Lecture 7

Graphs

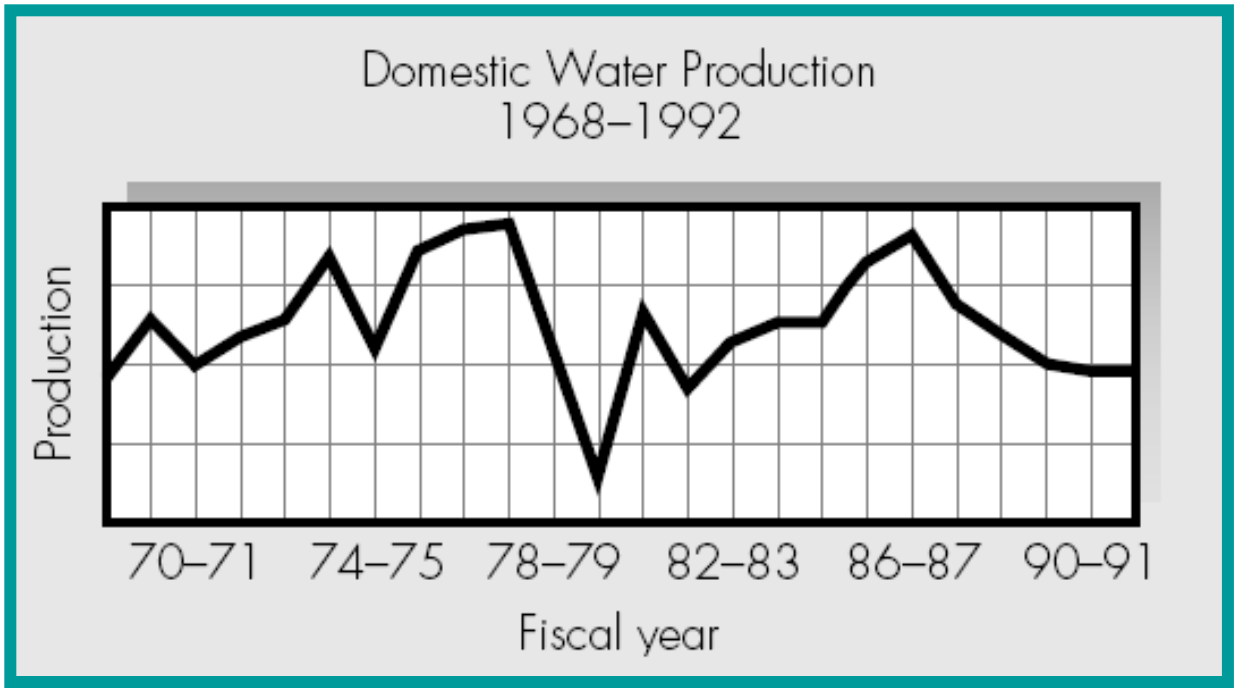
Thought Question 1:

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.

Thought Question 2:

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



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?

Thought Question 4:

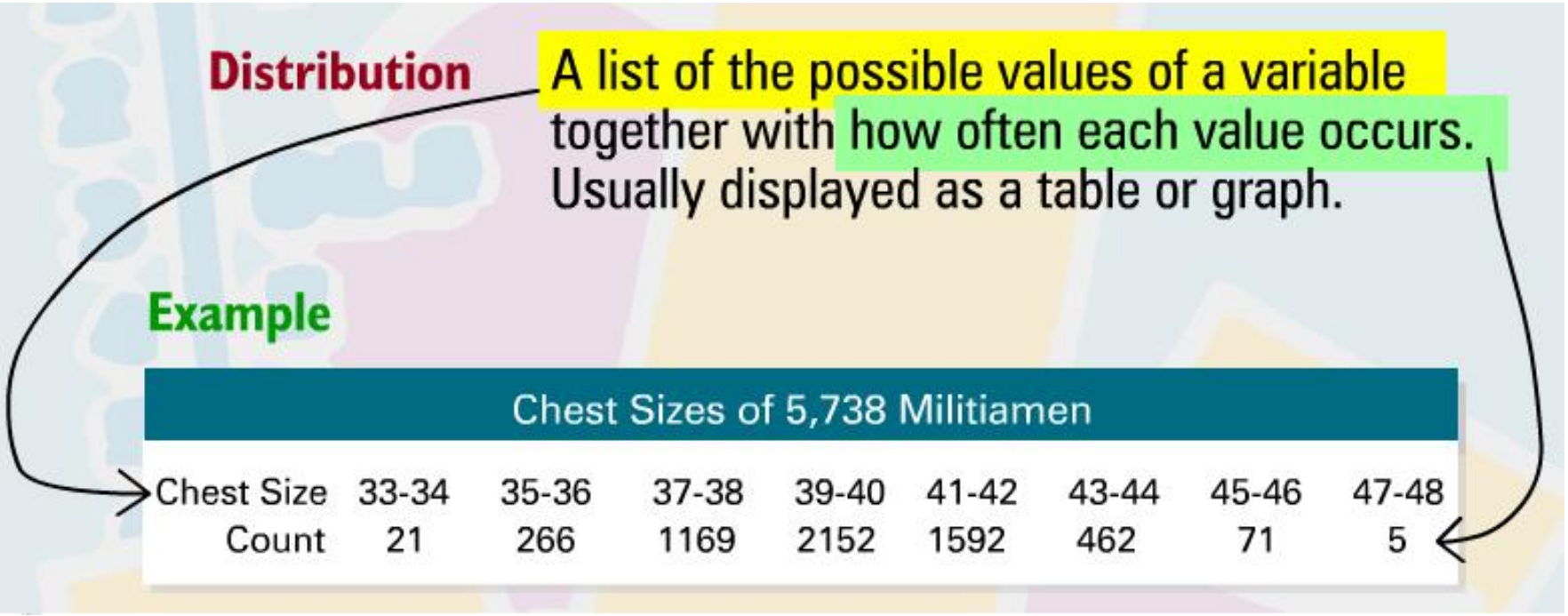
Suppose you own a company that produces candy bars and you want to display two graphs. One graph is for customers and shows the price of a candy bar for each of the past 10 years. The other graph is for stockholders and shows the amount the company was worth for each of the past 10 years. You decide 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.

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.

Distribution



- To graphically display the distribution of a categorical variable, use a pie chart or a bar graph. Summaries are counts or percentages.

How do we display categorical data? In a table

5. Did either of your natural parents die of a heart attack before age 60? (If your parents are younger than 60, mark no.)
   a. yes, one of them
   b. yes, both of them
   c. no
   d. not sure

Category	Count	Percentage (%)
a. yes, one of them	34	10.33
b. yes, both of them	16	4.86
c. no	268	81.46
d. not sure	11	3.34
Total	329	99.99

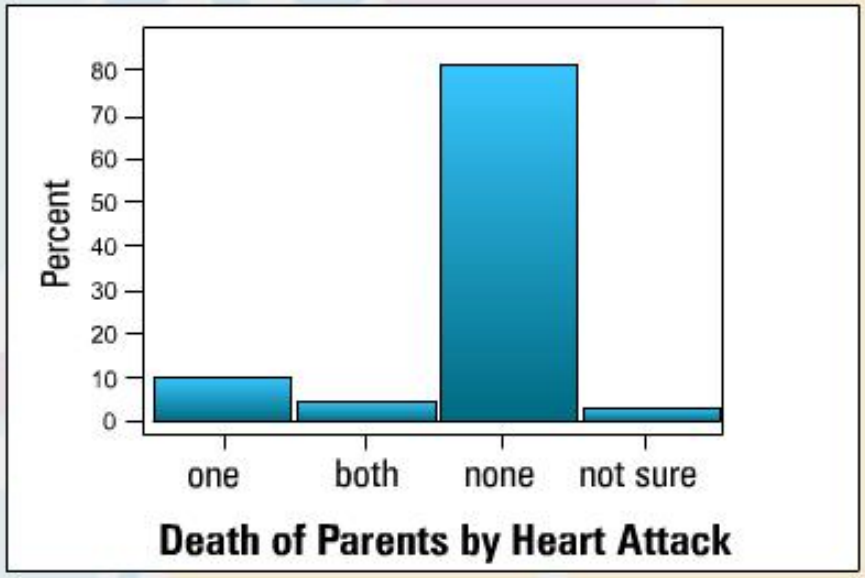
Note: Percentages do not sum to 100% due to round-off error.

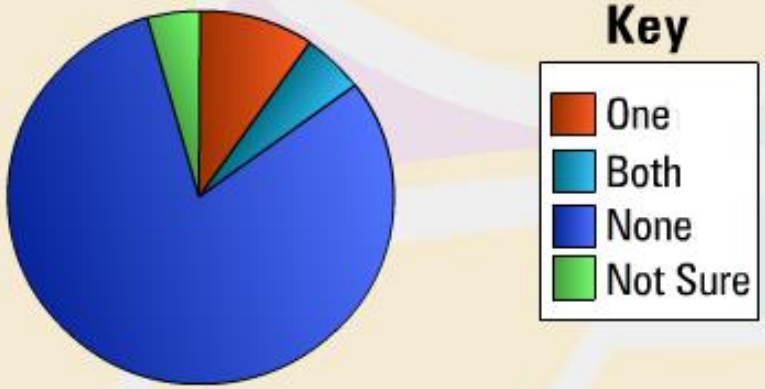
Displaying Categorical Data in a Graph

Which is better--Bar Graphs or Pie Charts?

5. Did either of your natural parents die of a heart attack before age 60? (If your parents are younger than 60, mark no.)

a. yes, one of them	**10.33%**
b. yes, both of them	**4.86%**
c. no	**81.46%**
d. not sure	**3.34%**





9.2 Pictures of Categorical Data

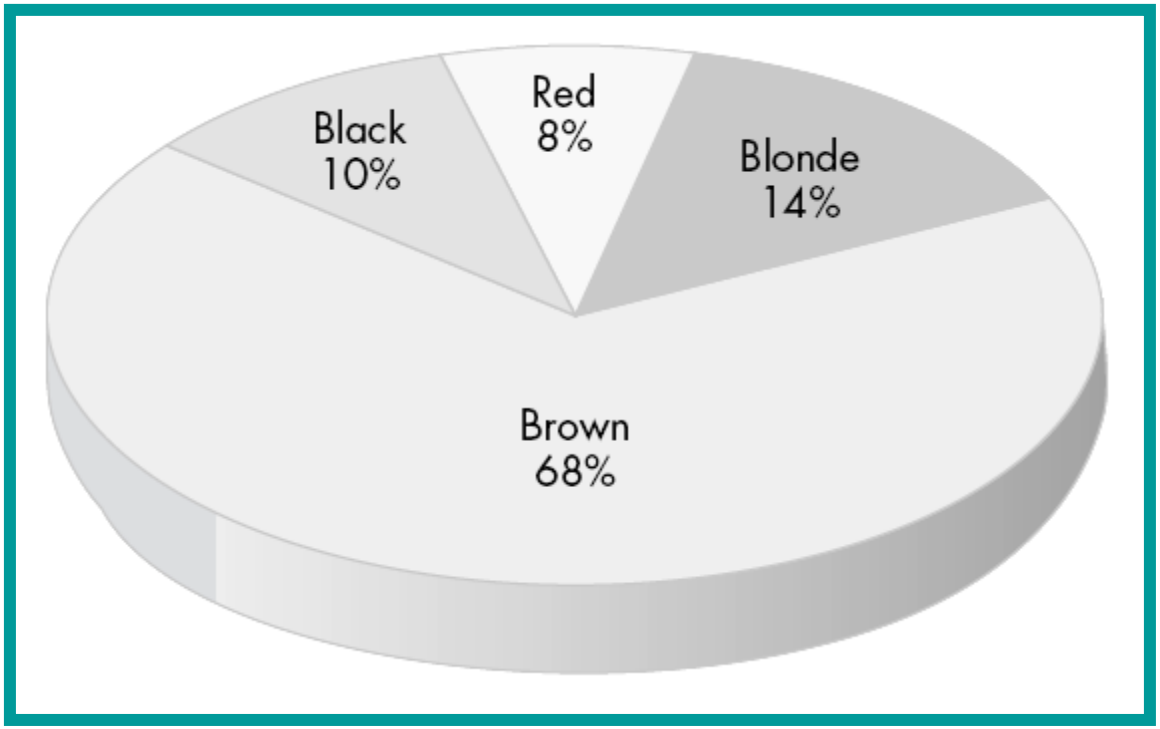
Three common pictures:

- Pie Charts
- Bar Graphs
- Pictograms

Pie Charts

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

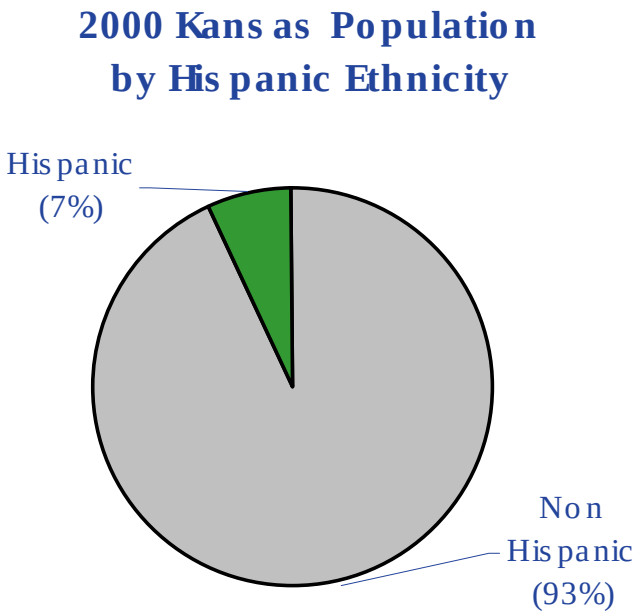
Pie chart of hair colors of white American children.



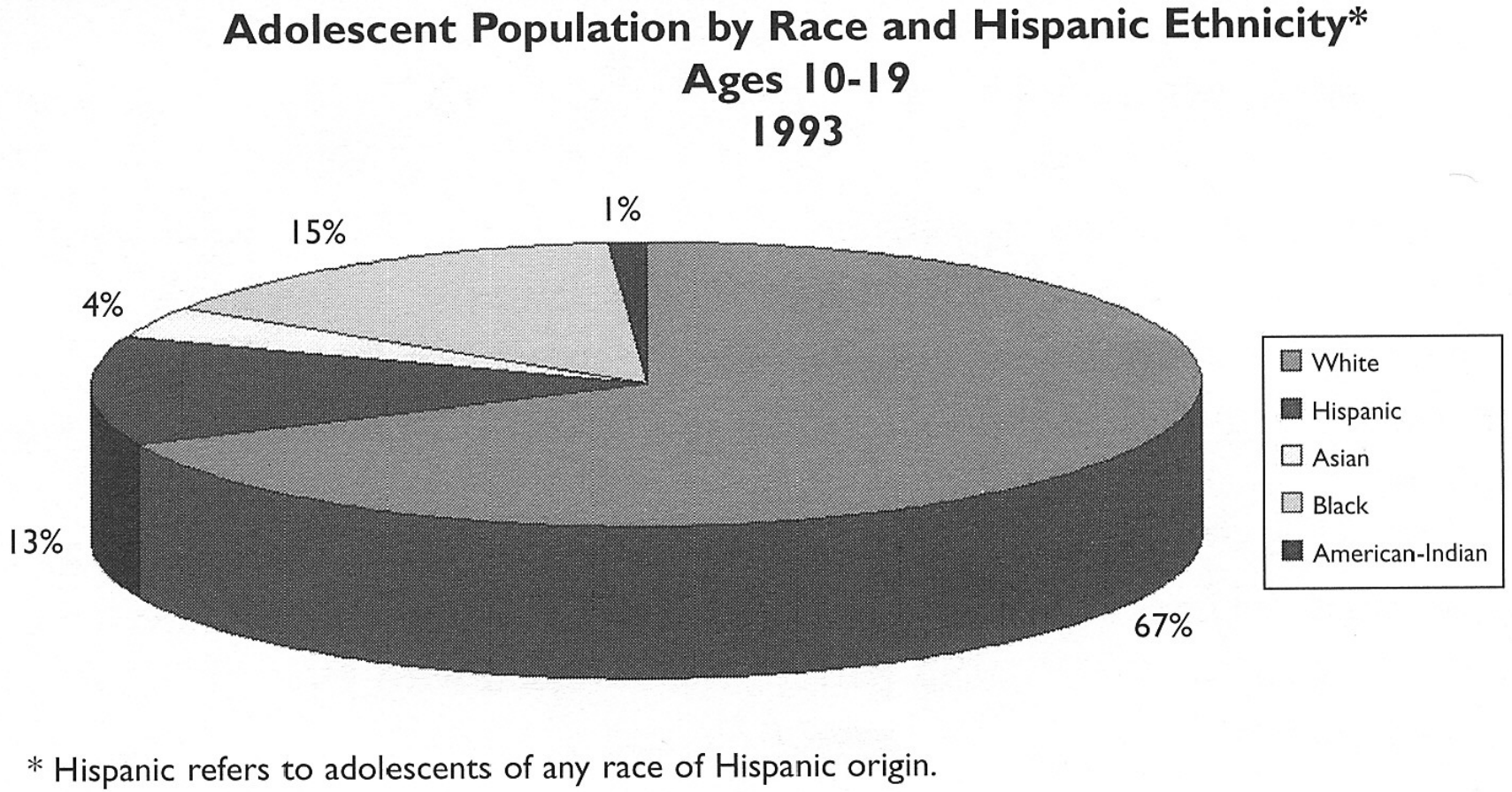
Source: Krantz, 1992, p. 188.

Pie Graph Guidelines

- Make sure data totals 100%, and there is NO duplicate data (e.g. race/ethnicity)
- Give percentages on the figure



Make sure data totals 100%

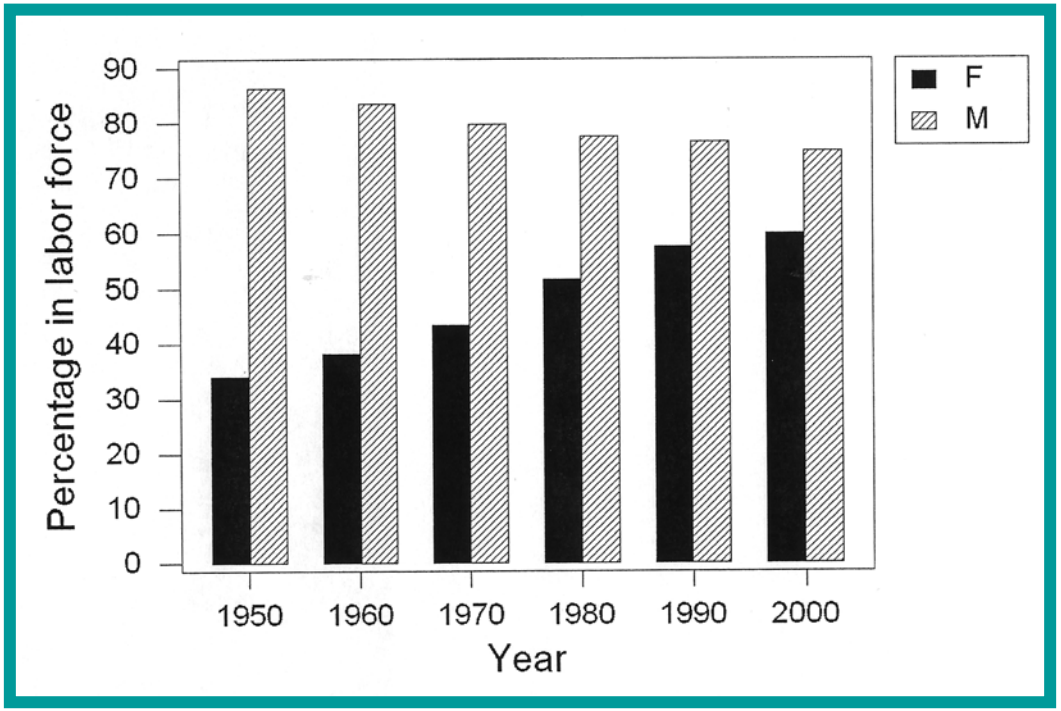


America's Adolescents: Are They Healthy? 1998

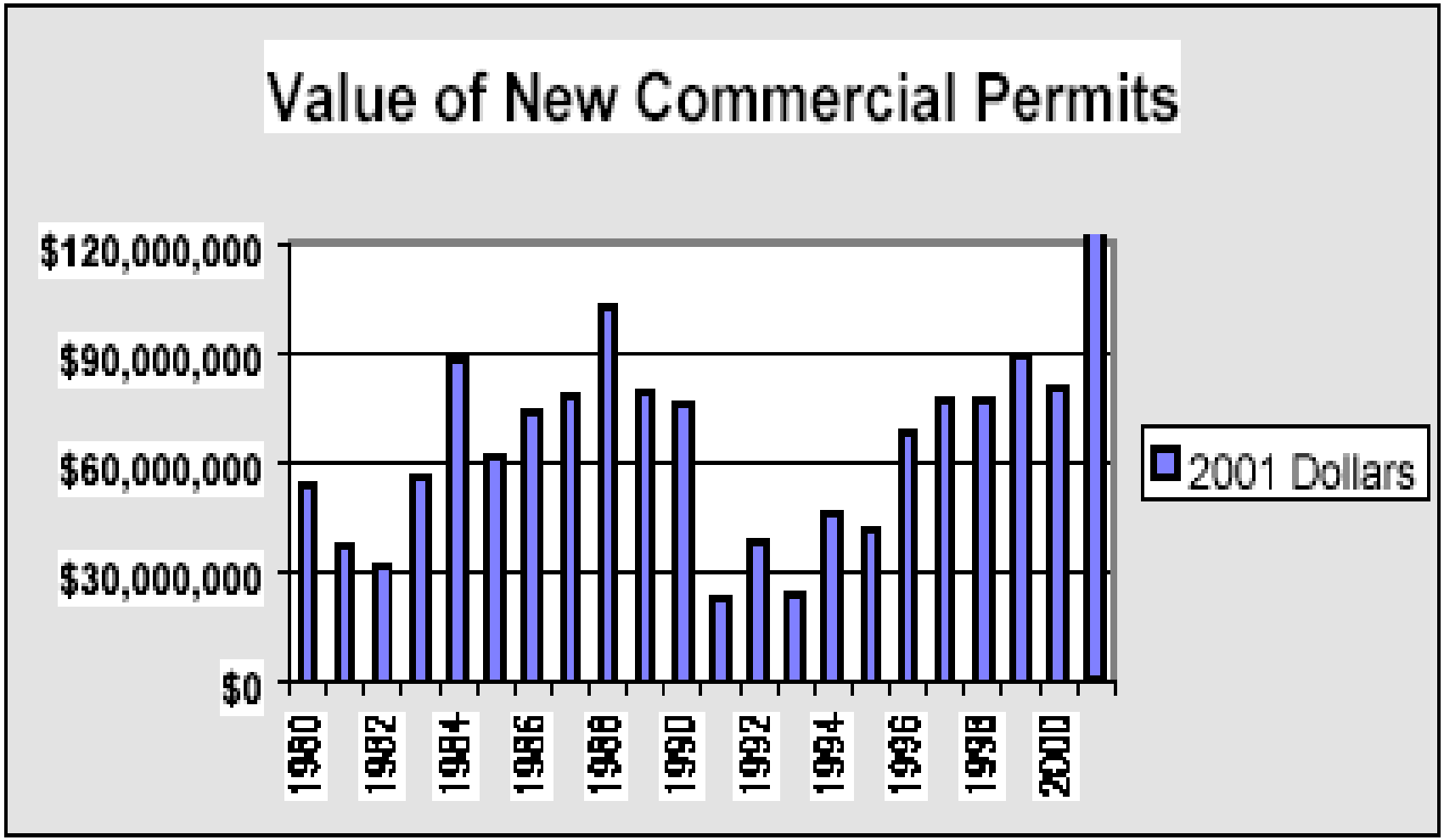
Bar Graphs

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

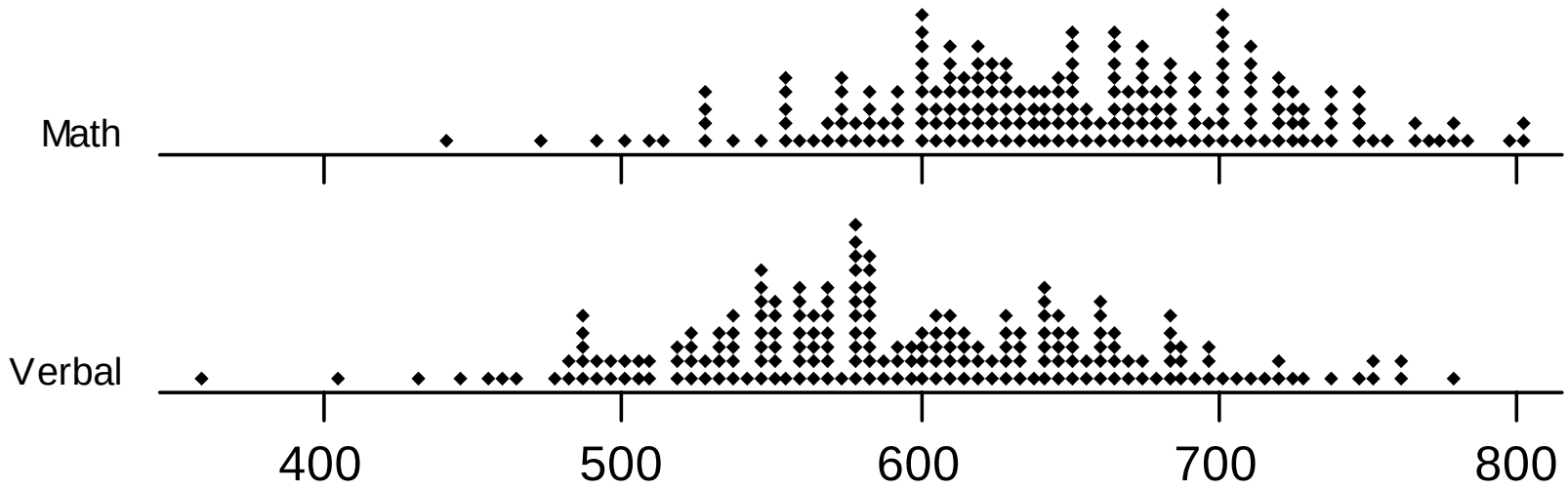
Percentage of men and women 16 and over in the labor force



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



Taken from the *Tallahassee Statistical Digest, 2001*


Math
Verbal
400
500
600
700
800

Bar Graph Guidelines

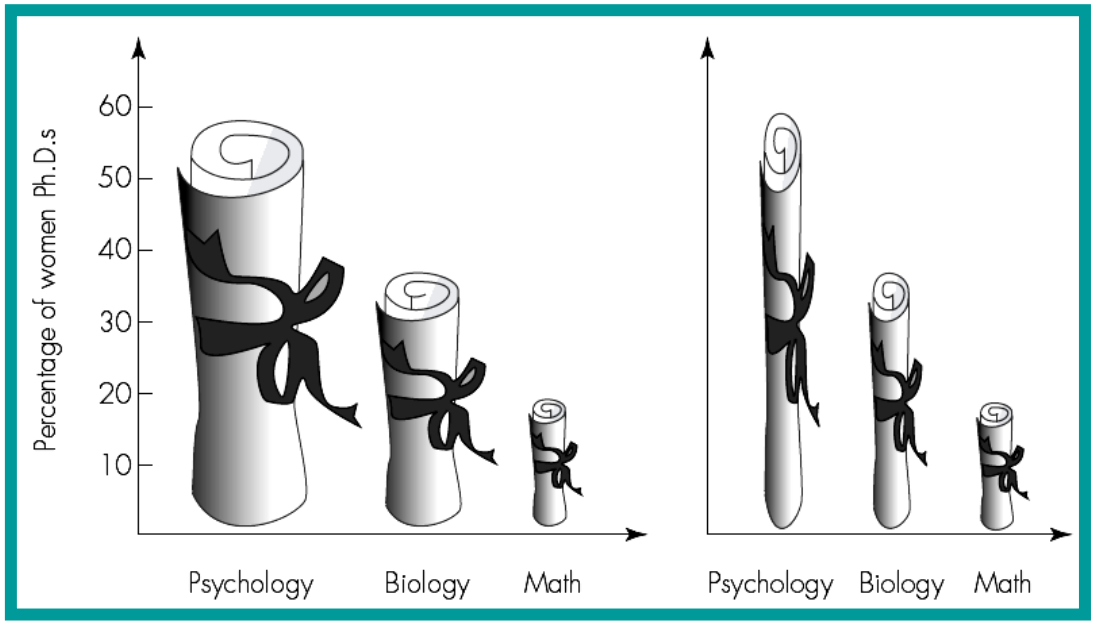
- Use for categorical data
- Vertical bars are better for visually comparing data, but using horizontal bars may allow for easier labeling

- Stacked bar graphs
  - Pie charts are better for part-to-whole or part-to-part comparisons.
  - Stacked bars allow for part-to-part comparisons across multiple categories
  - If used, make lower areas darkest and top areas lightest.

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.

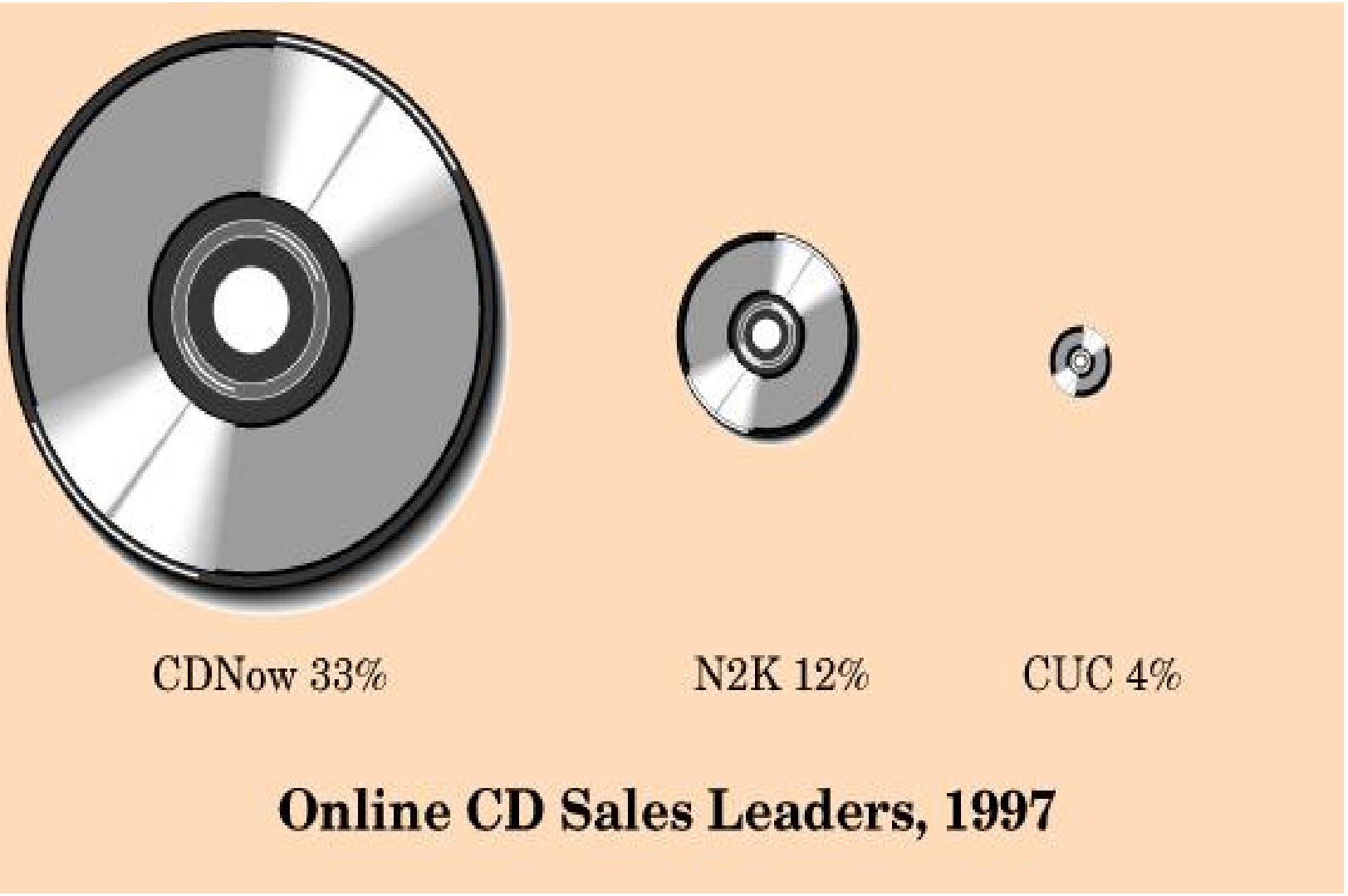
Source: Science (vol. 260, 16 April, 1993, p. 409).

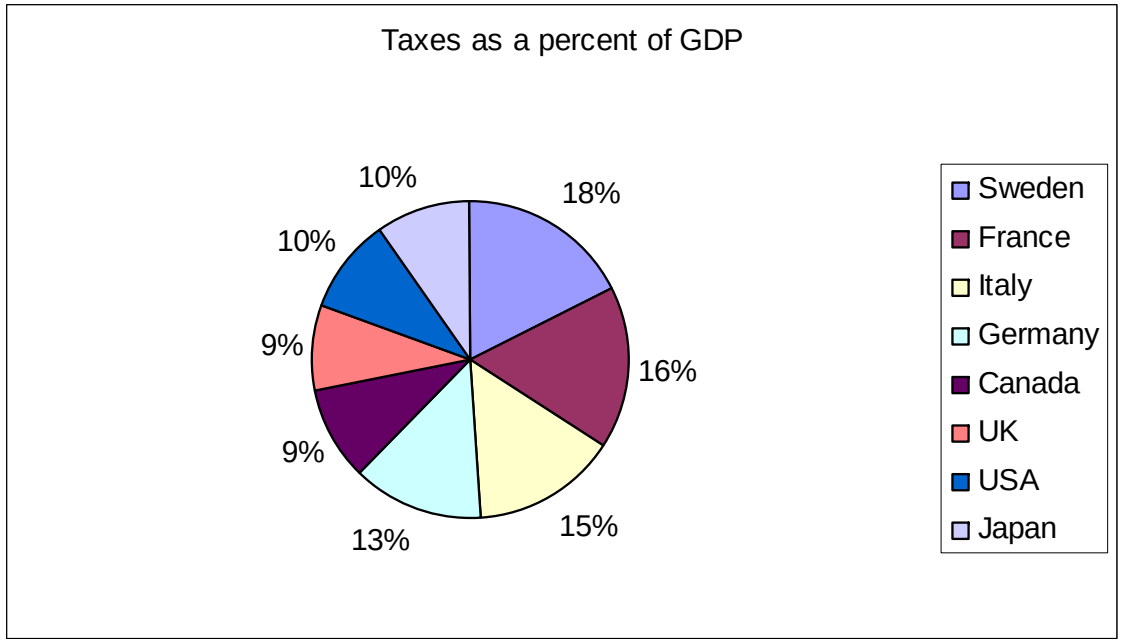
Pictograms

- Pictogram: is like a bar graph except that it uses pictures related to the topic of the graph

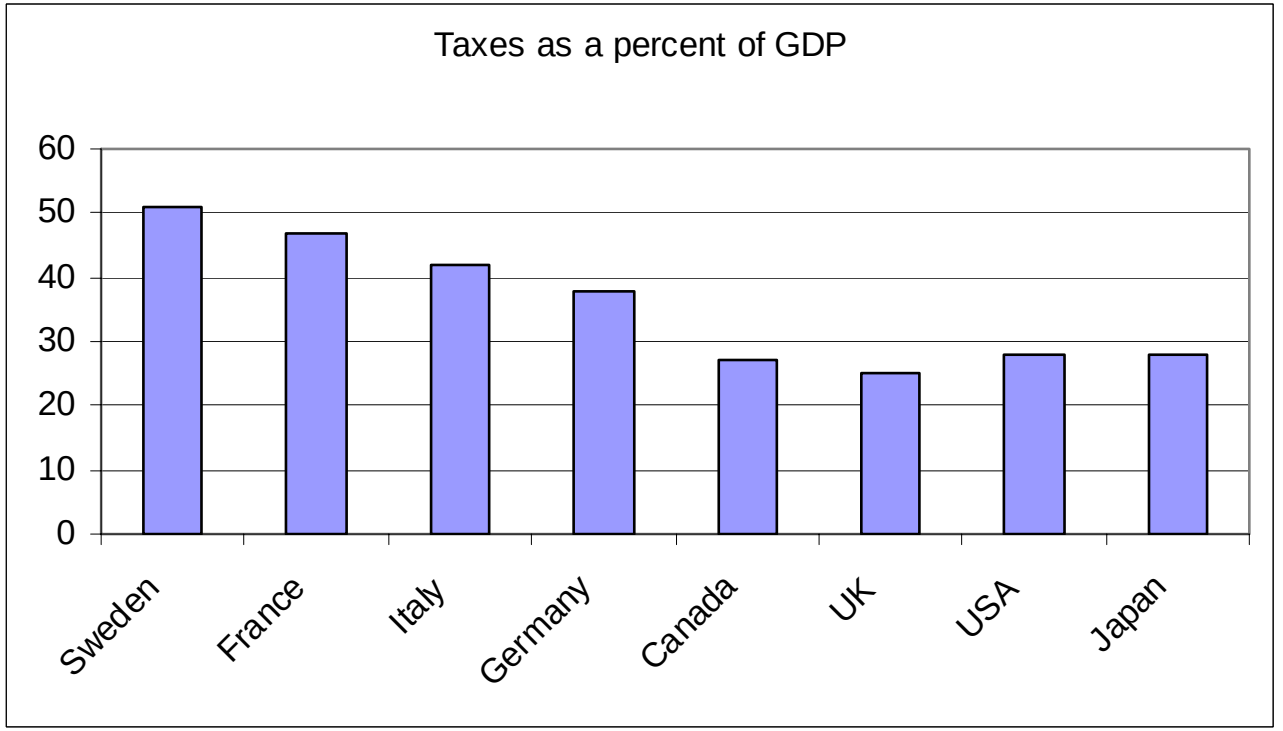
- It is easy to be misled by pictograms. One needs to be careful that the area of the pictures do not distort the true values.

Pictograms

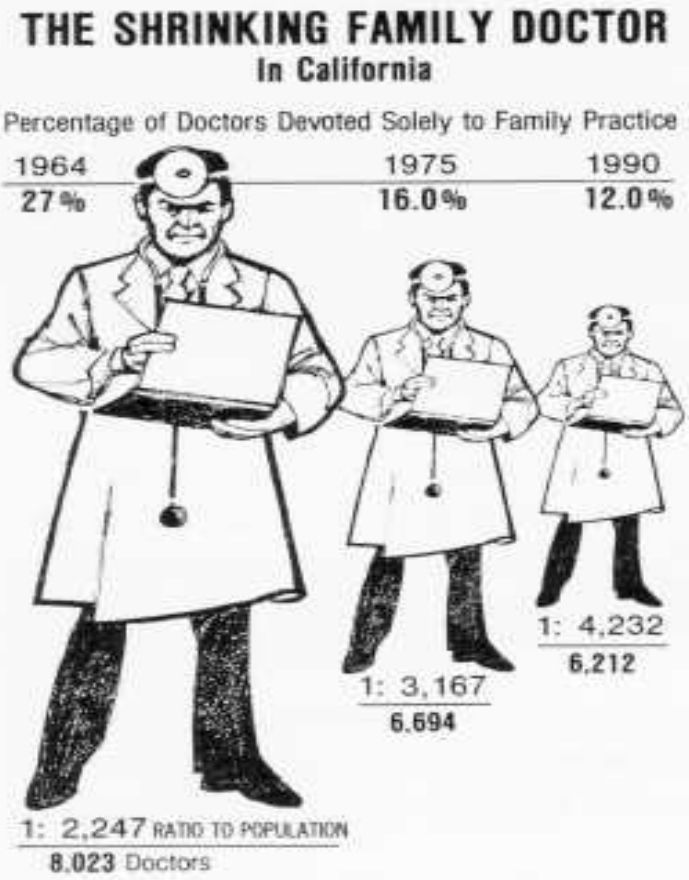




Which graph is more appropriate? Which represents the data on page 185 better?



How to Exaggerate with Graphs
from Tufte ’83



“Lie factor” = 2.8

Los Angeles Times, August 5, 1979, p. 3.

How to Exaggerate with Graphs
from Tufte ’83

Washington Post, October 25, 1978, p. 1.

9.3 Pictures of Measurement Variables

Single Variable Pictures:

- Stemplots
- Histograms

Displaying Relationships:

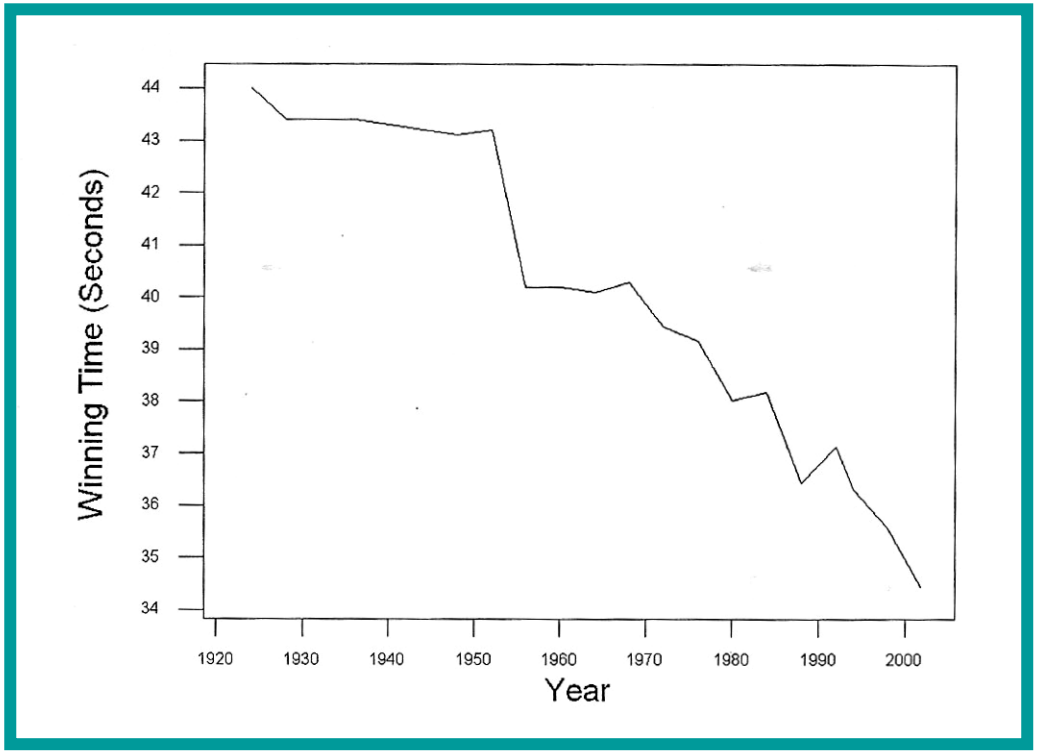
- Line Graphs
- Scatterplots

Line Graphs

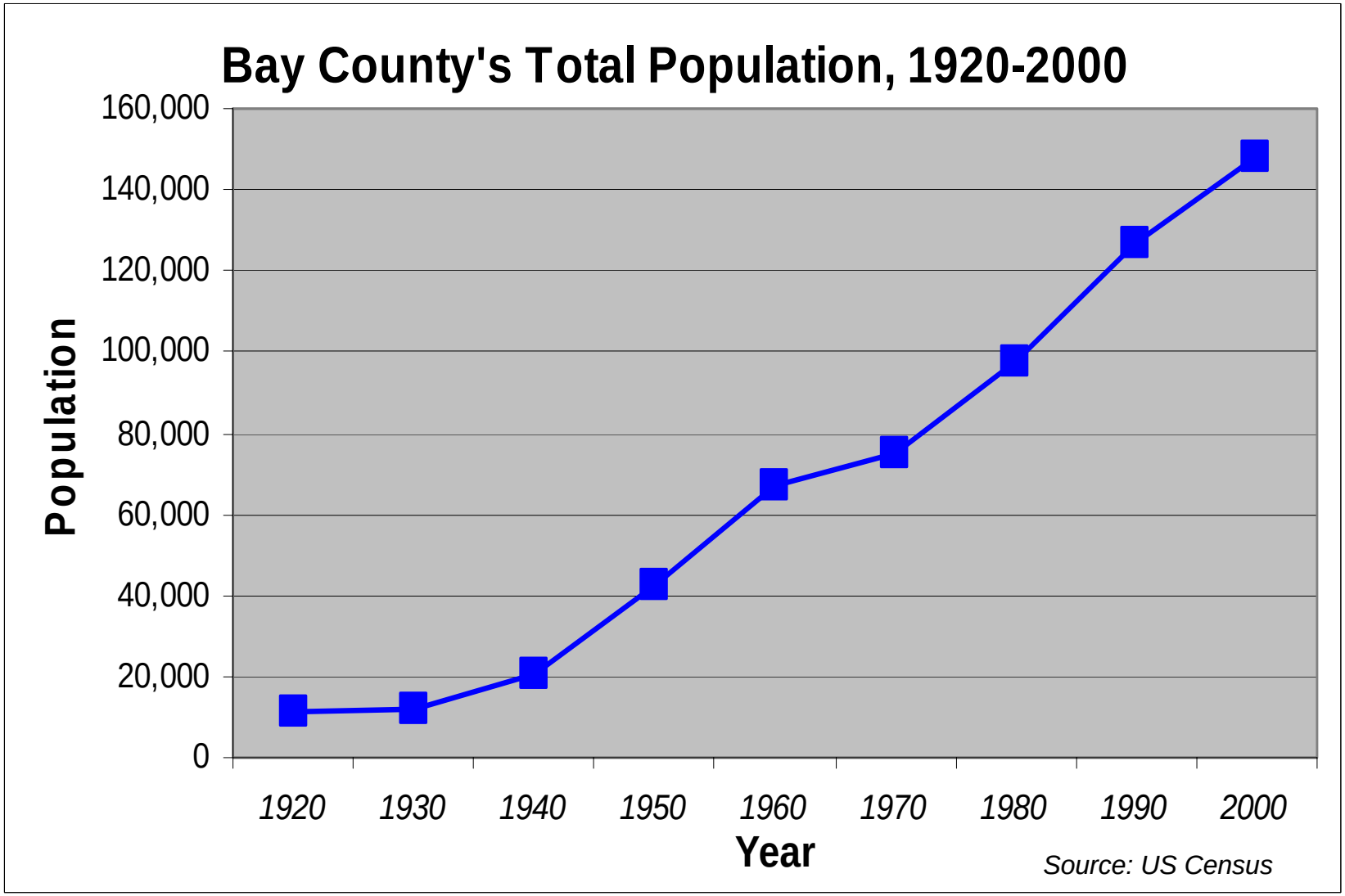
Displays a variable over time.

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

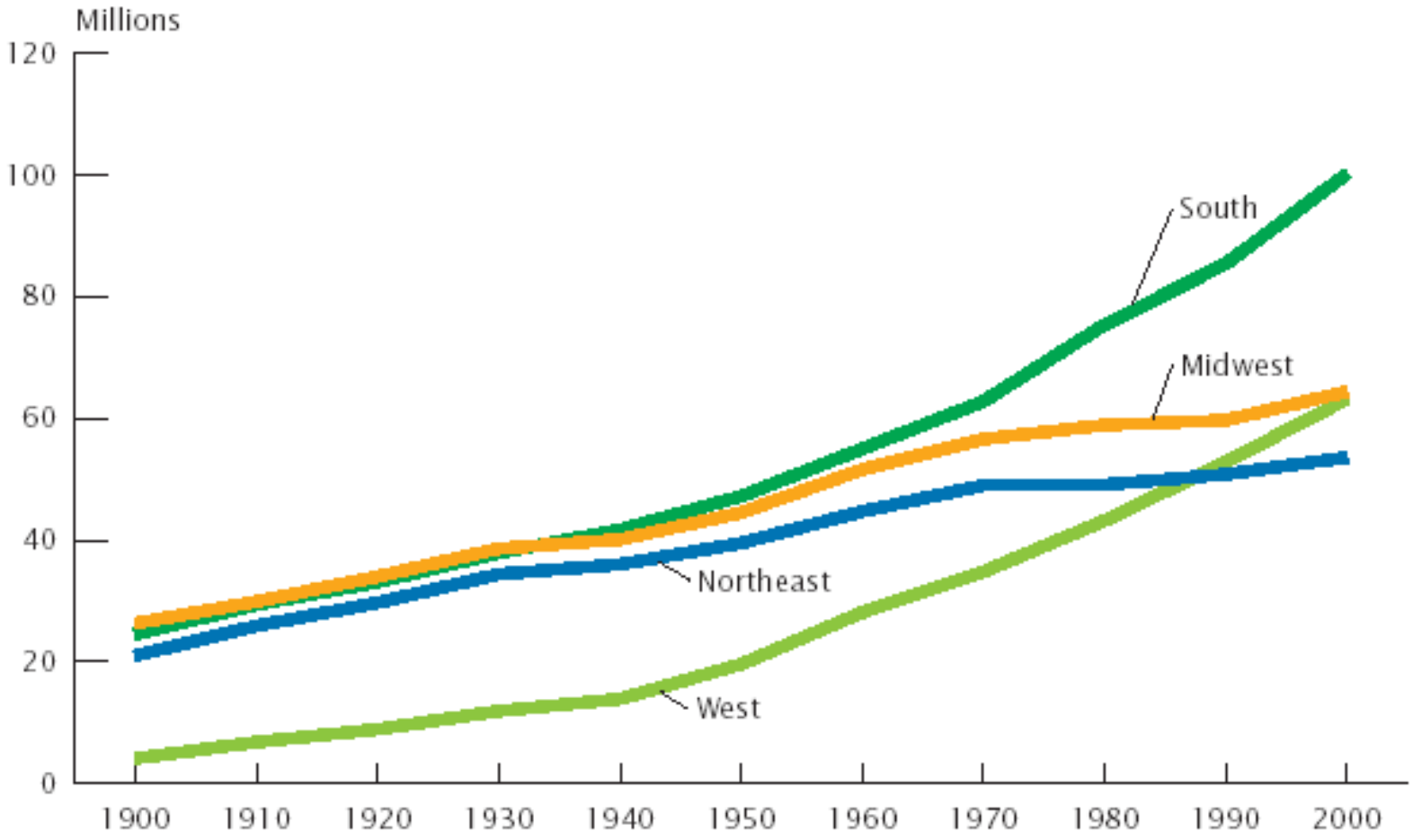
Overall downward trend with a larger drop in 1952-1956.



Source: http://sportsillustrated.cnn.com


Bay County's Total Population, 1920-2000
Population
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
1920
1930
1940
1950
1960
1970
1980
1990
2000
Year
Source: US Census

Total Population by Region: 1900 to 2000



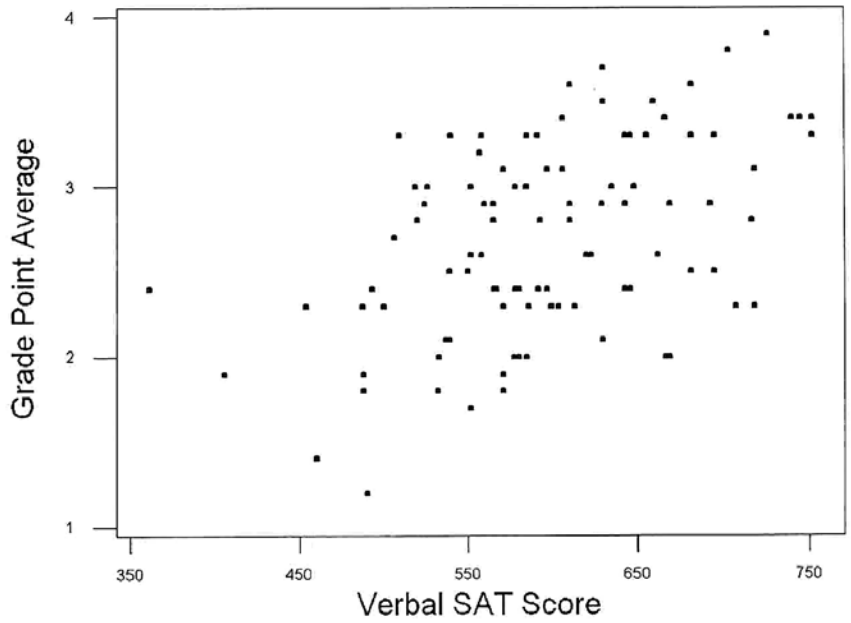
Source: U.S. Census Bureau, decennial census of population, 1900 to 2000.

Scatterplots

Displays relationship between two measurement variables.

Scatterplot of GPA and verbal SAT score.

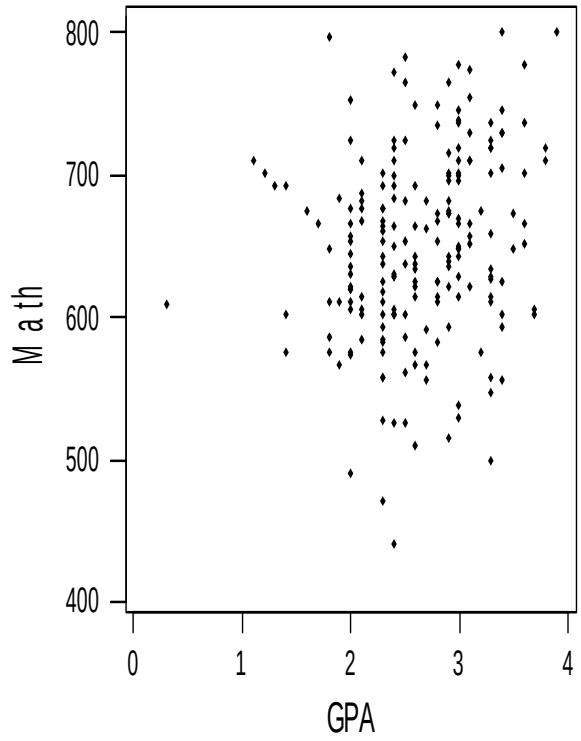
Overall increasing trend but still variability in GPAs at each level of verbal SAT scores.



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

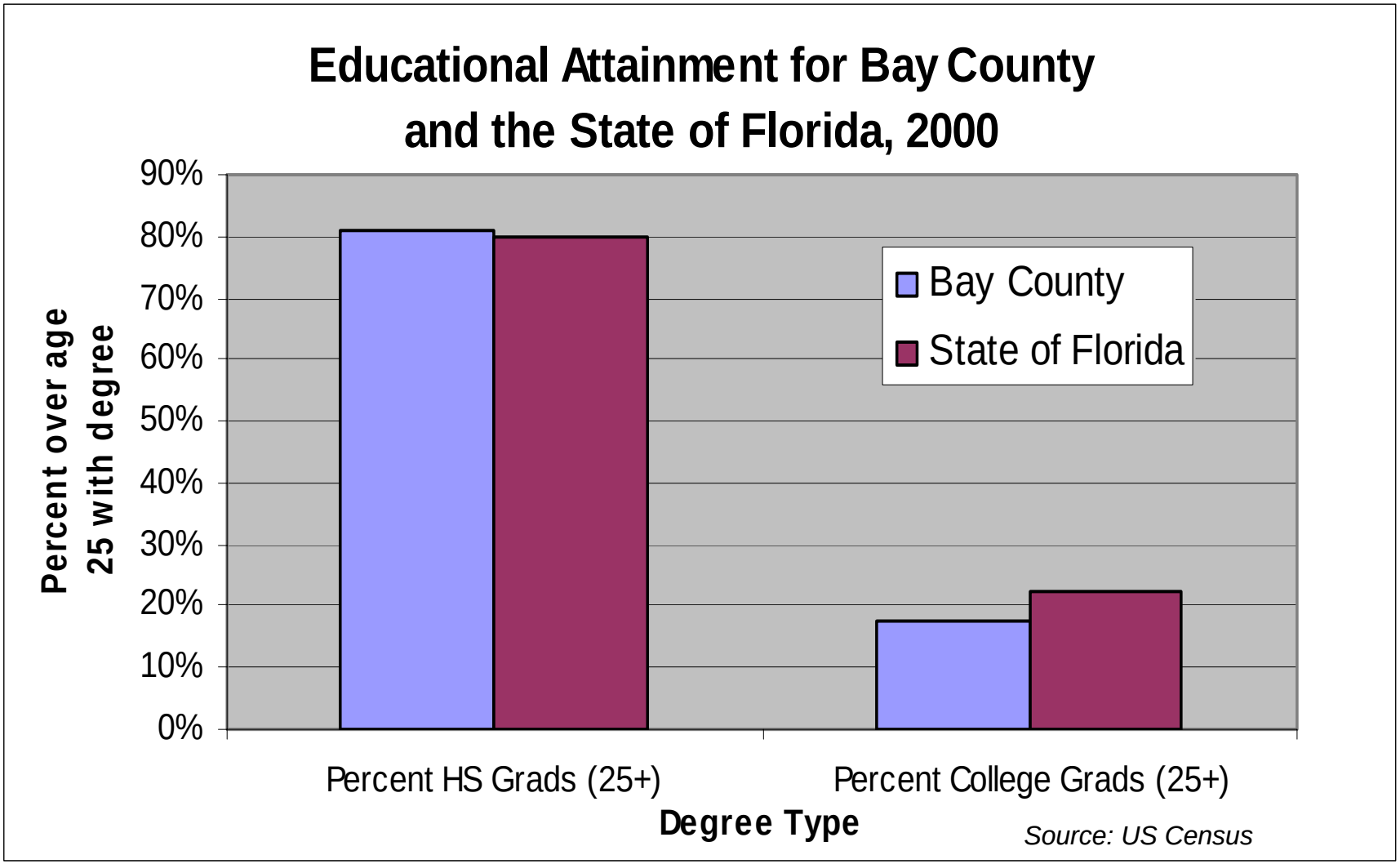
SCATTER DIAGRAMS

- Each point determined by two values
- Should provide a quick visual impression of the data
- Transformations sometimes performed for skewed data

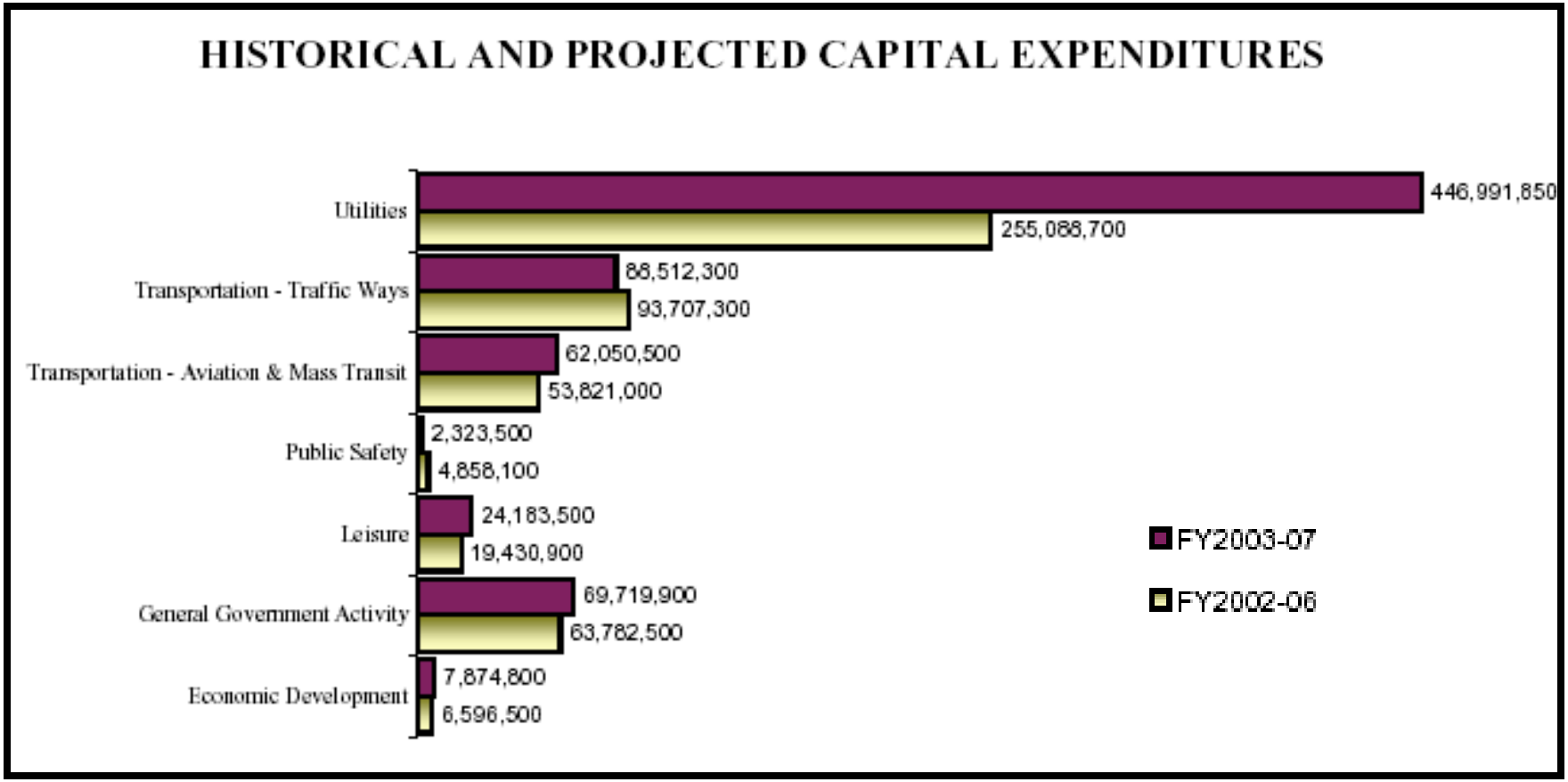


Complex Graphs: Combination Charts

- There are times when you want to present two or more variables in the same chart.
- ***Combination charts*** allow the user to combine two different variables in the same chart in order to illustrate the relationships between these variables.
- The choice of presentation method is dictated by the variables and the nature of the relationship between the variables.

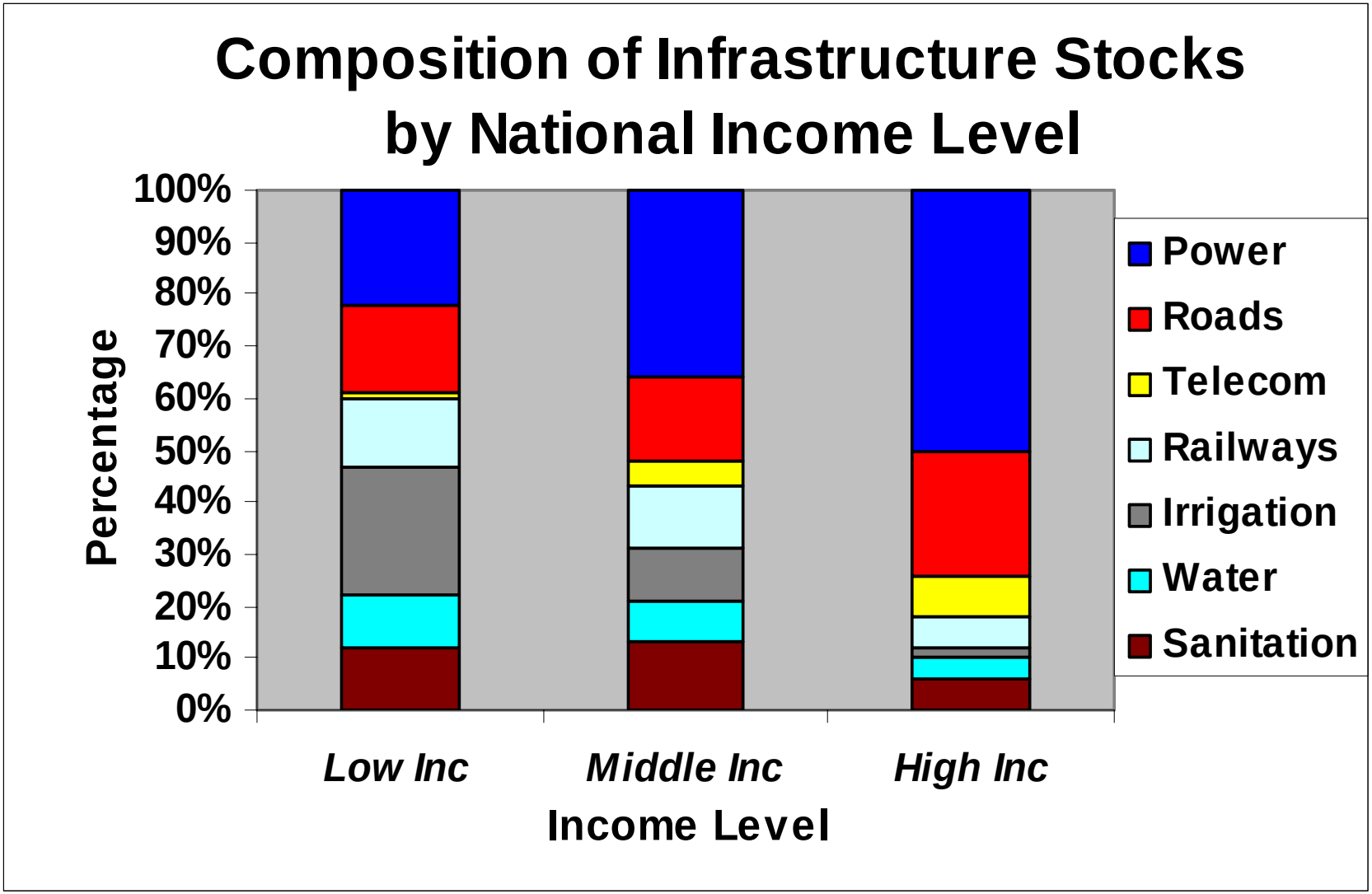


Clustered Column Chart Example

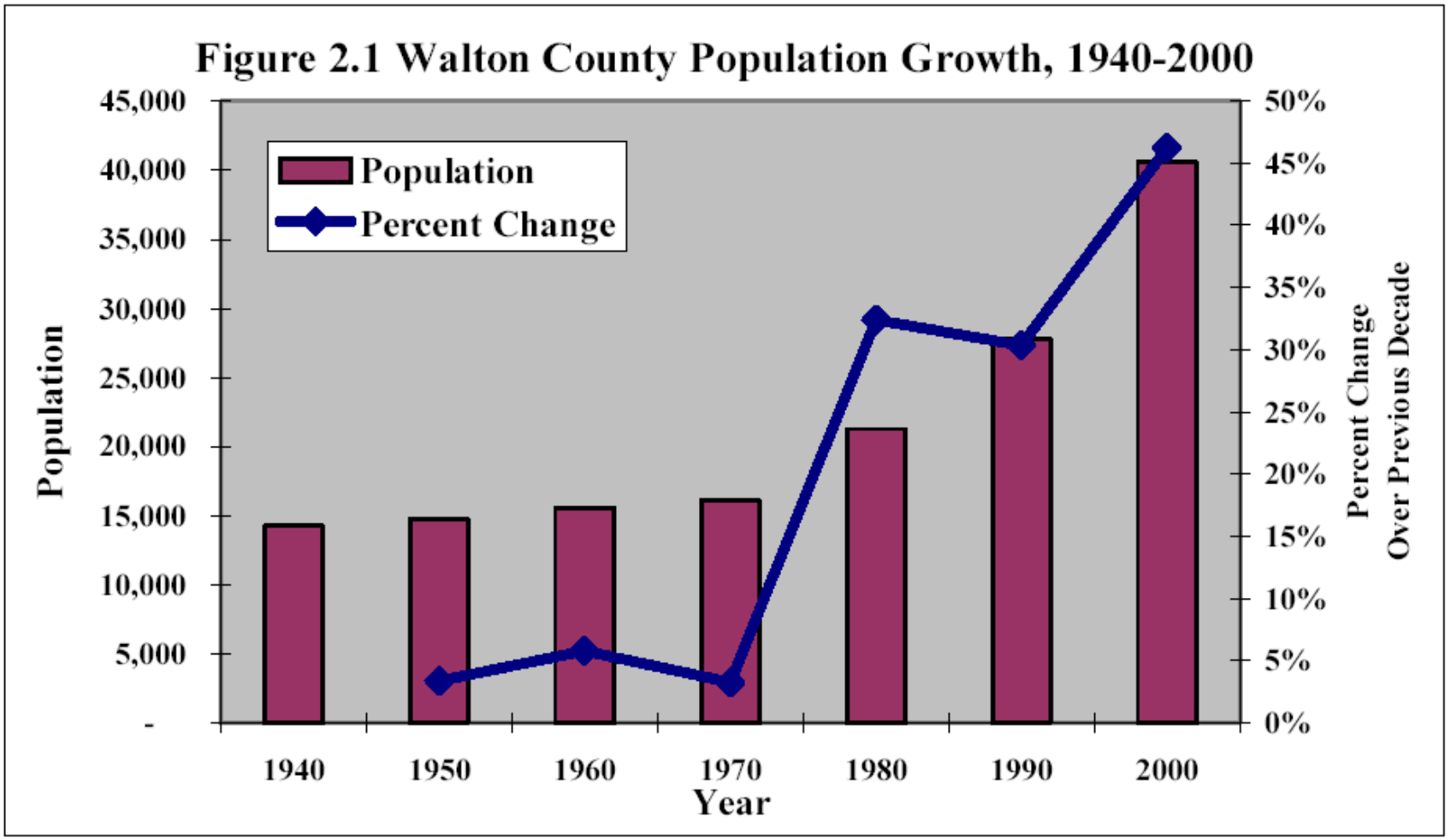


Source: Tallahassee 2003 CIP

Clustered Bar Chart Example



Stacked Column Chart Example



Line-Column Chart Example

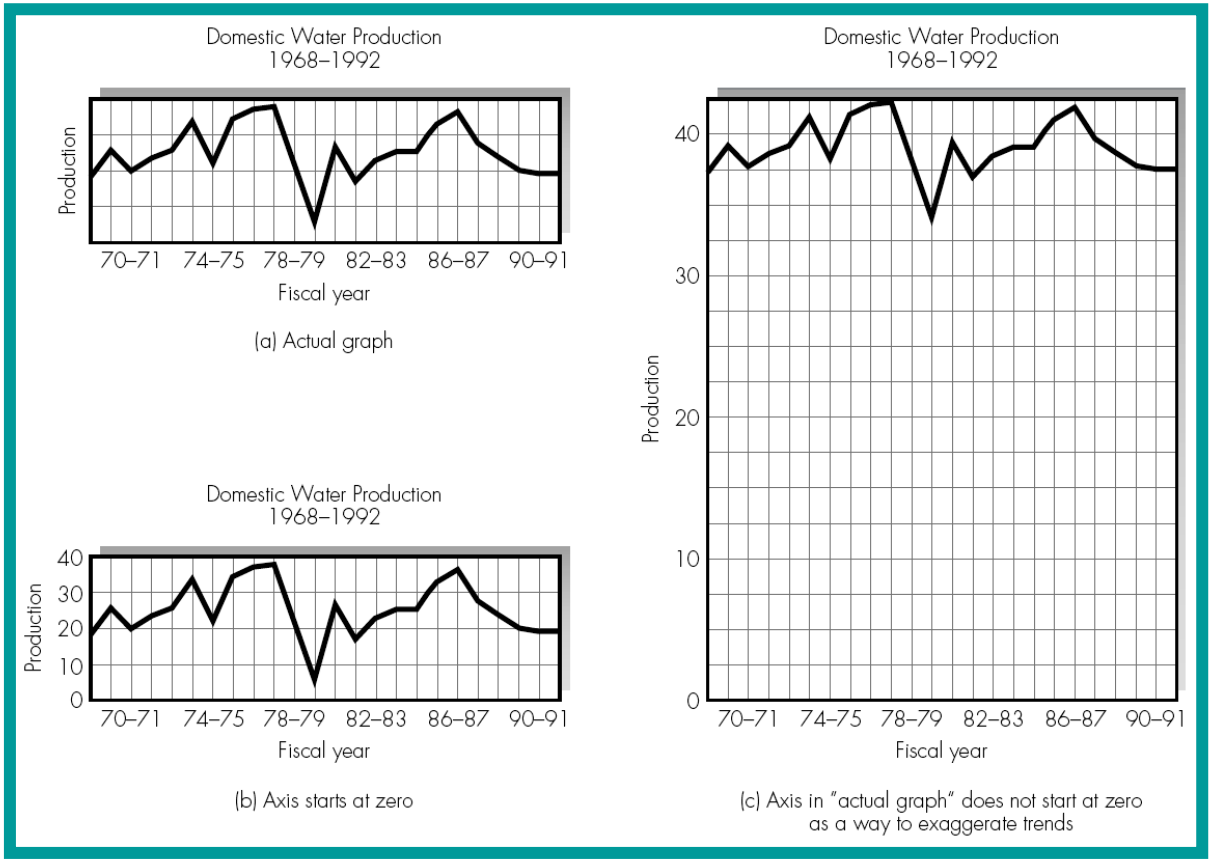
9.4 Difficulties and Disasters in Plots, Graphs, and Pictures

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**

No Labeling on One or More Axes

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

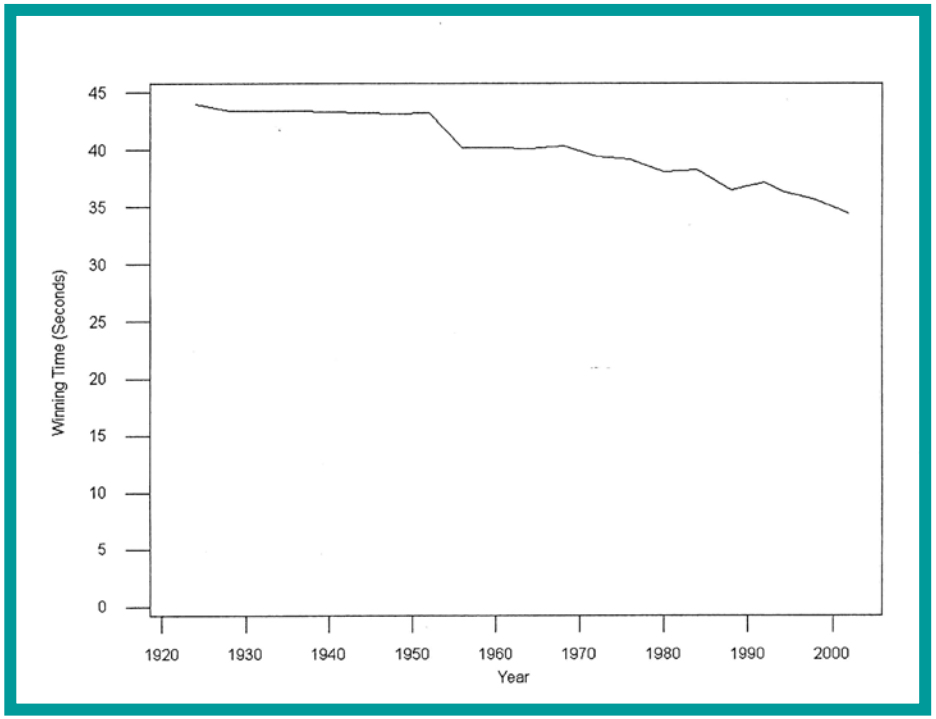


Source: Insert in the *California Aggie* (UC Davis), 30 May 1993.

Not Starting at Zero

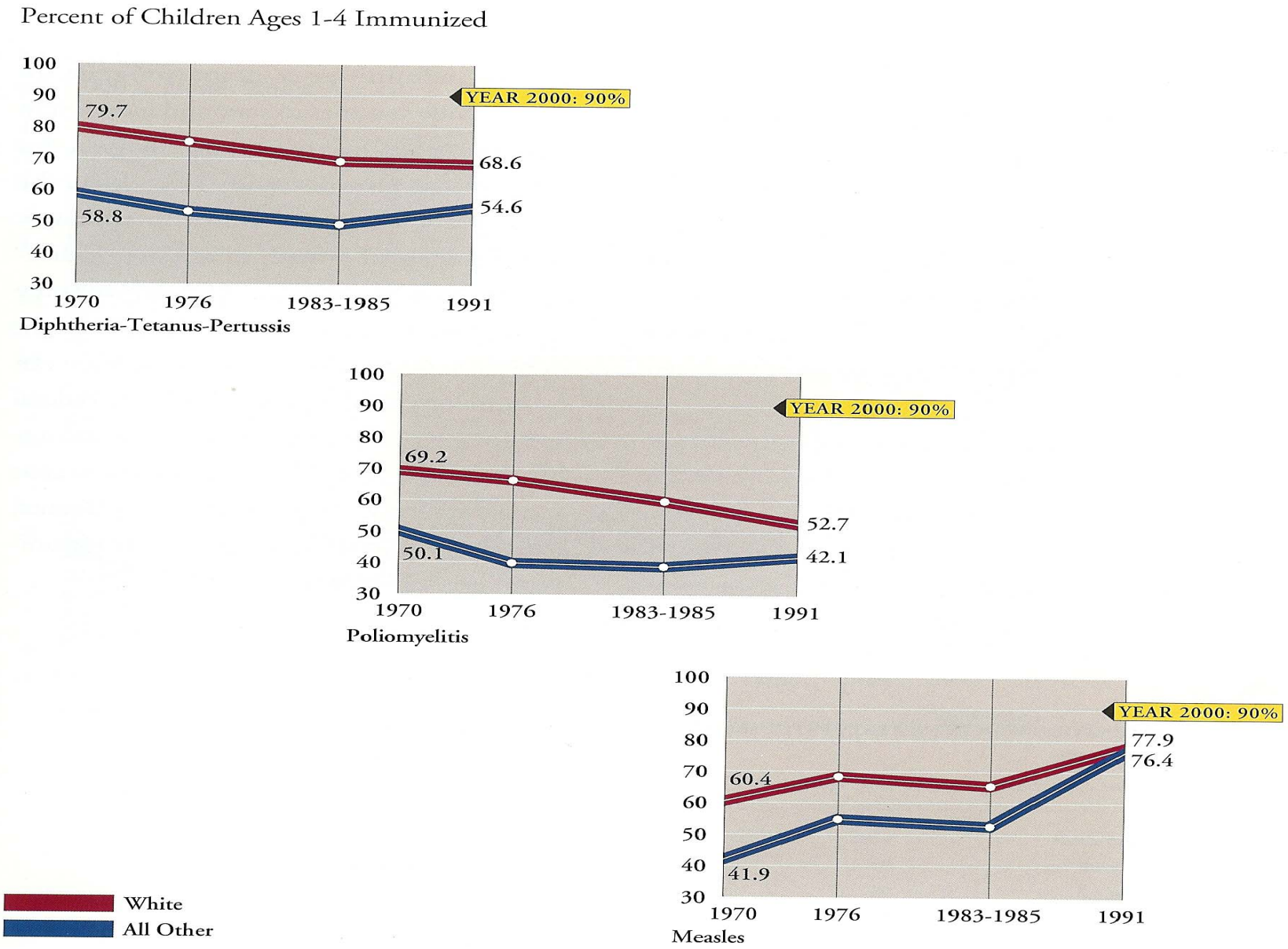
Example:
Winning times for Olympic speed skating data with vertical axis starting at 0.

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.

Include zero on the axis for line and bar graphs.



Access to Health Care: Key Indicators for Policy, November 1993

Graphical Deceit?

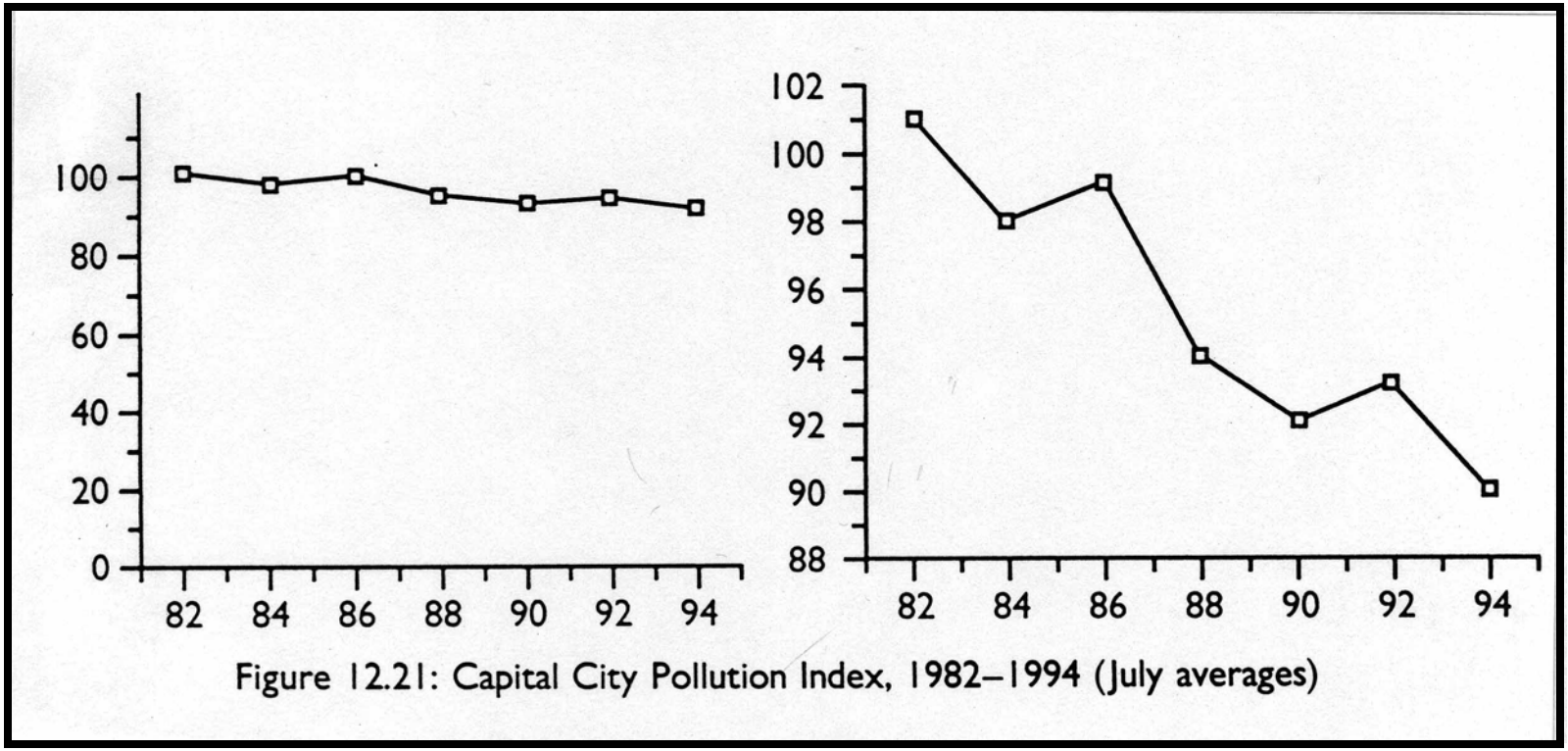
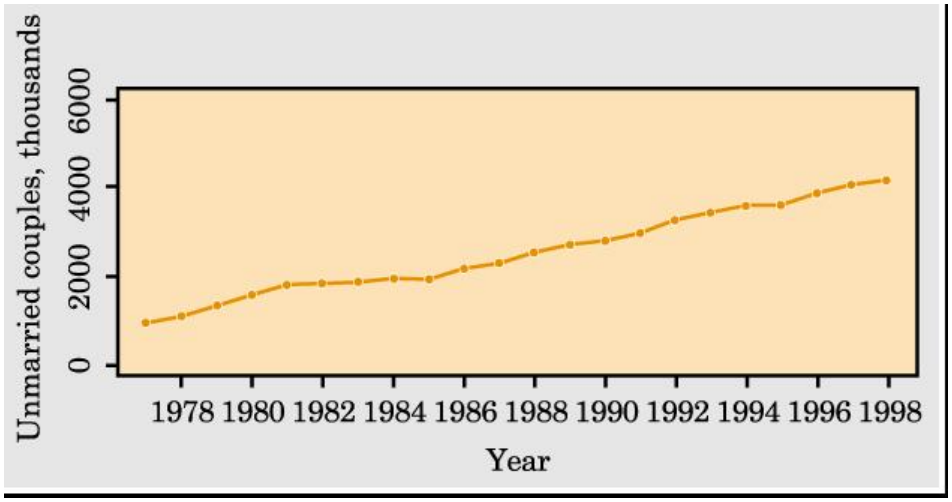
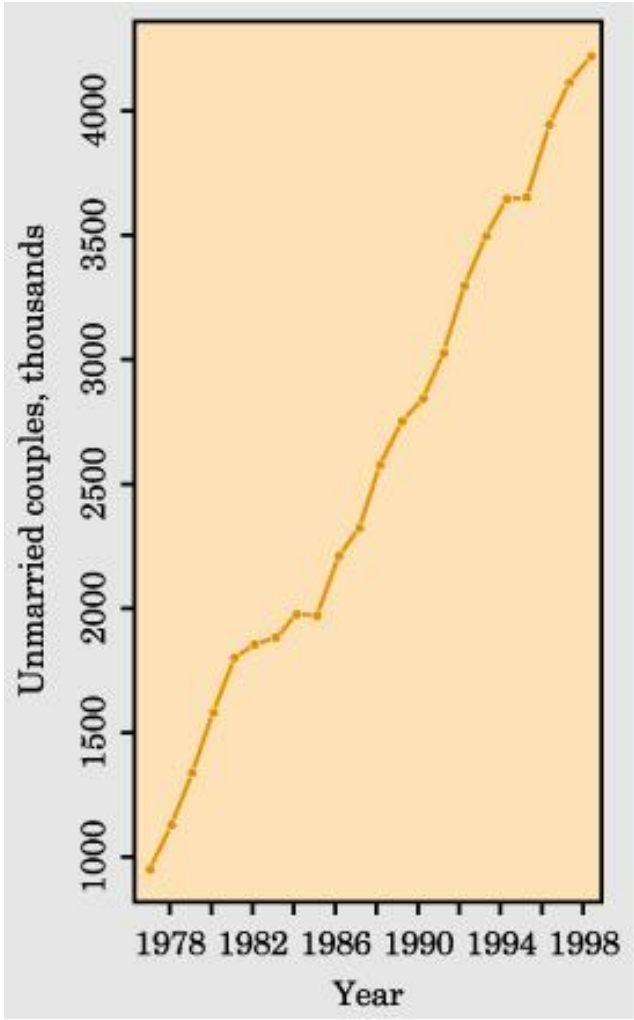


Figure 12.21: Capital City Pollution Index, 1982–1994 (July averages)

If you were running for Mayor in Capital City, which graphic would you use to illustrate your success?



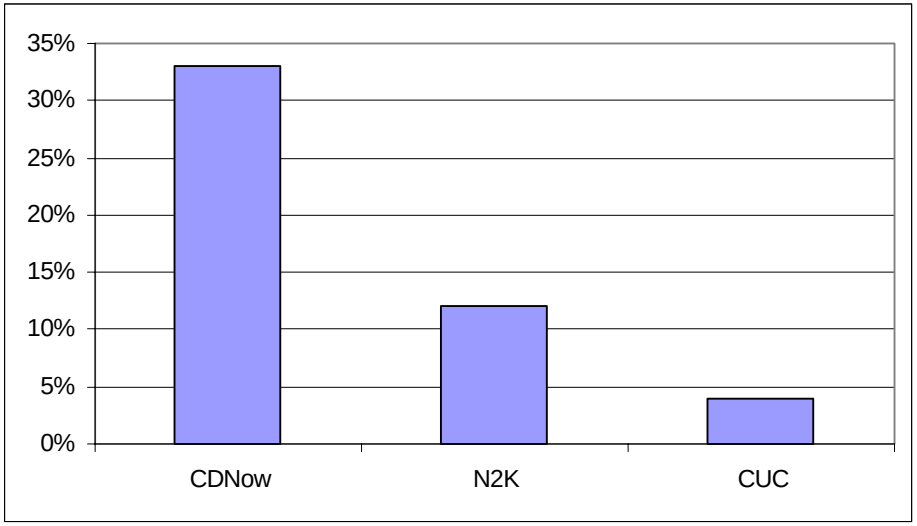


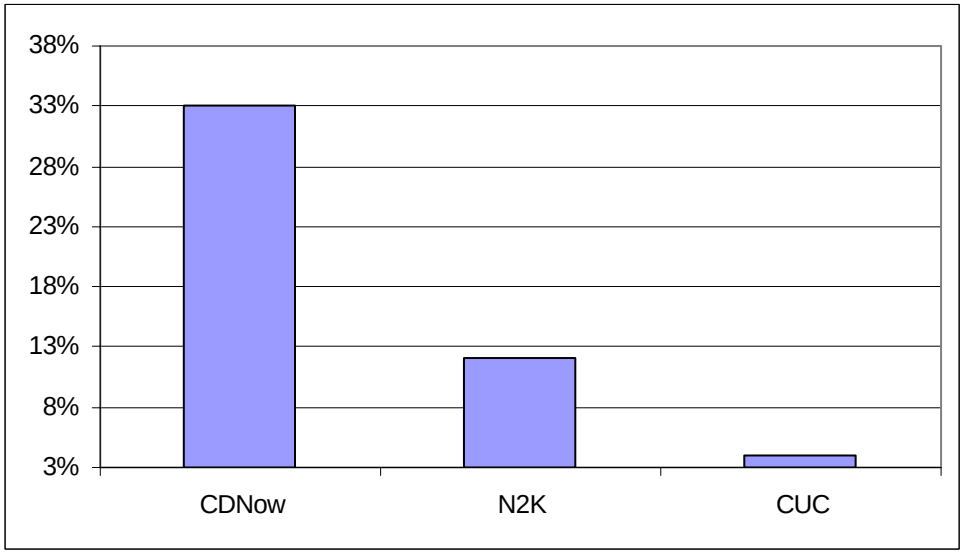
There is no one “right” scale for a line graph.

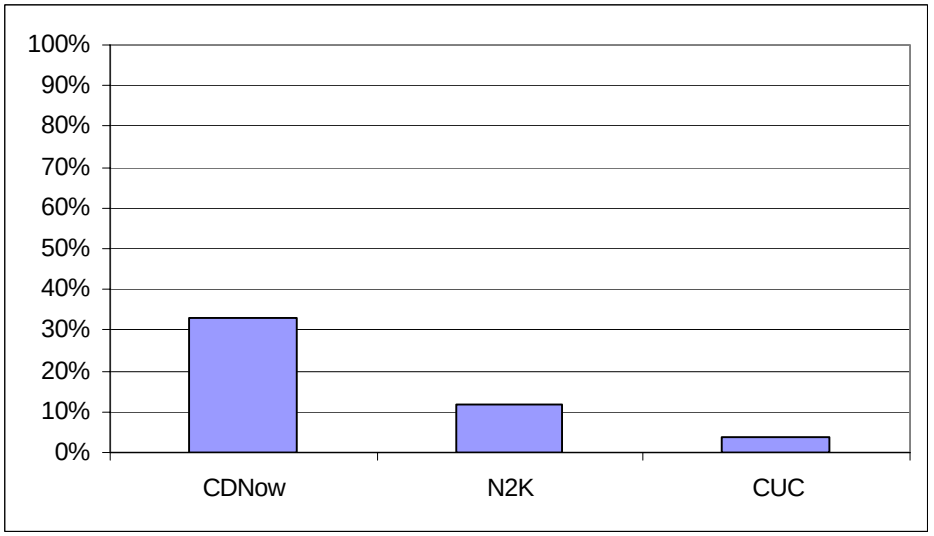
Correct graphs can give different impressions by their choices of scale.

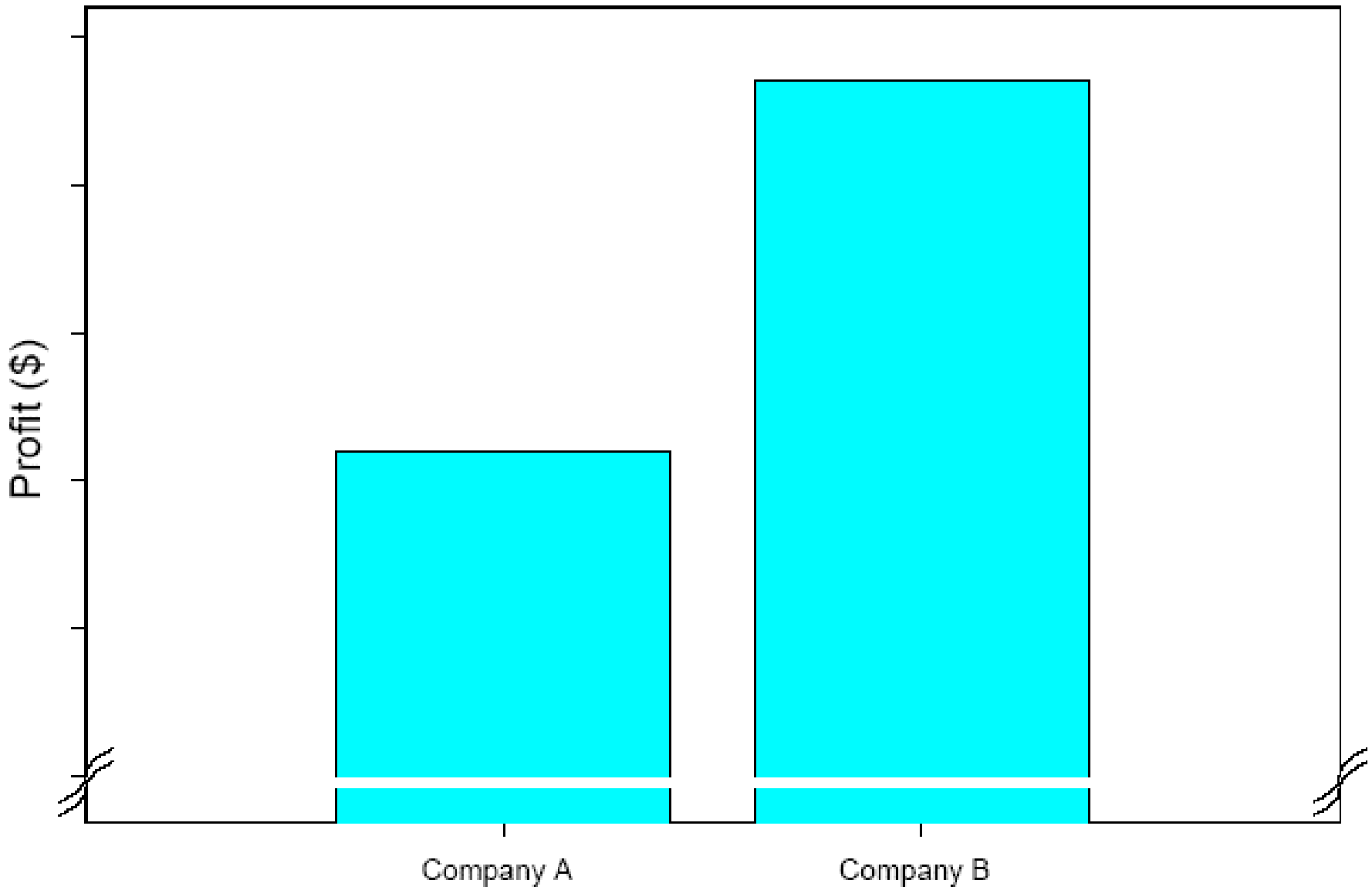
The careful reader of a graph looks closely at the scales marked on the axes.

Include zero? Include 100%?





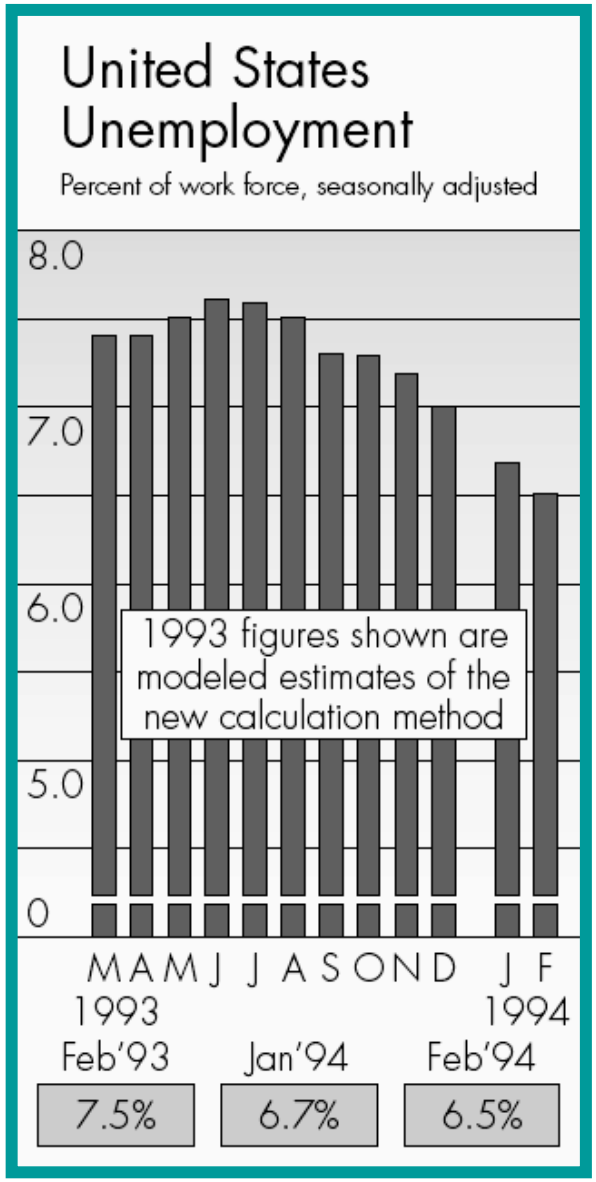



Profit ($)
Company A
Company B

Changes in Labeling on One or More Axes

Example:

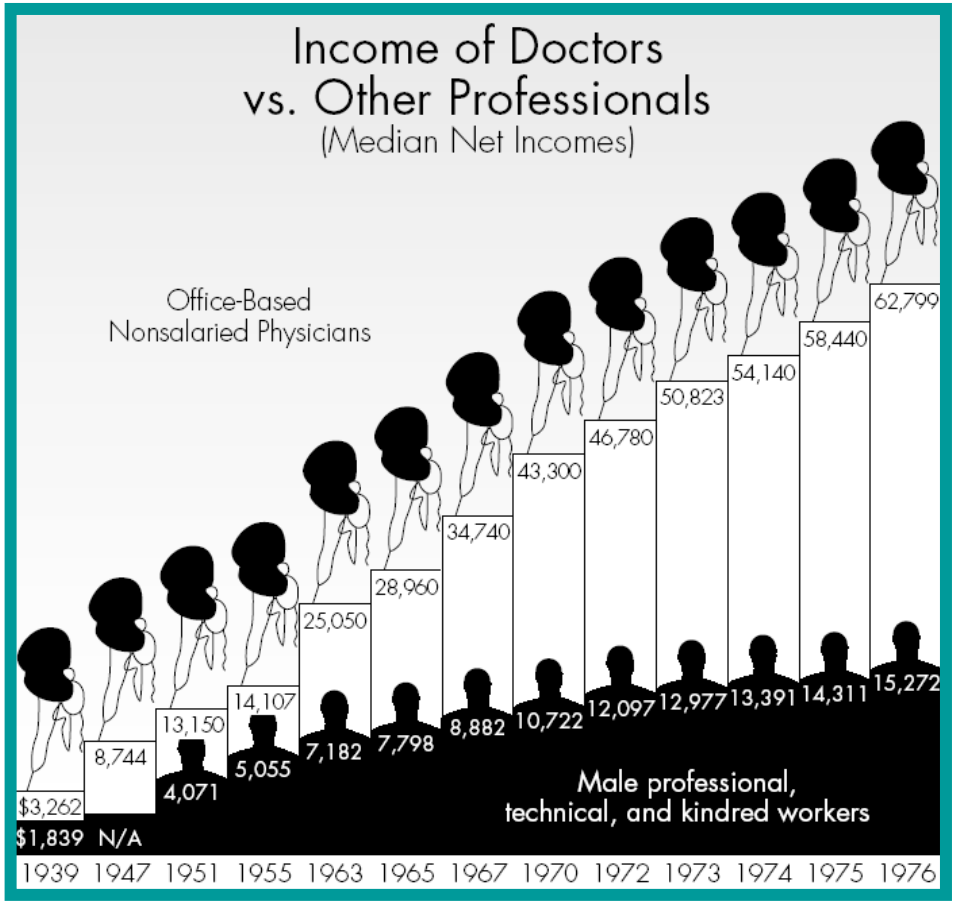
A bar graph with gap in labeling. At first look, seems vertical axis starts at 0, but bottom of the graph actually corresponds to 4.0%



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

Changes in Labeling on One or More Axes

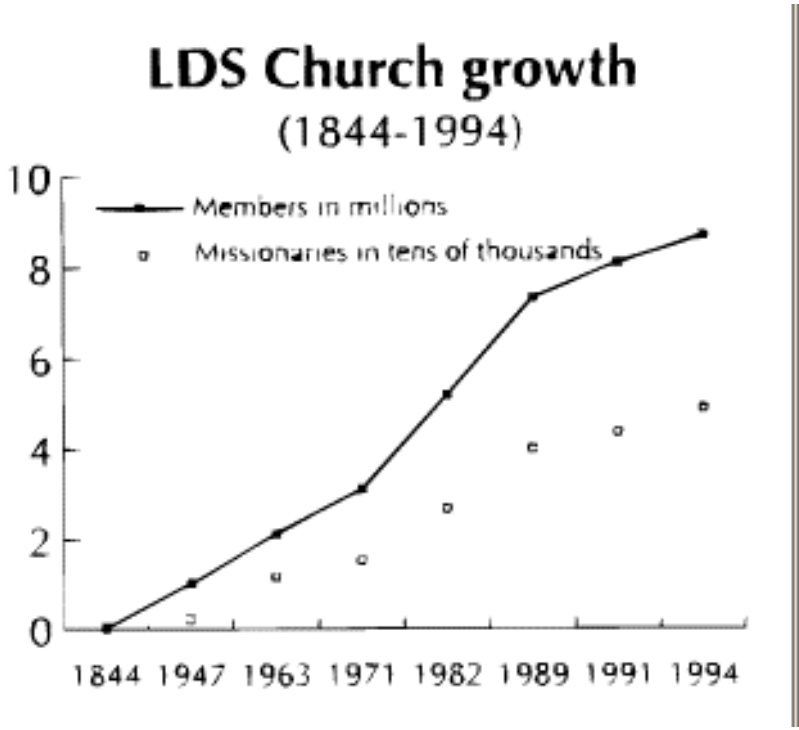
Example:
Distance between successive bars on horizontal axis keeps changing.

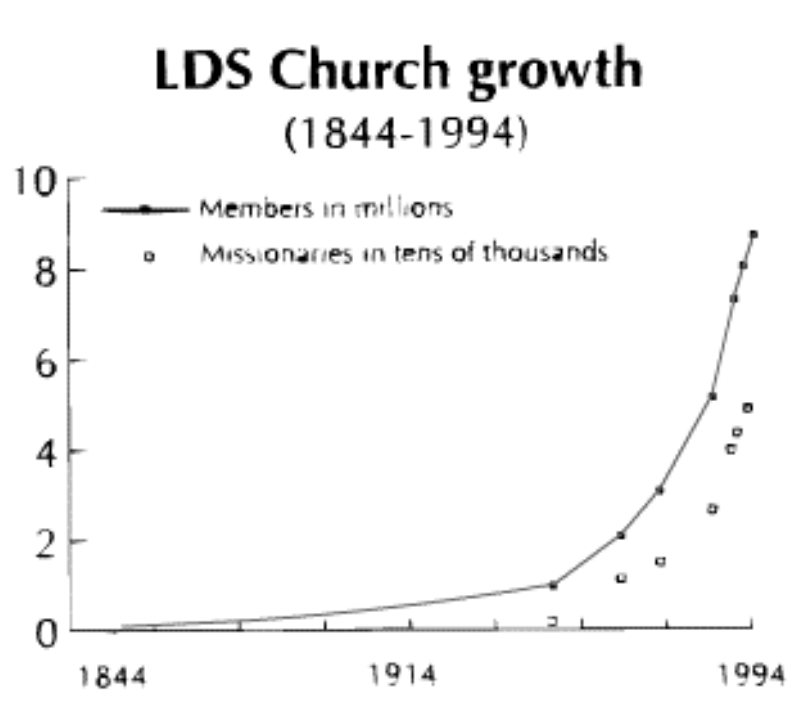


Source:
Washington Post graph reprinted in Wainer, 1984.

Changing in Labeling

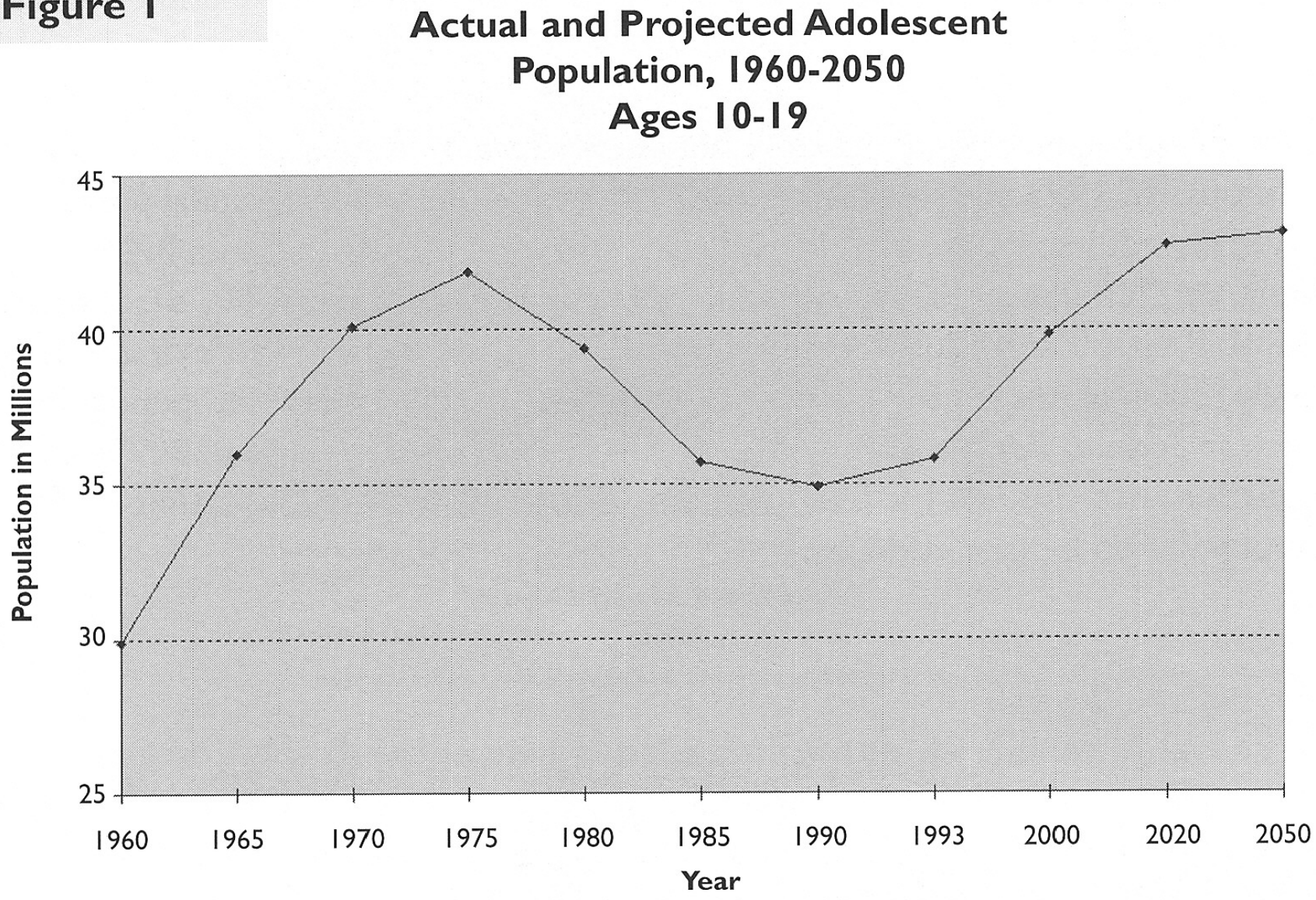
- Watch those scales.
- Because graphs speak strongly, they can mislead.
- Look closely on the scales marked on the axis.





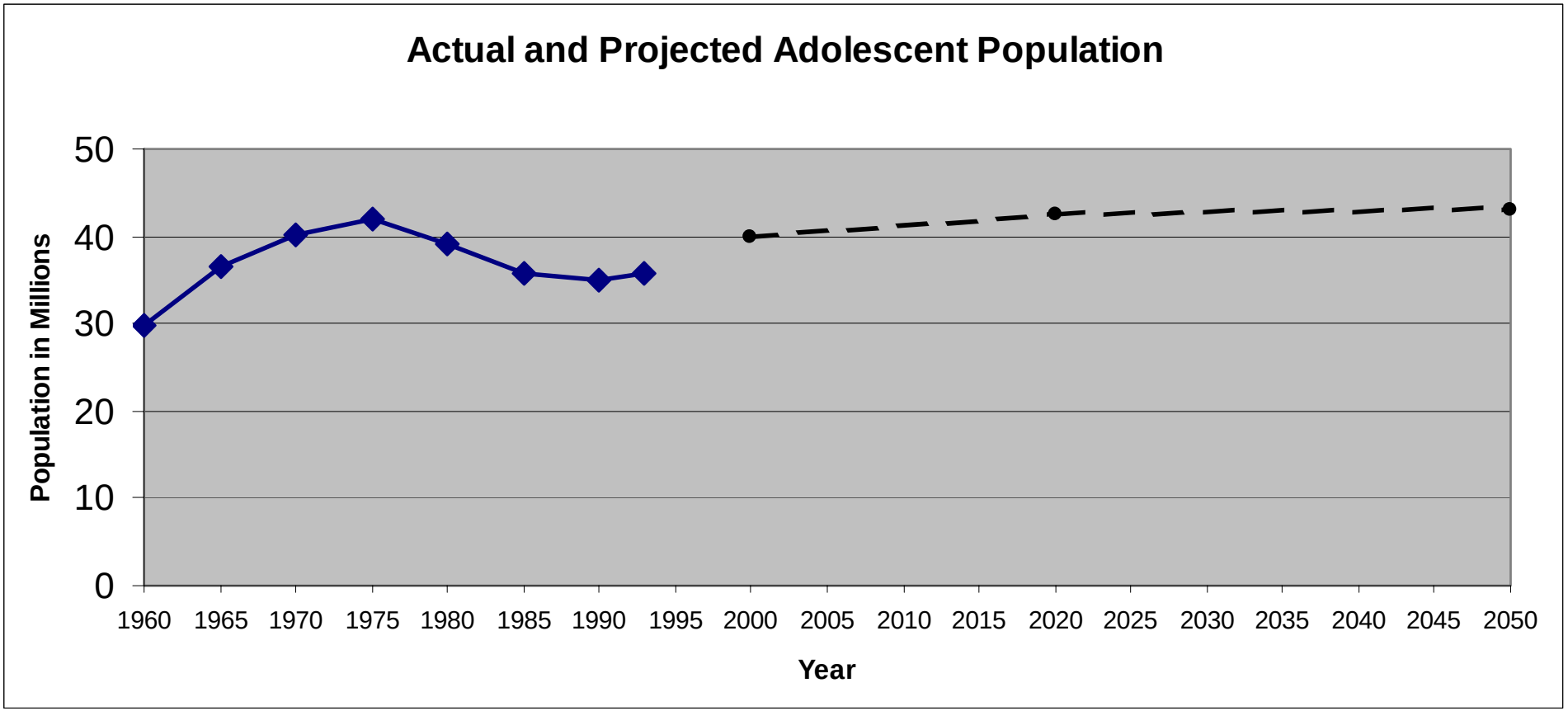
Use a consistent scale within the same graph.

Figure 1



America's Adolescents: Are They Healthy? 1998

Use a consistent scale within the same graph.

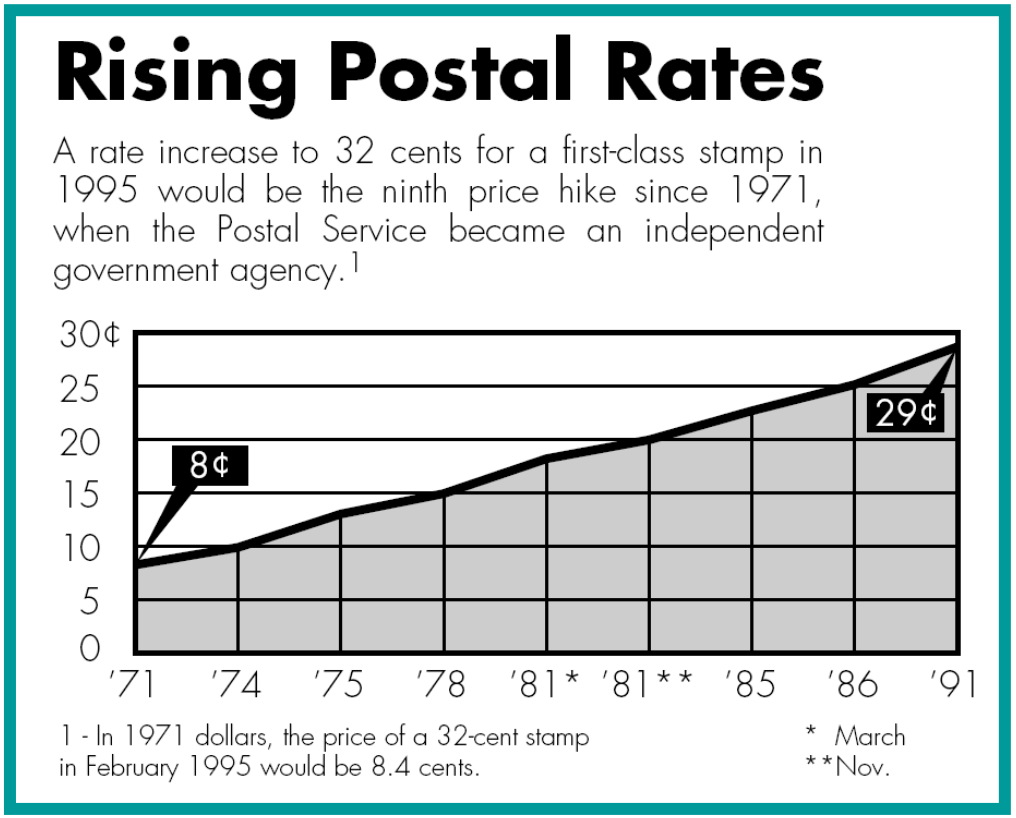


Misleading Units of Measurement

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.

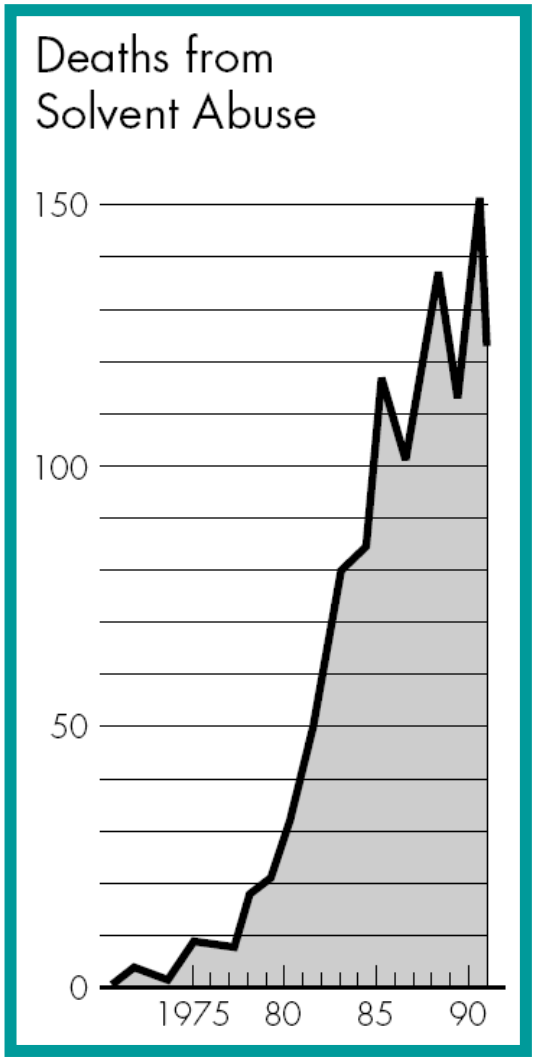
Using Poor Information

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.



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?

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?

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

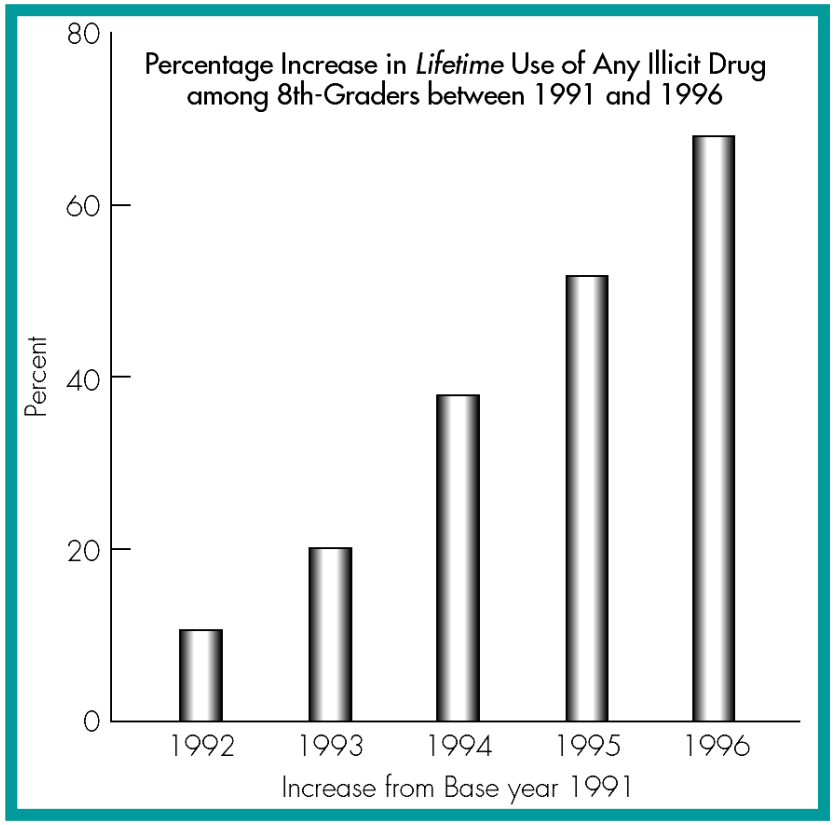
Headline: "*Emergency Situation among Our Youth.*"

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%

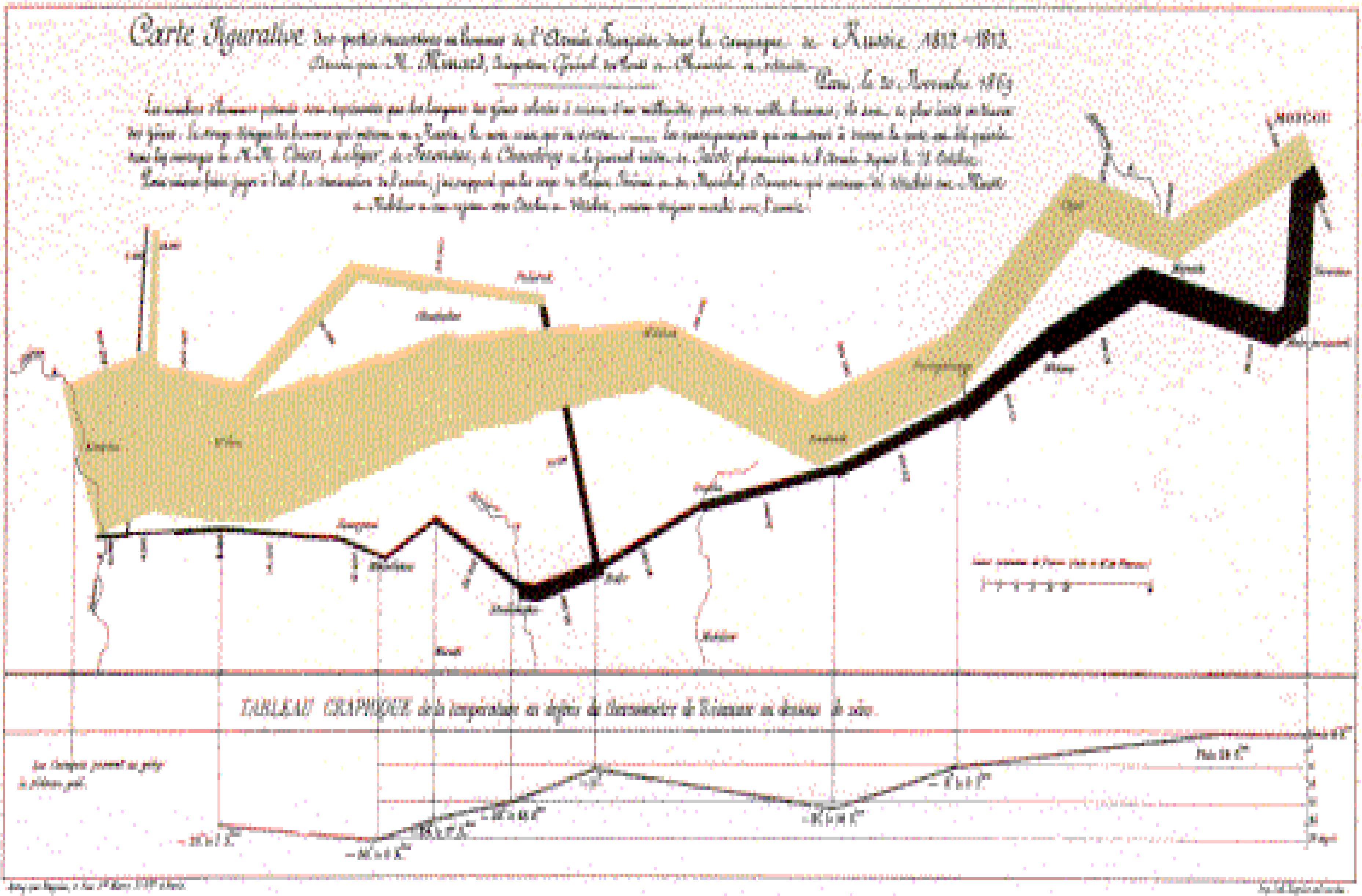
Source: U.S. Department of Justice



Insights from Tufte

“If the statistics are boring,

then you’ve got the wrong data.”


Carte Figurative
Russie 1812-1813.
MOSCOU
TABLEAU GRAPHIQUE

Dimensions are

- time (dates shown in vertical lines)
- geography/space (2 dimensional),
- size of army (thickness of line)
- direction (color),
- temperature (chart at bottom)

Why Worry About Graphics?

- "Data graphics are mainly devices for showing the obvious to the ignorant"
- "They have to be **alive**, **communicatively dynamic**, **decorated**, and **exaggerated**; otherwise all the dullards will fall asleep in the face of those boring statistics"
- Graphics are instruments for reasoning about quantitative information… They reveal data." (Tufte 1983)

Principles of Graphic Excellence

- Graphical excellence…..
  - consists of ***complex ideas communicated with clarity, precision, and efficiency***.
  - is ***that which gives to the viewer the greatest number of ideas in the shortest time*** with the least ink in the smallest space.
  - is ***nearly always multivariate*** (illuminating relationships between numerous variables)
  - requires ***telling the truth*** about the data

Some Rules of Thumb

- Graphic displays should…..
  - ***show the data***
  - ***avoid distorting the data***
  - ***induce the viewer to think about the substance of the graphic*** rather than the methodology, graphic design, or something else
  - ***make large amounts of data coherent***
  - ***encourage the viewer to use the graphic as you intend***, e.g. make comparisons
  - ***be closely integrated with statistical and verbal descriptions of the data***
  - ***be as simple as possible***

Design Guidelines

- To enhance visual quality:
  - ***use a properly chosen format***
  - ***use words, numbers, and graphics together***
  - ***display an accessible complexity of detail***
  - ***have a story to tell about the data***
  - ***produce technical details with care***
  - ***avoid chartjunk***

Tufte Principle

Maximize the data-ink ratio:

$$\text{Data-ink ratio} = \frac{\text{data ink}}{\text{total ink used in graphic}}$$

Avoid “chart junk”

Challenger Exercise (from Tufte)

Challenger Exercise

- On January 28, 1986 the decision was made to launch the space shuttle challenger. Two rubber O-rings leaked and the shuttle exploded. Air temperature that day was about 30F.
- The following data presentations were used in making the launch decision that morning. Can you suggest what information was missing from these presentations. (All the data is here to diagnose the problem!) How might you redesign this data to make the correct decision more apparent?

History of O-Ring Damage on Field Joints

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View: Erosion Depth (in.)	Cross Sectional View: Perimeter Affected (deg)	Cross Sectional View: Nominal Dia. (in.)	Top View: Length Of Max Erosion (in.)	Top View: Total Heat Affected Length (in.)	Clocking Location (deg)
61A LH Center Field**	22A	None	None	0.280	None	None	36° --66°
61A LH ~~CENTER~~ FIELD**	22A	NONE	NONE	0.280	NONE	NONE	338°-18°
51C LH Forward Field**	15A	0.010	154.0	0.280	4.25	5.25	163
51C RH Center Field (prim)***	15B	0.038	130.0	0.280	12.50	58.75	354
51C RH Center Field (sec)***	15B	None	45.0	0.280	None	29.50	354
41D RH Forward Field	13B	0.028	110.0	0.280	3.00	None	275
41C LH Aft Field*	11A	None	None	0.280	None	None	--
41B LH Forward Field	10A	0.040	217.0	0.280	3.00	14.50	351
STS-2 RH Aft Field	2B	0.053	116.0	0.280	--	--	90

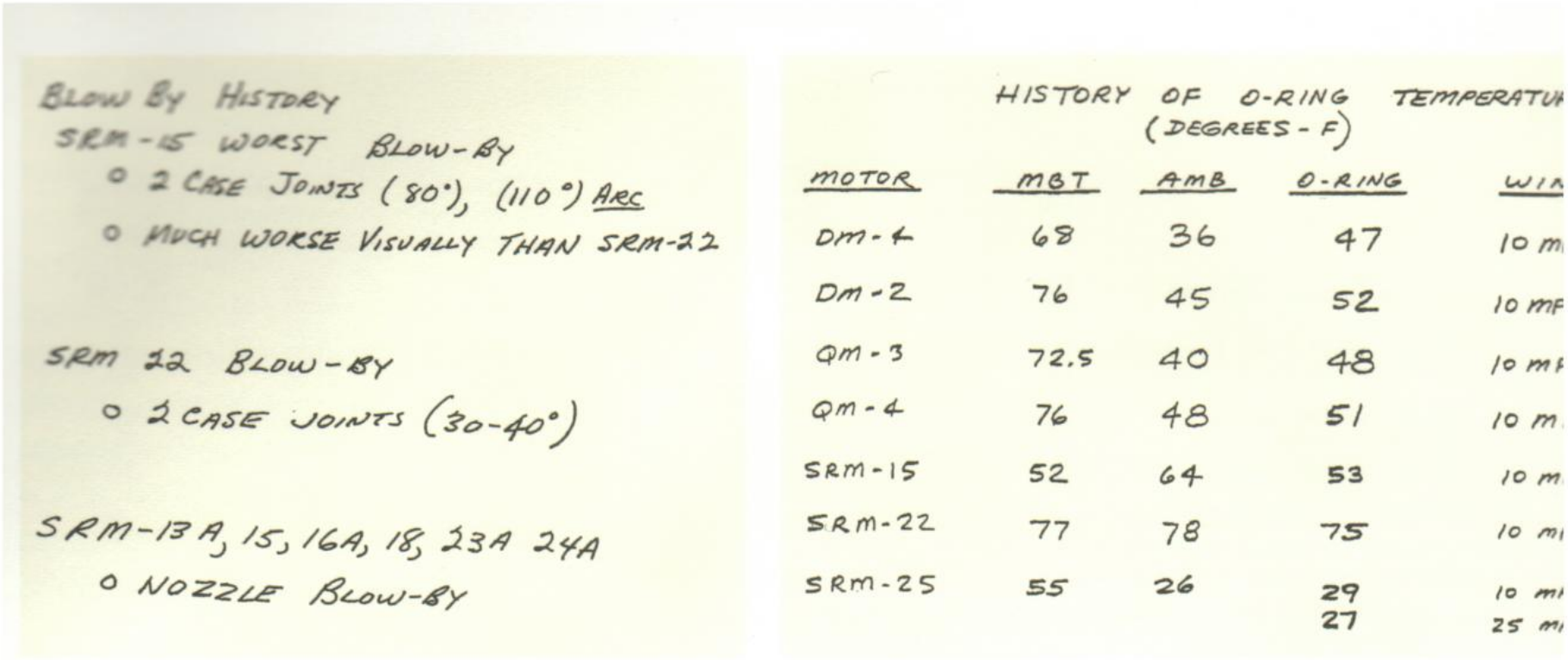
*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port - 0 deg.

OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.

SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO SOOT BLOWBY. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY.

SRM # = launch number of shuttle



Blow By History

SRM-15 worst Blow-By

- 2 case joints (80°), (110°) arc
- much worse visually than SRM-22

SRM 22 Blow-By

- 2 case joints (30-40°)

SRM-13A, 15, 16A, 18, 23A 24A

- nozzle Blow-By

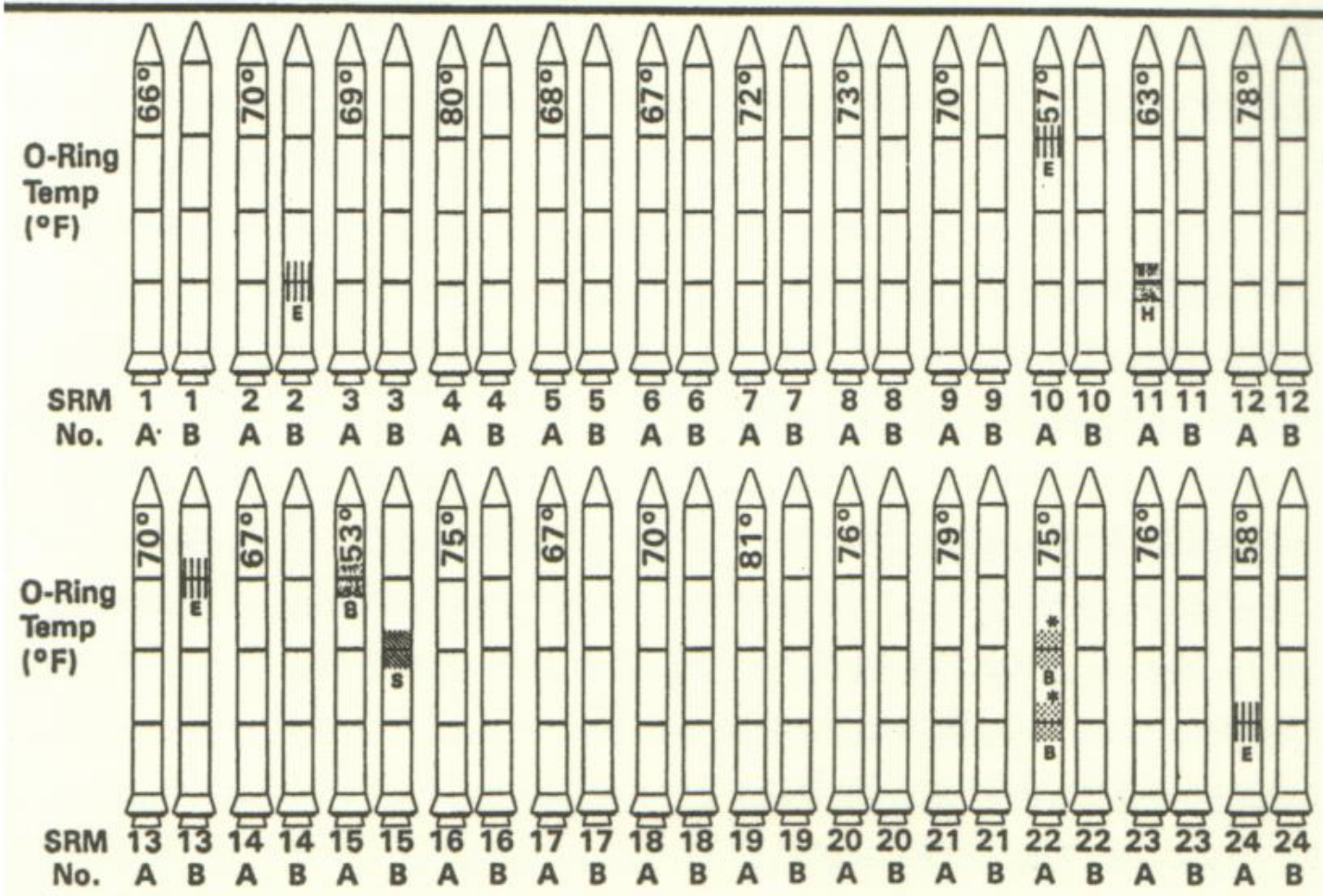
History of O-ring Temperatu
(degrees-F)

motor	MBT	AMB	O-ring	Win
DM-1	68	36	47	10 m
DM-2	76	45	52	10 mF
QM-3	72.5	40	48	10 mF
QM-4	76	48	51	10 m
SRM-15	52	64	53	10 m
SRM-22	77	78	75	10 m
SRM-25	55	26	29 27	10 m 25 m

AMB = ambient temperature

O-ring = temperature at the O-ring itself

History of O-Ring Damage in Field Joints (Cont)



MORTON THIOKOL, INC.
Wasatch Operations

* No Erosion

INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

Challenger story

Summary book review:

http://www.statview.com/support/techsup/faq/Tufte/tufte.shtml

Decision needed to be made on whether or not to have the launch.

The Morton Thiokol engineers (makers of the rocket) argued against it. They lost the argument.

Tufte describes factors that were obvious that made the presentation weak, including

lack of authorship

showing relationships

Adaptation of Tufte chart