

## 2.E: Descriptive Statistics (Exercises)

### 2.2: Stem-and-Leaf Graphs (Stemplots), Line Graphs, and Bar Graphs

#### Q 2.2.1

Student grades on a chemistry exam were: 77, 78, 76, 81, 86, 51, 79, 82, 84, 99

- Construct a stem-and-leaf plot of the data.
- Are there any potential outliers? If so, which scores are they? Why do you consider them outliers?

#### Q 2.2.2

The table below contains the 2010 obesity rates in U.S. states and Washington, DC.

State	Percent (%)	State	Percent (%)	State	Percent (%)
Alabama	32.2	Kentucky	31.3	North Dakota	27.2
Alaska	24.5	Louisiana	31.0	Ohio	29.2
Arizona	24.3	Maine	26.8	Oklahoma	30.4
Arkansas	30.1	Maryland	27.1	Oregon	26.8
California	24.0	Massachusetts	23.0	Pennsylvania	28.6
Colorado	21.0	Michigan	30.9	Rhode Island	25.5
Connecticut	22.5	Minnesota	24.8	South Carolina	31.5
Delaware	28.0	Mississippi	34.0	South Dakota	27.3
Washington, DC	22.2	Missouri	30.5	Tennessee	30.8
Florida	26.6	Montana	23.0	Texas	31.0
Georgia	29.6	Nebraska	26.9	Utah	22.5
Hawaii	22.7	Nevada	22.4	Vermont	23.2
Idaho	26.5	New Hampshire	25.0	Virginia	26.0
Illinois	28.2	New Jersey	23.8	Washington	25.5
Indiana	29.6	New Mexico	25.1	West Virginia	32.5
Iowa	28.4	New York	23.9	Wisconsin	26.3
Kansas	29.4	North Carolina	27.8	Wyoming	25.1

- Use a random number generator to randomly pick eight states. Construct a bar graph of the obesity rates of those eight states.
- Construct a bar graph for all the states beginning with the letter "A."
- Construct a bar graph for all the states beginning with the letter "M."

#### S 2.2.2

- Example solution for using the random number generator for the TI-84+ to generate a simple random sample of 8 states. Instructions are as follows.
  - Number the entries in the table 1–51 (Includes Washington, DC; Numbered vertically)
  - Press MATH
  - Arrow over to PRB
  - Press 5:randInt(
  - Enter 51,1,8)

Eight numbers are generated (use the right arrow key to scroll through the numbers). The numbers correspond to the numbered states (for this example: {47 21 9 23 51 13 25 4}). If any numbers are repeated, generate a different number by using 5:randInt(51,1)). Here, the states (and Washington DC) are {Arkansas, Washington DC, Idaho, Maryland, Michigan, Mississippi, Virginia, Wyoming}.

Corresponding percents are {30.1, 22.2, 26.5, 27.1, 30.9, 34.0, 26.0, 25.1}.

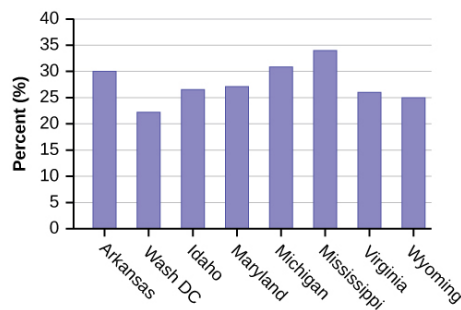


Figure 2.E. 1: (a)

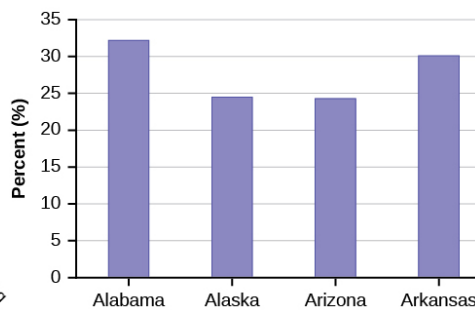


Figure 2.E. 1: (b)

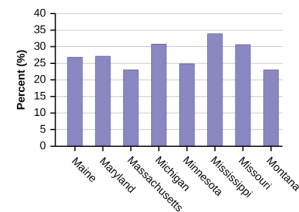


Figure 2.E. 1: (c)

For each of the following data sets, create a stem plot and identify any outliers.

### Exercise 2.2.7

The miles per gallon rating for 30 cars are shown below (lowest to highest).

19, 19, 19, 20, 21, 21, 25, 25, 25, 26, 26, 28, 29, 31, 31, 32, 32, 33, 34, 35, 36, 37, 37, 38, 38, 38, 38, 41, 43, 43

**Answer**

Stem	Leaf
1	9 9 9
2	0 1 1 5 5 5 6 6 8 9
3	1 1 2 2 3 4 5 6 7 7 8 8 8 8
4	1 3 3

The height in feet of 25 trees is shown below (lowest to highest).

25, 27, 33, 34, 34, 34, 35, 37, 37, 38, 39, 39, 39, 40, 41, 45, 46, 47, 49, 50, 50, 53, 53, 54, 54

The data are the prices of different laptops at an electronics store. Round each value to the nearest ten.

249, 249, 260, 265, 265, 280, 299, 299, 309, 319, 325, 326, 350, 350, 350, 365, 369, 389, 409, 459, 489, 559, 569, 570, 610

**Answer**

Stem	Leaf
2	5 5 6 7 7 8
3	0 0 1 2 3 3 5 5 5 7 7 9
4	1 6 9
5	6 7 7
6	1

The data are daily high temperatures in a town for one month.

61, 61, 62, 64, 66, 67, 67, 67, 68, 69, 70, 70, 70, 71, 71, 72, 74, 74, 74, 75, 75, 75, 76, 76, 77, 78, 78, 79, 79, 95

For the next three exercises, use the data to construct a line graph.

### Exercise 2.2.8

In a survey, 40 people were asked how many times they visited a store before making a major purchase. The results are shown in the Table below.

Number of times in store	Frequency
1	4
2	10
3	16
4	6
5	4

Answer



Figure 2.E. 2:

### Exercise 2.2.9

In a survey, several people were asked how many years it has been since they purchased a mattress. The results are shown in Table.

Years since last purchase	Frequency
0	2
1	8
2	13
3	22
4	16
5	9

### Exercise 2.2.10

Several children were asked how many TV shows they watch each day. The results of the survey are shown in the Table below.

Number of TV Shows	Frequency

Number of TV Shows	Frequency
0	12
1	18
2	36
3	7
4	2

**Answer**

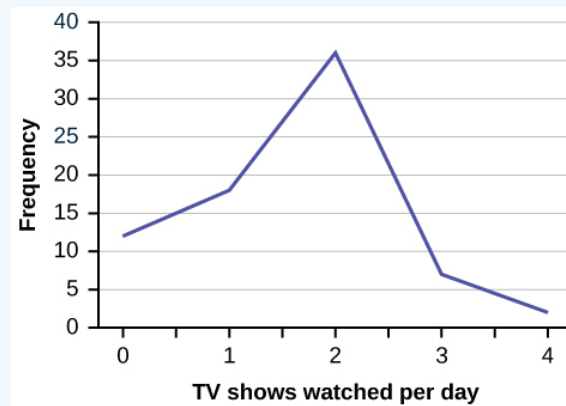


Figure 2.E. 3:

### Exercise 2.2.11

The students in Ms. Ramirez's math class have birthdays in each of the four seasons. Table shows the four seasons, the number of students who have birthdays in each season, and the percentage (%) of students in each group. Construct a bar graph showing the number of students.

Seasons	Number of students	Proportion of population
Spring	8	24%
Summer	9	26%
Autumn	11	32%
Winter	6	18%

Using the data from Mrs. Ramirez's math class supplied in the table above, construct a bar graph showing the percentages.

**Answer**

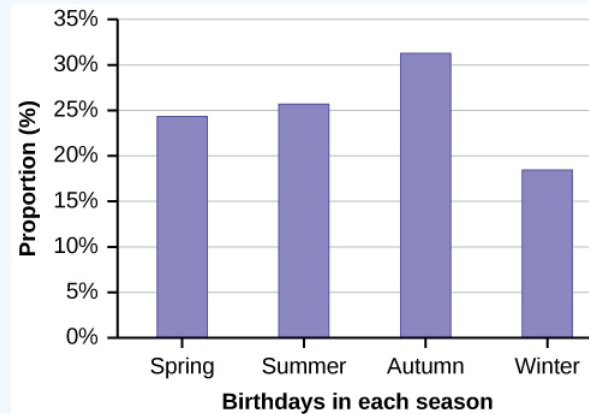


Figure 2.E. 4:

### Exercise 2.2.12

David County has six high schools. Each school sent students to participate in a county-wide science competition. Table shows the percentage breakdown of competitors from each school, and the percentage of the entire student population of the county that goes to each school. Construct a bar graph that shows the population percentage of competitors from each school.

High School	Science competition population	Overall student population
Alabaster	28.9%	8.6%
Concordia	7.6%	23.2%
Genoa	12.1%	15.0%
Mocksville	18.5%	14.3%
Tynneson	24.2%	10.1%
West End	8.7%	28.8%

Use the data from the David County science competition supplied in Exercise. Construct a bar graph that shows the county-wide population percentage of students at each school.

**Answer**

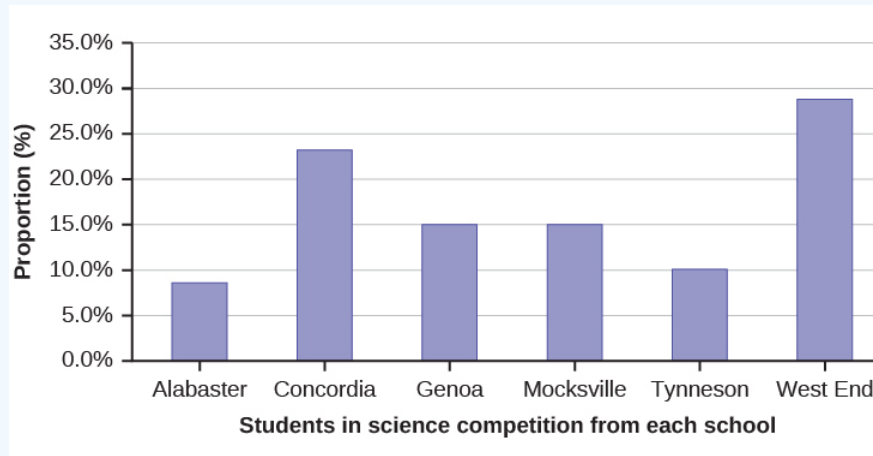


Figure 2.E. 5:

## 2.3: Histograms, Frequency, Polygons, and Time Series Graphs

### Q 2.3.1

Suppose that three book publishers were interested in the number of fiction paperbacks adult consumers purchase per month. Each publisher conducted a survey. In the survey, adult consumers were asked the number of fiction paperbacks they had purchased the previous month. The results are as follows:

Publisher A

# of books	Freq.	Rel. Freq.
0	10	
1	12	
2	16	
3	12	
4	8	
5	6	
6	2	
8	2	

Publisher B

# of books	Freq.	Rel. Freq.
0	18	
1	24	

# of books	Freq.	Rel. Freq.
2	24	
3	22	
4	15	
5	10	
7	5	
9	1	

Publisher C

# of books	Freq.	Rel. Freq.
0–1	20	
2–3	35	
4–5	12	
6–7	2	
8–9	1	

- Find the relative frequencies for each survey. Write them in the charts.
- Using either a graphing calculator, computer, or by hand, use the frequency column to construct a histogram for each publisher's survey. For Publishers A and B, make bar widths of one. For Publisher C, make bar widths of two.
- In complete sentences, give two reasons why the graphs for Publishers A and B are not identical.
- Would you have expected the graph for Publisher C to look like the other two graphs? Why or why not?
- Make new histograms for Publisher A and Publisher B. This time, make bar widths of two.
- Now, compare the graph for Publisher C to the new graphs for Publishers A and B. Are the graphs more similar or more different? Explain your answer.

### Q 2.3.2

Often, cruise ships conduct all on-board transactions, with the exception of gambling, on a cashless basis. At the end of the cruise, guests pay one bill that covers all onboard transactions. Suppose that 60 single travelers and 70 couples were surveyed as to their on-board bills for a seven-day cruise from Los Angeles to the Mexican Riviera. Following is a summary of the bills for each group.

Singles

Amount(\$)	Frequency	Rel. Frequency
51–100	5	
101–150	10	
151–200	15	
201–250	15	
251–300	10	
301–350	5	

Couples

Amount(\$)	Frequency	Rel. Frequency
100–150	5	

Amount(\$)	Frequency	Rel. Frequency
201–250	5	
251–300	5	
301–350	5	
351–400	10	
401–450	10	
451–500	10	
501–550	10	
551–600	5	
601–650	5	

- Fill in the relative frequency for each group.
- Construct a histogram for the singles group. Scale the  $x$ -axis by \$50 widths. Use relative frequency on the  $y$ -axis.
- Construct a histogram for the couples group. Scale the  $x$ -axis by \$50 widths. Use relative frequency on the  $y$ -axis.
- Compare the two graphs:
  - List two similarities between the graphs.
  - List two differences between the graphs.
  - Overall, are the graphs more similar or different?
- Construct a new graph for the couples by hand. Since each couple is paying for two individuals, instead of scaling the  $x$ -axis by \$50, scale it by \$100. Use relative frequency on the  $y$ -axis.
- Compare the graph for the singles with the new graph for the couples:
  - List two similarities between the graphs.
  - Overall, are the graphs more similar or different?
- How did scaling the couples graph differently change the way you compared it to the singles graph?
- Based on the graphs, do you think that individuals spend the same amount, more or less, as singles as they do person by person as a couple? Explain why in one or two complete sentences.

### S 2.3.2

Singles

Amount(\$)	Frequency	Relative Frequency
51–100	5	0.08
101–150	10	0.17
151–200	15	0.25
201–250	15	0.25
251–300	10	0.17
301–350	5	0.08

Couples

Amount(\$)	Frequency	Relative Frequency
100–150	5	0.07
201–250	5	0.07
251–300	5	0.07



Amount(\$)	Frequency	Relative Frequency
301–350	5	0.07
351–400	10	0.14
401–450	10	0.14
451–500	10	0.14
501–550	10	0.14
551–600	5	0.07
601–650	5	0.07

- a. See the tables above
- b. In the following histogram data values that fall on the right boundary are counted in the class interval, while values that fall on the left boundary are not counted (with the exception of the first interval where both boundary values are included).

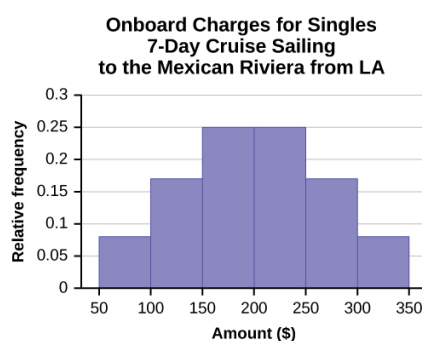


Figure 2.E. 6:

- c. In the following histogram, the data values that fall on the right boundary are counted in the class interval, while values that fall on the left boundary are not counted (with the exception of the first interval where values on both boundaries are included).

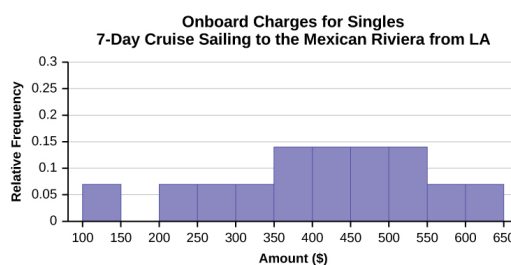


Figure 2.E. 7:

- d. Compare the two graphs:
- Answers may vary. Possible answers include:
    - Both graphs have a single peak.
    - Both graphs use class intervals with width equal to \$50.
  - Answers may vary. Possible answers include:
    - The couples graph has a class interval with no values.
    - It takes almost twice as many class intervals to display the data for couples.
  - Answers may vary. Possible answers include: The graphs are more similar than different because the overall patterns for the graphs are the same.
- e. Check student's solution.

- f. Compare the graph for the Singles with the new graph for the Couples:
- Both graphs have a single peak.
  - Both graphs display 6 class intervals.
  - Both graphs show the same general pattern.
- ii. Answers may vary. Possible answers include: Although the width of the class intervals for couples is double that of the class intervals for singles, the graphs are more similar than they are different.
- g. Answers may vary. Possible answers include: You are able to compare the graphs interval by interval. It is easier to compare the overall patterns with the new scale on the Couples graph. Because a couple represents two individuals, the new scale leads to a more accurate comparison.
- h. Answers may vary. Possible answers include: Based on the histograms, it seems that spending does not vary much from singles to individuals who are part of a couple. The overall patterns are the same. The range of spending for couples is approximately double the range for individuals.

### Q 2.3.3

Twenty-five randomly selected students were asked the number of movies they watched the previous week. The results are as follows.

# of movies	Frequency	Relative Frequency	Cumulative Relative Frequency
0	5		
1	9		
2	6		
3	4		
4	1		

- Construct a histogram of the data.
- Complete the columns of the chart.

Use the following information to answer the next two exercises: Suppose one hundred eleven people who shopped in a special t-shirt store were asked the number of t-shirts they own costing more than \$19 each.

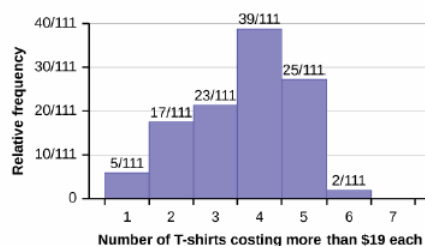


Figure 2.E. 8:

### Q 2.3.4

The percentage of people who own at most three t-shirts costing more than \$19 each is approximately:

- 21
- 59
- 41
- Cannot be determined

### S 2.3.4

c

### Q 2.3.5

If the data were collected by asking the first 111 people who entered the store, then the type of sampling is:

- cluster
- simple random
- stratified
- convenience

### Q 2.3.6

Following are the 2010 obesity rates by U.S. states and Washington, DC.

State	Percent (%)	State	Percent (%)	State	Percent (%)
Alabama	32.2	Kentucky	31.3	North Dakota	27.2
Alaska	24.5	Louisiana	31.0	Ohio	29.2
Arizona	24.3	Maine	26.8	Oklahoma	30.4
Arkansas	30.1	Maryland	27.1	Oregon	26.8
California	24.0	Massachusetts	23.0	Pennsylvania	28.6
Colorado	21.0	Michigan	30.9	Rhode Island	25.5
Connecticut	22.5	Minnesota	24.8	South Carolina	31.5
Delaware	28.0	Mississippi	34.0	South Dakota	27.3
Washington, DC	22.2	Missouri	30.5	Tennessee	30.8
Florida	26.6	Montana	23.0	Texas	31.0
Georgia	29.6	Nebraska	26.9	Utah	22.5
Hawaii	22.7	Nevada	22.4	Vermont	23.2
Idaho	26.5	New Hampshire	25.0	Virginia	26.0
Illinois	28.2	New Jersey	23.8	Washington	25.5
Indiana	29.6	New Mexico	25.1	West Virginia	32.5
Iowa	28.4	New York	23.9	Wisconsin	26.3
Kansas	29.4	North Carolina	27.8	Wyoming	25.1

Construct a bar graph of obesity rates of your state and the four states closest to your state. Hint: Label the  $x$ -axis with the states.

### S 2.3.7

Answers will vary.

### Exercise 2.3.6

Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars. Complete the table.

Data Value (# cars)	Frequency	Relative Frequency	Cumulative Relative Frequency

Data Value (# cars)	Frequency	Relative Frequency	Cumulative Relative Frequency

### Exercise 2.3.7

What does the frequency column in the Table above sum to? Why?

**Answer**

65

### Exercise 2.3.8

What does the relative frequency column in the Table above sum to? Why?

### Exercise 2.3.9

What is the difference between relative frequency and frequency for each data value in the Table above?

**Answer**

The relative frequency shows the *proportion* of data points that have each value. The frequency tells the *number* of data points that have each value.

### Exercise 2.3.10

What is the difference between cumulative relative frequency and relative frequency for each data value?

### Exercise 2.3.11

To construct the histogram for the data in the Table above, determine appropriate minimum and maximum  $x$  and  $y$  values and the scaling. Sketch the histogram. Label the horizontal and vertical axes with words. Include numerical scaling.

Figure 2.E. 9:

**Answer**

Answers will vary. One possible histogram is shown:

Figure 2.E. 10:

### Exercise 2.3.12

Construct a frequency polygon for the following:

a.	Pulse Rates for Women	Frequency
	60–69	12
	70–79	14
	80–89	11
	90–99	1
	100–109	1

Pulse Rates for Women	Frequency
110–119	0
120–129	1

b.

Actual Speed in a 30 MPH Zone	Frequency
42–45	25
46–49	14
50–53	7
54–57	3
58–61	1

c.

Tar (mg) in Nonfiltered Cigarettes	Frequency
10–13	1
14–17	0
18–21	15
22–25	7
26–29	2

### Exercise 2.3.13

Construct a frequency polygon from the frequency distribution for the 50 highest ranked countries for depth of hunger.

Depth of Hunger	Frequency
230–259	21
260–289	13
290–319	5
320–349	7
350–379	1
380–409	1
410–439	1

#### Answer

Find the midpoint for each class. These will be graphed on the x-axis. The frequency values will be graphed on the y-axis values.

Figure 2.E. 11:

### Exercise 2.3.14

Use the two frequency tables to compare the life expectancy of men and women from 20 randomly selected countries. Include an overlaid frequency polygon and discuss the shapes of the distributions, the center, the spread, and any outliers. What can we conclude about the life expectancy of women compared to men?

Life Expectancy at Birth – Women	Frequency
49–55	3
56–62	3
63–69	1
70–76	3
77–83	8
84–90	2

Life Expectancy at Birth – Men	Frequency
49–55	3
56–62	3
63–69	1
70–76	1
77–83	7
84–90	5

### Exercise 2.3.15

Construct a times series graph for (a) the number of male births, (b) the number of female births, and (c) the total number of births.

Sex/Year	1855	1856	1857	1858	1859	1860	1861
Female	45,545	49,582	50,257	50,324	51,915	51,220	52,403
Male	47,804	52,239	53,158	53,694	54,628	54,409	54,606
Total	93,349	101,821	103,415	104,018	106,543	105,629	107,009

Sex/Year	1862	1863	1864	1865	1866	1867	1868	1869
Female	51,812	53,115	54,959	54,850	55,307	55,527	56,292	55,033
Male	55,257	56,226	57,374	58,220	58,360	58,517	59,222	58,321
Total	107,069	109,341	112,333	113,070	113,667	114,044	115,514	113,354

Sex/Year	1871	1870	1872	1871	1872	1872	1874	1875
Female	56,099	56,431	57,472	56,099	57,472	58,233	60,109	60,146
Male	60,029	58,959	61,293	60,029	61,293	61,467	63,602	63,432
Total	116,128	115,390	118,765	116,128	118,765	119,700	123,711	123,578

Answer

Figure 2.E. 12:

### Exercise 2.3.16

The following data sets list full time police per 100,000 citizens along with homicides per 100,000 citizens for the city of Detroit, Michigan during the period from 1961 to 1973.

Year	1961	1962	1963	1964	1965	1966	1967
Police	260.35	269.8	272.04	272.96	272.51	261.34	268.89
Homicides	8.6	8.9	8.52	8.89	13.07	14.57	21.36

Year	1968	1969	1970	1971	1972	1973
Police	295.99	319.87	341.43	356.59	376.69	390.19
Homicides	28.03	31.49	37.39	46.26	47.24	52.33

- Construct a double time series graph using a common x-axis for both sets of data.
- Which variable increased the fastest? Explain.
- Did Detroit's increase in police officers have an impact on the murder rate? Explain.

## 2.4: Measures of the Location of the Data

### Q 2.4.1

The median age for U.S. blacks currently is 30.9 years; for U.S. whites it is 42.3 years.

- Based upon this information, give two reasons why the black median age could be lower than the white median age.
- Does the lower median age for blacks necessarily mean that blacks die younger than whites? Why or why not?
- How might it be possible for blacks and whites to die at approximately the same age, but for the median age for whites to be higher?

### Q 2.4.2

Six hundred adult Americans were asked by telephone poll, "What do you think constitutes a middle-class income?" The results are in the Table below. Also, include left endpoint, but not the right endpoint.

Salary (\$)	Relative Frequency
< 20,000	0.02
20,000–25,000	0.09
25,000–30,000	0.19
30,000–40,000	0.26
40,000–50,000	0.18
50,000–75,000	0.17
75,000–99,999	0.02
100,000+	0.01

- What percentage of the survey answered "not sure"?
- What percentage think that middle-class is from \$25,000 to \$50,000?
- Construct a histogram of the data.
  - Should all bars have the same width, based on the data? Why or why not?
  - How should the <20,000 and the 100,000+ intervals be handled? Why?
- Find the 40<sup>th</sup> and 80<sup>th</sup> percentiles
- Construct a bar graph of the data

### S 2.4.2

- $1 - (0.02 + 0.09 + 0.19 + 0.26 + 0.18 + 0.17 + 0.02 + 0.01) = 0.06$
- $0.19 + 0.26 + 0.18 = 0.63$
- Check student's solution.
- 40<sup>th</sup> percentile will fall between 30,000 and 40,000

80<sup>th</sup> percentile will fall between 50,000 and 75,000

- Check student's solution.

### Q 2.4.3

Given the following box plot:

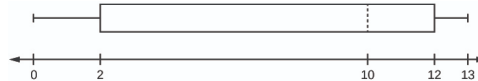


Figure 2.E. 13:

- which quarter has the smallest spread of data? What is that spread?
- which quarter has the largest spread of data? What is that spread?
- find the interquartile range (*IQR*).
- are there more data in the interval 5–10 or in the interval 10–13? How do you know this?
- which interval has the fewest data in it? How do you know this?
  - 0–2
  - 2–4
  - 10–12
  - 12–13
  - need more information

### Q 2.4.4

The following box plot shows the U.S. population for 1990, the latest available year.

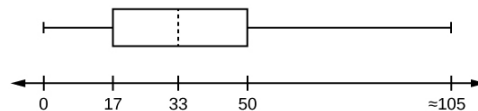


Figure 2.E. 14:

- Are there fewer or more children (age 17 and under) than senior citizens (age 65 and over)? How do you know?
- 12.6% are age 65 and over. Approximately what percentage of the population are working age adults (above age 17 to age 65)?

### S 2.4.4

- more children; the left whisker shows that 25% of the population are children 17 and younger. The right whisker shows that 25% of the population are adults 50 and older, so adults 65 and over represent less than 25%.
- 62.4%

## 2.5: Box Plots

### Q 2.5.1

In a survey of 20-year-olds in China, Germany, and the United States, people were asked the number of foreign countries they had visited in their lifetime. The following box plots display the results.



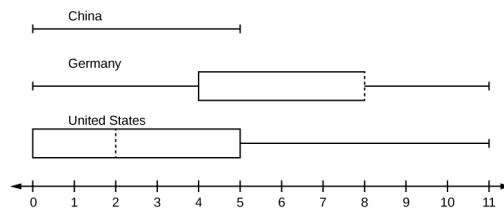


Figure 2.E. 15:

- In complete sentences, describe what the shape of each box plot implies about the distribution of the data collected.
- Have more Americans or more Germans surveyed been to over eight foreign countries?
- Compare the three box plots. What do they imply about the foreign travel of 20-year-old residents of the three countries when compared to each other?

### Q 2.5.2

Given the following box plot, answer the questions.

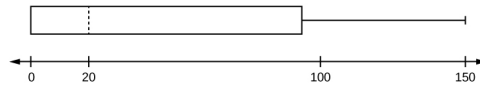


Figure 2.E. 16:

- Think of an example (in words) where the data might fit into the above box plot. In 2–5 sentences, write down the example.
- What does it mean to have the first and second quartiles so close together, while the second to third quartiles are far apart?

### S 2.5.2

- Answers will vary. Possible answer: State University conducted a survey to see how involved its students are in community service. The box plot shows the number of community service hours logged by participants over the past year.
- Because the first and second quartiles are close, the data in this quarter is very similar. There is not much variation in the values. The data in the third quarter is much more variable, or spread out. This is clear because the second quartile is so far away from the third quartile.

### Q 2.5.3

Given the following box plots, answer the questions.

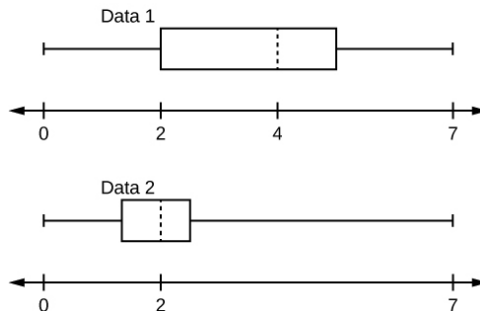


Figure 2.E. 17:

- In complete sentences, explain why each statement is false.
  - Data 1** has more data values above two than **Data 2** has above two.
  - The data sets cannot have the same mode.
  - For **Data 1**, there are more data values below four than there are above four.
- For which group, Data 1 or Data 2, is the value of “7” more likely to be an outlier? Explain why in complete sentences.

### Q 2.5.4

A survey was conducted of 130 purchasers of new BMW 3 series cars, 130 purchasers of new BMW 5 series cars, and 130 purchasers of new BMW 7 series cars. In it, people were asked the age they were when they purchased their car. The following box plots display the results.

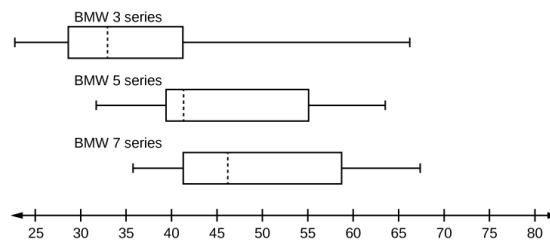


Figure 2.E. 18:

- In complete sentences, describe what the shape of each box plot implies about the distribution of the data collected for that car series.
- Which group is most likely to have an outlier? Explain how you determined that.
- Compare the three box plots. What do they imply about the age of purchasing a BMW from the series when compared to each other?
- Look at the BMW 5 series. Which quarter has the smallest spread of data? What is the spread?
- Look at the BMW 5 series. Which quarter has the largest spread of data? What is the spread?
- Look at the BMW 5 series. Estimate the interquartile range (IQR).
- Look at the BMW 5 series. Are there more data in the interval 31 to 38 or in the interval 45 to 55? How do you know this?
- Look at the BMW 5 series. Which interval has the fewest data in it? How do you know this?
  - 31–35
  - 38–41
  - 41–64

#### S 2.5.4

- Each box plot is spread out more in the greater values. Each plot is skewed to the right, so the ages of the top 50% of buyers are more variable than the ages of the lower 50%.
- The BMW 3 series is most likely to have an outlier. It has the longest whisker.
- Comparing the median ages, younger people tend to buy the BMW 3 series, while older people tend to buy the BMW 7 series. However, this is not a rule, because there is so much variability in each data set.
- The second quarter has the smallest spread. There seems to be only a three-year difference between the first quartile and the median.
- The third quarter has the largest spread. There seems to be approximately a 14-year difference between the median and the third quartile.
- $IQR \sim 17$  years
- There is not enough information to tell. Each interval lies within a quarter, so we cannot tell exactly where the data in that quarter is concentrated.
- The interval from 31 to 35 years has the fewest data values. Twenty-five percent of the values fall in the interval 38 to 41, and 25% fall between 41 and 64. Since 25% of values fall between 31 and 38, we know that fewer than 25% fall between 31 and 35.

#### Q 2.5.5

Twenty-five randomly selected students were asked the number of movies they watched the previous week. The results are as follows:

# of movies	Frequency
0	5
1	9
2	6
3	4
4	1

Construct a box plot of the data.

## 2.6: Measures of the Center of the Data

### Q 2.6.1

The most obese countries in the world have obesity rates that range from 11.4% to 74.6%. This data is summarized in the following table.

Percent of Population Obese	Number of Countries
11.4–20.45	29
20.45–29.45	13
29.45–38.45	4
38.45–47.45	0
47.45–56.45	2
56.45–65.45	1
65.45–74.45	0
74.45–83.45	1

- What is the best estimate of the average obesity percentage for these countries?
- The United States has an average obesity rate of 33.9%. Is this rate above average or below?
- How does the United States compare to other countries?

### Q 2.6.2

The table below gives the percent of children under five considered to be underweight. What is the best estimate for the mean percentage of underweight children?

Percent of Underweight Children	Number of Countries
16–21.45	23
21.45–26.9	4
26.9–32.35	9
32.35–37.8	7
37.8–43.25	6
43.25–48.7	1

### S 2.6.2

The mean percentage,  $\bar{x} = \frac{1328.65}{50} = 26.75$

## 2.7: Skewness and the Mean, Median, and Mode

### Q 2.7.1

The median age of the U.S. population in 1980 was 30.0 years. In 1991, the median age was 33.1 years.

- What does it mean for the median age to rise?
- Give two reasons why the median age could rise.
- For the median age to rise, is the actual number of children less in 1991 than it was in 1980? Why or why not?

## 2.8: Measures of the Spread of the Data

Use the following information to answer the next nine exercises: The population parameters below describe the full-time equivalent number of students (FTES) each year at Lake Tahoe Community College from 1976–1977 through 2004–2005.

- $\mu = 1000$  FTES
- median = 1,014 FTES
- $\sigma = 474$  FTES
- first quartile = 528.5 FTES
- third quartile = 1,447.5 FTES
- $n = 29$  years

#### Q 2.8.1

A sample of 11 years is taken. About how many are expected to have a FTES of 1014 or above? Explain how you determined your answer.

#### S 2.8.1

The median value is the middle value in the ordered list of data values. The median value of a set of 11 will be the 6th number in order. Six years will have totals at or below the median.

#### Q 2.8.2

75% of all years have an FTES:

- a. at or below: \_\_\_\_\_
- b. at or above: \_\_\_\_\_

#### Q 2.8.3

The population standard deviation = \_\_\_\_\_

#### S 2.8.3

474 FTES

#### Q 2.8.4

What percent of the FTES were from 528.5 to 1447.5? How do you know?

#### Q 2.8.5

What is the *IQR*? What does the *IQR* represent?

#### S 2.8.5

919

#### Q 2.8.6

How many standard deviations away from the mean is the median?

*Additional Information:* The population FTES for 2005–2006 through 2010–2011 was given in an updated report. The data are reported here.

Year	2005–06	2006–07	2007–08	2008–09	2009–10	2010–11
Total FTES	1,585	1,690	1,735	1,935	2,021	1,890

#### Q 2.8.7

Calculate the mean, median, standard deviation, the first quartile, the third quartile and the *IQR*. Round to one decimal place.

#### S 2.8.7

- mean = 1,809.3
- median = 1,812.5
- standard deviation = 151.2
- first quartile = 1,690
- third quartile = 1,935

- $IQR = 245$

### Q 2.8.8

Construct a box plot for the FTES for 2005–2006 through 2010–2011 and a box plot for the FTES for 1976–1977 through 2004–2005.

### Q 2.8.9

Compare the  $IQR$  for the FTES for 1976–77 through 2004–2005 with the  $IQR$  for the FTES for 2005–2006 through 2010–2011. Why do you suppose the  $IQR$ s are so different?

### S 2.8.10

Hint: Think about the number of years covered by each time period and what happened to higher education during those periods.

### Q 2.8.11

Three students were applying to the same graduate school. They came from schools with different grading systems. Which student had the best GPA when compared to other students at his school? Explain how you determined your answer.

Student	GPA	School Average GPA	School Standard Deviation
Thuy	2.7	3.2	0.8
Vichet	87	75	20
Kamala	8.6	8	0.4

### Q 2.8.12

A music school has budgeted to purchase three musical instruments. They plan to purchase a piano costing \$3,000, a guitar costing \$550, and a drum set costing \$600. The mean cost for a piano is \$4,000 with a standard deviation of \$2,500. The mean cost for a guitar is \$500 with a standard deviation of \$200. The mean cost for drums is \$700 with a standard deviation of \$100. Which cost is the lowest, when compared to other instruments of the same type? Which cost is the highest when compared to other instruments of the same type. Justify your answer.

### S 2.8.12

For pianos, the cost of the piano is 0.4 standard deviations BELOW the mean. For guitars, the cost of the guitar is 0.25 standard deviations ABOVE the mean. For drums, the cost of the drum set is 1.0 standard deviations BELOW the mean. Of the three, the drums cost the lowest in comparison to the cost of other instruments of the same type. The guitar costs the most in comparison to the cost of other instruments of the same type.

### Q 2.8.13

An elementary school class ran one mile with a mean of 11 minutes and a standard deviation of three minutes. Rachel, a student in the class, ran one mile in eight minutes. A junior high school class ran one mile with a mean of nine minutes and a standard deviation of two minutes. Kenji, a student in the class, ran 1 mile in 8.5 minutes. A high school class ran one mile with a mean of seven minutes and a standard deviation of four minutes. Nedda, a student in the class, ran one mile in eight minutes.

- Why is Kenji considered a better runner than Nedda, even though Nedda ran faster than he?
- Who is the fastest runner with respect to his or her class? Explain why.

### Q 2.8.14

The most obese countries in the world have obesity rates that range from 11.4% to 74.6%. This data is summarized in the table below

Percent of Population Obese	Number of Countries
11.4–20.45	29
20.45–29.45	13

Percent of Population Obese	Number of Countries
29.45–38.45	4
38.45–47.45	0
47.45–56.45	2
56.45–65.45	1
65.45–74.45	0
74.45–83.45	1

What is the best estimate of the average obesity percentage for these countries? What is the standard deviation for the listed obesity rates? The United States has an average obesity rate of 33.9%. Is this rate above average or below? How “unusual” is the United States’ obesity rate compared to the average rate? Explain.

#### S 2.8.14

- $\bar{x} = 23.32$
- Using the TI 83/84, we obtain a standard deviation of:  $s_x = 12.95$ .
- The obesity rate of the United States is 10.58% higher than the average obesity rate.
- Since the standard deviation is 12.95, we see that  $23.32 + 12.95 = 36.27$  is the obesity percentage that is one standard deviation from the mean. The United States obesity rate is slightly less than one standard deviation from the mean. Therefore, we can assume that the United States, while 34% obese, does not have an unusually high percentage of obese people.

#### Q 2.8.15

The Table below gives the percent of children under five considered to be underweight.

Percent of Underweight Children	Number of Countries
16–21.45	23
21.45–26.9	4
26.9–32.35	9
32.35–37.8	7
37.8–43.25	6
43.25–48.7	1

What is the best estimate for the mean percentage of underweight children? What is the standard deviation? Which interval(s) could be considered unusual? Explain.

This page titled [2.E: Descriptive Statistics \(Exercises\)](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.