Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

SKU: 9780321782595-SOLUTIONS

Chapter 2
Organizing and Summarizing Data

Overview

Remember, this course is divided into four parts that correspond to the four parts of our definition of Statistics. Chapters 2 through 4 represent the organizing and summarizing portion of the definition - descriptive statistics.

Chapter 2 discusses describing data through tables and graphs. We summarize raw qualitative data in Section 2.1 and raw quantitative data in Sections 2.2 and 2.3. Section 2.4 provides the opportunity to present misleading graphs.

What to Emphasize

The material in this chapter is elementary. Many of your students have likely seen much of this content in prior classes. Therefore, do not get bogged down in the details of construction of the graphs. Instead, focus on proper graphics construction and how to display qualitative and quantitative data in a fashion that clearly tells the story of the data.

- **Summarizing Qualitative Data -** We begin with a discussion of summarizing qualitative data in tables. This should not be a challenge for the students, so don't spend a lot of class time on this material. Ideally, you will use software such as StatCrunch to build frequency and relative frequency tables. Same goes for the construction of bar graphs and pie charts. Do not emphasize by-hand construction - nobody would ever be expected to construct one of these graphs by hand. Instead, emphasize when each type of graph should be constructed. Bar graphs are typically used when we wish to compare one value of a variable to another, while pie charts are useful when comparing a part to the whole. For example, what proportion of taxes collected by the federal government are from personal income taxes? This is easier to see from a pie chart than a bar graph. Bar graphs can be used to display ordinal data (Poor, Fair, Good, Excellent). Can the same information be displayed effectively in a pie chart?

  Also, emphasize side-by-side bar graphs and the fact that they should be constructed using relative frequencies (since it is not "fair" to compare frequencies when the sample/population sizes differ). Plus, a version of side-by-side bar graphs will be used in Chapters 4 and 12 when we study conditional distributions.
- **Summarizing Quantitative Data -** We decided to segment summarizing quantitative data into two sections - the popular displays (Section 2.2) and other graphs (Section 2.3). Section 2.3 is optional and can be skipped without loss of continuity. When considering which topics you might skip, ask yourself if the topic will be needed or revisited later in the course. For example, graphs such as frequency polygons are not utilized later in the course, so you might consider skipping this topic. That said, some graphs allow for interesting results, such as time series plots. So, judgment should be used when deciding what to skip.

Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

In Section 2.2, do not get bogged down on by-hand creation of frequency or relative frequency distributions or the construction of histograms. Allow technology to do the work so that you may focus on the fact that there are many acceptable class widths that provide a nice summary of the data. However, be sure to emphasize that some class widths are really poor. In fact, spend time making sure students understand the results from *Activity 1: Choosing Class Width* in Section 2.2. In addition, spend time on distribution shape. Require students to justify their conclusions regarding shape, rather than simply claiming a distribution is skewed right.

Section 2.3 is entirely optional and may be skipped without loss of continuity. That said, you may consider at least requiring that students review time series graphs since they are popular in the media, and time series data is discussed at various points in the course (especially when we present correlation versus causation).

- **Graphical Misrepresentations of Data** - While this is an optional section, it does have merit. The media is full of examples where graphs mislead or misrepresent data. Be sure to alert students to be on the lookout for poor graphics and require students to clearly label each graph they create.

Ideas for Traditional/Online/Blended/Flipped

Again, real data should be utilized to illustrate concepts.

Whether you are in a traditional or online setting, require students to justify results. For example, create four histograms of the same set of data and ask students (either through discussion groups or small in-class groups) to rank the graphs from best to worst. Change things up by not including titles or labels. Ask students how a graphic might be improved. Activities such as these not only require students to know how the graphics are constructed, but also get them thinking about appropriate techniques. Another idea is to give various data sets to students and ask them to summarize the information. Don't tell them what type of graphic to create - let them decide. Mix the data up between qualitative and quantitative. For example, include a small data set with discrete data (collect data from your class on number of siblings, which is likely best summarized using a dot plot) along with larger data sets that might be better summarized using histograms (use data from the 2014 World Cup at http://www.statcrunch.com/app/index.php?dataid=1130049). If a student or group chooses a histogram, ask for justification of the choice. In a classroom, perhaps the students could be required to present the graphic to the class and explain what message the graphic conveys.

A great **applet** that may be used to emphasize that there is no such thing as the "correct" class width when constructing a histogram is the Histogram with Sliders applet in StatCrunch. Use a data set to build the applet and allow students to experiment with various starting points and class widths.

Don't forget about the **supplemental exercises** located in MyStatLab. Again, these exercises may be used as homework assignments or form the basis for classroom discussion/online discussion. In particular, consider Problem 11 in Section 2.1; Problem 9 in Section 2.2; Problem 14 in Section 2.2; Problem 4 in Section 2.4; and Problem 8 in Section 2.4.

Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

Classroom Examples

Section 2.1

1. The 2010 Census results include a summary of the racial composition of the population. The races reported by the populations of the United States and the State of California are summarized below. The data are given in millions. (Source: census.gov)

Race	U.S.	California
White	223.6	21.5
Black or African American	38.9	2.3
American Indian and Alaska Native	2.9	0.4
Asian	14.7	4.9
Native Hawaiian and Other Pacific Islander	0.5	0.1
Some Other Race	19.1	6.3
Two or More Races	9	1.8

a. Construct a relative frequency distribution for races in the U.S.

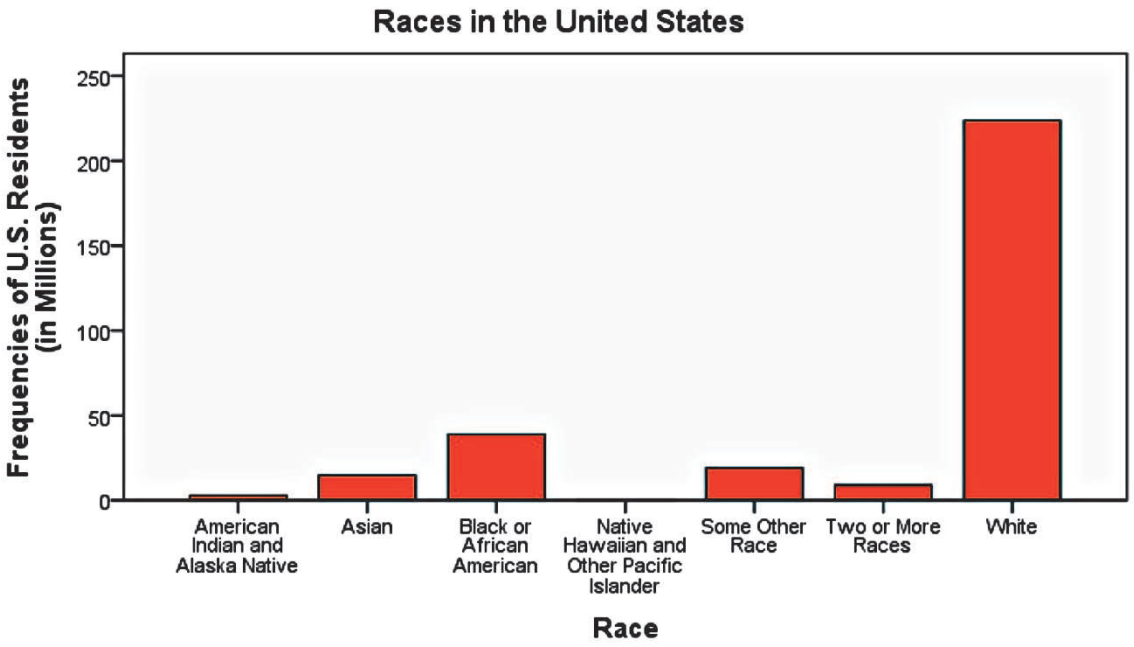
Race	Relative Frequency (U.S.)
White	0.7241
Black or African American	0.1261
American Indian and Alaska Native	0.0095
Asian	0.0475
Native Hawaiian and Other Pacific Islander	0.0017
Some Other Race	0.0619
Two or More Races	0.0292

b. What percentage of U.S. residents claims two or more races? 2.92%
c. What percentage of U.S. residents is not Asian? 95.25%

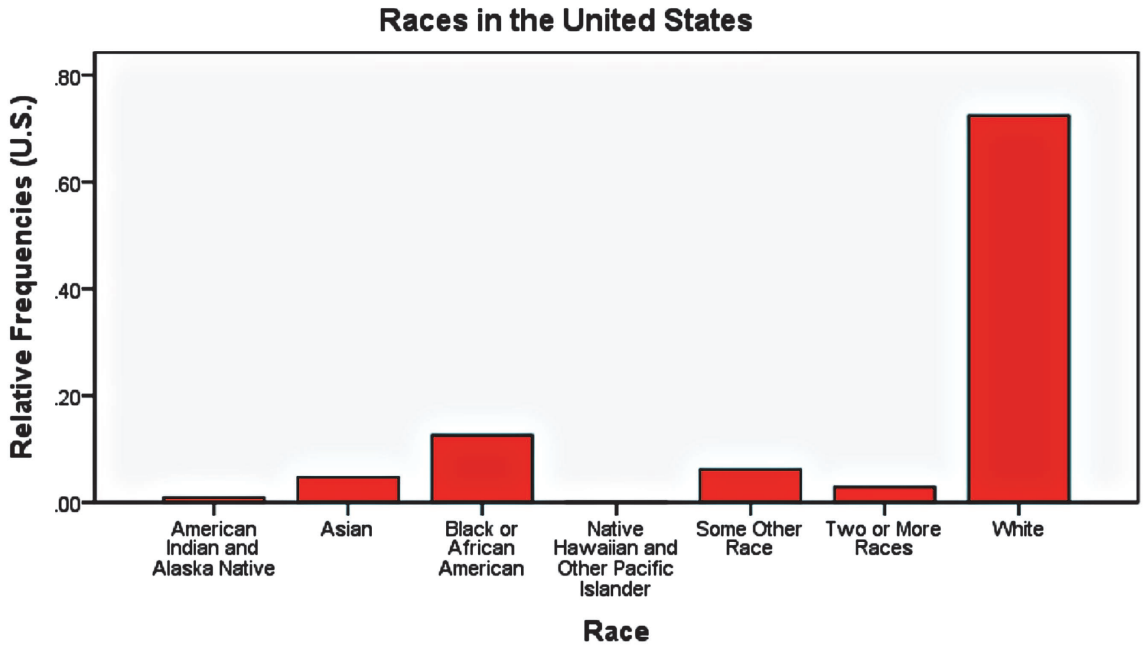
Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

d. Construct a frequency bar graph for races in the U.S.



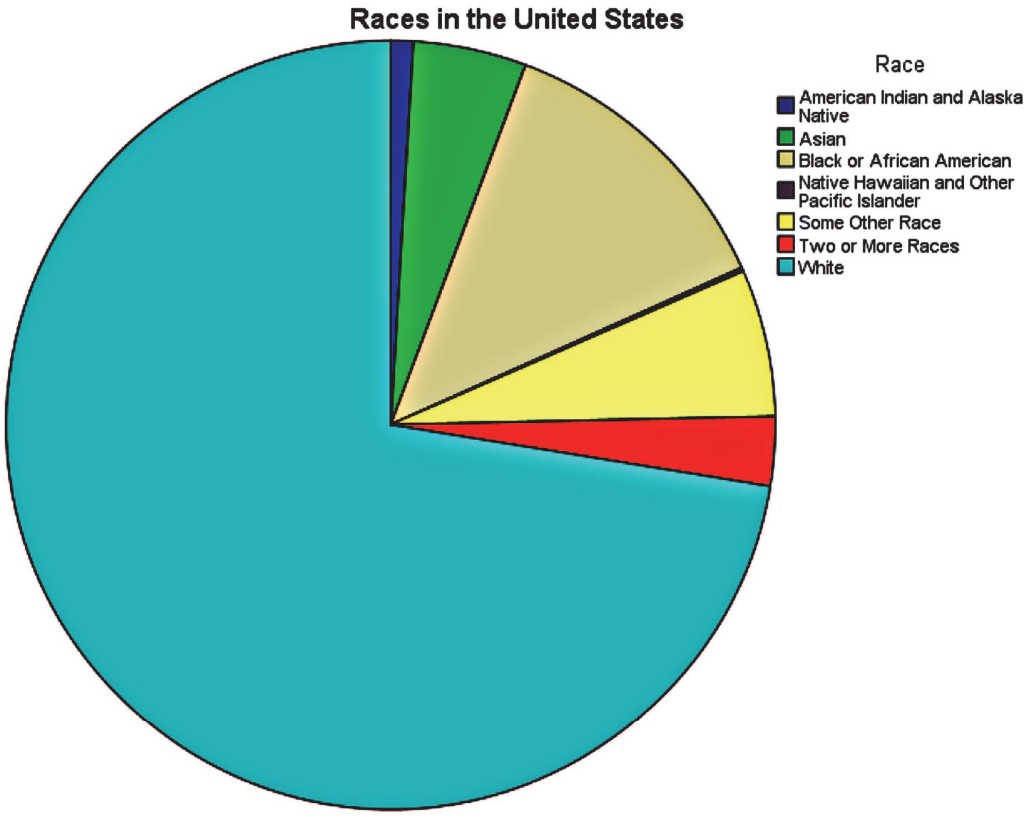
e. Construct a relative frequency bar graph for races in the U.S.



Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

f. Construct a pie chart for races in the U.S.



2. Use the data in Problem 1 to answer this question.
a. Construct a relative frequency distribution for the races in California.

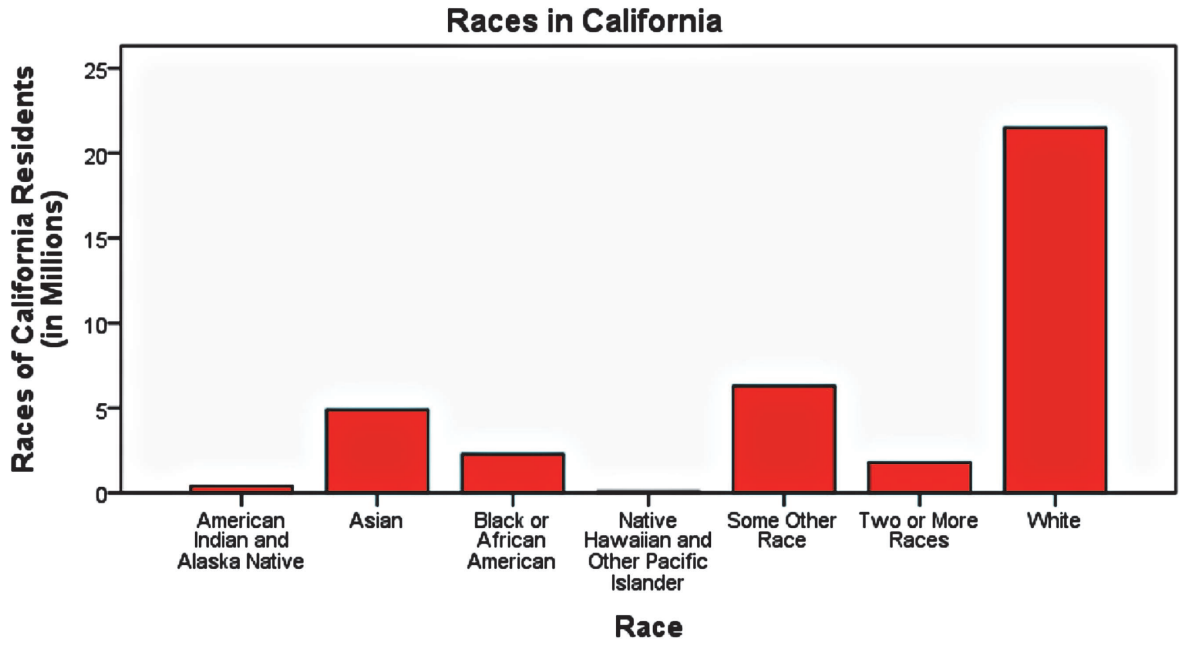
Race	Relative Frequency (California)
White	0.5759
Black or African American	0.0617
American Indian and Alaska Native	0.0097
Asian	0.1305
Native Hawaiian and Other Pacific Islander	0.0039
Some Other Race	0.1696
Two or More Races	0.0487

b. What percentage of California residents claims two or more races? 4.87%
c. What percentage of U.S. residents is not Asian? 86.95%

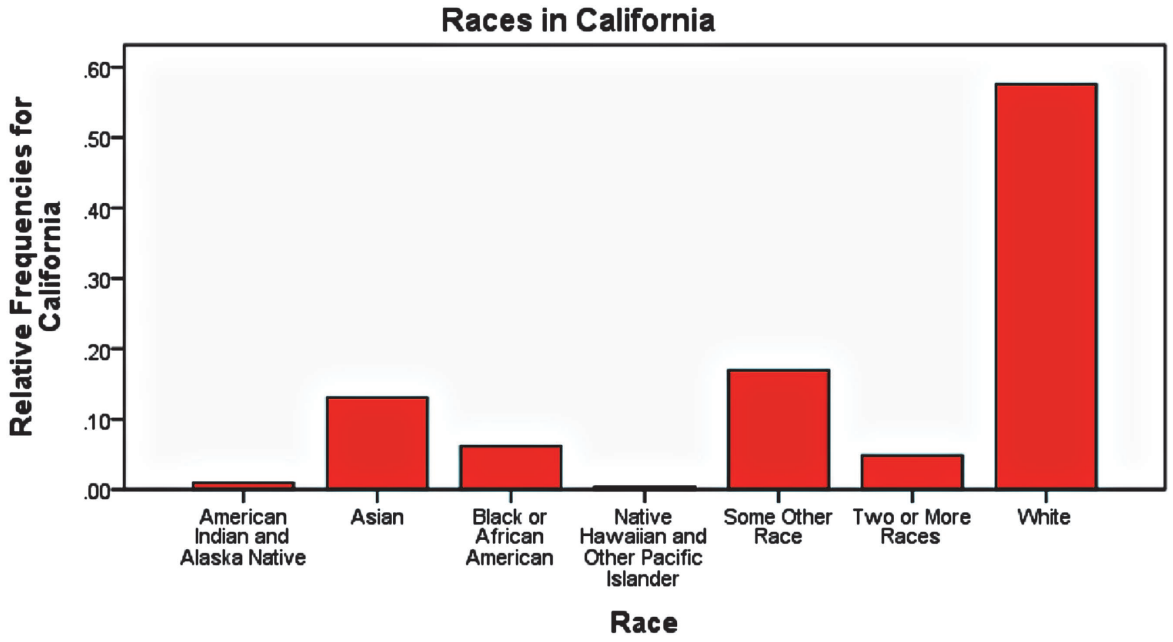
Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

d. Construct a frequency bar graph for California.



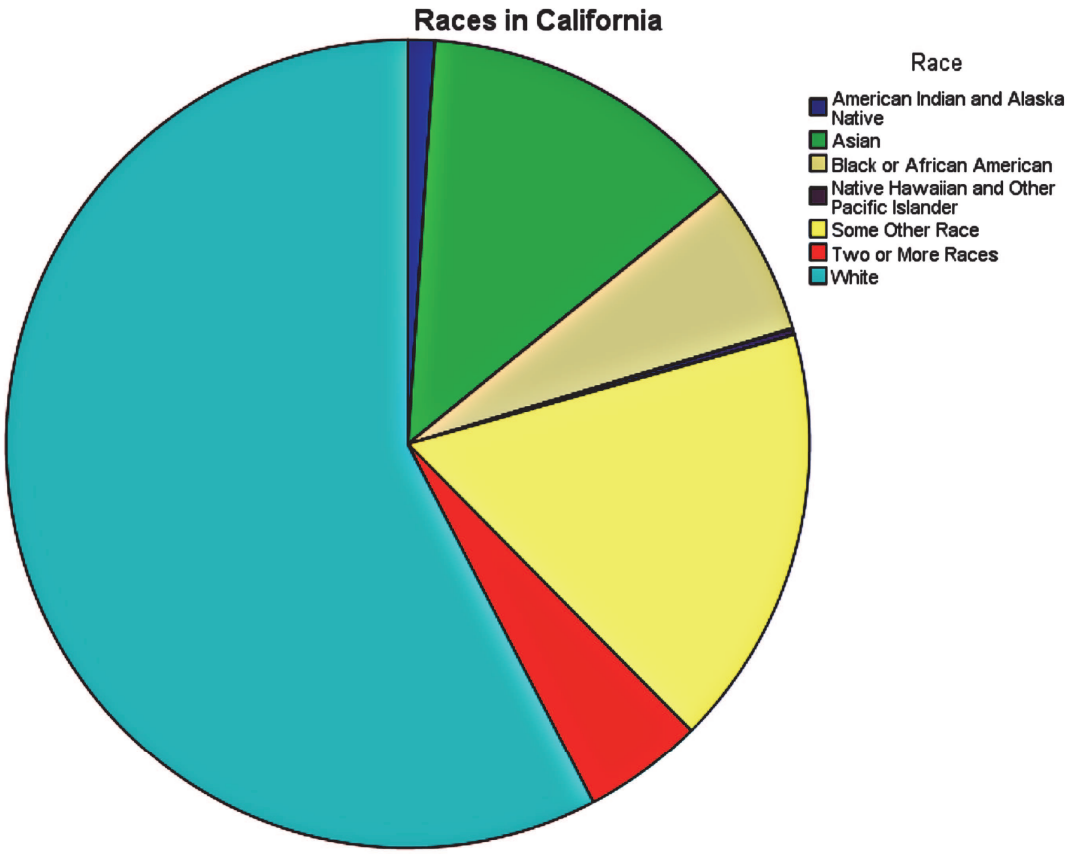
e. Construct a relative frequency bar graph for California.



Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

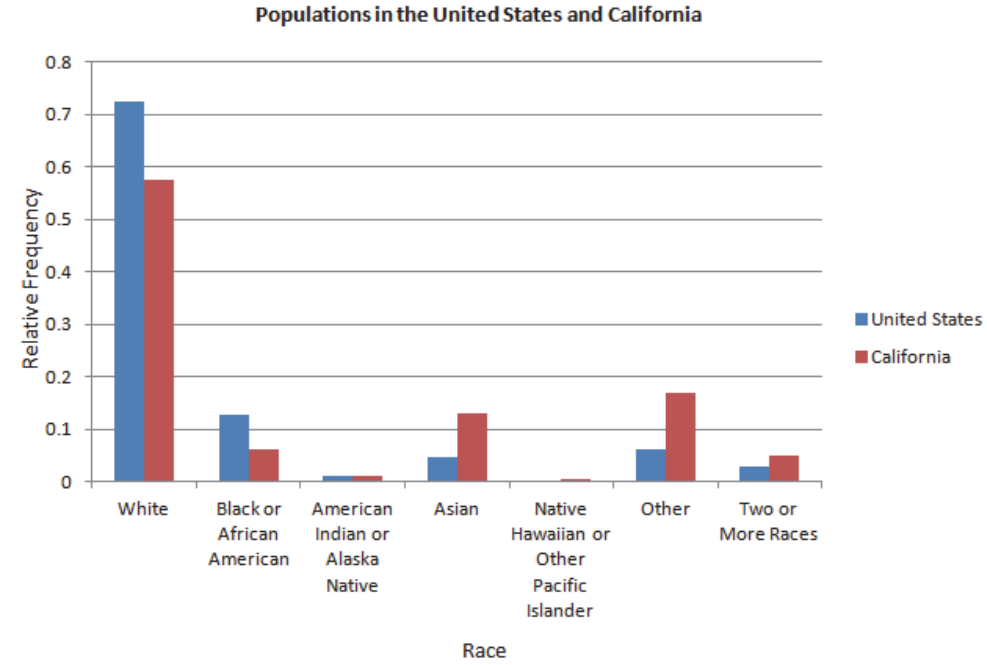
f. Construct a pie chart for California.



Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

3. Draw a side-by-side relative frequency bar graph of the populations of the United States and California. What do you notice?



4. In-Class Activity (Requires one bag of multicolored candy, such as M&M's or Skittles for each student.) Ask the students bring one bag of candy, such as M&M's or Skittles to class. Have them create a frequency distribution of the counts of each color of candy in their bag. Have the students create a bar graph, relative frequency bar graph, and a pie chart illustrating the number of each type they observed.

Section 2.2

1. The following data represent the graduation rate for a random sample of 60 colleges and universities in the United States. Data from www.payscale.com.

86	59	65	55	59	40
90	48	37	65	67	67
39	53	82	57	71	46
83	77	40	52	92	38
56	57	36	61	34	54
69	35	73	29	92	39
24	48	41	46	79	41
44	43	28	61	49	65
48	42	72	35	58	39
80	75	44	52	52	47

Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

With a first class having lower class limit of 20 and a class width of 10:

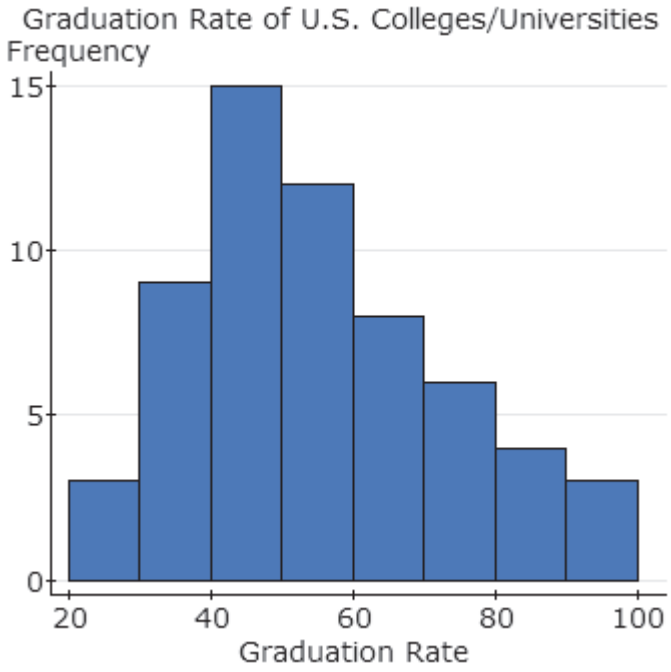
a. Construct a frequency distribution.

Graduation Rate	Frequency
20 - 29	3
30 - 39	9
40 - 49	15
50 - 59	12
60 - 69	8
70 - 79	6
80 - 89	4
90 - 99	3

b. Construct a relative frequency distribution.

Graduation Rate	Relative Frequency
20 - 29	0.05
30 - 39	0.15
40 - 49	0.25
50 - 59	0.2
60 - 69	0.1333
70 - 79	0.1
80 - 89	0.0667
90 - 99	0.05

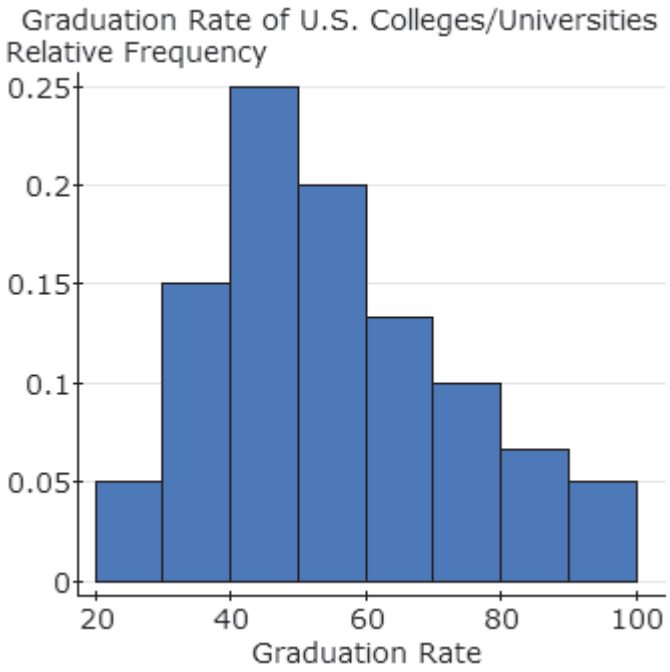
c. Construct a frequency histogram of the data.



Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

d. Construct a relative frequency histogram of the data.



e. Describe the shape of the distribution. The distribution is skewed right.

2. Draw a stem-and-leaf plot of the data from Problem 1.

```
2 : 489
3 : 455678999
4 : 001123446678889
5 : 222345677899
6 : 11555779
7 : 123579
8 : 0236
9 : 022
```

Legend 2|4 represents 24

3. Go to http://en.wikipedia.org/wiki/United_States_congressional_apportionment. The site includes data on the number of representatives in the House of Representatives for each census year. Create a frequency and relative frequency histogram of the data for the latest census. Draw a dot plot of the data. Note: The data can be easily extracted using StatCrunch This! in StatCrunch.

4. Ask the students to compute their heights (in inches). Randomly select some students to report their heights. Create a histogram of the students' heights. It would be good to make a histogram for men and a histogram for women. Ask the students to identify the shape of the distribution.

Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

Section 2.3

1. A simple random sample of 4,000 women who gave birth was collected. The following table summarizes the mother's ages at the time they gave birth. (Based on data at: www.infoplease.com/ipa/A0005074.html)

Age of Mother	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
Frequency	9	401	1006	1094	919	467	101	6

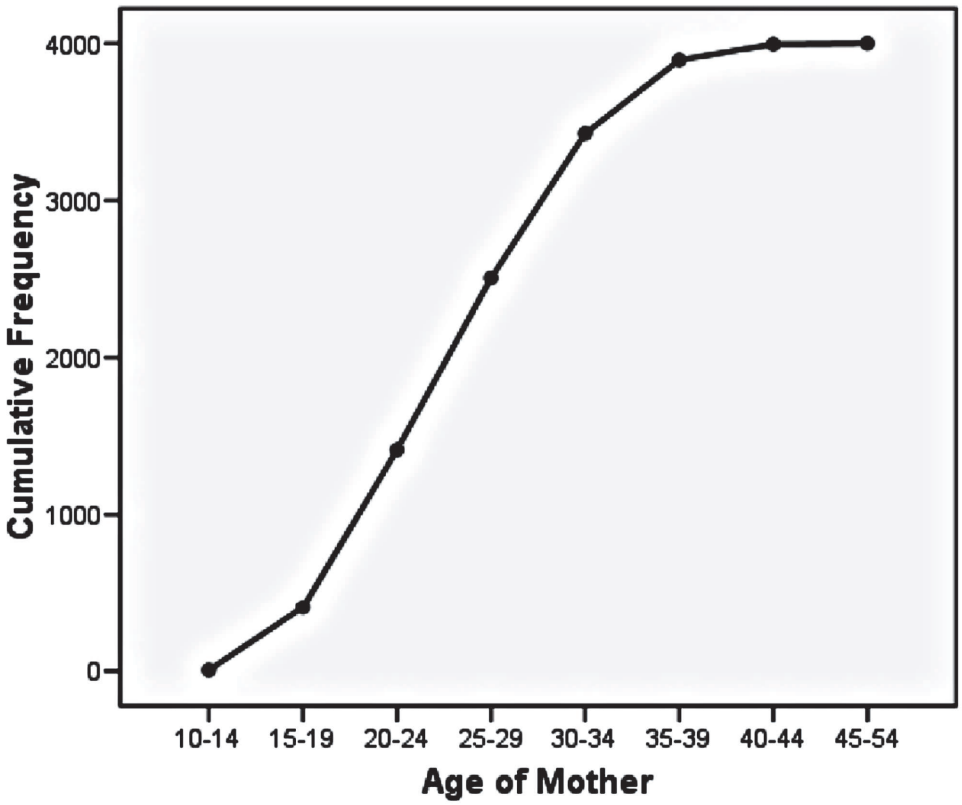
a. Create a cumulative frequency table summarizing the data.
b. Create a relative frequency table of the data. (See above)
c. Create a cumulative relative frequency table for the data. (Note: Answers may vary slightly due to rounding.)

Age of Mother	Frequency	Cumulative Frequency	Relative Frequency	Cumulative Relative Frequency
10-14	6	6	0.0015	0.0015
15-19	401	407	0.1003	0.1018
20-24	1006	1413	0.2515	0.3533
25-29	1094	2507	0.2735	0.6268
30-34	919	3426	0.2298	0.8565
35-39	467	3893	0.1168	0.9733
40-44	101	3994	0.0253	0.9985
45-54	6	4000	0.0015	1.0000

Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

d. Construct a frequency ogive for the data.



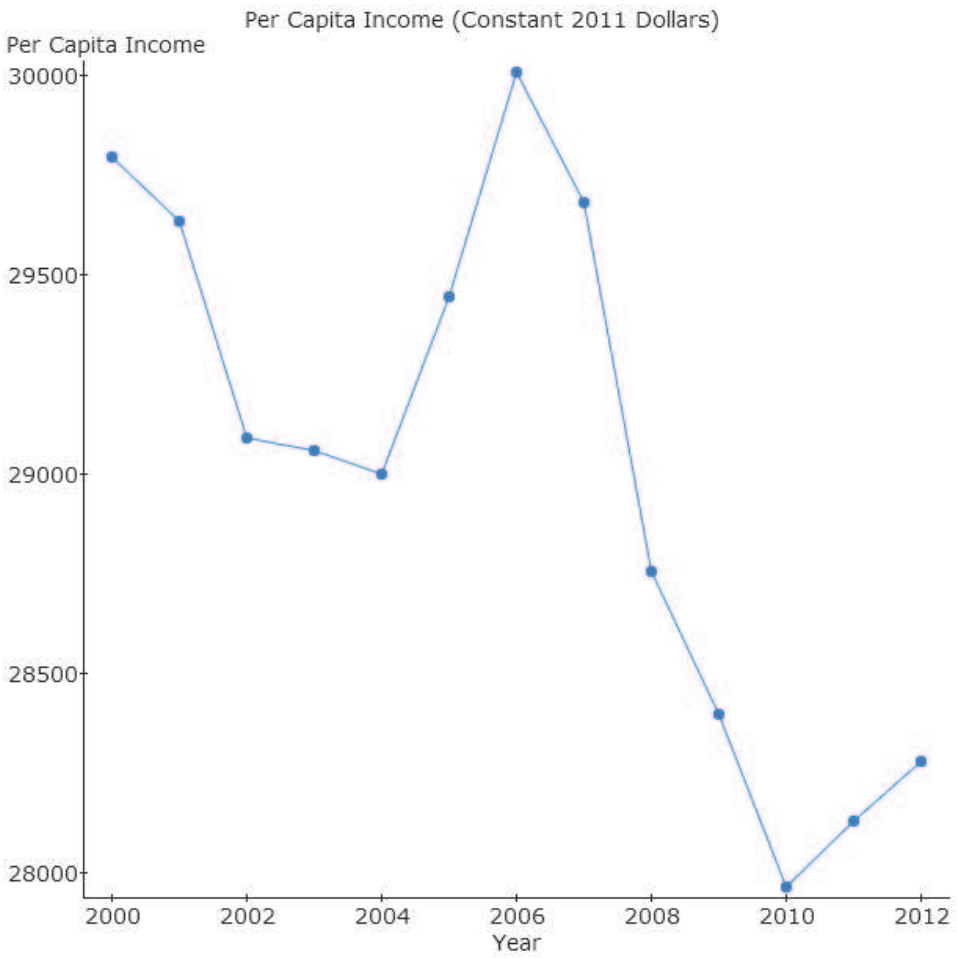
2. The following data represent the per capita income in the United States from 2000 - 2012 in constant 2011 dollars (that is, adjusted for inflation). Draw a time series graph of the data.

Year	Per Capita Income (2011 Dollars)
2000	29,795
2001	29,636
2002	29,092
2003	29,058
2004	29,000
2005	29,446
2006	30,010
2007	29,682
2008	28,755
2009	28,400
2010	27,968
2011	28,130
2012	28,281

Source: United States Census Bureau

Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual



Section 2.4

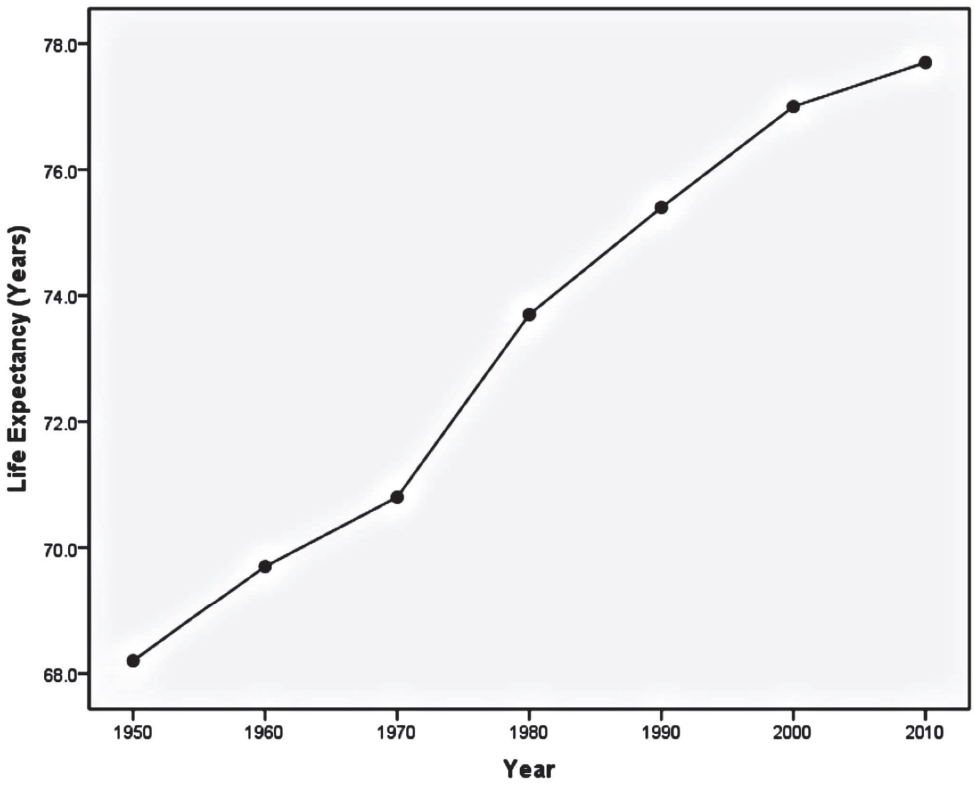
1. The data in the table below represent the historical life expectancies (in years) of residents of the United States. (Source: National Center for Health Statistics, *National Vital Statistics Reports,* www.cdc.gov/nchs and http://www.cdc.gov/nchs/data/hus/hus10.pdf#022)

Year, x	**Life Expectancy, y**
1950	68.2
1960	69.7
1970	70.8
1980	73.7
1990	75.4
2000	77.0
2010	77.7

Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

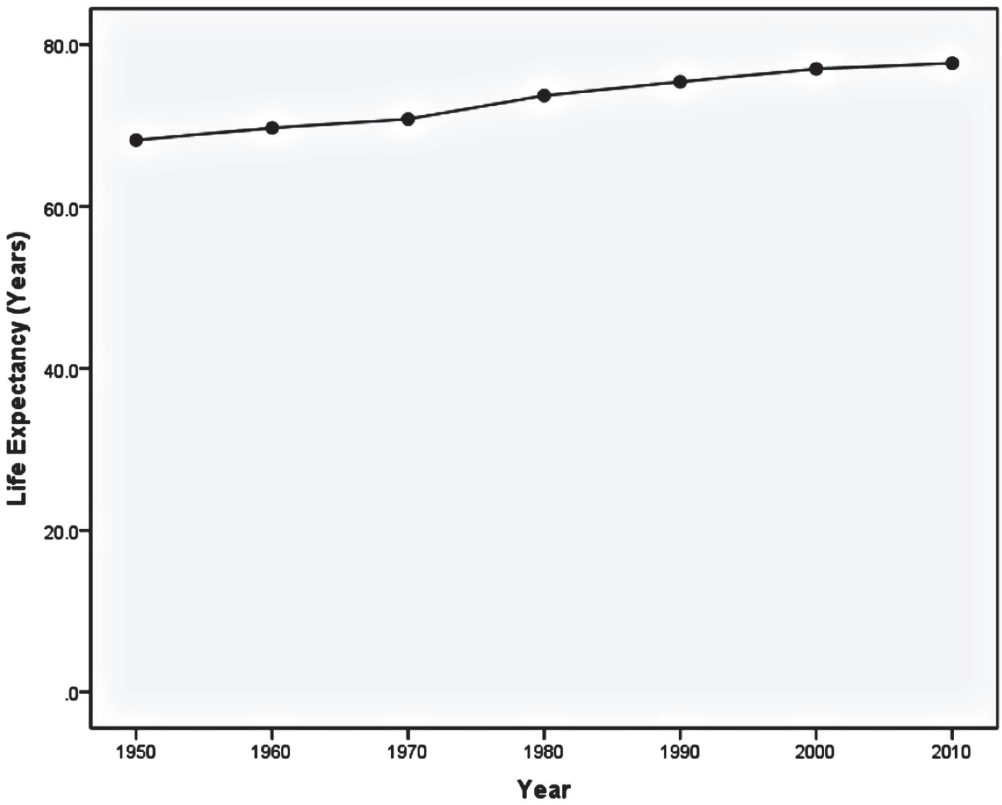
a. Construct a misleading time-series plot that indicates that the life expectancy has risen sharply over time.



Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

b. Construct a time-series plot that is not misleading.



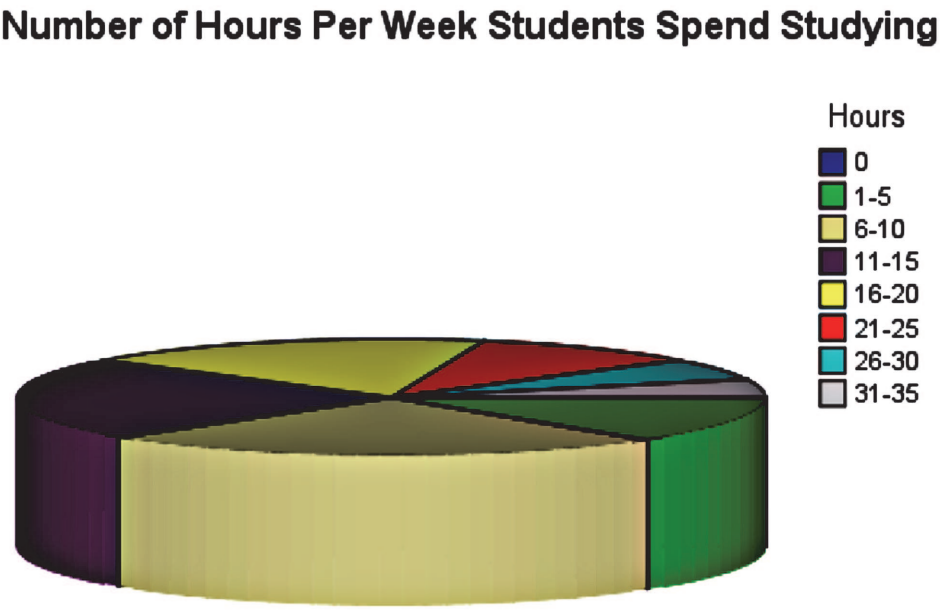
2. The National Survey of Student Engagement is a survey that (among other things) asks first year students at liberal arts colleges how much time they spend preparing for class each week. The results are summarized below. (Source: NSSE)

Hours	0	1-5	6-10	11-15	16-20	21-25	26-30	31-35
Percentage of 1st year students	0%	13%	25%	23%	18%	10%	6%	5%

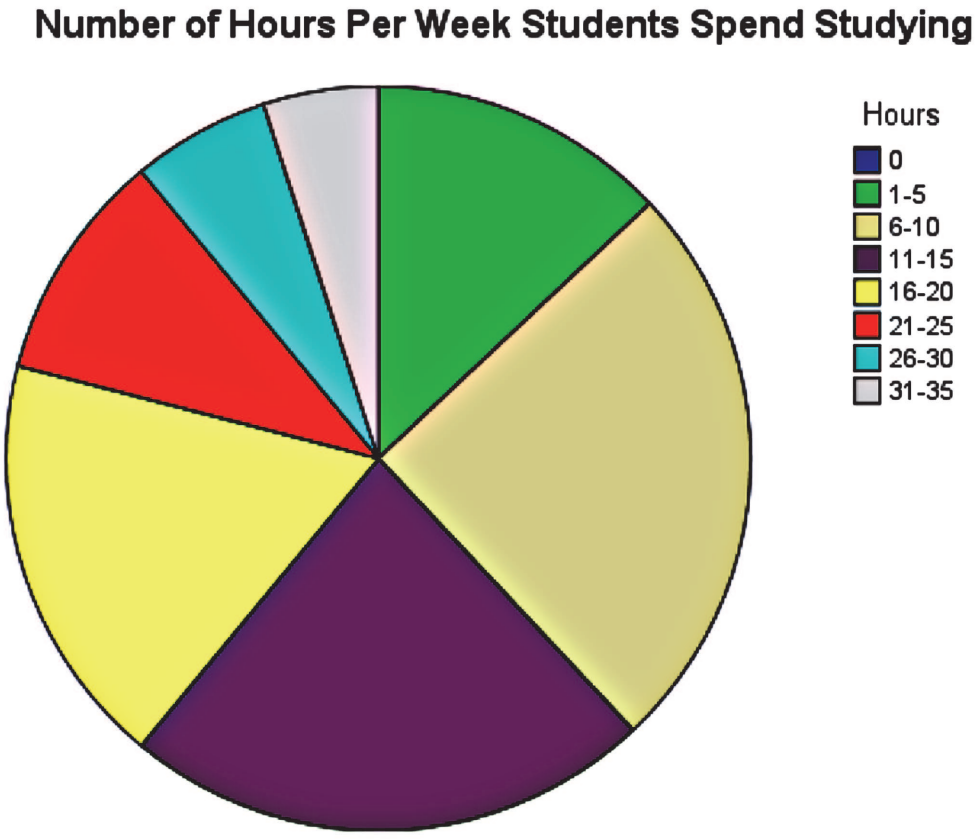
Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual

a. Construct a pie chart that exaggerates the percentage of students who spend between 6 and 10 hours preparing for class each week.



b. Construct a pie chart that is not misleading.



Note: If you use Excel to create 3-D pie charts, you can rotate the graph so students may visualize the distortion created by the three-dimensional art.

Copyright © 2016 Pearson Education, Inc.

Interactive Statistics Informed Decisions Using Data 1st edition Sullivan Solutions Manual