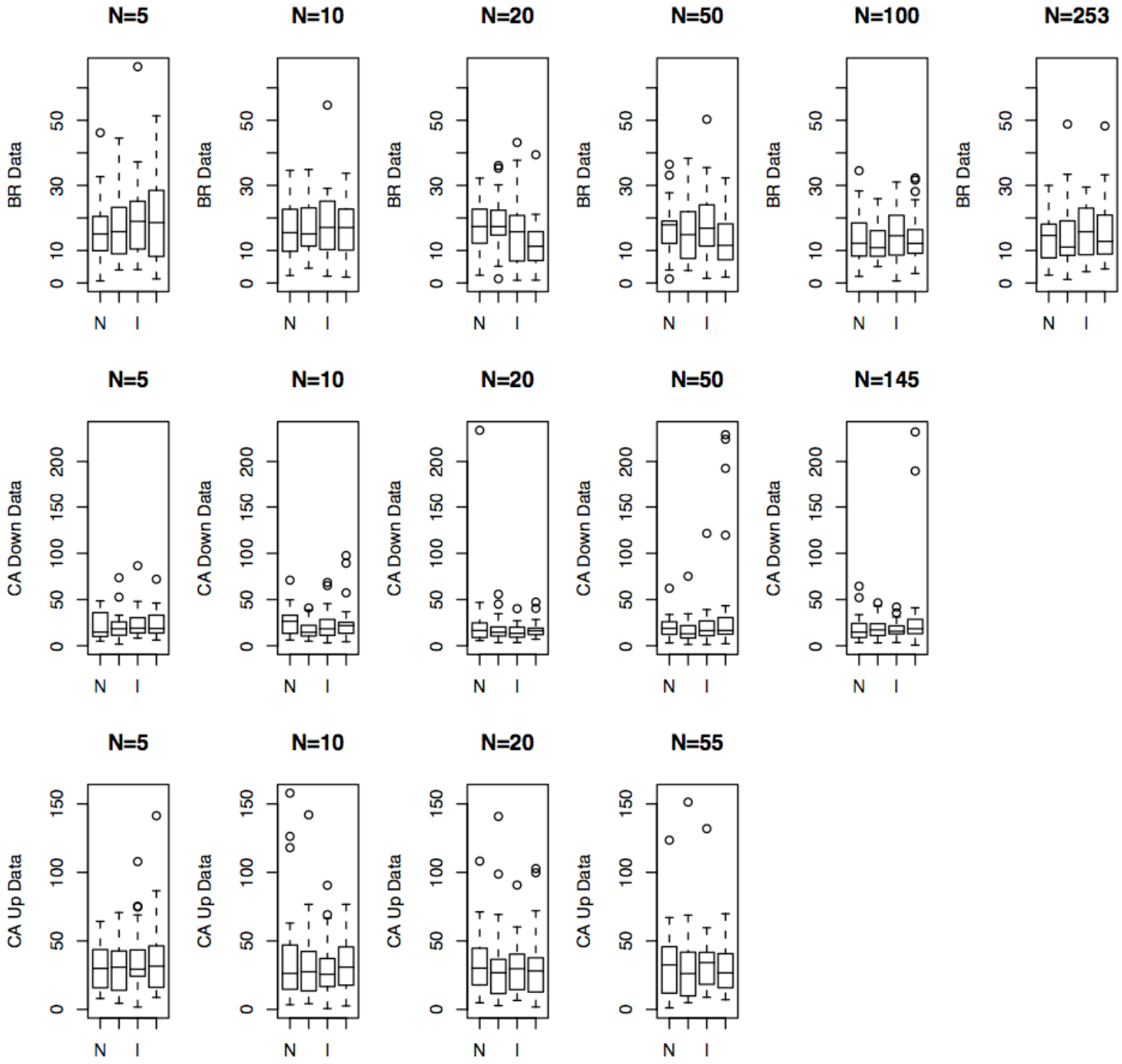
Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.



Copyright ©2008 Brooks/Cole, a division of Thomson Learning, Inc.



Copyright ©2005 Brool


Corridor Effect: None (N), Main (M), Interaction (I), Both (B), Error in Feet
N=5
N=10
N=20
N=50
N=100
N=253
BR Data
N=5
N=10
N=20
N=50
N=145
CA Down Data
N=5
N=10
N=20
N=55
CA Up Data
N
I

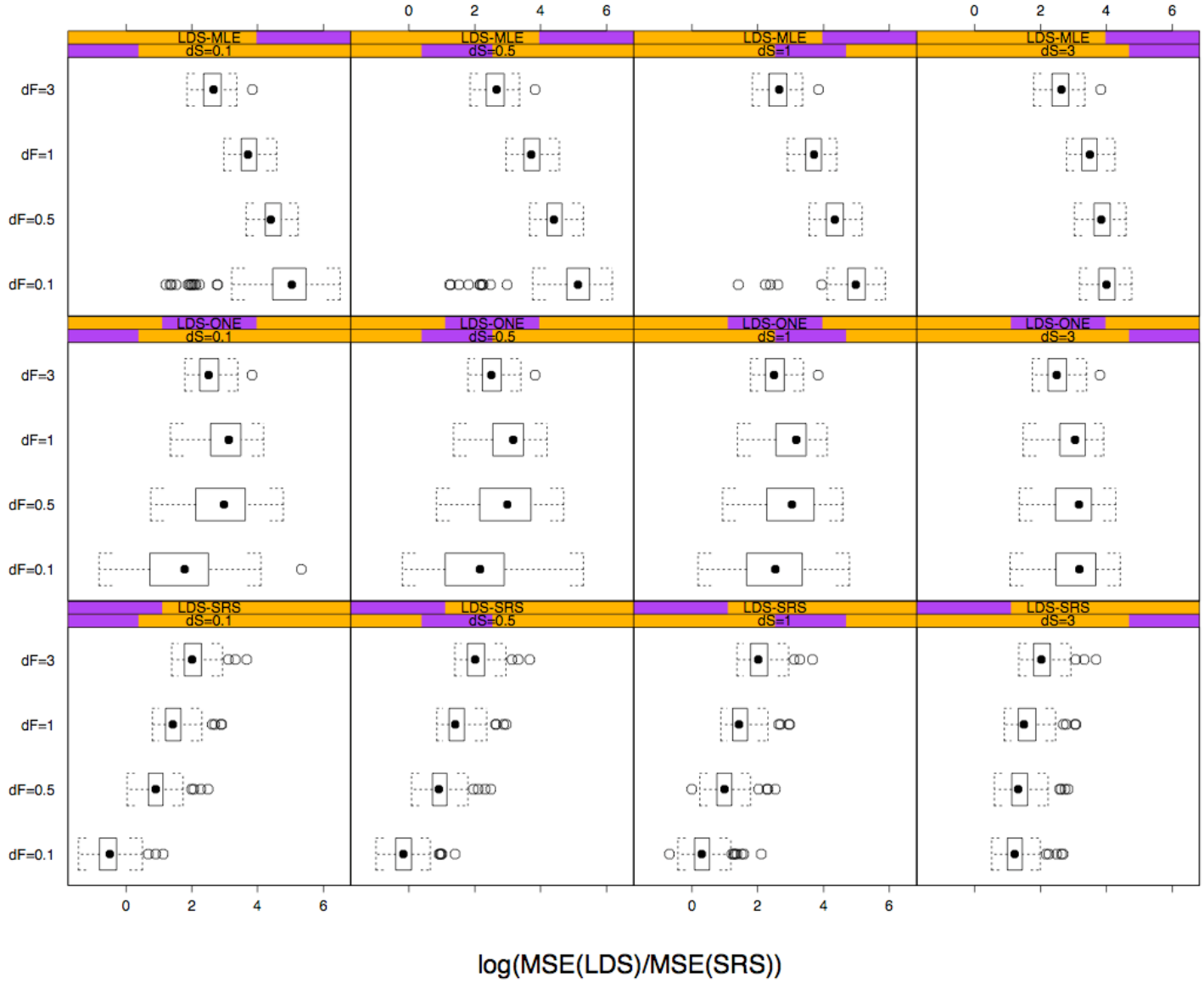
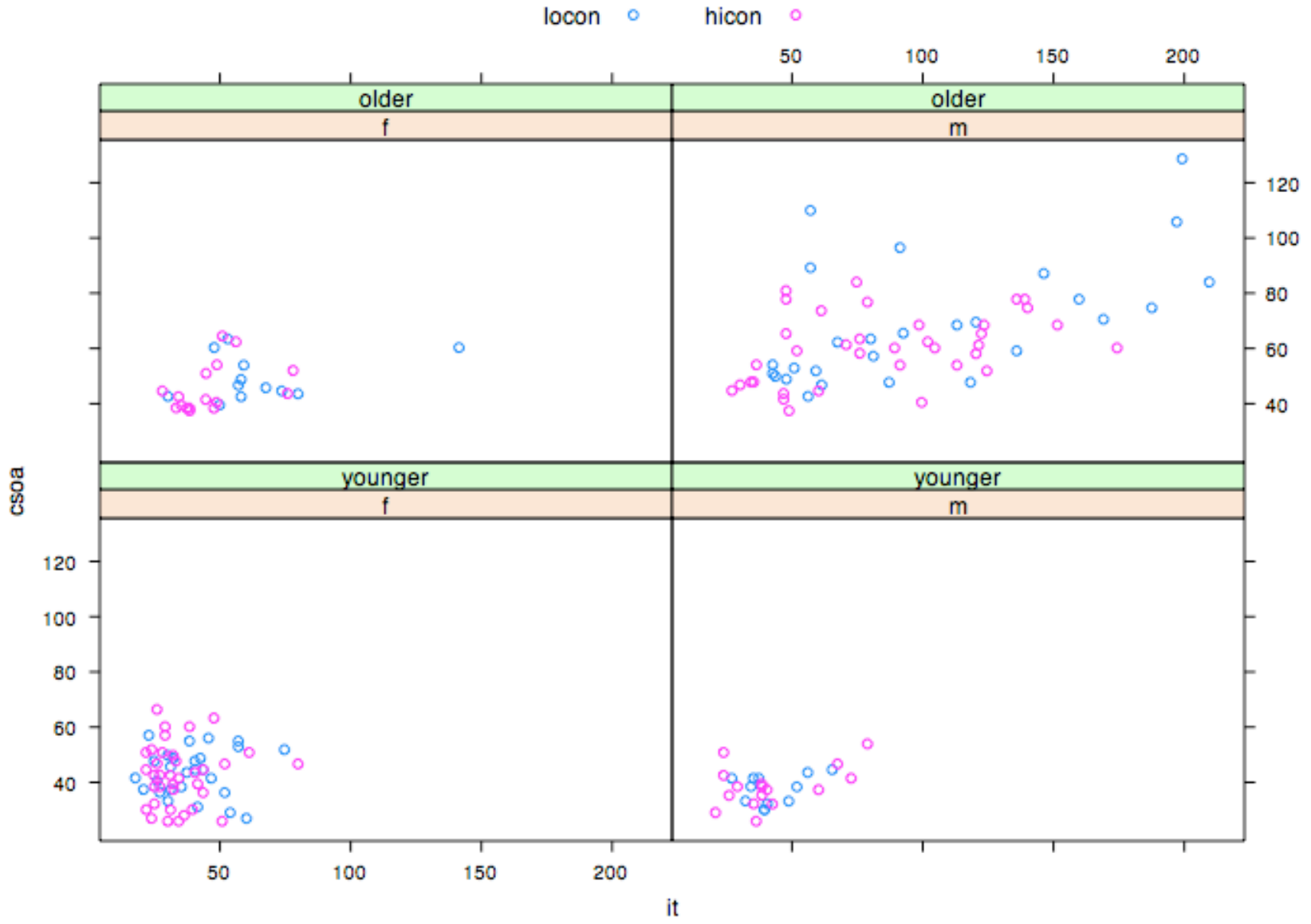


Figure 4: *Small Scale Simulation Results. Each boxplot shows a particular setting of* $\check{\theta}$, d_F, *and* d_S. *The horizontal axes show the log-ratio of the mean square error from random sampling to the mean square error from LDS.*

Tinting

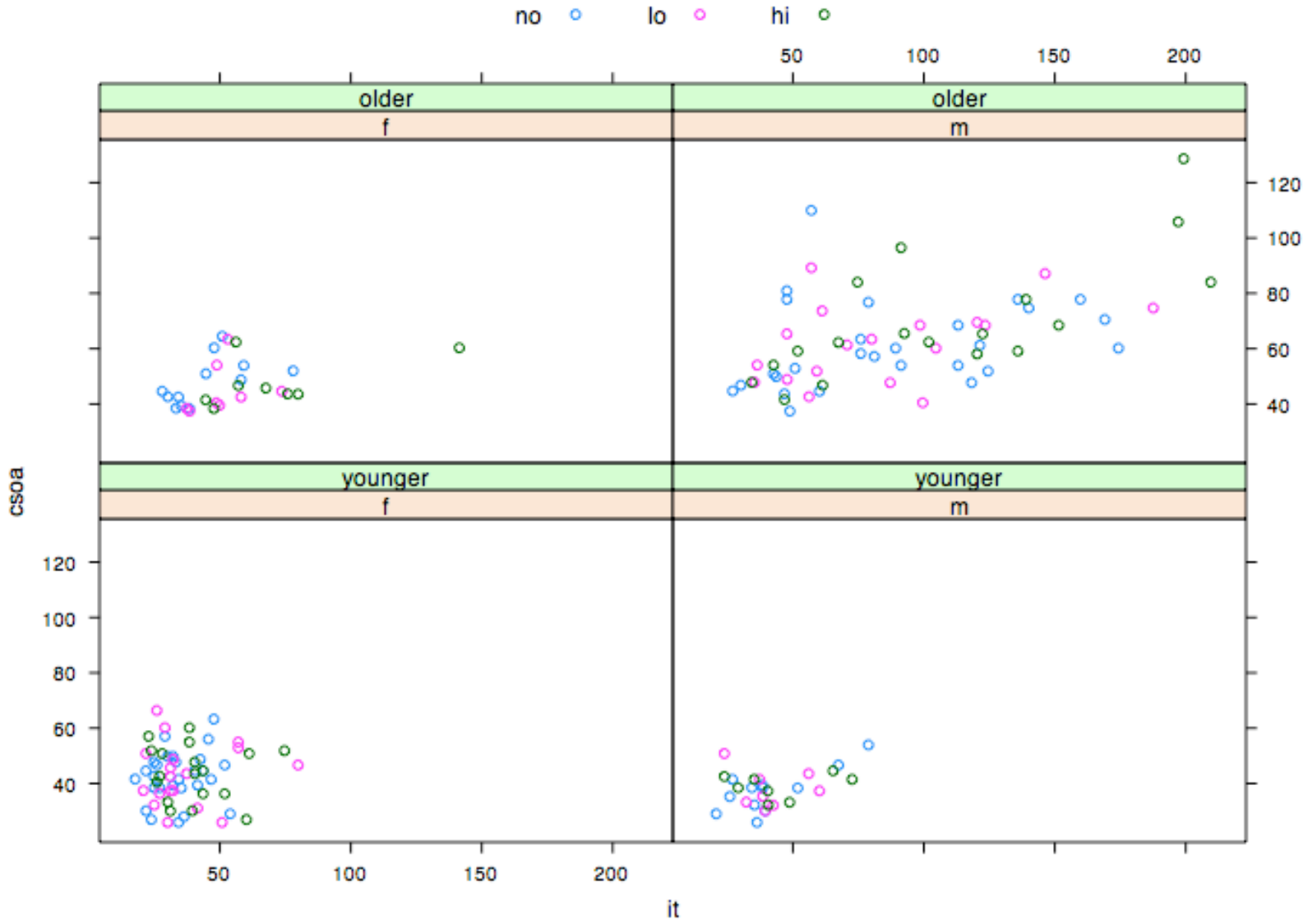
- Experiment to model the effects of car window tinting on visual performance
- csoa: critical stimulus onset asynchrony (time to recognize an alphanumeric target)
- it: inspection time (time required for a simple discrimination task)
- age, tint (no,lo,hi), target (locon,hicon), sex

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.



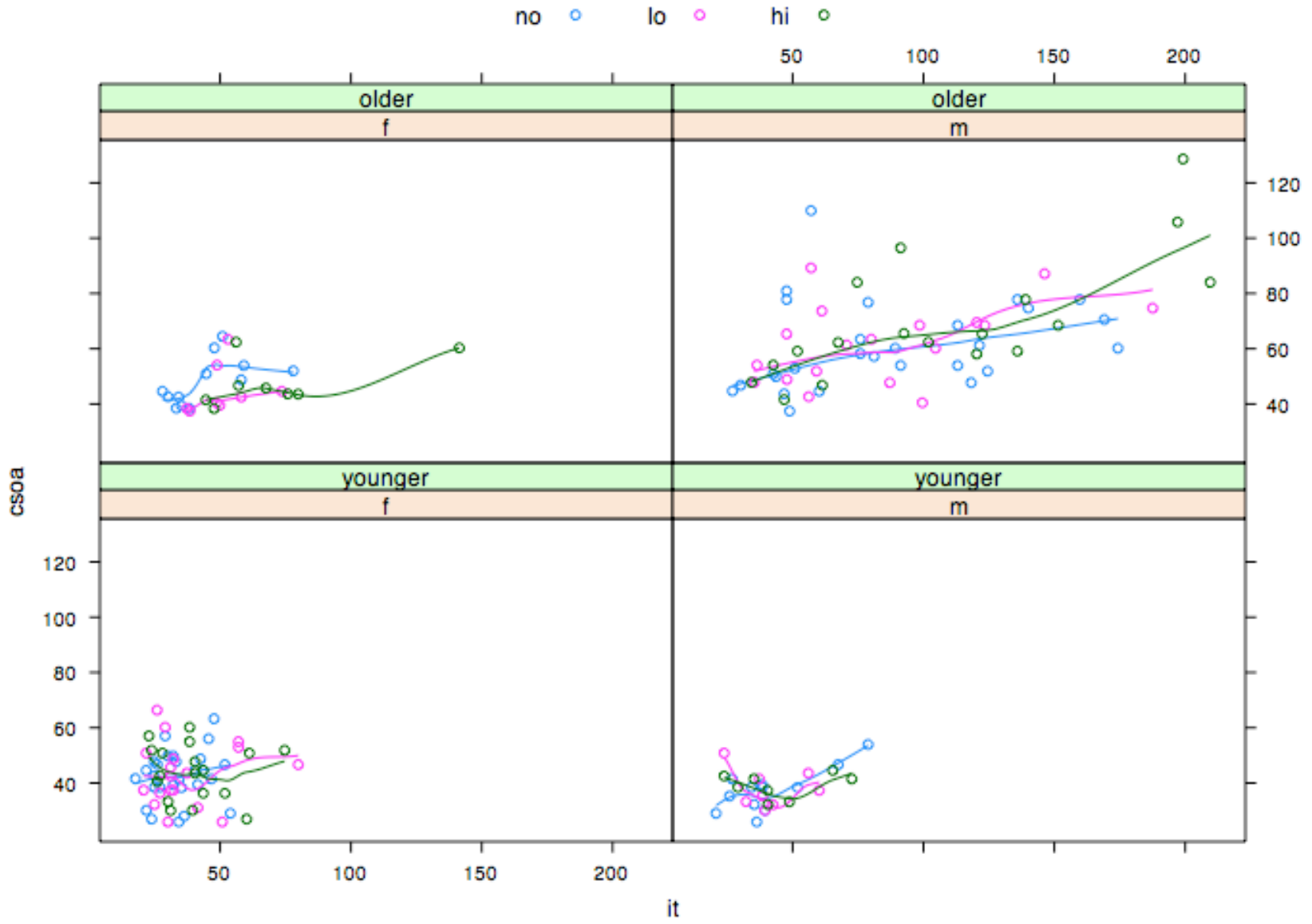
xyplot(csoa~it | sex*agegp, data=tinting, groups=target, auto.key=list(columns=2))

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

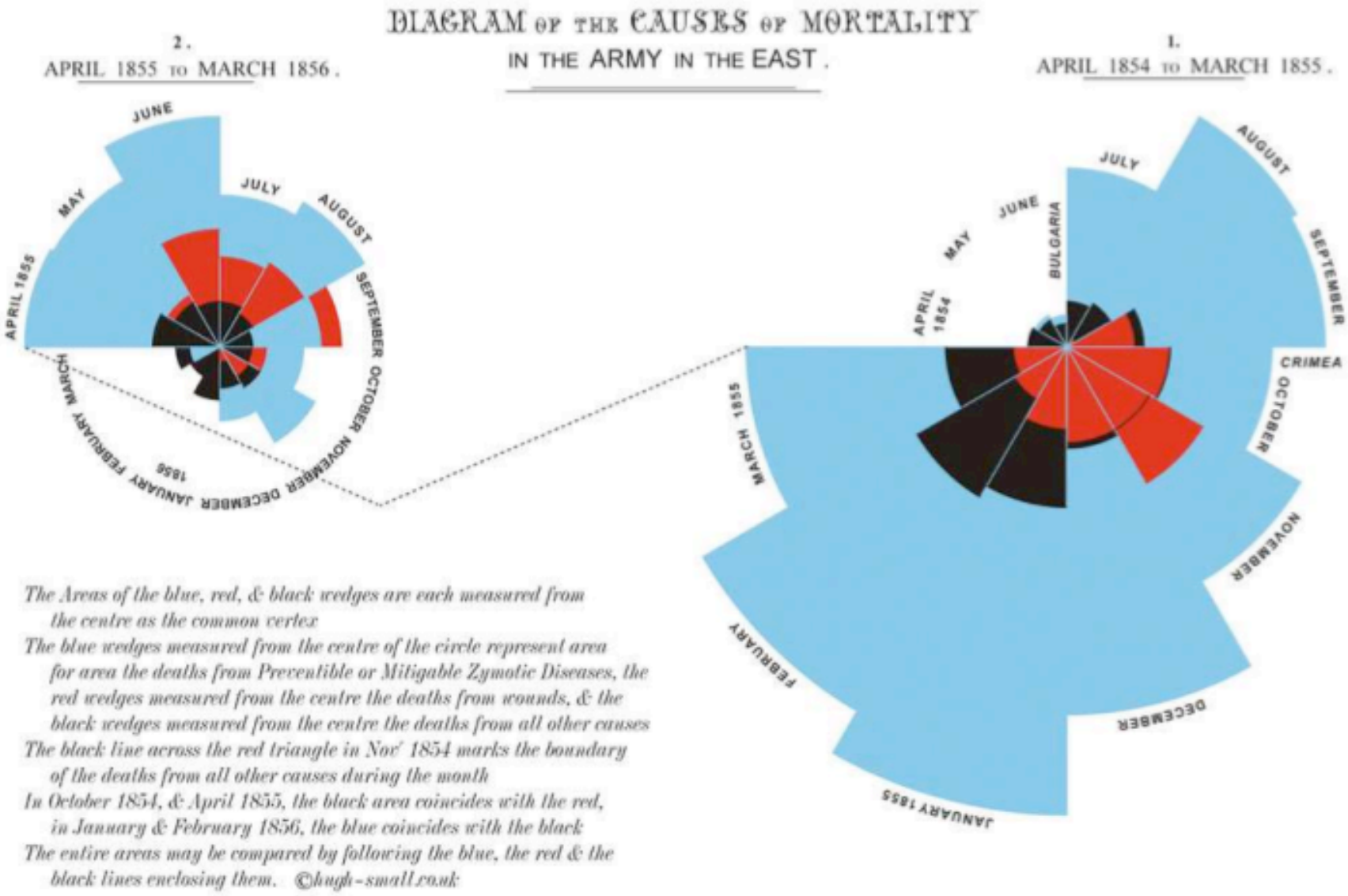


xyplot(csoa~it | sex*agegp, data=tinting, groups=tint, auto.key=list(columns=3))

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.



Copyright (c) 2005 Brooks/Cole, a division of Thomson Learning, Inc.



Source: Michael Friendly

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Tufte:

Graphical excellence is the well-designed presentation of interesting data - a matter of substance, of statistics, and of design.

Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency.

Graphical excellence is that which gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.

Graphical excellence is nearly always multivariate.

And graphical excellence requires telling the truth about the data.

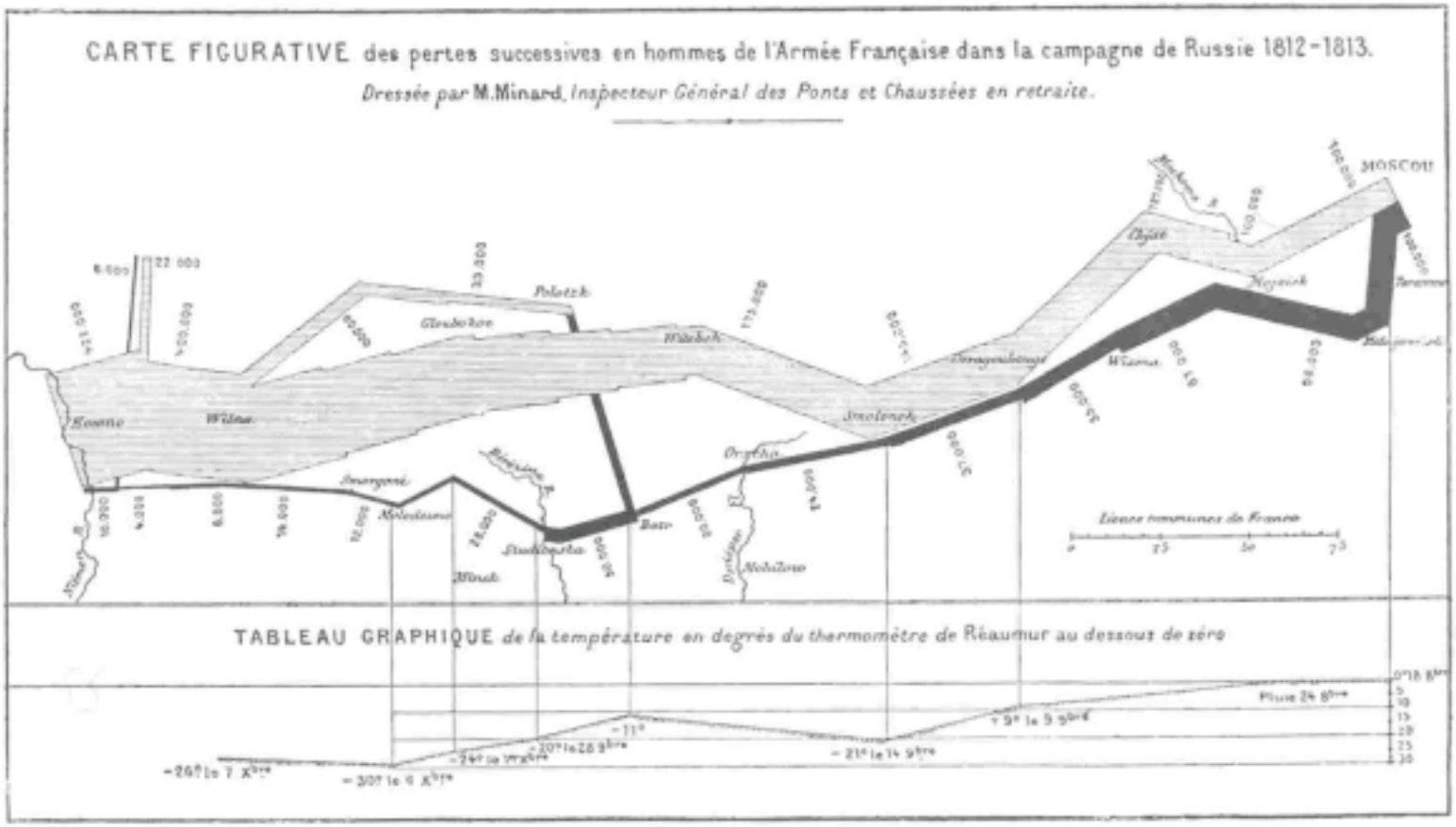
Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Tufte also insists that graphical displays should:

induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production or something else

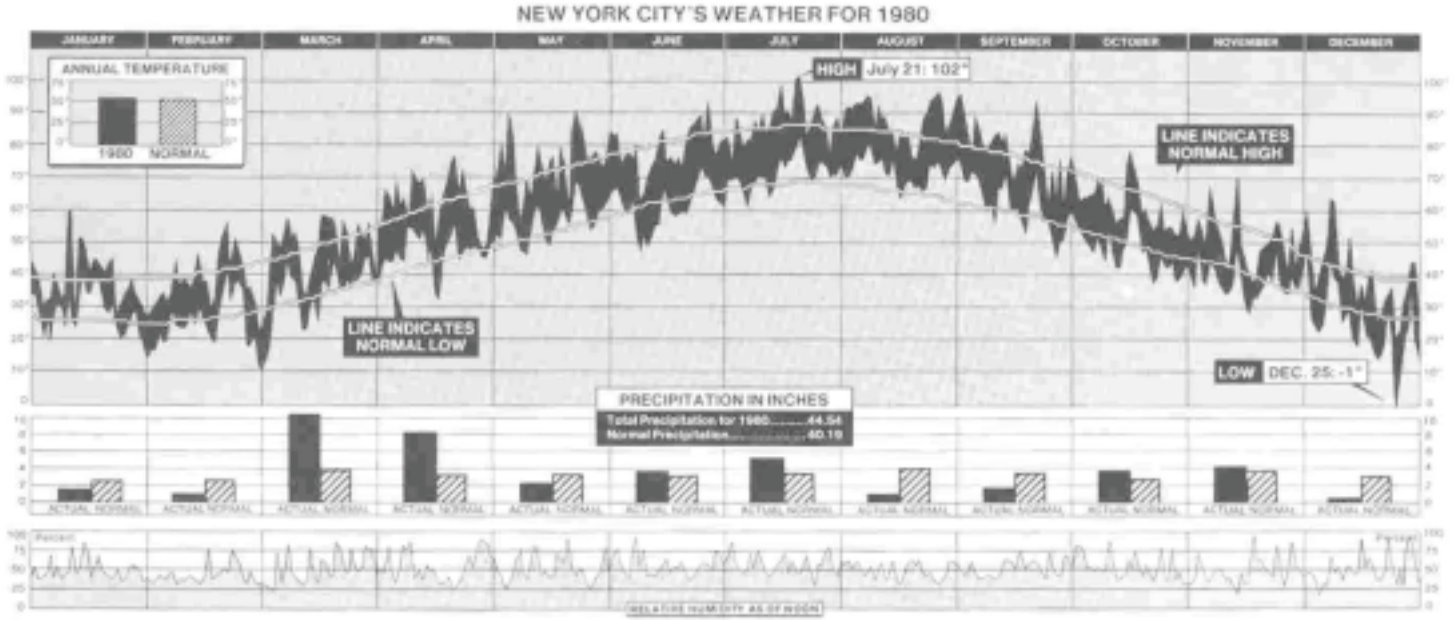
reveal the data at several levels of detail, from a broad overview to the fine structure

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.



Xbre = December 9bre = November 8bre = October

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.



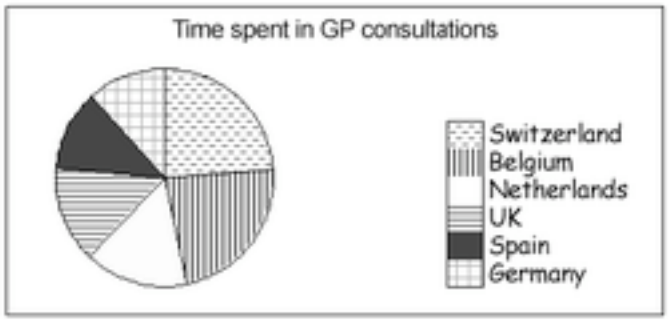
New York Times, January 11, 1981, p. 32.

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

In the following example, from The Times of Saturday 1/2/3 is a superb example of this form of abuse. The the two shells supposedly represent two quantities in the ratio 500 to 364, so the first should be 500/364 or 1.374 times bigger than the second, representing a 37.4% increase. But their lengths are in the ratio 10cm to 6.5cm, making the first 1.56 times as long as the second, and giving it a volume greater than that of the second by a factor of 1.569 cubed, or 3.864. This gives a shocking lie factor of 3.864/1.374 or 2.8 times!

Copyright © 2005 Brooks/Cole, a division of Thomson Learning, Inc.

This pie chart shows time spent with doctors. Use it to answer questions 4 to 7.



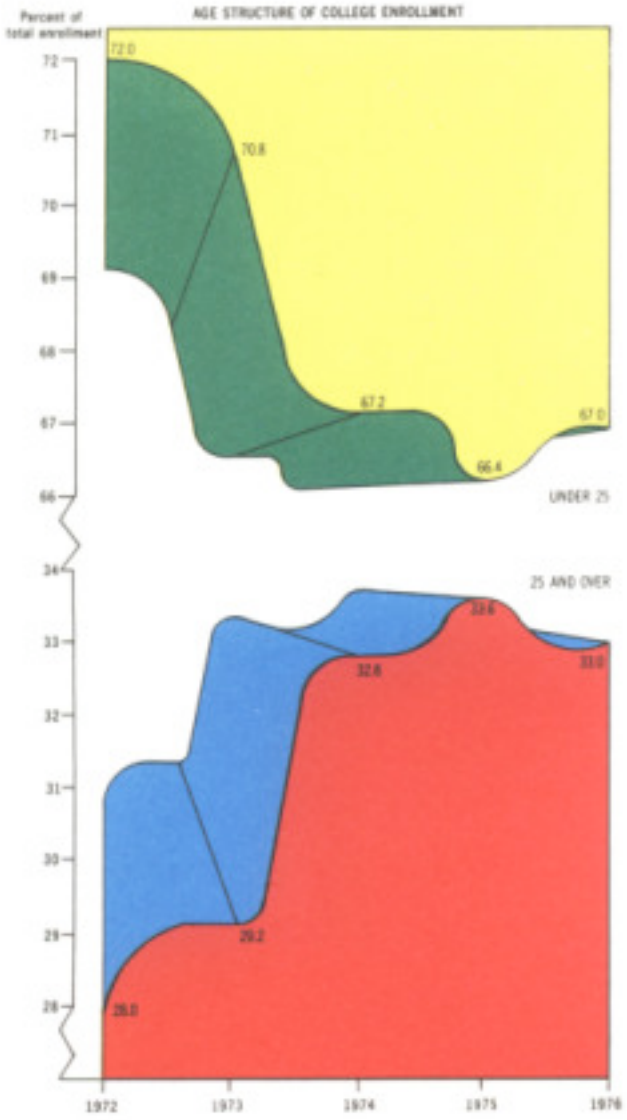
4) Which two countries give their patients the most time?
5) Which two countries give their patients the least time?
6) What colour is the UK slice?
7) Which country gives their patients about the same amount of time as the UK?

Now check your answers with those on the answer sheet.

BBC http://www.bbc.co.uk/skillswise ©

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Tufte's worst graphic ever!



Copyright ©2005 Brooks/C

http://www.ted.com/talks/view/id/92

ggobi

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

9.1 Well-Designed Statistical Pictures

Basic Characteristics:

1. Data should stand out clearly from background.
2. Clear labeling that indicates
   a. title or purpose of picture.
   b. what each axis, bar, pie segment, …, denotes.
   c. scale of each axis, including starting points.
3. Source for the data.
4. As little "chart junk" (extraneous material) as possible.

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

9.2 Pictures of Categorical Data

Three common pictures:

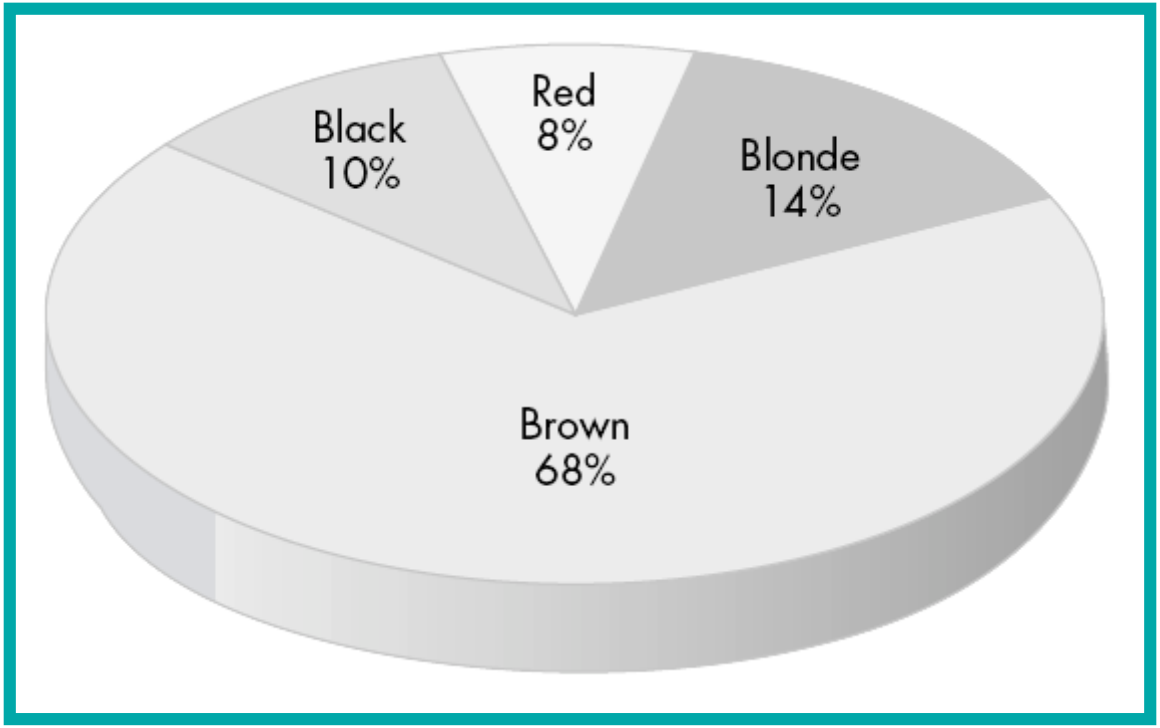
- Pie Charts
- Bar Graphs
- Pictograms

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Pie Charts

Show what percentage of the whole fall into each category for a single variable.

Pie chart of hair colors of white American children.



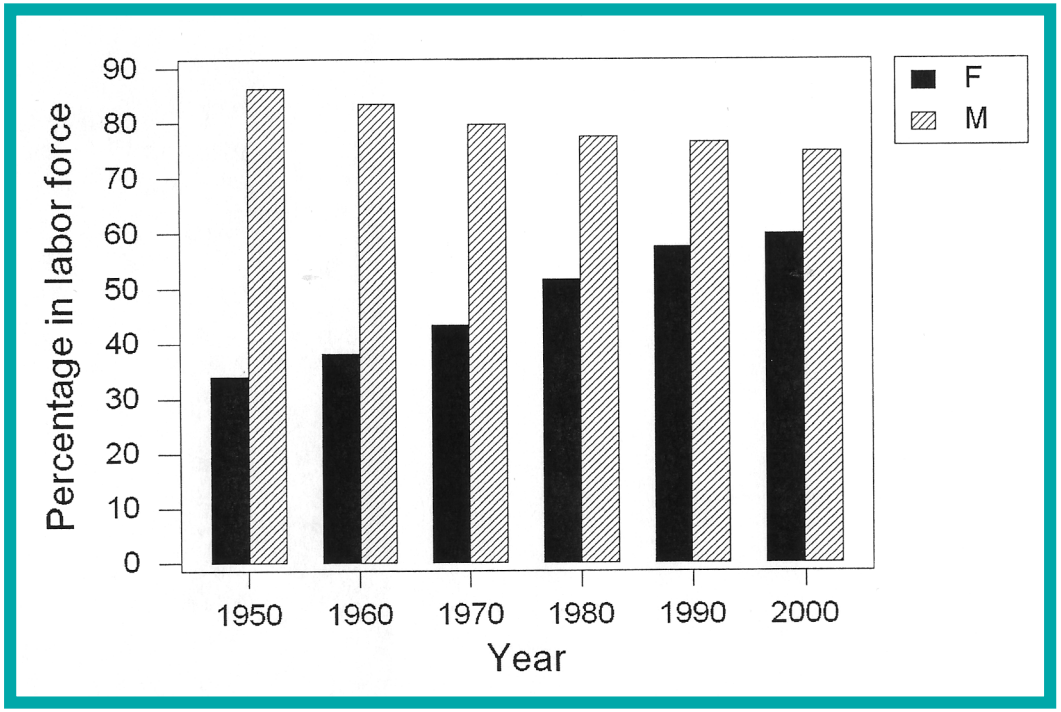
Source: Krantz, 1992, p. 188.

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Bar Graphs

Show what percentage or frequency of the whole fall into each category – can be used for two or three variables simultaneously.

Percentage of men and women 16 and over in the labor force



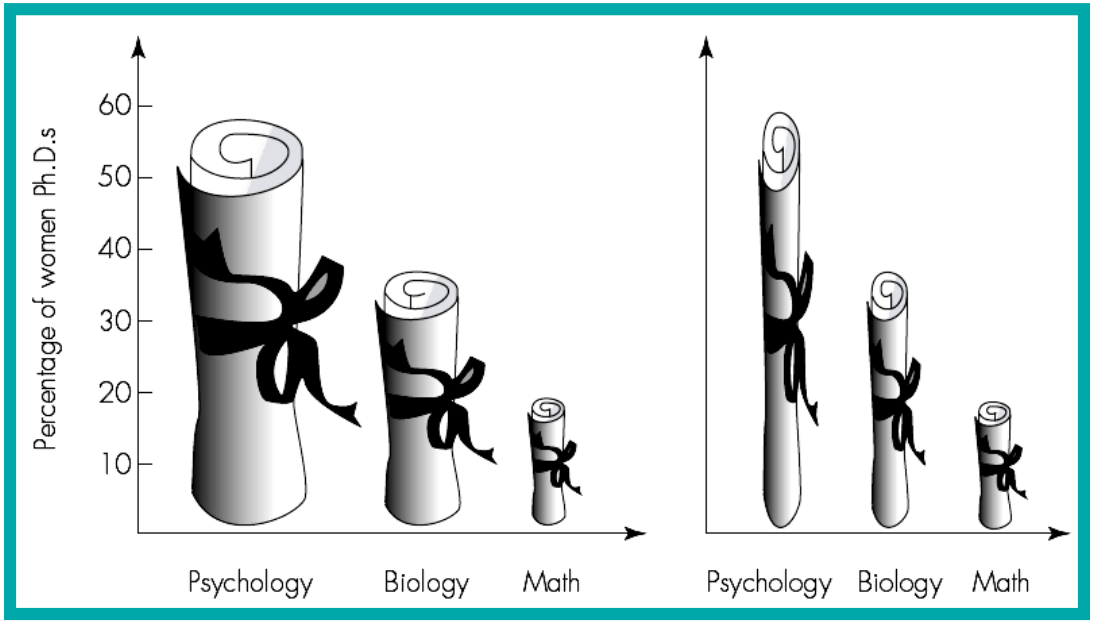
Source: U.S. Dept. of Labor, BLS, *Current Population survey.*

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Pictograms

Bar graph that uses pictures related to topic.

Percentage of Ph.D.s earned by women.



Left pictogram: Misleading because eye focuses on ***area*** rather than just height.

Right pictogram: Visually more accurate, but less appealing.

Source: Science (vol. 260, 16 April, 1993, p. 409).

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

9.3 Pictures of Measurement Variables

Single Variable Pictures:

- Stemplots
- Histograms

Displaying Relationships:

- Line Graphs
- Scatterplots

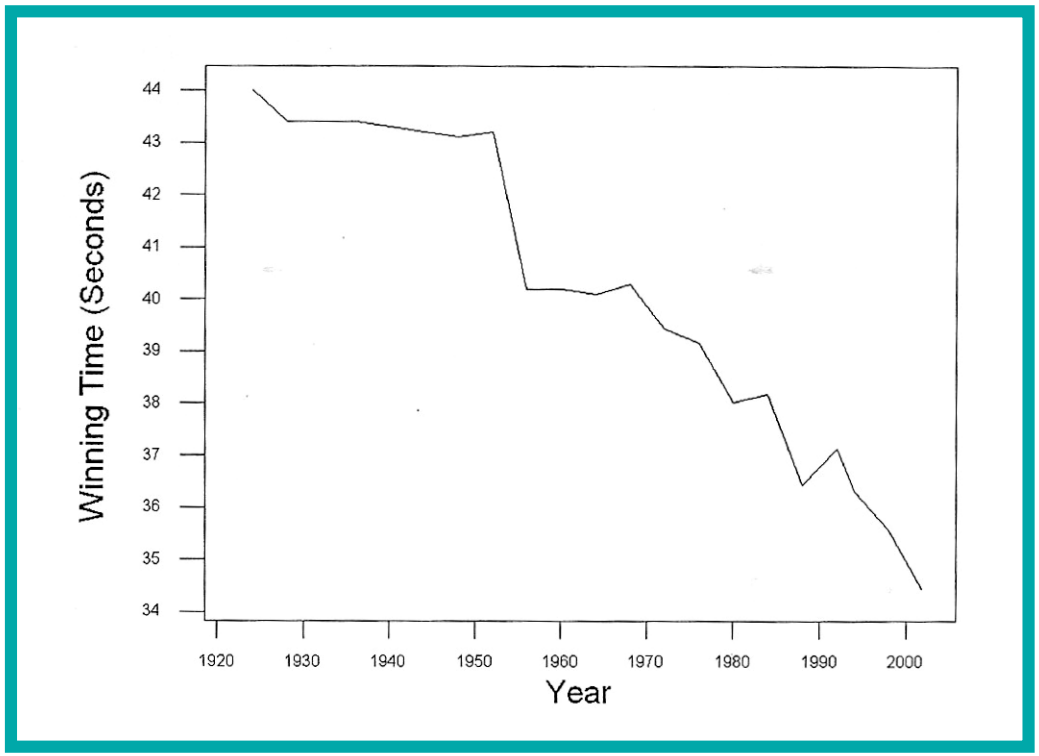
Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Line Graphs

Displays a variable over time.

Line graph of winning times for men's 500-meter speed skating in Winter Olympics 1924 to 2002

Overall downward trend with a larger drop in 1952-1956.



Source: http://sportsillustrated.cnn.com

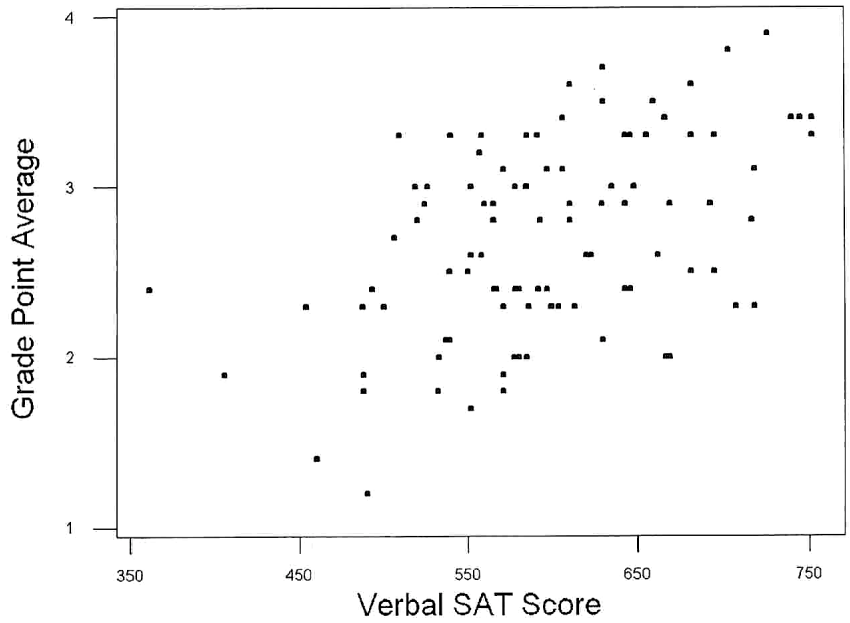
Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Scatterplots

Displays relationship between two measurement variables.

Scatterplot of GPA and verbal SAT score.

Overall increasing trend but still variability in GPAs at each level of verbal SAT scores.



Source: Ryan, Joiner, and Ryan, 1985, pp. 309-312.

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

9.4 Difficulties and Disasters in Plots, Graphs, and Pictures

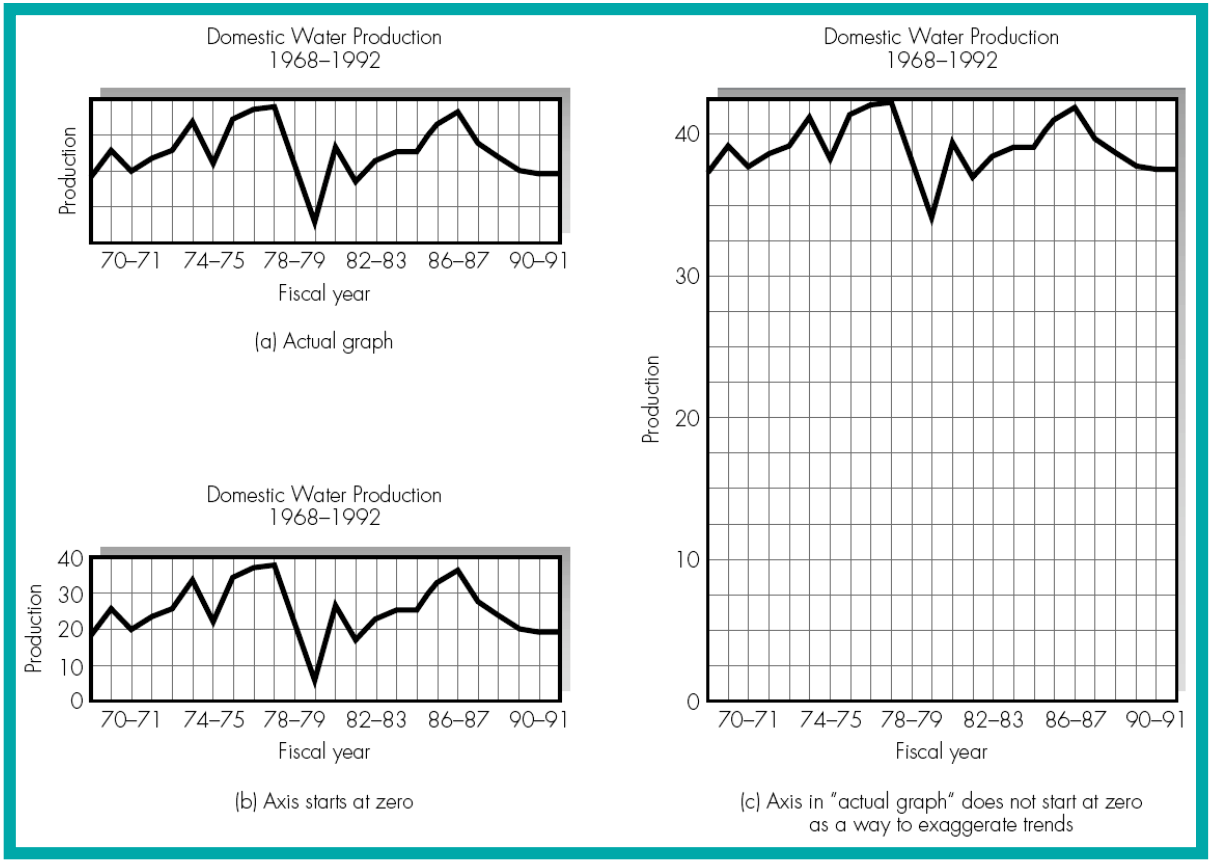
Most Common Problems:

1. **No labeling** on one or more axes
2. **Not starting at zero** as a way to exaggerate trends
3. **Change(s) in labeling** on one or more axes
4. **Misleading units** of measurement
5. **Using poor information**

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

No Labeling on One or More Axes

Example:
Graph with no labeling (a) and possible interpretations (b and c)



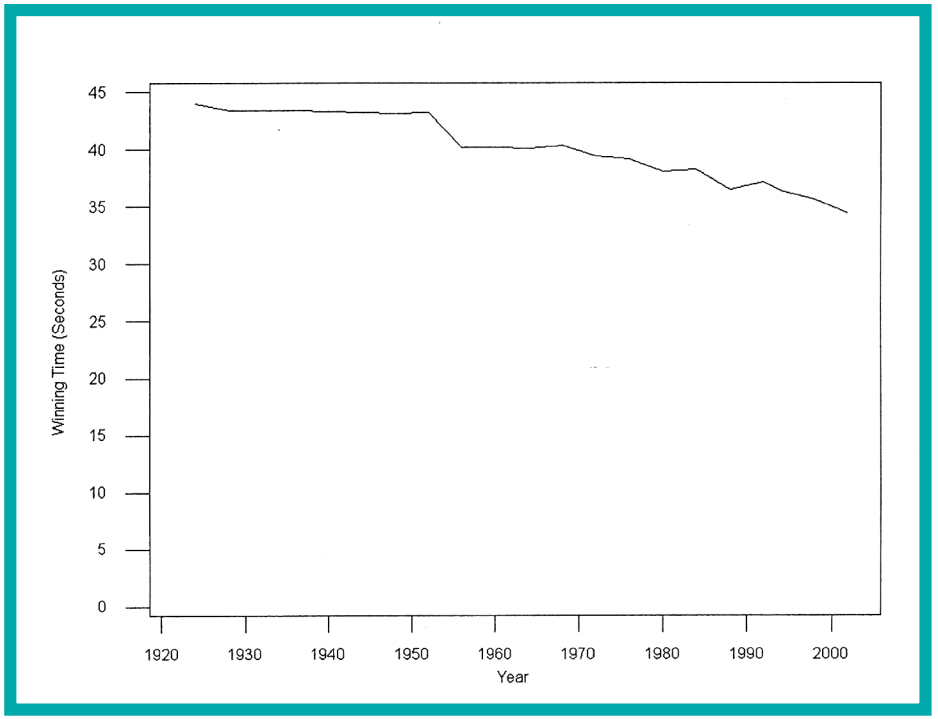
Source: Insert in the *California Aggie* (UC Davis), 30 May 1993.

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Not Starting at Zero

Example:
Winning times for Olympic speed skating data with vertical axis starting at 0.

Drop is ***not*** as dramatic.

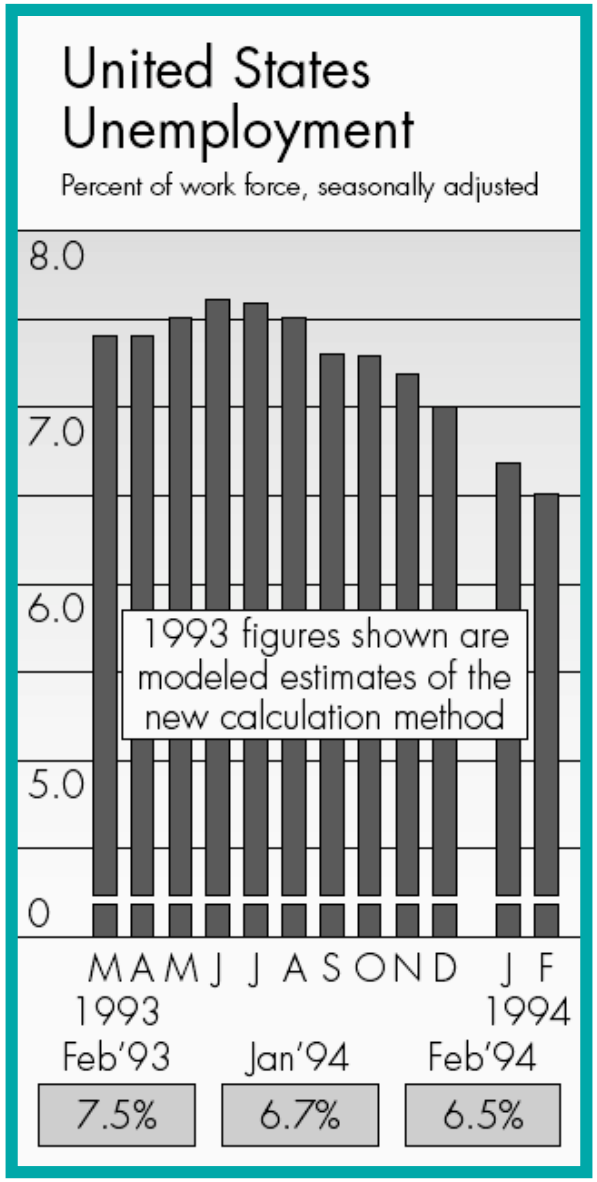


Note: For some variables, graphs should not start at zero.
e.g. SAT scores with range from 350 to 800.

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Changes in Labeling on One or More Axes

Example:
A bar graph with gap in labeling. At first look, seems vertical axis starts at 0, but bottom of the graph actually corresponds to 4.0%

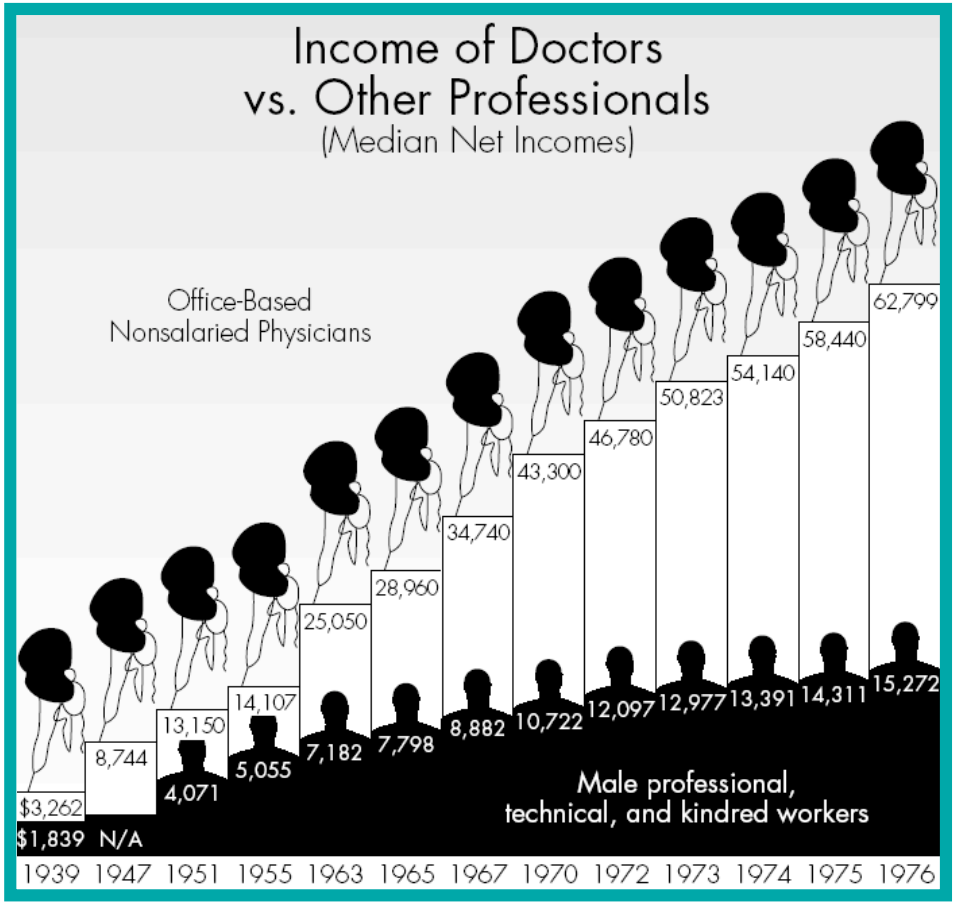


Source: *Davis* (CA) *Enterprise,* 4 March 1994, p. A-7.

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Changes in Labeling on One or More Axes

Example:
Distance between successive bars on horizontal axis keeps changing.



Source:
Washington Post graph reprinted in Wainer, 1984.

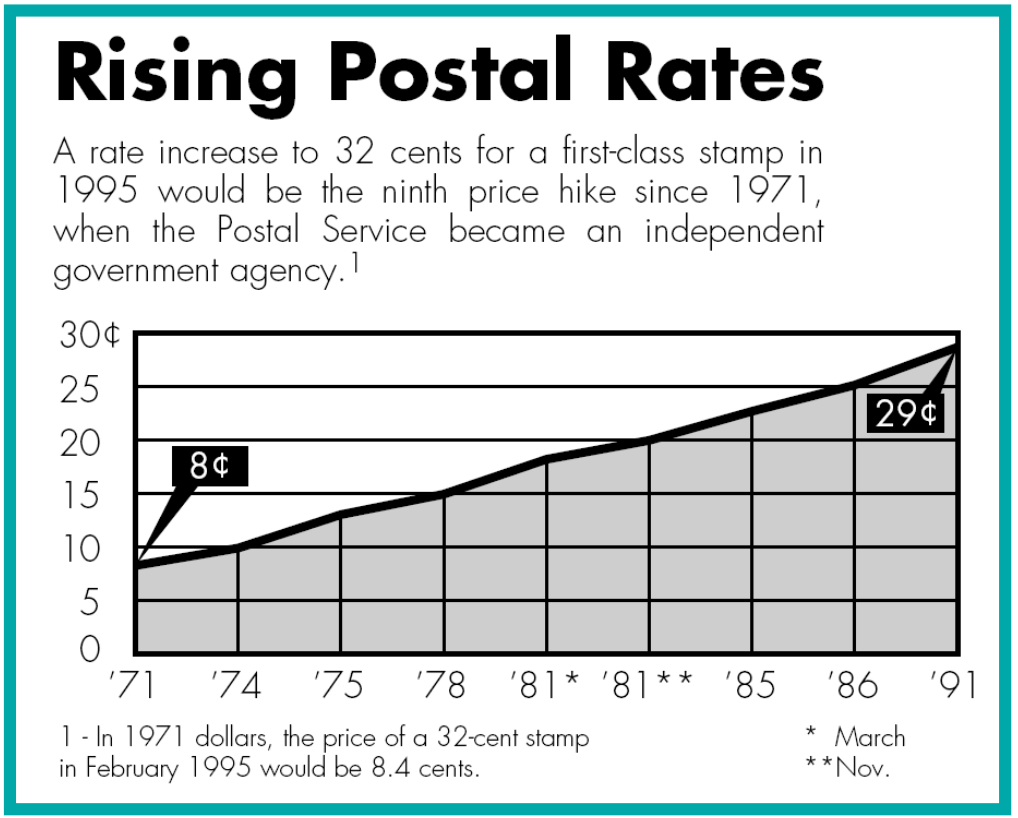
Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Misleading Units of Measurement

Units can be different from those important to the reader.

Fine print: "In 1971 dollars, the price of a 32-cent stamp in February 1995 would be 8.4 cents."

More truthful picture: show changing price of a first-class stamp ***adjusted for inflation***.



Source: USA Today, 7 March 1994, p. 13A.

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

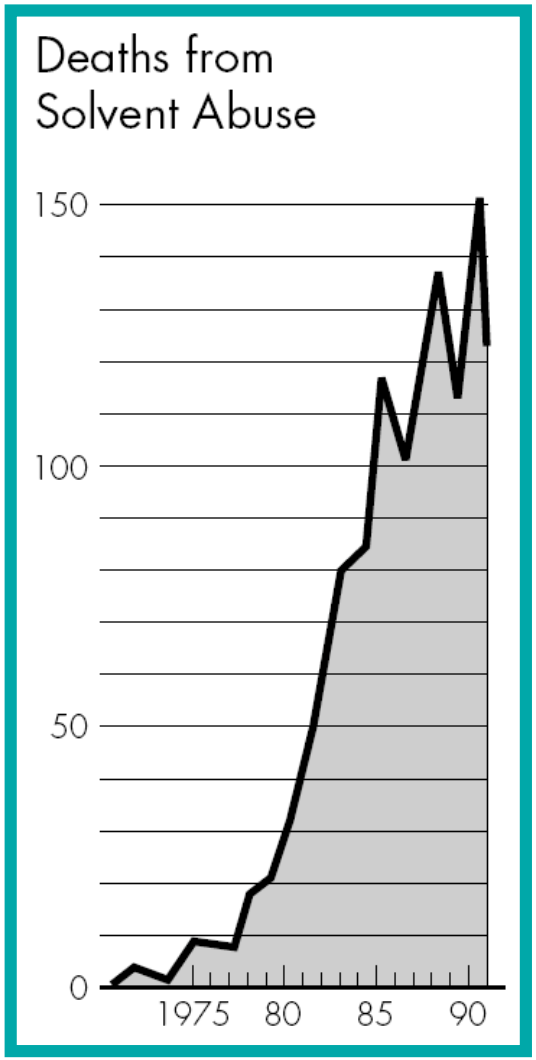
Using Poor Information

Picture only as accurate as the information used to design it.

Graph appears to show **very few deaths** from solvent abuse **before late 1970's**.

Article quote: "It's only since we have started **collecting accurate data since 1982** that we have begun to discover the real scale of the problem" (p.5).

Source: The Independent on Sunday (London), 13 March 1994.



Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

9.5 A Checklist for Statistical Pictures

Ten questions to ask before interpreting:

1. Does the **message** of interest stand out clearly?
2. Is the **purpose or title** of the picture evident?
3. Is a **source** given for the data, either with the picture or in an accompanying article?
4. Did the information in the picture come from a **reliable, believable source**?
5. Is everything **clearly labeled**, leaving no ambiguity?

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

9.5 A Checklist for Statistical Pictures

Ten questions to ask before interpreting:

6. Do the axes **start at zero** or not?
7. Do the axes maintain a **constant scale**?
8. Are there **any breaks** in the numbers on the axes that may be easy to miss?
9. For financial data, have the numbers been **adjusted for inflation**?
10. Is there **information cluttering** the picture or **misleading** the eye?

Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.

Case Study 9.1: Time to Panic about Illicit Drug Use?

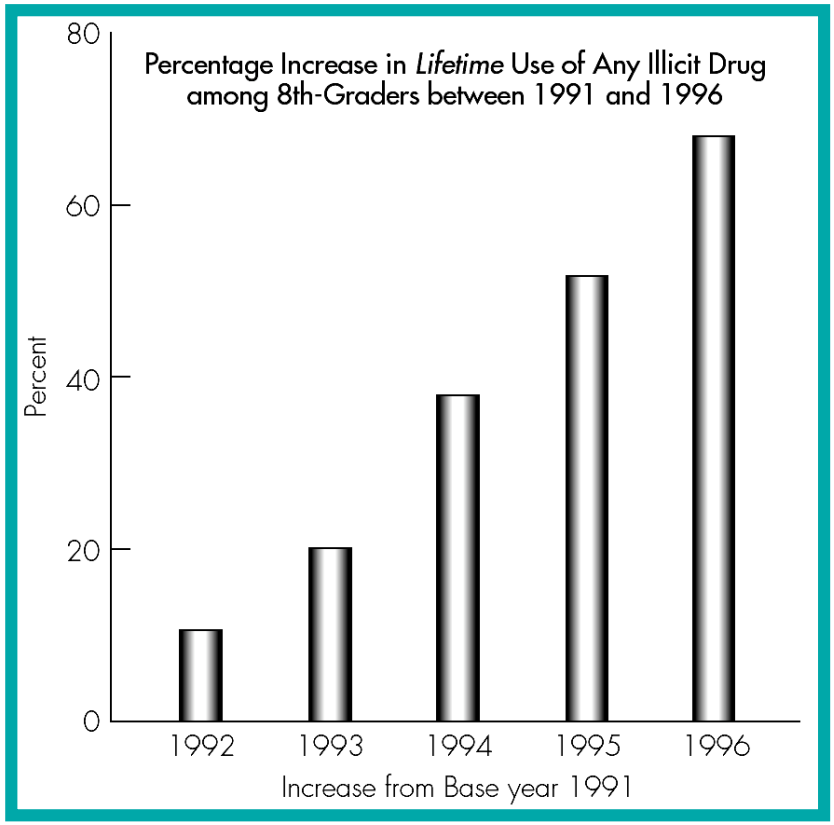
Headline: "*Emergency Situation among Our Youth.*"

First look: seems 80% used drugs in 1996 vs. 10% in 1992.

Careful reading: In 1996, the rate of use was 80% higher, or 1.8 times what it was in 1991. Actual rate of use *not* provided.

1991 rate was 11%, so in 1996 it was 1.8(11) = 19.8%

Source: U.S. Department of Justice



Copyright ©2005 Brooks/Cole, a division of Thomson Learning, Inc.