Lecture Notes for Chapter 2

Tony Baker
MATK123 Elementary Statistics Fall 2022 9-12-2022
Three Rivers Community College
Introductory Statistics 10th ed. by Weiss.

Key Topics:
2.1 Variables and Data
2.2 Organizing Qualitative Data
2.3 Organizing Quantitative Data
2.4 Distribution Shapes
2.5 Misleading Graphs

Definition 2.1: Variables

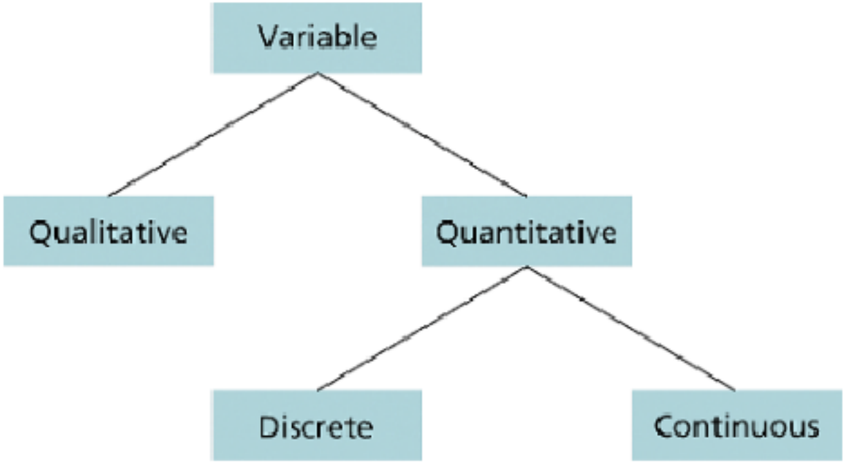
Variable: A characteristic that varies from one person or thing to another.

Qualitative variable: A nonnumerically valued variable.

Quantitative variable: A numerically valued variable.

Discrete variable: A quantitative variable whose possible values can be listed. In particular, a quantitative variable with only a finite number of possible values is a discrete variable.

Continuous variable: A quantitative variable whose possible values form some interval of numbers.



Definition 2.2: Data

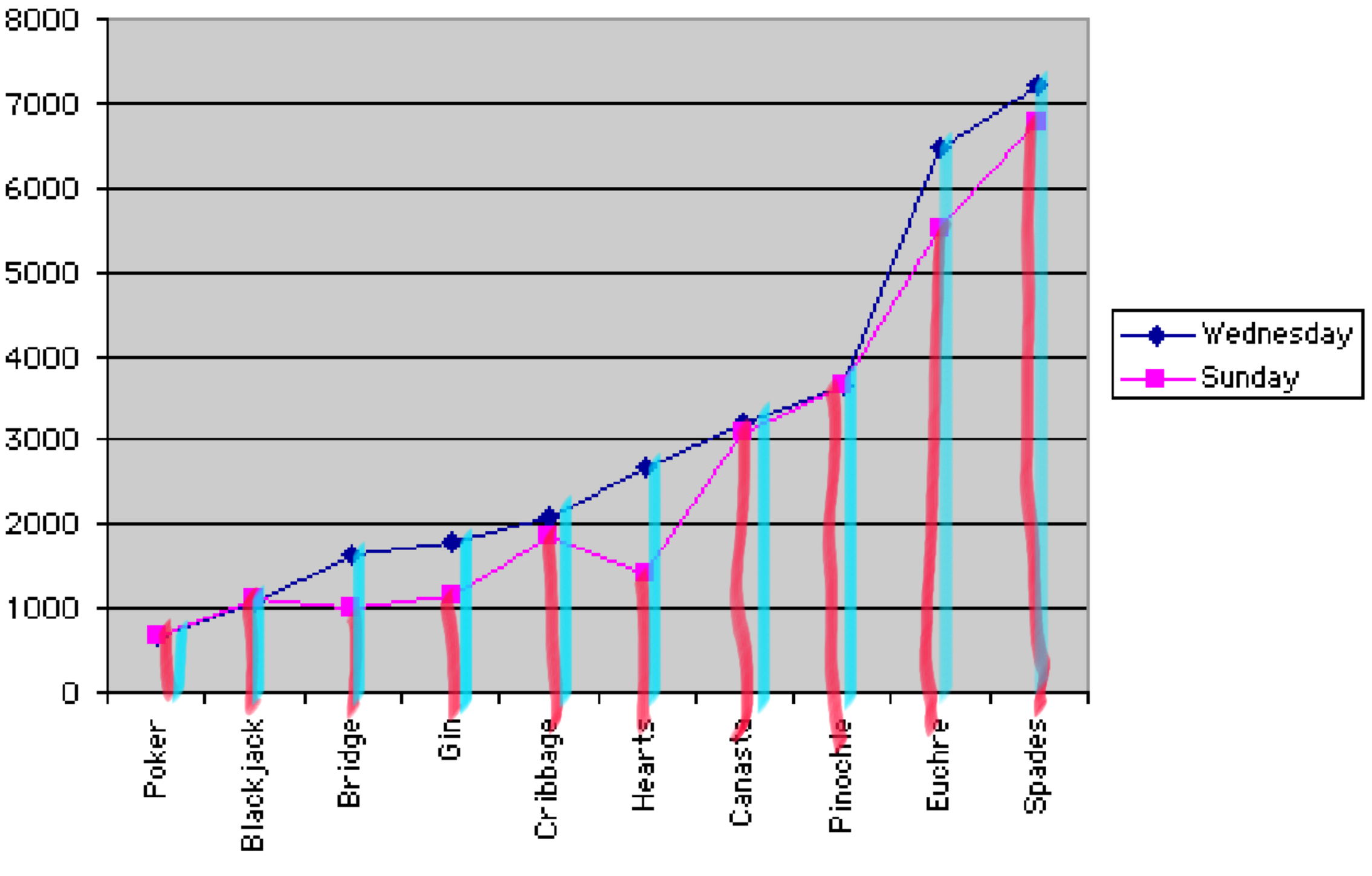
Data: Values of a variable.

Qualitative data: Values of a qualitative variable.

Quantitative data: Values of a quantitative variable.

Discrete data: Values of a discrete variable.

Continuous data: Values of a continuous variable.


8000
7000
6000
5000
4000
3000
2000
1000
0
Poker
Blackjack
Bridge
Gin
Cribbage
Hearts
Canasta
Pinochle
Euchre
Spades
Wednesday
Sunday

Human Blood Types Human beings have one of four blood types: A, B, AB, or O. What kind of data do you receive when you are told your blood type?

Household Size The U.S. Census Bureau collects data on household size and publishes the information in *Current Population Reports*. What kind of data is the number of people in your household?

The World's Highest Waterfalls The *Information Please Almanac* lists the world's highest waterfalls. The list shows that Angel Falls in Venezuela is 3281 feet high, or more than twice as high as Ribbon Falls in Yosemite, California, which is 1612 feet high. What kind of data are these heights?

Definition 2.3: Frequency Distribution of Qualitative Data

A **frequency distribution** of qualitative data is a listing of the distinct values and their frequencies.

To Construct a Frequency Distribution of Qualitative Data

Step 1 **List the distinct values of the observations in the data set in the first column of a table.**

Step 2 **For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct value.**

Step 3 **Count the tallies for each distinct value and record the totals in the third column of the table.**

Table 2.1 **Political party affiliations of the students in introductory statistics**

Democratic	Other	Democratic	Other	Democratic
Republican	Republican	Other	Other	Republican
Republican	Republican	Republican	Democratic	Republican
Republican	Democratic	Democratic	Other	Republican
Democratic	Democratic	Republican	Democratic	Democratic
Republican	Republican	Other	Other	Democratic
Republican	Democratic	Republican	Other	Other
Republican	Republican	Republican	Democratic	Republican

D → 13

R → 18

O → 9

Definition 2.4: Relative-Frequency Distribution of Qualitative Data

A **relative-frequency distribution** of qualitative data is a listing of the distinct values and their relative frequencies.

To Construct a Relative-Frequency Distribution of Qualitative Data

Step 1 **Obtain a frequency distribution of the data.**

Step 2 **Divide each frequency by the total number of observations.**

Political Party Affiliations Refer to Example 2.5. Construct a relative-frequency distribution of the political party affiliations of the students in Professor Weiss's introductory statistics class presented in Table 2.1.

D → 13/40 = 0.325 → 32.5%

R → 18/40 = 0.45 → 45%

O → 9/40 = 0.225 → 22.5%

Total 1

Definition 2.5: Pie Chart

A **pie chart** is a disk divided into wedge-shaped pieces proportional to the relative frequencies of the qualitative data.

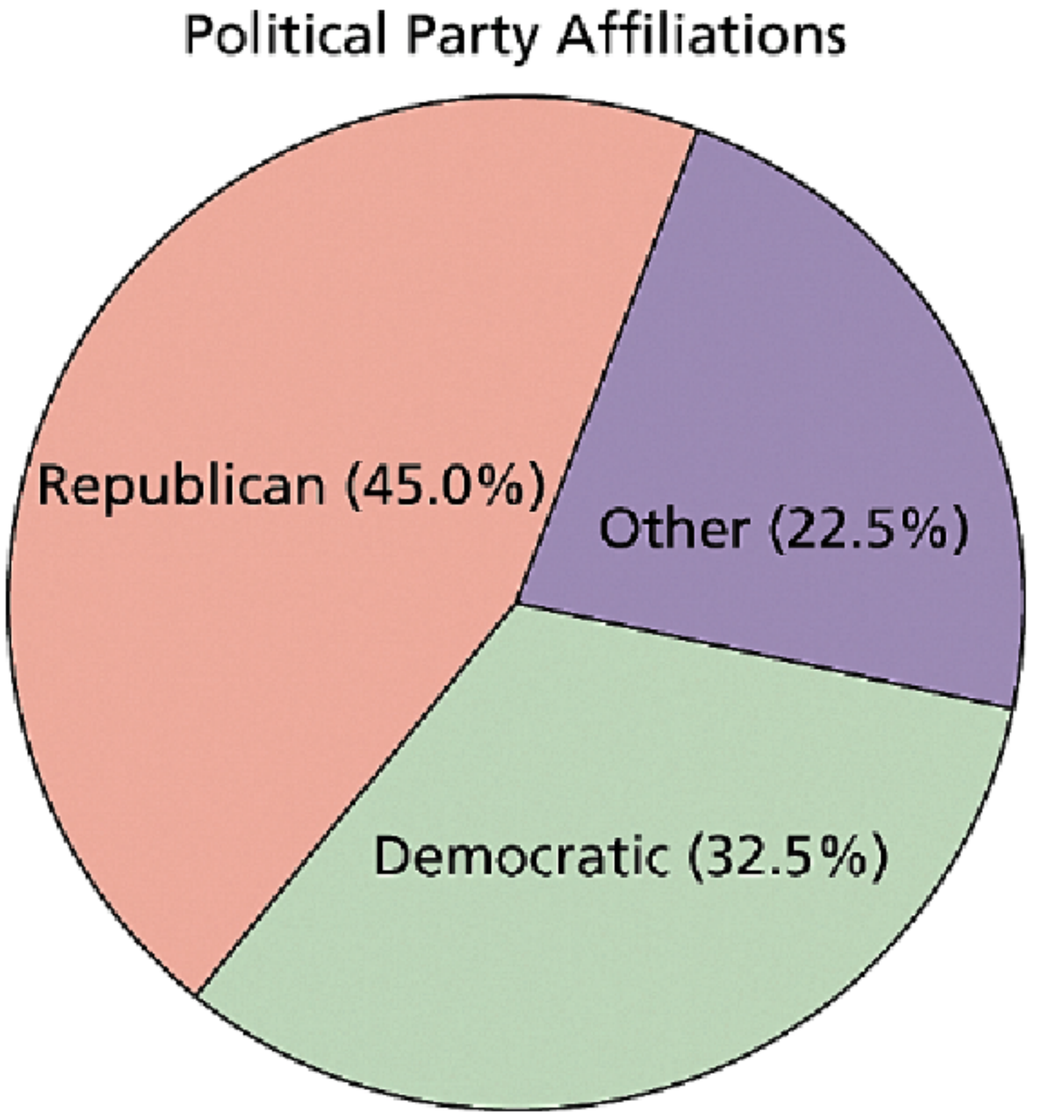
To Construct a Pie Chart

Step 1 **Obtain a relative-frequency distribution of the data by applying** Procedure 2.2.

Step 2 **Divide a disk into wedge-shaped pieces proportional to the relative frequencies.**

Step 3 **Label the slices with the distinct values and their relative frequencies.**

Political Party Affiliations Construct a pie chart of the political party affiliations of the students in Professor Weiss's introductory statistics class presented in Table 2.1.



Definition 2.6: Bar Chart

A **bar chart** displays the distinct values of the qualitative data on a horizontal axis and the relative frequencies (or frequencies or percents) of those values on a vertical axis. The relative frequency of each distinct value is represented by a vertical bar whose height is equal to the relative frequency of that value. The bars should be positioned so that they do not touch each other.

To Construct a Bar Chart

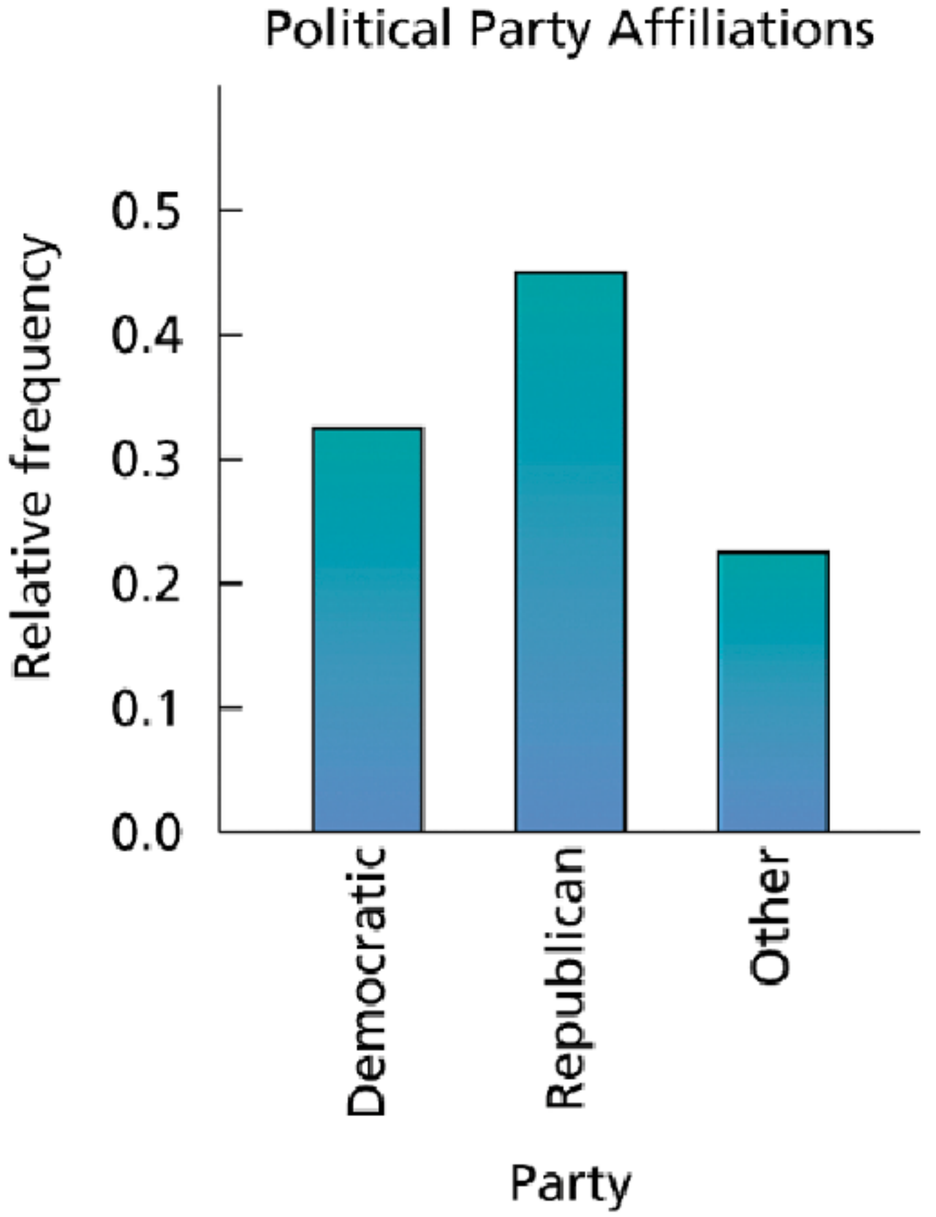
Step 1 **Obtain a relative-frequency distribution of the data by applying** Procedure 2.2.

Step 2 **Draw a horizontal axis on which to place the bars and a vertical axis on which to display the relative frequencies.**

Step 3 **For each distinct value, construct a vertical bar whose height equals the relative frequency of that value.**

Step 4 **Label the bars with the distinct values, the horizontal axis with the name of the variable, and the vertical axis with "Relative frequency."**

Political Party Affiliations Construct a bar chart of the political party affiliations of the students in Professor Weiss's introductory statistics class presented in Table 2.1.



Example 2.12: Single-Value Grouping

TVs per Household The Television Bureau of Advertising publishes information on television ownership in *Trends in Television*. Table 2.4 gives the number of TV sets per household for 50 randomly selected households. Use single-value grouping to organize these data into frequency and relative-frequency distributions.

Table 2.4 **Number of TV sets in each of 50 randomly selected households**

1	1	1	2	6	3	3	4	2	4
3	2	1	5	2	1	3	6	2	2
3	1	1	4	3	2	2	2	2	3
0	3	1	2	1	2	3	1	1	3
3	2	1	2	1	1	3	1	5	1

1 → 1 → .02
1 → 16 → .32
2 → 14 → .28
3 → 12 → .24
4 → 3 → .06
5 → 2 → .04
6 → 2 → .04

Example 2.13: Limit Grouping

Days to Maturity for Short-Term Investments Table 2.6 displays the number of days to maturity for 40 short-term investments. The data are from *BARRON'S* magazine. Use limit grouping, with grouping by 10s, to organize these data into frequency and relative-frequency distributions.

Table 2.6 **Days to maturity for 40 short-term investments**

70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	88	95	86
57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70

.075 ← 36-39 → 3

.025 ← 40-49 → 1

.20 ← 50-59 → 8

.25 ← 60-69 → 10

.175 ← 70-79 → 7

.175 ← 80-89 → 7

.10 ← 90-99 → 4

Definition 2.7: Terms Used in Limit Grouping

Lower class limit: The smallest value that could go in a class.

Upper class limit: The largest value that could go in a class.

Class width: The difference between the lower limit of a class and the lower limit of the next-higher class.

Class mark: The average of the two class limits of a class.

129.2	185.3	218.1	182.5	142.8
155.2	170.0	151.3	187.5	145.6
167.3	161.0	178.7	165.0	172.5
191.1	150.7	187.0	173.7	178.2
161.7	170.1	165.8	214.6	136.7
278.8	175.6	188.7	132.1	158.5
146.4	209.1	175.4	182.0	173.6
149.9	158.6			

Weight (lb)	Frequency	Relative frequency
120–under 140	3	0.081
140–under 160	9	0.243
160–under 180	14	0.378
180–under 200	7	0.189
200–under 220	3	0.081
220–under 240	0	0.000
240–under 260	0	0.000
260–under 280	1	0.027
	37	0.999

Definition 2.8: Terms Used in Cutpoint Grouping

Lower class cutpoint: The smallest value that could go in a class.

Upper class cutpoint: The smallest value that could go in the next-higher class (equivalent to the lower cutpoint of the next-higher class).

Class width: The difference between the cutpoints of a class.

Class midpoint: The average of the two cutpoints of a class.

Choosing the Grouping Method

We have now discussed three methods for grouping quantitative data: single-value grouping, limit grouping, and cutpoint grouping. The following table provides guidelines for deciding which grouping method should be used.

Grouping method	When to use
Single-value grouping	Use with discrete data in which there are only a small number of distinct values.
Limit grouping	Use when the data are expressed as whole numbers and there are too many distinct values to employ single-value grouping.
Cutpoint grouping	Use when the data are continuous and are expressed with decimals.

A **histogram** displays the classes of the quantitative data on a horizontal axis and the frequencies (relative frequencies, percents) of those classes on a vertical axis. The frequency (relative frequency, percent) of each class is represented by a vertical bar whose height is equal to the frequency (relative frequency, percent) of that class. The bars should be positioned so that they touch each other.

- For single-value grouping, we use the distinct values of the observations to label the bars, with each such value centered under its bar.
- For limit grouping or cutpoint grouping, we use the lower class limits (or, equivalently, lower class cutpoints) to label the bars. *Note:* Some statisticians and technologies use class marks or class midpoints centered under the bars

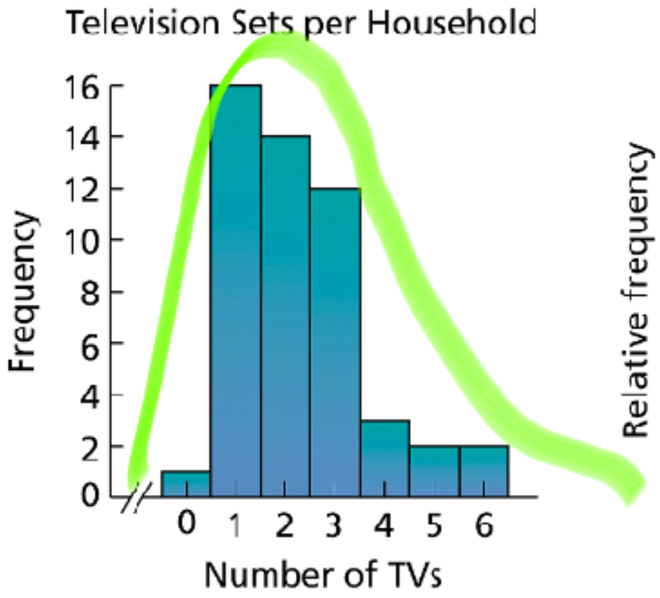
To Construct a Histogram

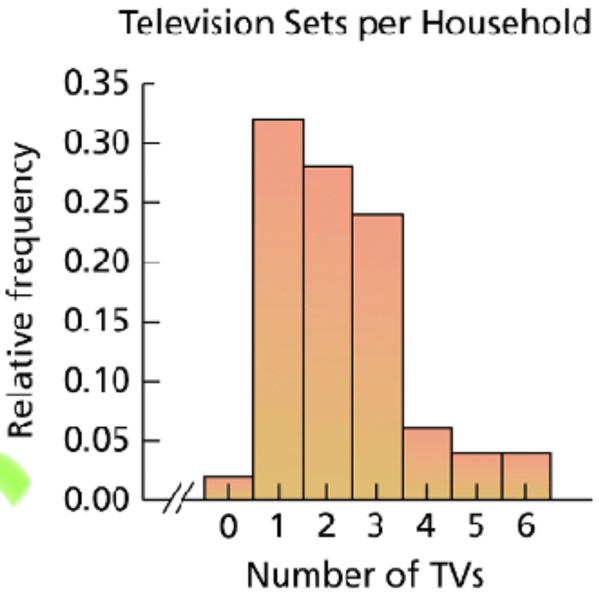
Step 1 **Obtain a frequency (relative-frequency, percent) distribution of the data.**

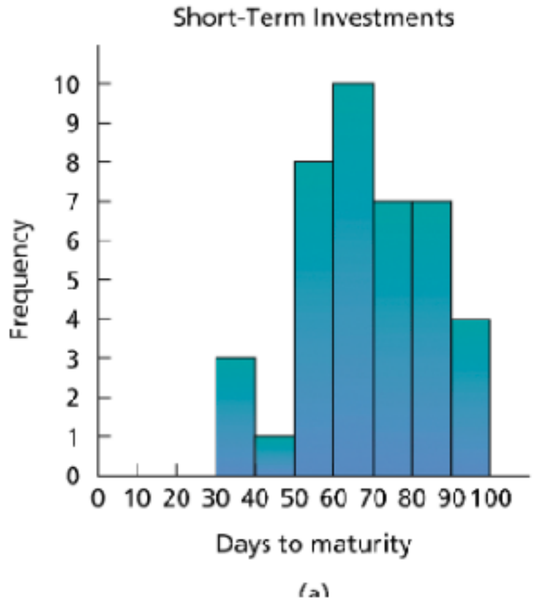
Step 2 **Draw a horizontal axis on which to place the bars and a vertical axis on which to display the frequencies (relative frequencies, percents).**

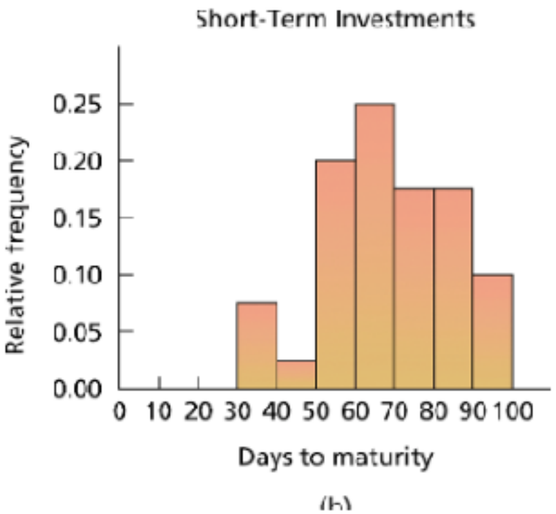
Step 3 **For each class, construct a vertical bar whose height equals the frequency (relative frequency, percent) of that class.**

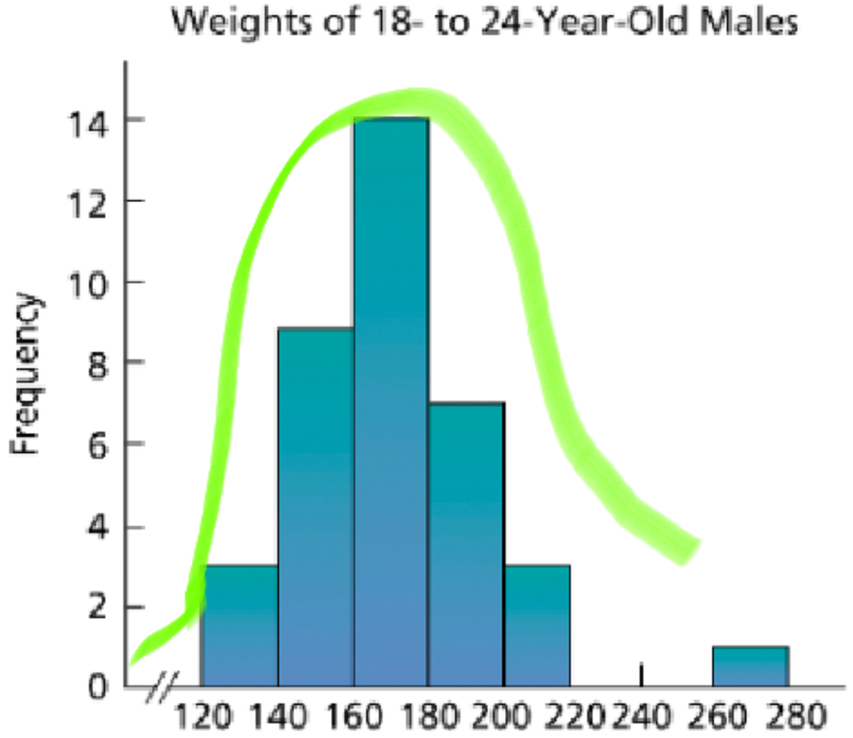
Step 4 **Label the bars with the classes, as explained in** Definition 2.9, **the horizontal axis with the name of the variable, and the vertical axis with "Frequency" ("Relative frequency," "Percent").**

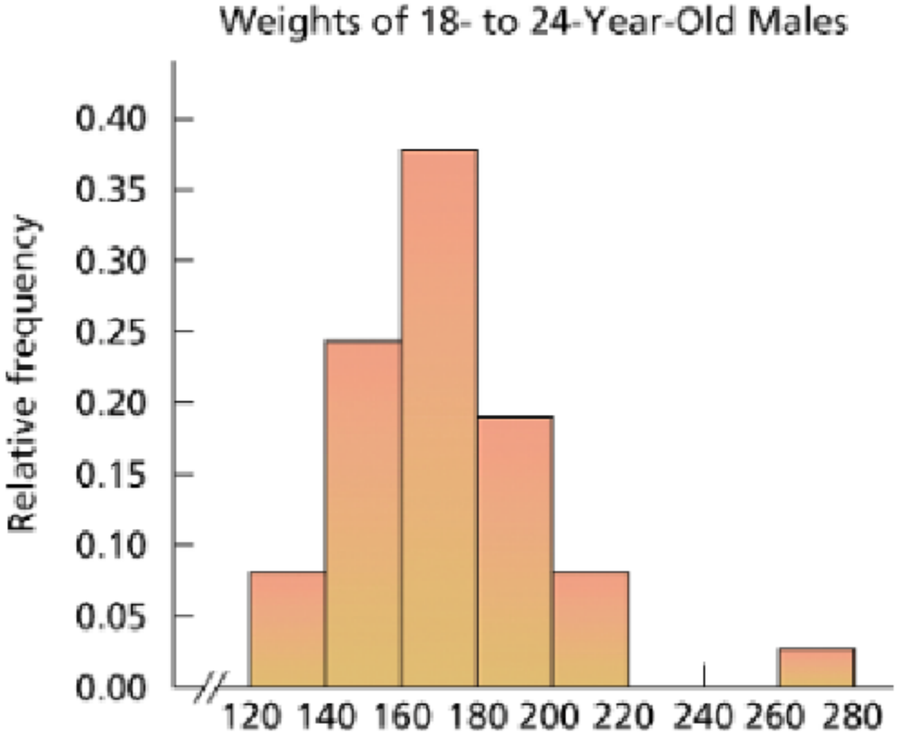

Television Sets per Household
Frequency
Number of TVs


Television Sets per Household
Relative frequency
Number of TVs


Short-Term Investments
Frequency
Days to maturity


Short-Term Investments
Relative frequency
Days to maturity


Weights of 18- to 24-Year-Old Males
Frequency


Weights of 18- to 24-Year-Old Males
Relative frequency

Definition 2.10: Dotplot

A **dotplot** is a graph in which each observation is plotted as a dot at an appropriate place above a horizontal axis. Observations having equal values are stacked vertically.

To Construct a Dotplot

Step 1 **Draw a horizontal axis that displays the possible values of the quantitative data.**

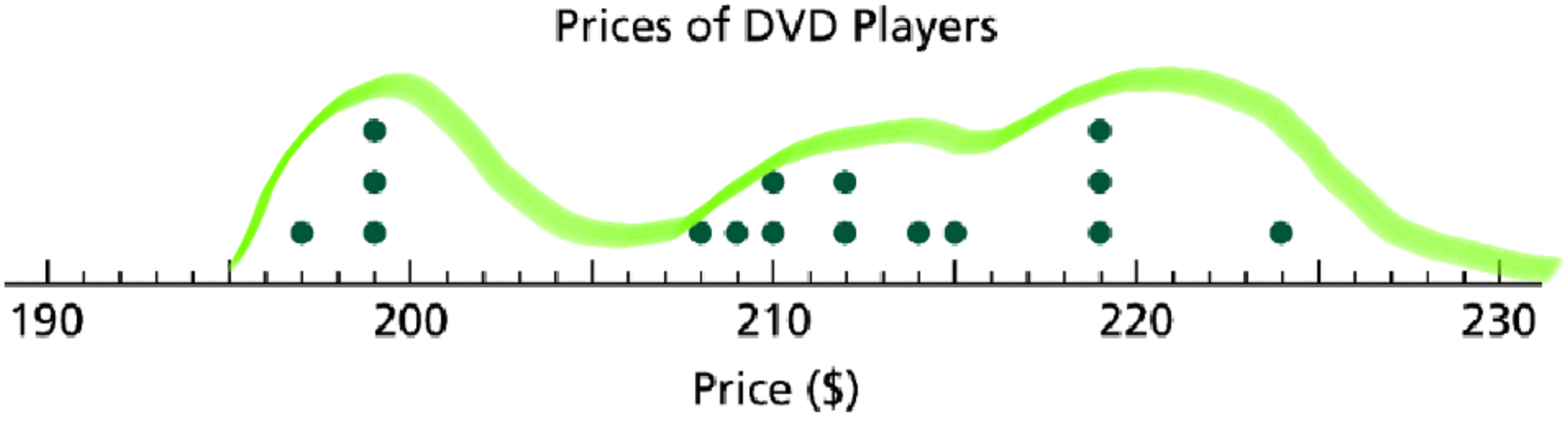
Step 2 **Record each observation by placing a dot over the appropriate value on the horizontal axis.**

Step 3 **Label the horizontal axis with the name of the variable.**

Prices of DVD Players One of Professor Weiss's sons wanted to add a new DVD player to his home theater system. He used the Internet to shop and went to pricewatch.com. There he found 16 quotes on different brands and styles of DVD players. Table 2.11 lists the prices, in dollars. Construct a dotplot for these data.

Table 2.11 Prices, in dollars, of 16 DVD players

210	219	214	197
224	219	199	199
208	209	215	199
212	212	219	210



Definition 2.11: Stem-and-Leaf Diagram

In a **stem-and-leaf diagram** (or **stemplot**), each observation is separated into two parts, namely, a **stem**—consisting of all but the rightmost digit—and a **leaf**, the rightmost digit.

To Construct a Stem-and-Leaf Diagram

Step 1 Think of each observation as a stem—consisting of all but the rightmost digit—and a leaf, the rightmost digit.

Step 2 Write the stems from smallest to largest in a vertical column to the left of a vertical rule.

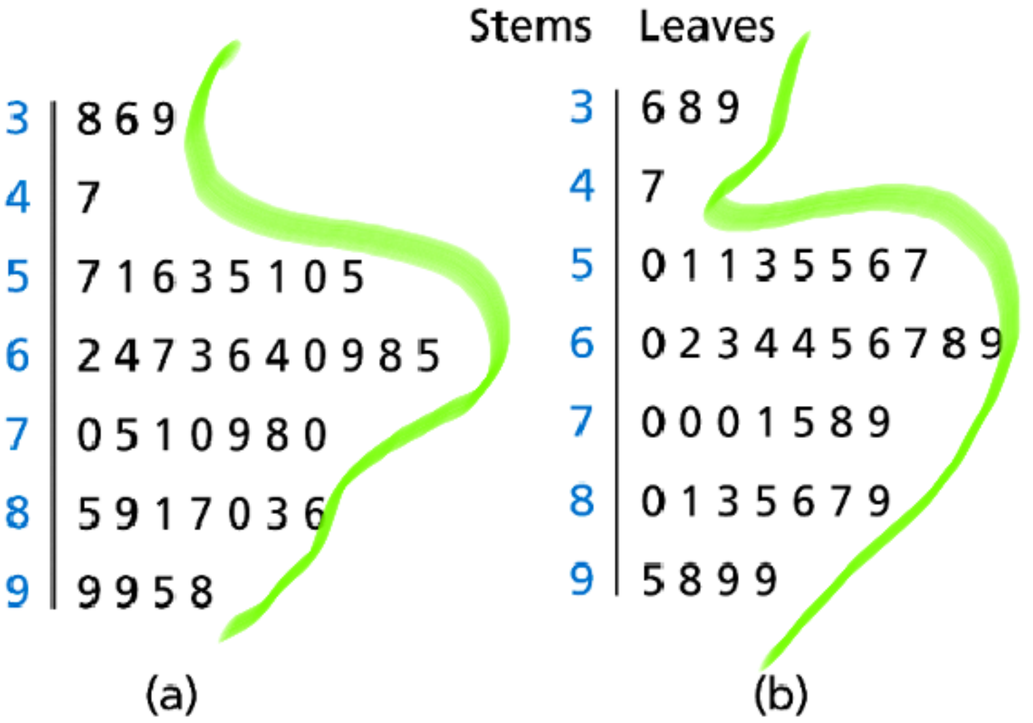
Step 3 Write each leaf to the right of the vertical rule in the row that contains the appropriate stem.

Step 4 Arrange the leaves in each row in ascending order.

Days to Maturity for Short-Term Investments Table 2.12 repeats the data on the number of days to maturity for 40 short-term investments. Previously, we grouped these data with a frequency distribution (Table 2.7) and graphed them with a frequency histogram (Fig. 2.5(a)). Now let's construct a stem and leaf diagram, which simultaneously groups the data and provides a graphical display similar to a histogram.

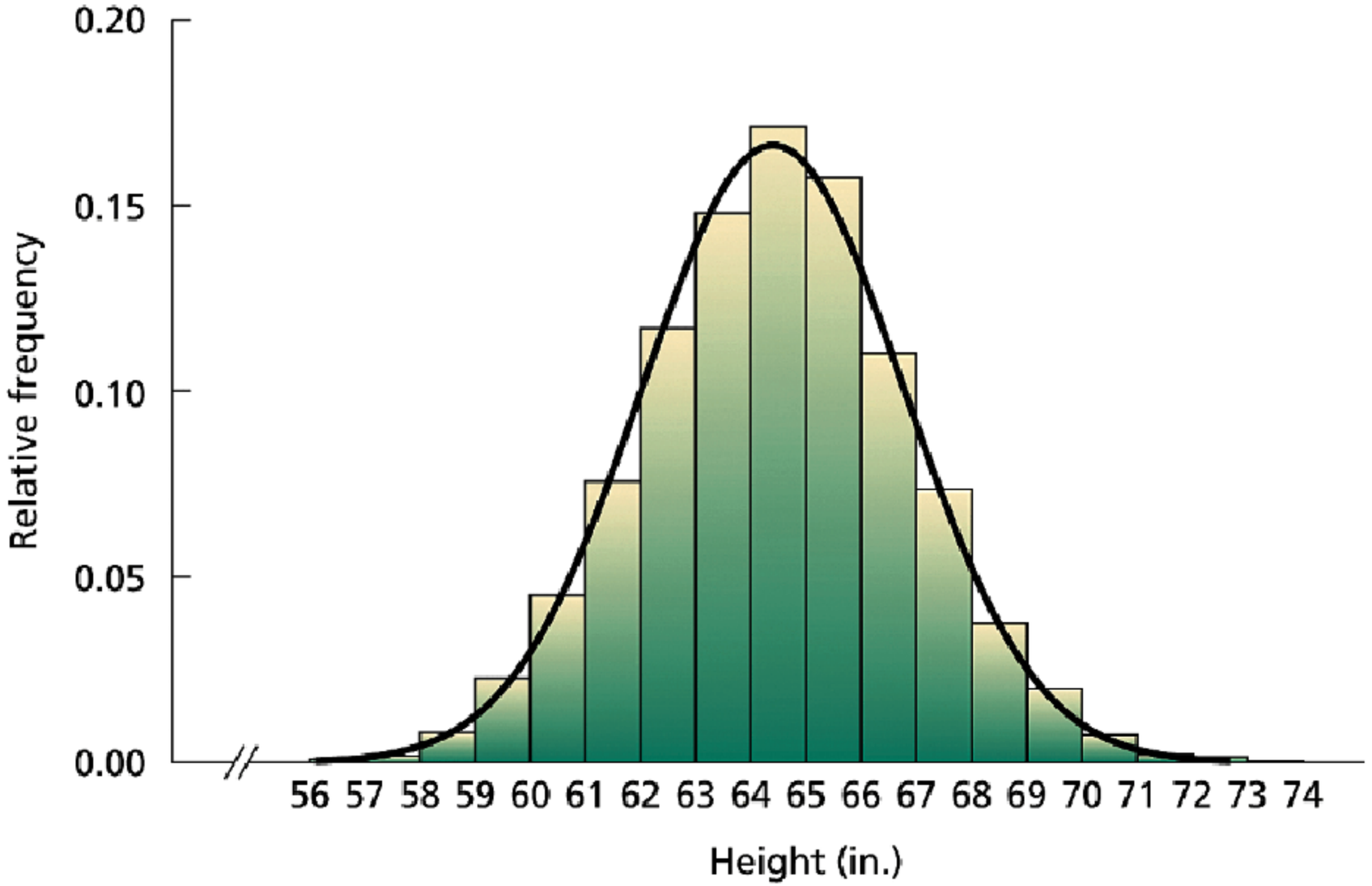
Table 2.12 Days to maturity for 40 short-term investments

70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86
57	53	47	50	55	81	80	98
51	36	83	66	85	79	83	70



Definition 2.12: Distribution of a Data Set

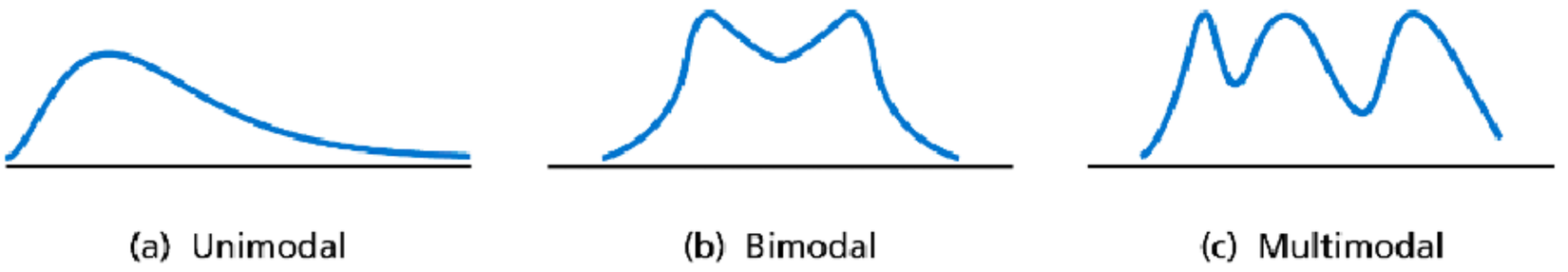
The **distribution of a data set** is a table, graph, or formula that provides the values of the observations and how often they occur.



Modality

When considering the shape of a distribution, you should observe its number of peaks (highest points). A distribution is **unimodal** if it has one peak, **bimodal** if it has two peaks, and **multimodal** if it has three or more peaks. Figure 2.11 shows examples of unimodal, bimodal, and multimodal distributions.

Figure 2.11



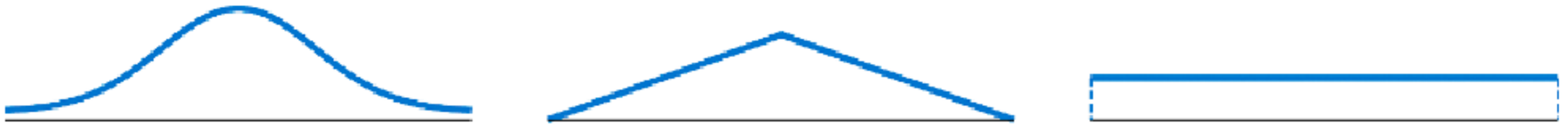
Examples of (a) unimodal, (b) bimodal, and (c) multimodal distributions

Symmetry

Another important consideration when examining the shape of a distribution is *symmetry*. A distribution that can be divided into two pieces that are mirror images of one another is called **symmetric.** The distributions in Figs. 2.10 and 2.11(b) are symmetric.

The three distributions in Fig. 2.12, called **bell shaped**, **triangular**, and **uniform** (or **rectangular**), are specific categories of symmetric distributions. Note that the distribution of heights in Fig. 2.10 is bell shaped.

Figure 2.12



Examples of symmetric distributions: (a) bell shaped, (b) triangular, and (c) uniform

Skewness

A unimodal distribution that is not symmetric is either *right skewed* or *left skewed*. A **right-skewed** distribution rises to its peak rapidly and comes back toward the horizontal axis more slowly—its "right tail" is longer than its "left tail." A **left-skewed** distribution rises to its peak slowly and comes back toward the horizontal axis more rapidly—its "left tail" is longer than its "right tail." Figure 2.13 shows generic right-skewed and left-skewed distributions.

Figure 2.13

Generic skewed distributions: (a) right skewed and (b) left skewed

Definition 2.13: Population and Sample Data

Population data: The values of a variable for the entire population.

Sample data: The values of a variable for a sample of the population.

Definition 2.14: Population and Sample Distributions; Distribution of a Variable

The distribution of population data is called the **population distribution**, or the **distribution of the variable.**

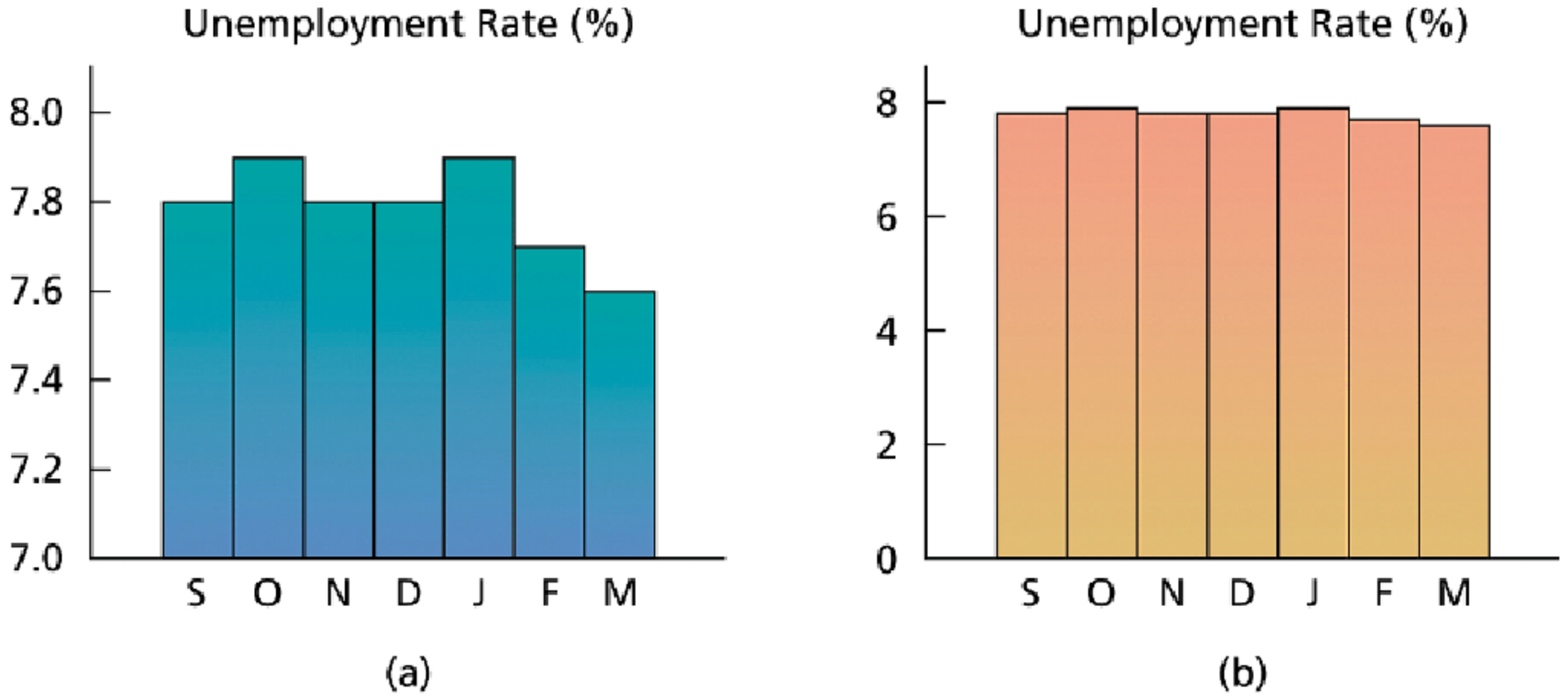
The distribution of sample data is called a **sample distribution.**

2.5 Misleading Graphs

Graphs and charts are frequently misleading, sometimes intentionally and sometimes inadvertently. Regardless of intent, we need to read and interpret graphs and charts with a great deal of care. In this section, we examine some misleading graphs and charts.

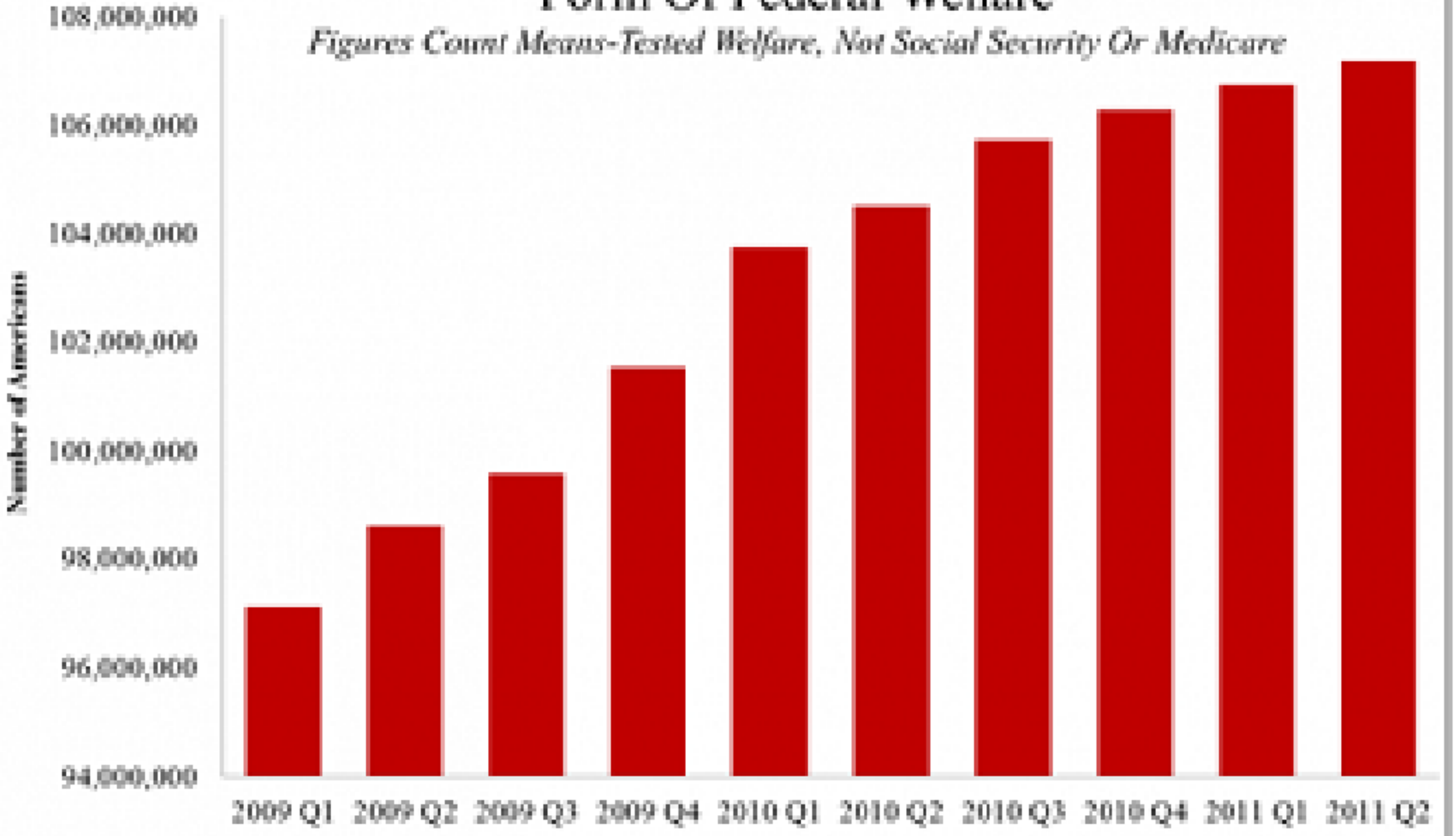
Unemployment Rates The Bureau of Labor Statistics collects data on unemployment rates and publishes its finding in *Labor Force Statistics*. Figure 2.17(a) shows a bar chart from an article in a major metropolitan newspaper. The graph displays the unemployment rates in the United States from September of one year through March of the next year.

Figure 2.17



Over 100 Million People In U.S. Now Receiving Some Form Of Federal Welfare

Figures Count Means-Tested Welfare, Not Social Security Or Medicare

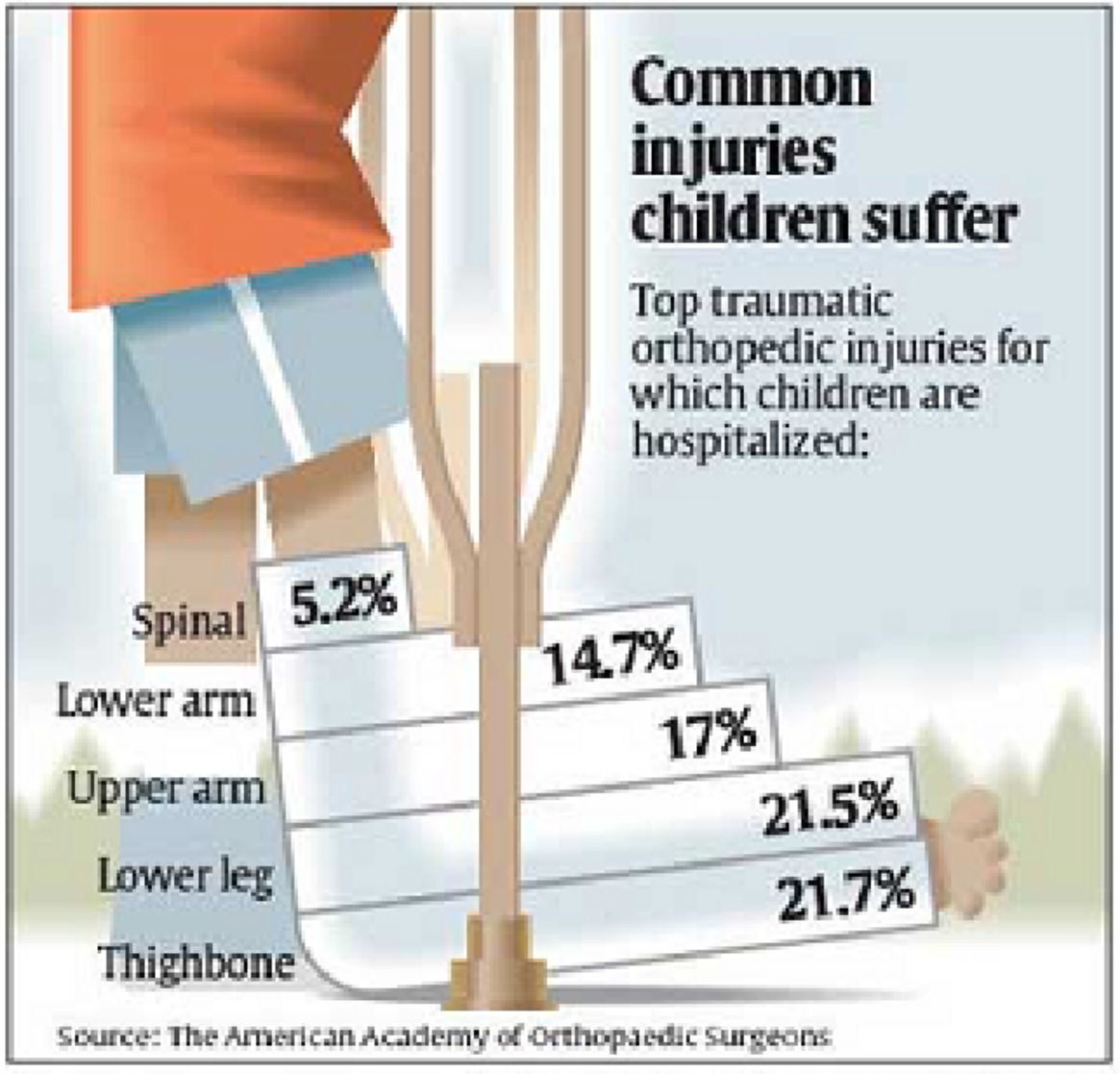


Figures include anyone residing in a household in which at least one person received a program benefit.

Source: U.S. Census' Survey of Income and Program Participation.

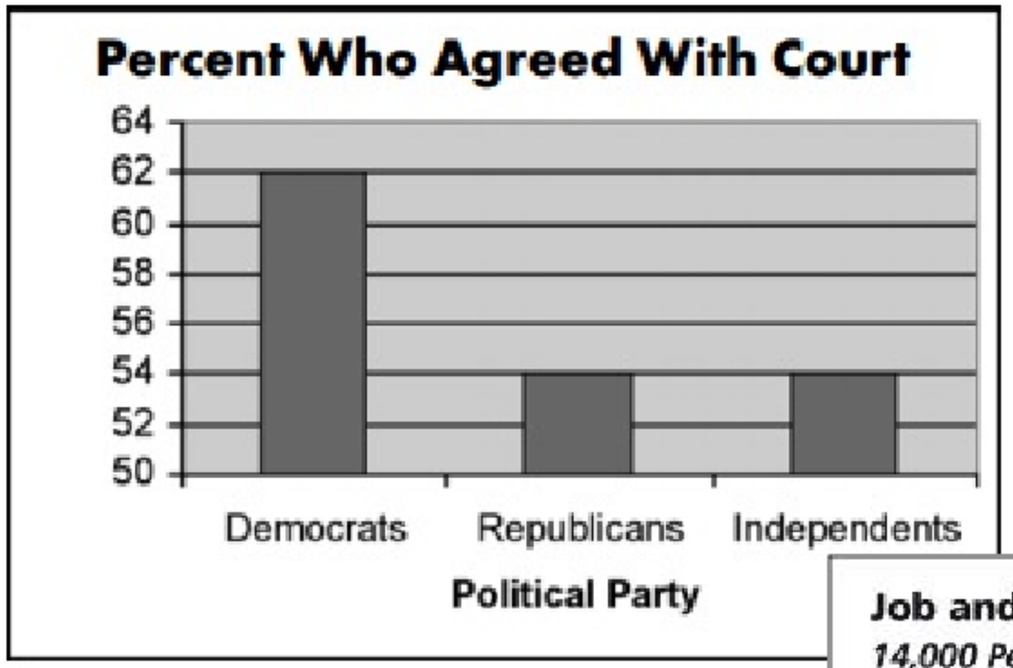
Produced by Senate Budget Committee Republican staff, Ranking Member Jeff Sessions || http://budget.senate.gov/republican

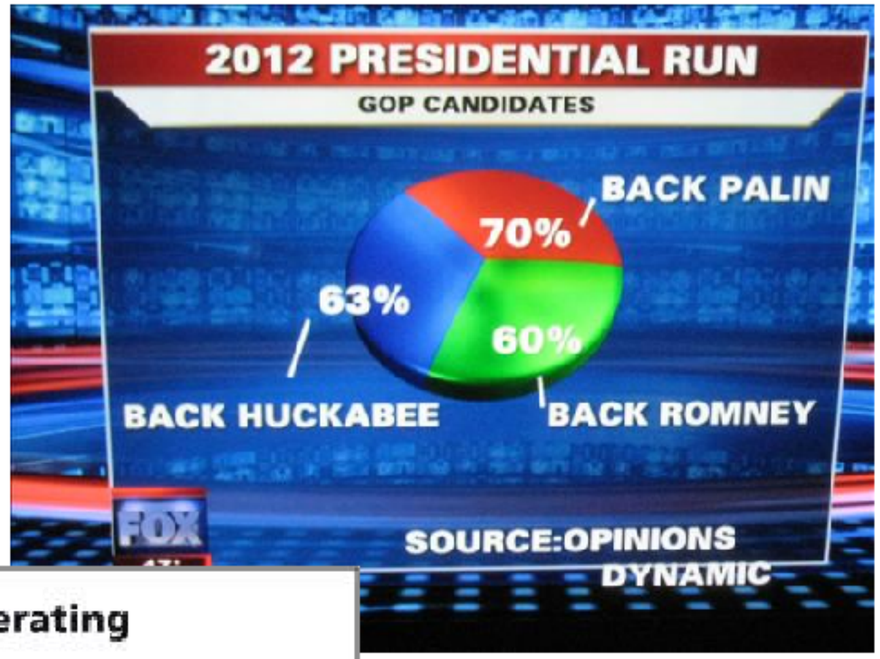
USA TODAY Snapshots™

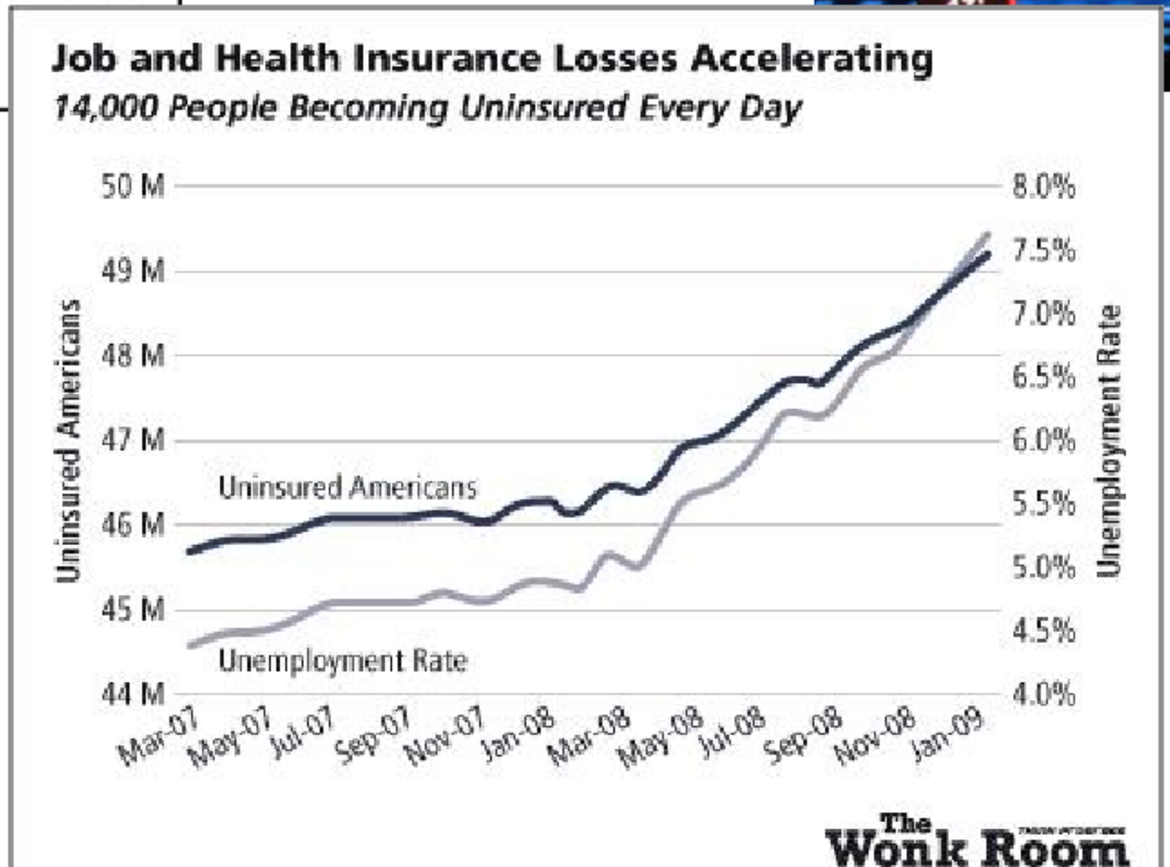


By Shannon Reilly and Frank Pompa, USA TODAY

For the record, the real figure is about .0000003% (based on 2000 injuries per year out of a population of around 74,000,000).


Percent Who Agreed With Court
64
62
60
58
56
54
52
50
Democrats
Republicans
Independents
Political Party


2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
FOX
SOURCE:OPINIONS DYNAMIC


Job and Health Insurance Losses Accelerating
14,000 People Becoming Uninsured Every Day
50 M
49 M
48 M
47 M
46 M
45 M
44 M
Uninsured Americans
8.0%
7.5%
7.0%
6.5%
6.0%
5.5%
5.0%
4.5%
4.0%
Unemployment Rate
Uninsured Americans
Unemployment Rate
Mar-07
May-07
Jul-07
Sep-07
Nov-07
Jan-08
Mar-08
May-08
Jul-08
Sep-08
Nov-08
Jan-09
The Wonk Room