Good and Bad Graphs

- ***Example 1:*** **Look at the graph comparing the interest rates of several lenders.**
- **What is the perception that the reader has initially?**
- **How well or poorly does this perception reflect the truth of the numerical comparison?**

Types of Data

- ***Categorical Variables*** **place individuals into one of several categories.**
- ***Quantitative*** **or** ***(Numerical) Variables*** **measure a characteristic of an individual with a (mathematically meaningful) number.**
- **It makes sense to add, or average, the values of a numerical variable.**
- **Variables measured on college applicants: SAT score, high school GPA, race, gender, number of AP courses taken.**
- **Which are categorical? Which are quantitative? (Clicker quiz next)**

Clicker Quiz 1

Variables measured on college applicants: SAT score, high school GPA, race, gender, number of AP courses taken.

Which are categorical? Which are quantitative?

A. ***Categorical:*** **race, gender, number of AP courses taken.** ***Quantitative:*** **SAT score, high school GPA.**

B. ***Categorical:*** **gender.** ***Quantitative:*** **race, SAT score, high school GPA, number of AP courses taken.**

C. ***Categorical:*** **race, gender, high school GPA.** ***Quantitative:*** **SAT score, number of AP courses taken.**

D. ***Categorical:*** **race, gender.** ***Quantitative:*** **SAT score, high school GPA, number of AP courses taken.**

Data Tables

- **A *data table* is often a reasonable way to summarize data, especially categorical data.**
- **A table is a non-graphical way to present the *distribution* of a variable.**
- **The *distribution* of a variable indicates two things: (1) What value(s) a variable can take, and (2) how often it takes those values.**
- **A value's *frequency* is a count of how many times it has occurred in a data set.**
- **A value's *relative frequency* is the proportion of times it has occurred in a data set.**

An Example Data Table

Table 1: Counts for students in an introductory statistics class, Summer 2002.

Year in School	Count	Percent
Freshman	18	41.9%
Sophomore	10	23.3%
Junior	6	14.0%
Senior	9	20.9%
Total	43	100%

Rates vs. Counts (Again)

- **Previous table presents the *distribution* of the variable "Year in School."**
- **Note that *rates* may be more useful/informative in a data table than raw counts.**
- **Especially true when there are a huge number of individuals in the overall data set.**
- **When percentages in a table are rounded off, total percentage may not add to 100% due to *roundoff error*.**
- **Check previous table: 41.9 + 23.3 + 14.0 + 20.9 = ?**

Pie Charts and Bar Graphs

- ***Pie Charts*** **and** ***Bar Graphs*** **are common ways to present a graphical summary of** ***categorical*** **data.**
- ***Pie Charts*** **present proportions as wedges in a circle of various sizes.**
- **Visually we immediately see "pieces of a whole."**
- ***Bar graphs*** **present proportions as vertical bars of various heights.**
- **Easy to immediately compare categories to see which is biggest, smallest, etc.**
- **See example** ***Pie Chart*** **and** ***Bar Graph*** **for school-year data.**

Notes and Cautions: Pie Charts and Bar Graphs

- ***Pie Charts*** **and** ***Bar Graphs*** **are not appropriate for summarizing** ***quantitative*** **data (unless we define certain numerical classes for the data).**
- ***Bar Graphs*** **are easier to construct by hand than** ***Pie Charts*****.**
- **These types of graphs are usually generated with the help of software.**
- ***Bar Graphs*** **can also graph data that are not part of a whole (see example with Tax Data for Countries).**

More on Pie Charts and Bar Graphs

- **From the help file of a statistical software package: "Pie charts are a very bad way of displaying information. The eye is good at judging linear measures and bad at judging relative areas."**
- **Based on empirical investigations of Cleveland (1985) and McGill as well as investigations by perceptual psychologists.**
- **Make sure the numerical axis of the bar graph is not misleading (recall Example 1!).**

Clicker Quiz 2

What type of variable's distribution could most easily be represented by a pie chart or bar graph?

A. Height of basketball players

B. 100-meter dash time of sprinters

C. Party of members of Congress

D. Weight of oranges

Pictograms and their Drawbacks

- ***Pictograms*** **are like** ***Bar Graphs*****, but use a picture rather than a simple bar to represent proportions.**
- **The width of the picture is often increased with the height so the pictures aren't distorted.**
- **Our eye perceives the whole area of the picture, rather than just the height.**
- **This makes comparing categories quite deceptive.**
- **See example showing building permit data.**

Line Graphs

- ***Line Graphs*** **show the pattern of variation of some variable** ***over time*****.**
- **Time (in years? months? days? hours?) is always plotted on the horizontal axis.**
- **The value of the variable is plotted on the vertical axis.**
- **Look for** ***overall pattern*****: Increasing trend? Decreasing trend? Constant trend?**
- **Look for times when the pattern shows** ***deviation*** **from the overall trend.**
- **Look for** ***seasonal variation*****, especially common when data are plotted monthly or weekly (gas prices example).**
- **See example showing freshman percentage over several years.**

Clicker Quiz 3

If we plotted the annual proportion (over a couple of decades) of eligible citizens who voted on Election Day, what type of pattern would we likely see?

A. Seasonal variation

B. Increasing trend

C. Decreasing trend

D. Completely haphazard variation

More on Line Graphs

- **Advanced statistical methods can adjust for (remove) seasonal variation from data.**
- **If a variable is known vary seasonally, it's more instructive to study the pattern *after removing* the seasonal variation.**
- ***Example:* Unemployment rate known to always rise in January. Why?**
- **Very important to use the *appropriate scales* for the axes.**
- **The appearance of line graphs can change greatly if the scales are changed.**
- **Could lead to deceptive conclusions.**
- **See Example: Data on unmarried couples.**

Clicker Quiz 4

What is another aspect of the unmarried couples plot that could use improvement?

A. A different color should be used for each decade (1980s, 1990s, etc.)

B. Couples (rather than thousands of couples) should be plotted on the vertical axis.

C. Instead of the count of unmarried couples being plotted, a rate should be used to account for population changes.

More Graphs: Good, Not-so-good, Terrible

- **Compare the two graphs (on page 204 of book) showing stock performances from 1970-2003**
- **What is top graph showing? What is bottom graph showing?**
- **Which shows the pattern of variation of stock prices better?**
- **Another Example: State and Local Property Taxes Graph**
- **How would you change the scales on these axes?**

Good Graphical Practices

- **Use *labels* and *legends* to say what variables are plotted, units, source of data, etc.**
- **Make the data stand out, not fancy artistic flourishes.**
- **Have the information reflect what the eye initially sees.**
- **Avoid cluttering the plot with needless visual elements.**
- **Be wary of misleading graphs that you may encounter!**