# Chapter 2 – Organizing and Summarizing Data

## OUTLINE

* 1. Organizing Qualitative Data
  2. Organizing Quantitative Data: The Popular Displays
  3. Additional Displays of Quantitative Data
  4. Graphical Misrepresentations of Data

## Putting It Together

Chapter 1 discussed how to identify the research objective and collect data. We learned that data can be obtained from either observational studies or designed experiments. When data are obtained, they are referred to as **raw data**.

The purpose of this chapter is to learn how to organize raw data into a meaningful form so that we can understand what the data are telling us. The first step in determining how to organize raw data is to determine whether the data is qualitative or quantitative.

## Section 2.1 Organizing Qualitative Data

### Objectives

1. Organize Qualitative Data in Tables
2. Construct Bar Graphs
3. Construct Pie Charts

#### Objective 1: Organize Qualitative Data in Tables

Objective 1, Page 1

1. What is used to list each category of data and the number of occurrences for each category of data?

Objective 1, Page 2

**Example 1 *Organizing Qualitative Data into a Frequency Distribution***

A physical therapist wants to determine types of rehabilitation required by her patients. To do so, she obtains a simple random sample of 30 of her patients and records the body part requiring rehabilitation. (See Table 1.) Construct a frequency distribution of location of injury.

**Table 1**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Back | Back | Hand | Wrist | Back | Back |
| Groin | Elbow | Back | Back | Back | Groin |
| Shoulder | Shoulder | Hip | Knee | Hip | Shoulder |
| Neck | Knee | Knee | Shoulder | Shoulder | Neck |
| Back | Back | Back | Back | Knee | Back |

Data from Krystal Catton, student at Joliet Junior College

OBJECTIVE 1, PAGE 3

1. In any frequency distribution, it is a good idea to add up the frequency column. What should the total be equal to?

Objective 1, Page 6

1. Define the relative frequency of a category.
2. What is a relative frequency distribution?

Objective 1, Page 7

**Example 2 *Constructing a Relative Frequency Distribution of Qualitative Data***

Using the summarized data in Table 2, construct a relative frequency distribution.

**Table 2**

| **Body Part** | **Frequency** |
| --- | --- |
| Back | 12 |
| Hand | 2 |
| Wrist | 2 |
| Groin | 1 |
| Elbow | 1 |
| Shoulder | 4 |
| Hip | 2 |
| Knee | 5 |
| Neck | 1 |

Objective 1, Page 8

1. When working with a relative frequency distribution, what should the total of the relative frequencies be equal to? Why?

#### Objective 2: Construct Bar Graphs

Objective 2, Page 1

1. Explain how a bar graph is constructed. What do the heights of each rectangle represent?

Objective 2, Page 2

**Example 3 *Constructing a Frequency and Relative Frequency Bar Graph***

Use the data summarized in Table 3 to construct a frequency bar graph and relative frequency bar graph.

**Table 3**

| **Body Part** | **Frequency** | **Relative Frequency** |
| --- | --- | --- |
| Back | 12 | 0.4 |
| Hand | 2 | 0.0667 |
| Wrist | 2 | 0.0667 |
| Groin | 1 | 0.0333 |
| Elbow | 1 | 0.0333 |
| Shoulder | 4 | 0.1333 |
| Hip | 2 | 0.0667 |
| Knee | 5 | 0.1667 |
| Neck | 1 | 0.0333 |

Objective 2, Page 4

1. What is a Pareto chart?

Objective 2, Page 5

1. Explain why it is best to use relative frequencies when comparing data sets.

Objective 2, Page 6

**Example 4 *Comparing Two Data Sets***

The frequency data in Table 4 represent the educational attainment (level of education) in 1990 and 2016 of adults 25 years and older who are U.S. residents. The data are in thousands. So 39,344 represents 39,344,000.

**Table 4**

| **Educational Attainment** | **1990** | **2016** |
| --- | --- | --- |
| Not a high school graduate | 39,344 | 23,453 |
| High school diploma | 47,643 | 62,002 |
| Some college, no degree | 29,780 | 36,003 |
| Associate’s degree | 9792 | 21,657 |
| Bachelor’s degree | 20,833 | 44,778 |
| Graduate or professional degree | 11,478 | 27,122 |
| Totals | 158,870 | 215,015 |

1. Draw a side-by-side relative frequency bar graph of the data.

OBJECTIVE 2, PAGE 6 (CONTINUED)

1. The side-by-side relative frequency bar graph shows additional information that was not easy to identify from the frequency table in Table 4. Comment on the interesting features of the side-by-side relative frequency bar graph.

Objective 2, Page 8

1. Explain when it would be preferable to use horizontal bars rather than vertical bars when constructing a bar graph.

#### Objective 3: Construct Pie Charts

Objective 3, Page 1

1. What is a pie chart?

OBJECTIVE 3, PAGE 2

**Example 5 *Constructing a Pie Chart***

The frequency data presented in Table 6 represent the educational attainment of U.S. residents 25 years and older in 2016. The data are in thousands so 23,453 represents 23,453,000. Construct a pie chart of the data.

**Table.6**

| **Educational Attainment** | **2016** |
| --- | --- |
| Not a high school graduate | 23,453 |
| Some college, no degree | 36,003 |
| Associate's degree | 21,657 |
| Bachelor's degree | 44,778 |
| Graduate or professional degree | 27,122 |
| Totals | 215,015 |

Objective 3, Page 5

*Answer the following after watching the video.*

1. Which graph, a pie chart or a bar graph, is better at comparing one category to another category?
2. Which graph, a pie chart or a bar graph, is better at comparing one category to the whole?

## Section 2.2 Organizing Quantitative Data: The Popular Displays

### Objectives

1. Organize Discrete Data in Tables
2. Construct Histograms of Discrete Data
3. Organize Continuous Data in Tables
4. Construct Histograms of Continuous Data
5. Draw Dot Plots
6. Identify the Shape of a Distribution

#### Objective 1: Organize Discrete Data in Tables

Objective 1, Page 1

1. What do we use to create the classes when the number of distinct data values of a discrete variable is small?

Objective 1, Page 2

**Example 1 *Constructing Frequency and Relative Frequency Distributions from Discrete Data***

The manager of a Wendy’s® fast-food restaurant wants to know the typical number of customers who arrive during the lunch hour. The data represent the number of customers who arrive at Wendy’s for 40 randomly selected 15-minute intervals of time during lunch. Construct a frequency and relative frequency distribution.

**Number of Arrivals at Wendy’s**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 7 | 6 | 6 | 6 | 4 | 6 | 2 | 6 |
| 5 | 6 | 6 | 11 | 4 | 5 | 7 | 6 |
| 2 | 7 | 1 | 2 | 4 | 8 | 2 | 6 |
| 6 | 5 | 5 | 3 | 7 | 5 | 4 | 6 |
| 2 | 2 | 9 | 7 | 5 | 9 | 8 | 5 |

#### Objective 2: Construct Histograms of Discrete Data

Objective 2, Page 1

1. Explain how a histogram is constructed.

Objective 2, Page 2

**Example 2 *Drawing a Histogram of Discrete Data***

Construct a frequency histogram and a relative frequency histogram using the data in Table 9. Recall that this table summarizes the data for the number of customers who arrive at Wendy's for 40 randomly selected 15-minute intervals of time during lunch.

**Table 9**

| **Number of Customers** | **Frequency** | **Relative Frequency** |
| --- | --- | --- |
| 1 | 1 | 0.25 |
| 2 | 6 | 0.15 |
| 3 | 1 | 0.025 |
| 4 | 4 | 0.1 |
| 5 | 7 | 0.175 |
| 6 | 11 | 0.275 |
| 7 | 5 | 0.125 |
| 8 | 2 | 0.05 |
| 9 | 2 | 0.05 |
| 10 | 0 | 0.0 |
| 11 | 1 | 0.025 |

OBJECTIVE 2, PAGE 2 (CONTINUED)

#### Objective 3: Organize Continuous Data in Tables

Objective 3, Page 1

**Note:** When a data set consists of a large number of different discrete data values or when a data set consists of continuous data, we must create classes by using intervals of numbers.

Define the following terms.

1. Lower class limit
2. Upper class limit
3. Class width
4. When creating classes for a frequency distribution, the classes must not \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

Objective 3, Page 2

1. What is an open-ended table?

Objective 3, Page 3

**Example 3 *Organizing Continuous Data into a Frequency and Relative Frequency Distribution***

Suppose you are considering investing in a Roth IRA. You collect the data in Table 12, which represent the five-year rate of return (in percent, adjusted for sales charges) for a simple random sample of 40 large-blend mutual funds. Construct a frequency and relative frequency distribution of the data.

**Table 12 Five-Year Rate of Return of Mutual Funds (in percent)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 10.94 | 14.60 | 12.80 | 16.00 | 11.93 | 15.68 | 9.03 | 13.40 |
| 10.53 | 13.98 | 13.86 | 12.36 | 13.54 | 9.94 | 13.94 | 13.63 |
| 14.12 | 14.88 | 14.77 | 13.13 | 8.28 | 19.43 | 12.98 | 13.16 |
| 12.26 | 14.20 | 14.80 | 13.26 | 13.67 | 10.08 | 14.86 | 8.71 |
| 12.17 | 10.26 | 15.22 | 13.26 | 13.55 | 13.90 | 15.64 | 12.80 |

Data from [Morningstar.com](http://morningstar.com/)

#### Objective 4: Construct Histograms of Continuous Data

Objective 4, Page 2

**Example 4 *Drawing a Histogram of Continuous Data***

Construct a frequency and relative frequency histogram of the five-year rate of return data discussed in Example 3.

**Table 12 Five-Year Rate of Return of Mutual Funds (in percent)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 10.94 | 14.60 | 12.80 | 16.00 | 11.93 | 15.68 | 09.03 | 13.40 |
| 10.53 | 13.98 | 13.86 | 12.36 | 13.54 | 09.94 | 13.93 | 13.63 |
| 14.12 | 14.88 | 14.77 | 13.13 | 08.28 | 19.43 | 12.98 | 13.16 |
| 12.26 | 14.20 | 14.80 | 13.26 | 13.67 | 10.08 | 14.86 | 8.71 |
| 12.17 | 10.26 | 15.22 | 13.36 | 13.55 | 13.90 | 15.64 | 12.80 |

Data from [Morningstar.com](http://morningstar.com/)

Objective 4, Page 4

There is no one correct frequency distribution for a particular set of data. However, some frequency distributions better illustrate patterns within the data than others. So constructing frequency distributions is somewhat of an art form. Use the distribution that seems to provide the best overall summary of the data.

Objective 4, Page 5

*Answer the following after using the applet in Activity 1: Choosing Class Width.*

1. What happens to the number of classes as the bin width increases?

Objective 4, Page 7

1. The number of classes in a frequency distribution is typically between what two numbers?
2. Explain how to choose the lower class limit of the first class in a frequency distribution.
3. Once you decide on the number of classes, explain how to determine the class width.

#### Objective 5: Draw Dot Plots

Objective 5, Page 1

1. Explain how to draw a dot plot.

Objective 5, Page 2

**Example 5 *Drawing a Dot Plot***

Draw a dot plot for the data from Table 8.

**Table 8 Number of Arrivals at Wendy’s**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 7 | 6 | 6 | 6 | 4 | 6 | 2 | 6 |
| 5 | 6 | 6 | 11 | 4 | 5 | 7 | 6 |
| 2 | 7 | 1 | 2 | 4 | 8 | 2 | 6 |
| 6 | 5 | 5 | 3 | 7 | 5 | 4 | 6 |
| 2 | 2 | 9 | 7 | 5 | 9 | 8 | 5 |

#### Objective 6: Identify the Shape of a Distribution

Objective 6, Page 1

1. Draw an example of a uniform distribution.
2. Draw an example of a bell-shaped distribution.
3. Draw an example of a distribution that is skewed right.

Objective 6, Page 1 (continued)

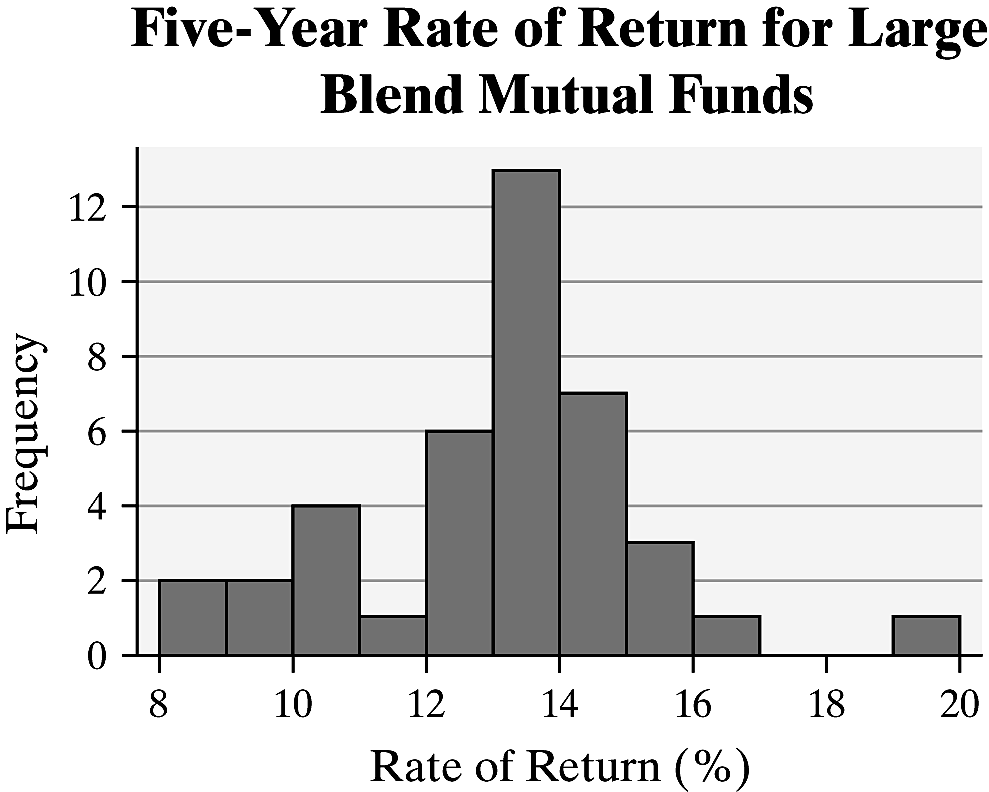
1. Draw an example of a distribution that is skewed left.

Objective 6, Page 2

**Example 6 *Identifying the Shape of a Distribution***

Figure 10 displays the histogram obtained in Example 4 for the five-year rate of return for large-blended mutual funds. Describe the shape of the distribution.

**Figure 10**



## Section 2.3 Additional Displays of Quantitative Data

### Objectives

* 1. Draw Stem-and-Leaf Plots
  2. Construct Frequency Polygons
  3. Create Cumulative Frequency and Relative Frequency Distributions
  4. Construct Frequency and Relative Frequency Ogives
  5. Draw Time-Series Graphs

#### Objective 1: Draw Stem-and-Leaf Plots

Objective 1, Page 1

1. In a stem-and-leaf plot, how are the stem and leaf identified?

Objective 1, Page 2

**Example 1 *Constructing a Stem-and-Leaf Plot***

The data in Table 14 represent the percentage of persons living in poverty, by state, in 2015. Draw a stem-and-leaf plot of the data.

Table 14 Percentage of People in Poverty by State Using Two-Year Averages: 2011–2012

| **State** | **Percent** |
| --- | --- |
| Alabama | 16.3 |
| Alaska | 9.2 |
| Arizona | 17.2 |
| Arkansas | 16.1 |
| California | 13.9 |
| Colorado | 9.9 |
| Connecticut | 9.1 |
| Delaware | 11.1 |
| District of Columbia | 16.6 |
| Florida | 16.2 |
| Georgia | 18.1 |
| Hawaii | 10.9 |
| Idaho | 12.3 |
| Illinois | 10.9 |
| Indiana | 13.5 |
| Iowa | 10.4 |
| Kansas | 14.2 |
| Kentucky | 19.5 |
| Louisiana | 18.6 |
| Maine | 12.3 |
| Maryland | 9.6 |
| Massachusetts | 11.5 |
| Michigan | 12.8 |
| Minnesota | 7.8 |
| Mississippi | 19.1 |
| Missouri | 9.8 |
| Montana | 11.9 |
| Nebraska | 10.3 |
| Nevada | 13.0 |
| New Hampshire | 7.3 |
| New Jersey | 11.2 |
| New Mexico | 19.7 |
| New York | 14.2 |
| North Carolina | 15.3 |
| North Dakota | 10.7 |
| Ohio | 13.6 |
| Oklahoma | 14.2 |
| Oregon | 11.9 |
| Pennsylvania | 12.3 |
| Rhode Island | 11.8 |
| South Carolina | 14.3 |
| South Dakota | 13.9 |
| Tennessee | 14.7 |
| Texas | 14.7 |
| Utah | 9.3 |
| Vermont | 10.7 |
| Virginia | 10.9 |
| Washington | 11.4 |
| West Virginia | 14.5 |
| Wyoming | 9.8 |

Data from united States Census Bureau

Objective 1, Page 3

1. List the four steps for constructing a stem-and-leaf plot.

Objective 1, Page 4

1. List an advantage that a stem-and-leaf plot has over frequency distributions and histograms.
2. Under what conditions do stem-and-leaf plots lose their usefulness?

Objective 1, Page 7

1. When constructing a stem-and-leaf plot, under what conditions is it advisable to use split stems?

#### Objective 2: Construct Frequency Polygons

Objective 2, Page 1

1. Explain how to construct a frequency polygon.

Objective 2, Page 2

**Example 2 *Constructing a Frequency Polygon***

Draw a frequency polygon of the five-year rate of return data listed in Table 16.

**Table 16 Five-Year Rate of Return of Mutual Funds (in percent)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 10.94 | 14.60 | 12.80 | 16.00 | 11.93 | 15.68 | 9.03 | 13.40 |
| 10.53 | 13.98 | 13.86 | 12.36 | 13.54 | 9.94 | 13.93 | 13.63 |
| 14.12 | 14.88 | 14.77 | 13.13 | 8.28 | 19.43 | 12.98 | 13.16 |
| 12.26 | 14.20 | 14.80 | 13.26 | 13.67 | 10.08 | 14.86 | 8.71 |
| 12.17 | 10.26 | 15.22 | 13.36 | 13.55 | 13.90 | 15.64 | 12.80 |

Data from [Morningstar.com](https://www.morningstar.com/)

#### Objective 3: Create Cumulative Frequency and Relative Frequency Distributions

Objective 3, Page 1

1. What does a cumulative frequency distribution display?
2. What does a cumulative relative frequency distribution display?
3. Explain how to find the cumulative frequency for the fifth class in a cumulative frequency distribution.

Objective 3, Page 2

**Example 3 *Constructing a Cumulative and Cumulative Relative Frequency Distribution***

Obtain a cumulative frequency distribution and cumulative relative frequency distribution for the five-year rate of return data listed in Table 13.

| **Class (5-year rate of return)** | **Frequency** | **Relative Frequency** |
| --- | --- | --- |
| 8–8.99 | 2 | 0.05 |
| 9–9.99 | 2 | 0.05 |
| 10-10.99 | 4 | 0.1 |
| 11–11.99 | 1 | 0.025 |
| 12–12.99 | 6 | 0.15 |
| 13–13.99 | 13 | 0.325 |
| 14–14.99 | 7 | 0.175 |
| 15–15.99 | 3 | 0.075 |
| 16–16.99 | 1 | 0.025 |
| 17–17.99 | 0 | 0 |
| 18–18.99 | 0 | 0 |
| 19–19.99 | 1 | 0.025 |

#### Objective 4: Construct Frequency and Relative Frequency Ogives

Objective 4, Page 1

1. What does an ogive represent?
2. Explain the difference between *x*-coordinates for a frequency polygon and a frequency ogive.
3. Explain the difference between *y*-coordinates for a frequency polygon and a frequency ogive.

Objective 4, Page 2

**Example 4 *Constructing Ogives***

Draw a relative frequency ogive of the five-year rate of return data listed in Table 17.

**Table 17**

| **Class (5-year rate of return)** | **Frequency** | **Relative Frequency** | **Cumulative Frequency** | **Cumulative Relative Frequency** |
| --- | --- | --- | --- | --- |
| 8–8.99 | 2 | 0.05 | 2 | 0.05 |
| 9–9.99 | 2 | 0.05 | 4 | 0.1 |
| 10–10.99 | 4 | 0.1 | 8 | 0.2 |
| 11–11.99 | 1 | 0.025 | 9 | 0.225 |
| 12–12.99 | 6 | 0.15 | 15 | 0.3.75 |
| 13–13.99 | 13 | 0.325 | 28 | 0.7 |
| 14–14.99 | 7 | 0.175 | 35 | 0.875 |
| 15–15.99 | 3 | 0.075 | 38 | 0.95 |
| 16–16.99 | 1 | 0.025 | 39 | 0.975 |
| 17–17.99 | 0 | 0 | 39 | 0.975 |
| 18–18.99 | 0 | 0 | 39 | 0.975 |
| 19–19.99 | 1 | 0.025 | 40 | 1 |

#### Objective 5: Draw Time-Series Graphs

Objective 5, Page 1

1. Define time-series data.

Objective 5, Page 1 (continued)

1. Explain how to create a time-series plot.

Objective 5, Page 2

**Example 5 *Drawing a Time-Series Plot***

The Partisan Conflict Index (PCI) tracks the degrees of political disagreement among U.S. politicians in the federal government. It is found by measuring the frequency of newspaper articles reporting disagreement in a given month. Higher values of the index suggest greater conflict among political parties, Congress, and the President. The data in Table 18 represent the PCI in March from 1999 to 2017. Construct a time-series plot of the data. In what year was the index highest? In what year was the index lowest?

**Table 18**

| **Year** | **Partisan Conflict Index (PCI)** |
| --- | --- |
| 1999 | 85.87 |
| 2000 | 94.67 |
| 2001 | 78.23 |
| 2002 | 86.67 |
| 2003 | 88.49 |
| 2004 | 98.55 |
| 2005 | 100.07 |
| 2006 | 91.49 |
| 2007 | 85.44 |
| 2008 | 90.87 |
| 2009 | 88.04 |
| 2010 | 142.42 |
| 2011 | 155.83 |
| 2012 | 154.18 |
| 2013 | 180.56 |
| 2014 | 131.4 |
| 2015 | 163.54 |
| 2016 | 173.88 |
| 2017 | 207.72 |

*Source*: Federal Reserve Bank of Philadelphia

## Section 2.4 Graphical Misrepresentations of Data

### Objective

1. Describe What Can Make a Graph Misleading or Deceptive

#### Objective 1: Describe What Can Make a Graph Misleading or Deceptive

Objective 1, Page 1

*Answer the following after watching the video.*

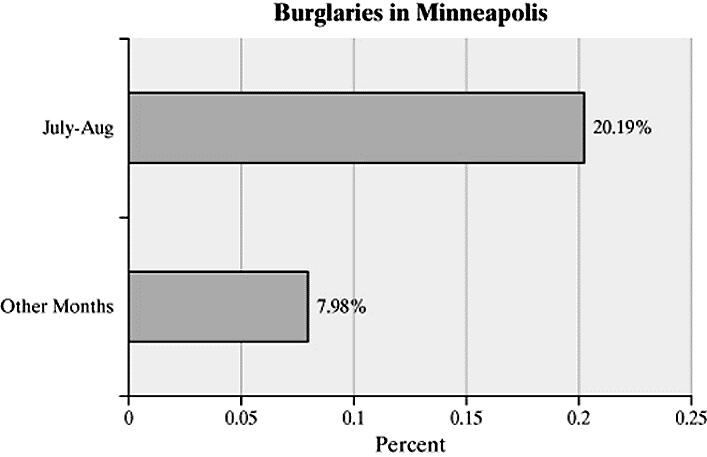
1. Explain the difference between a graph that is misleading and a graph that is deceiving.
2. List what the most common misrepresentations of data involve.

* Increments between tick marks should be consistent.
* Scales for comparative graphs should be the same.
* The baseline, or zero point, should be at the bottom of the graph.

Objective 1, Page 2

**Example 1 *Misrepresentations of Data***

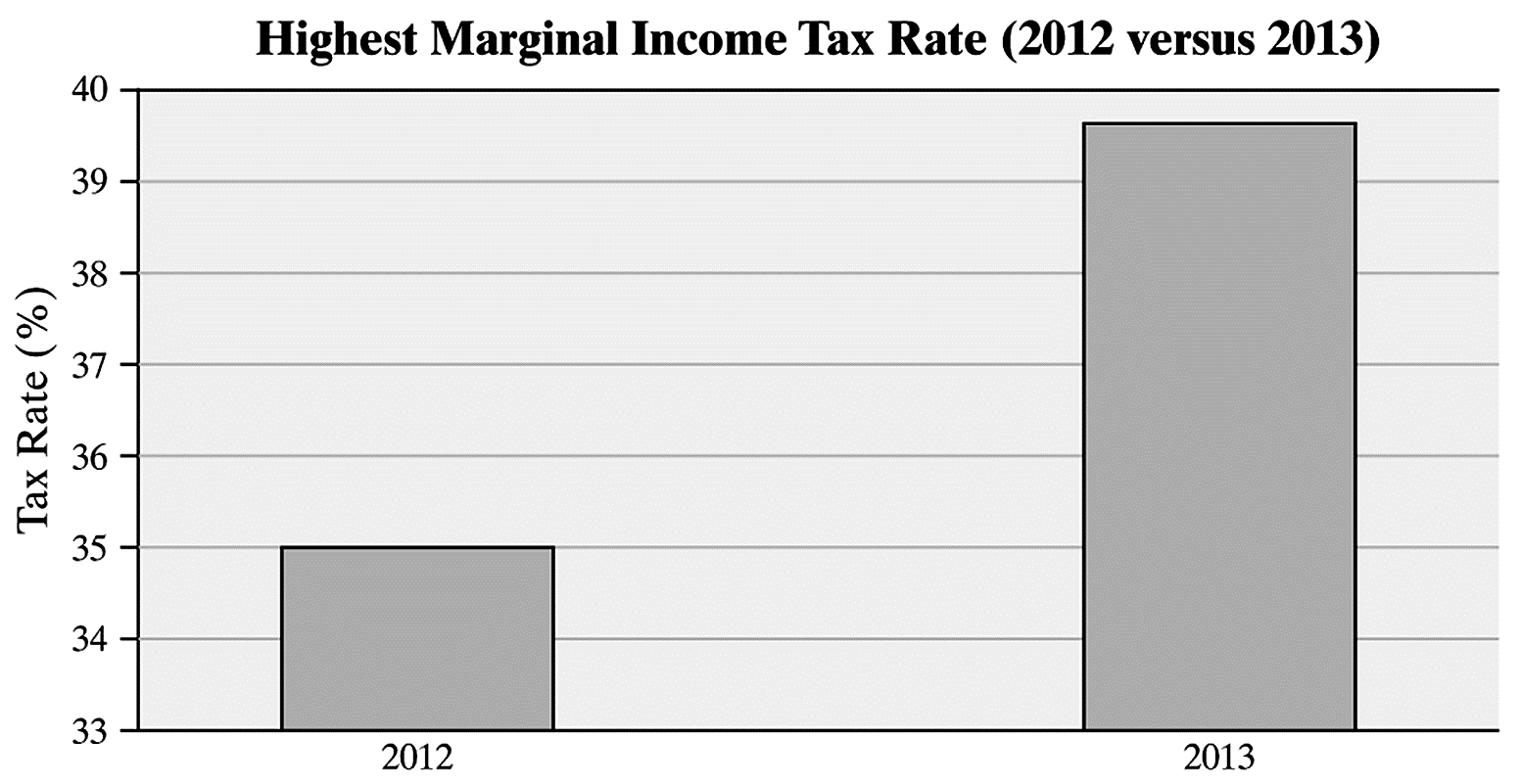
A home security company located in Minneapolis, Minnesota, develops a summer ad campaign with the slogan "When you leave for vacation, burglars leave for work." According to the city of Minneapolis, roughly 20% of home burglaries occur during the peak vacation months of July and August. The advertisement contains the graphic shown. Explain what is wrong with the graphic.



Objective 1, Page 3

**Example 2 *Misrepresentations of Data by Manipulating the Vertical Scale***

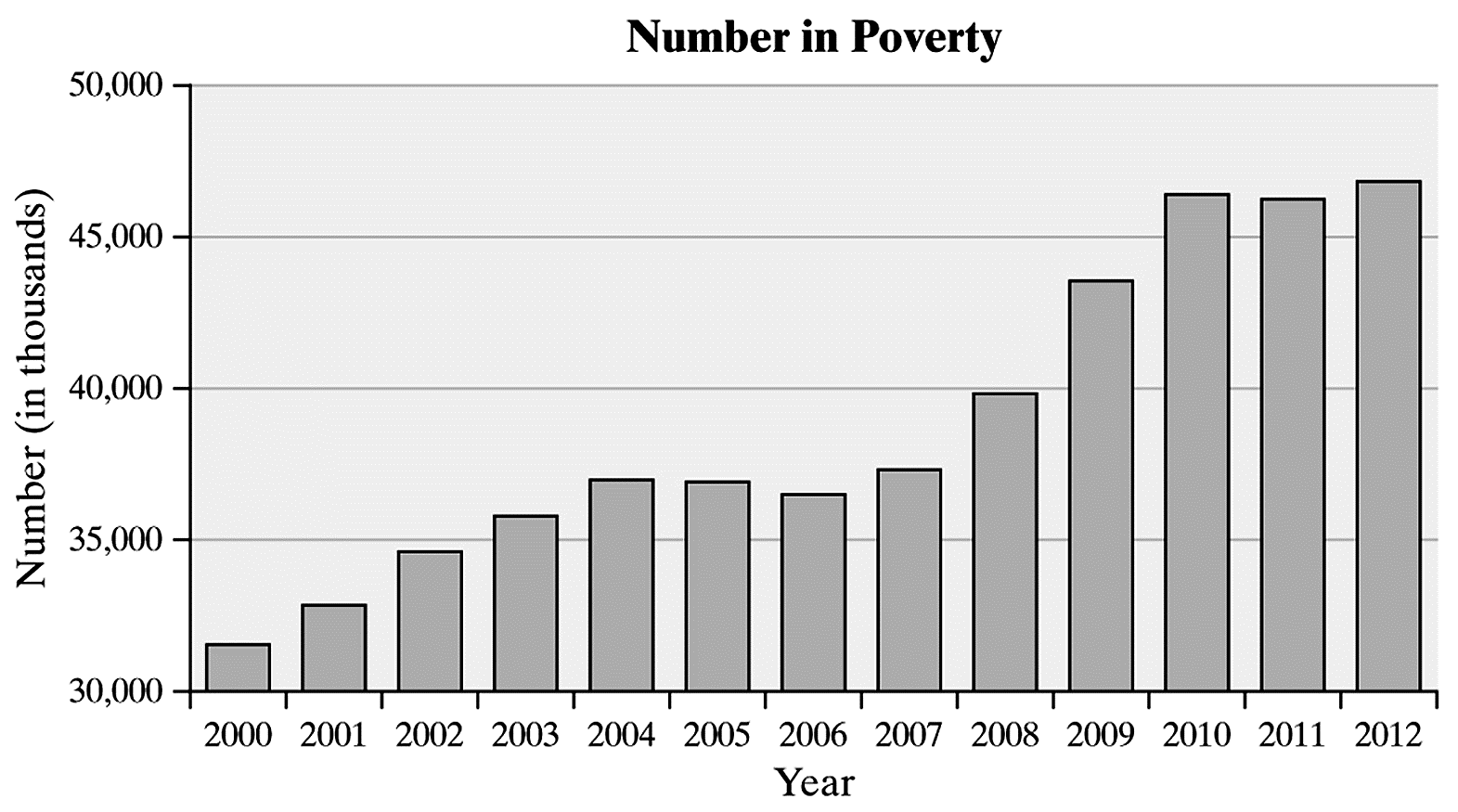
A national news organization developed the following graphic to illustrate the change in the highest marginal tax rate effective January 1, 2013. Why might this graph be considered misleading?



Objective 1, Page 5

**Example 3 *Misrepresentations of Data by Manipulating the Vertical Scale***

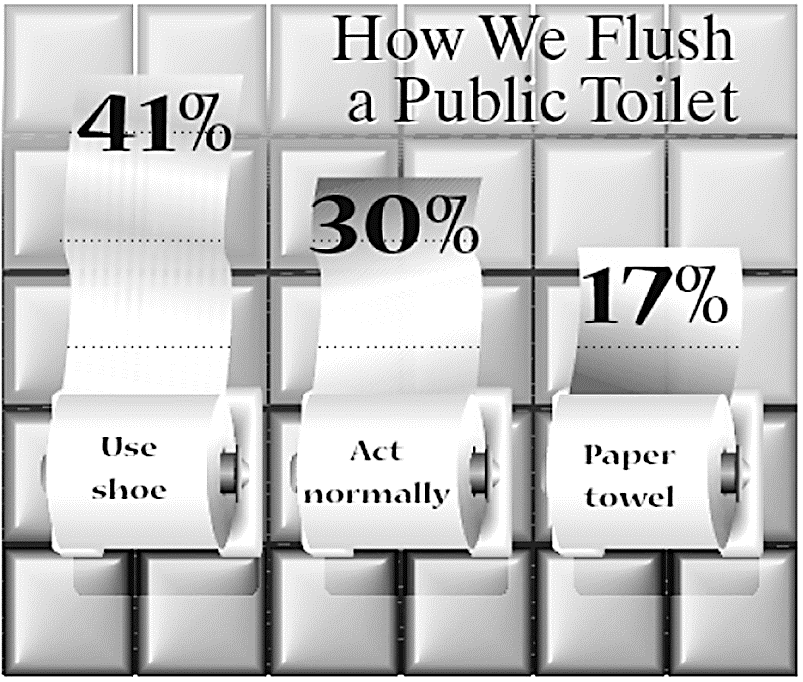
The graph depicts the number of residents in the United States living in poverty. Why might this graph be considered misrepresentative?



Objective 1, Page 7

**Example 4 *Misrepresentations of Data***

The bar graph shown is a *USA Today*-type graph. A survey was conducted by Impulse Research in which individuals were asked how they would flush a toilet when the facilities are not sanitary. What is wrong with the graphic?



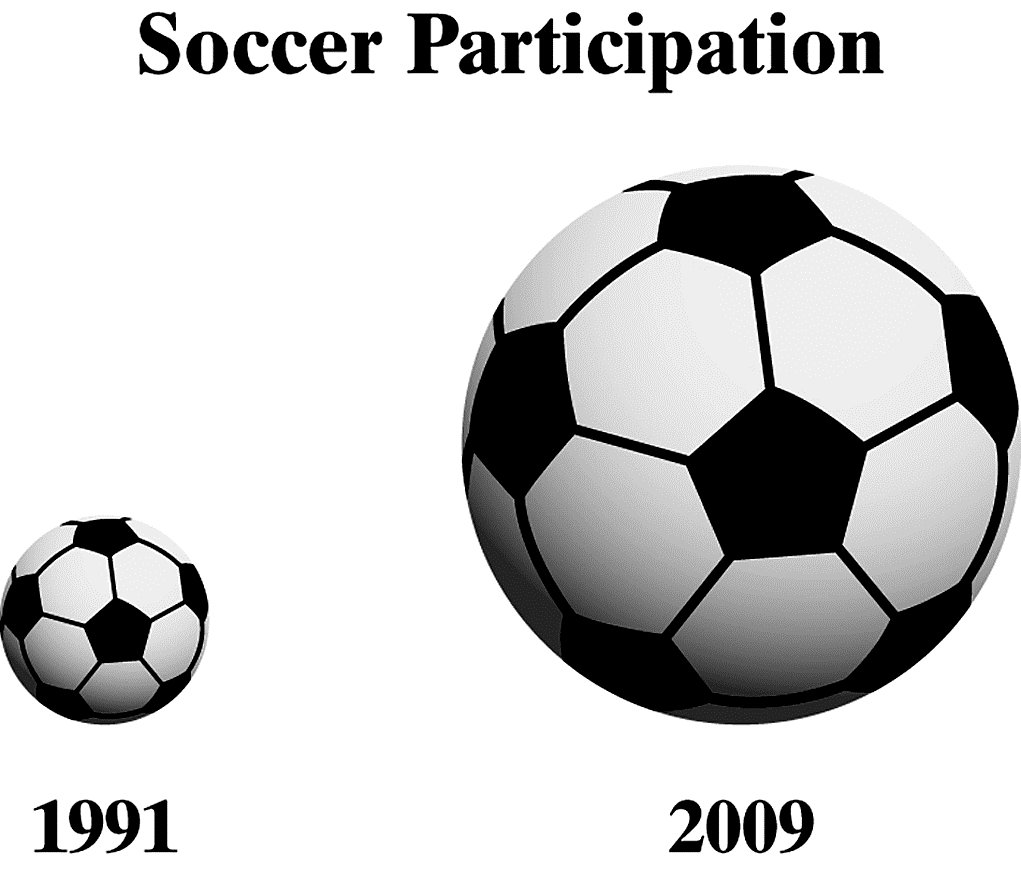
Objective 1, Page 9

1. Why is the use of 3-D effects strongly discouraged?
2. Why do we emphasize that the bars or classes should have the same width?

Objective 1, Page 10

**Example 5 *Misrepresentations of Data by Manipulating Dimension***

Soccer continues to grow in popularity as a sport in the United States. In 1991, there were approximately 10 million participants in the United States aged 7 years and older. By 2009, this number had climbed to 14 million. To illustrate this increase, we could create a graphic like the one shown below. Describe how the graph may be misleading. *Source:* U.S. Census Bureau; National Sporting Goods Association



Objective 1, Page 11

**Example 6 *Misrepresentations of Data: Three-Dimensional Scale***

The figure represents the educational attainment (level of education) in 2016 of adults 25 years and older who are U.S. residents. Why might this graph be considered misrepresentative?

