Chapter 10: Graphs, Good and Bad

Thought Question

What is confusing or misleading about the pie chart?

It doesn’t add up to 100%.

Thought Question

What is confusing or misleading about the bar graph?

There is no scale on the vertical axis.
Thought Question

What is confusing or misleading about the bar graph?

The scale on the vertical axis covers only a small range of values, making the difference look more significant than it really is.

Global Warming Out of Control!

What is confusing or misleading about the chart and line graph?

It only shows the first half of the year, so of course temperatures are rising dramatically.
Tables

How can a table provide a good summary of data?

Terminology

- **Distribution**: what values a variable takes and how often it takes those values.
- **Categorical Variable**: puts an individual into one of several categories.
- **Quantitative Variable**: takes numerical values.

Example: Education

![Table showing educational attainment]

<table>
<thead>
<tr>
<th>Level of education</th>
<th>Number of persons (thousands)</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than high school</td>
<td>26,415</td>
<td>13.3</td>
</tr>
<tr>
<td>High school graduate</td>
<td>61,626</td>
<td>31.1</td>
</tr>
<tr>
<td>Some college, no degree</td>
<td>33,832</td>
<td>17.1</td>
</tr>
<tr>
<td>Associate's degree</td>
<td>17,838</td>
<td>9.0</td>
</tr>
<tr>
<td>Bachelor's degree</td>
<td>37,635</td>
<td>19.0</td>
</tr>
<tr>
<td>Advanced degree</td>
<td>20,938</td>
<td>10.6</td>
</tr>
<tr>
<td>Total</td>
<td>198,285</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Source: Census Bureau, *Educational Attainment in the United States: 2011*
What makes this a clear table?

- The table is clearly labeled so that we can see the subject of the data at once.
- Labels within the table identify the variables and state the units in which they are measured.
- The source of the data appears at the bottom of the table.

Add up the total number of people in the second column of the table.

198,284 (thousands)

Why doesn’t your answer match the total in the table?

Each entry in the table is rounded to the nearest thousand, while the total is rounded separately.

This is an example of roundoff errors.
Pie Charts

How can we display the distribution of a categorical variable?

Example: Education

Distribution of the level of education among persons aged 25 years and over in 2011

What is the benefit of summarizing the data with a pie chart?

It allows us to see that the parts make a whole.

Do you think this is the best way to compare the sizes of the different percentages?

No, because angles are harder to compare than lengths.
Bar Graphs

How can we display the distribution of a categorical variable?

Example: Education

Distribution of the level of education among persons aged 25 years and over in 2011

What are the benefits of summarizing the data with a bar graph instead of the pie chart?

Taller bars clearly indicate a higher percentage, the bar graph is easier to draw, and we can order the amount of education on the horizontal axis.
Line Graphs

How can we display how quantitative variables change over time?

**Example: S&P 500**

![Real Price S&P500 Graph](fatasmihov.blogspot.com/2013_05_01_archive.html)

**Do you see any overall patterns or trends?**

It seems that the price of the S&P 500 is generally roughly increasing with time, but with some variation.

**Do you see any striking deviations that might indicate unusual occurrences? What were these occurrences?**

Yes, 2001-2002 (dot-com bubble) and 2008-2011 (recession).
Making Bad Graphs

Some more examples!

Example: Pictogram

Why is this pictogram misleading?

Because both the width and height of the image are increased. Doubling each dimension quadruples the area of the image.
Example: Chartjunk

Why is this pie chart confusing?

It’s hard to interpret the area of the pieces the way the pie chart is angled. The image is distracting.
And finally...
Making Good Graphs

*Best Practices:*

- Title your graph.

- Make sure labels and legends describe variables and their measurement units.

- Be careful with the scales used.

- Make the data stand out. Avoid distracting grids, artwork, etc.

- Pay attention to what the eye sees. Avoid pictograms and tacky effects.
Chapter 10 Exercises

1. In the U.S. Census Bureau document *America’s Families and Living Arrangements: 2011*, we find the following data on the marital status of American women aged 15 years and older as of 2011:

<table>
<thead>
<tr>
<th>Marital Status</th>
<th>Count (thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never married</td>
<td>34,963</td>
</tr>
<tr>
<td>Married</td>
<td>65,000</td>
</tr>
<tr>
<td>Widowed</td>
<td>11,306</td>
</tr>
<tr>
<td>Divorced</td>
<td>13,762</td>
</tr>
</tbody>
</table>

(a) How many women were not married in 2011?

The number of unmarried women is $34,963 + 11,306 + 13762 = 60,631$.

(b) Make a bar graph to show the distribution of marital status.

(c) Would it also be correct to use a pie chart? Explain.

A pie chart could be used because we are displaying the distribution of a categorical variable and we know the total for all categories.
2. The following figure compares the value of exports (in billions of dollars) in 2002 from the world’s leading exporters: Germany, Japan, and the United States.

(a) Explain why this is not an accurate graph.

This is a pictogram and it is misleadingly scaled.

(b) Make an accurate graph to compare the exports of these nations.
3. Here are data on births to unwed mothers as a percentage of births in the United States, from the *Statistical Abstract*. These data show a clear increasing trend over time. Make two line graphs of these data: one designed to show only a gradual increase over time, and a second designed to show a shockingly steep increase.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>% Unwed Mothers</td>
<td>5.3</td>
<td>7.7</td>
<td>10.7</td>
<td>14.2</td>
<td>18.4</td>
<td>22.0</td>
<td>28.0</td>
<td>32.2</td>
<td>33.2</td>
<td>36.9</td>
<td>40.7</td>
</tr>
</tbody>
</table>
4. Sketch line graphs of a series of observations over time having each of the following characteristics. Mark your time axis in years.

(a) A strong downward trend, but no seasonal variation.

(b) Seasonal variation each year, but no clear trend.

(c) A strong downward trend with seasonal variation.