Chapter 2
Displaying and Describing Data

Solutions to Class Examples:

1. Answers will vary according to your data.

2. Answers will vary according to your data.

3. See Class Example 3.

4. Answers will vary according to your data.

5. See Class Example 5.

Investigative Task

The task, Dollars for Students, examines education spending in the United States.

Statistics Quiz A—Chapter 2   *Name* _____

1. Students in a statistics course were asked to answer a variety of questions. One question asked for students to pick a favorite Superpower. Results are listed to the right.

   a. Create a visual display for these data.

Superpower	Num of Students
Fly	19
Freeze Time	11
Invisibility	7
Super strength	4
Telepathy	12

   b. Describe the distribution of the students' choices.

2. Which of the following variables would be appropriate to graph using a pie or bar graph?

   a. Annual income for 20 employees   Yes   or   No

   b. The favorite baseball team of 30 students   Yes   or   No

   c. The number of pages in 15 textbooks   Yes   or   No

   d. The country of origin of 25 immigrants   Yes   or   No

3. Describe why the area principle is important in making a bar graph. It might be fun to ask an artist to liven up a bar graph by turning the bars into images. But include in your explanation why this might be risky.

Statistics Quiz A—Chapter 2 KEY

1. Students in a statistics course were asked to answer a variety of questions. One question asked for students to pick a favorite Superpower. Results are listed to the right.

Superpower	Num of Students
Fly	19
Freeze Time	11
Invisibility	7
Super strength	4
Telepathy	12

   a. Create a visual display for these data.
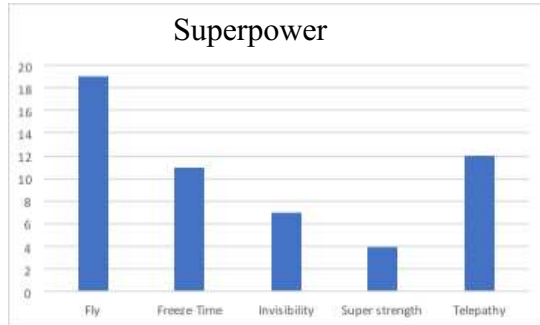
   b. Describe the distribution of the students' choices.

> The ability to fly is the students' most popular choice, followed by telepathy and freeze time. Super strength is the least popular, with invisibility also not very popular.

2. Which of the following variables would be appropriate to graph using a pie or bar graph?

   a. Annual income for 20 employees   Yes   or   **(No)**

   b. The favorite baseball team of 30 students   **(Yes)**   or   No

   c. The number of pages in 15 textbooks   Yes   or   **(No)**

   d. The country of origin of 25 immigrants   **(Yes)**   or   No

3. Describe why the area principle is important in making a bar graph. It might be fun to ask an artist to liven up a bar graph by turning the bars into images. But include in your explanation why this might be risky.

> If the reader of the graphs is going to understand the relationship between the groups, it is important that the area allotted to each group match the proportion of group in the sample. If an artist turns the bars into some kind of image, he might be not make each picture the proper area, and thus distort the actual percentage breakdown of each group.

Statistics Quiz B—Chapter 2   *Name* _____

1. A survey conducted in a college intro stats class asked students about the number of credit hours they were taking that quarter. The number of credit hours for a random sample of 16 students is given in the table.

10	10	12	14	15	15	15	15
17	17	19	20	20	20	20	22

    a. Sketch a histogram of these data

    b. Find the mean and standard deviation for the number of credit hours.

    c. Find the median and IQR for the number of credit hours.

    d. Is it more appropriate to use the mean and standard deviation or the median and IQR to summarize theses data? Explain.

2. Suppose that the student taking 22 credit hours in the data set in the previous question was actually taking 28 credit hours instead of 22 (so we would replace the 22 in the data set with 28). Indicate whether changing the number of credit hours for that student would make each of the following summary statistics increase, decrease, or stay about the same:

    a. mean   _____

    b. median   _____

    c. range   _____

    d. IQR   _____

    e. standard deviation   _____
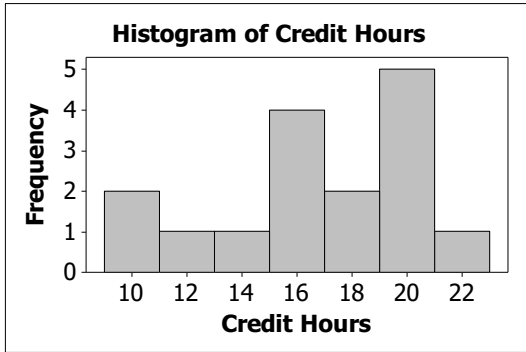
Statistics Quiz B—Chapter 2—KEY   *Name* _____

1. A survey conducted in a college intro stats class
    asked students about the number of credit hours
    they were taking that quarter. The number of credit
    hours for a random sample of 16 students is given
    in the table.

10	10	12	14	15	15	15	15
17	17	19	20	20	20	20	22

a. Sketch a histogram of these data

b. Find the mean and standard deviation
   for the number of credit hours.

$\bar{x} = 16.3$ credit hours; $s = 3.7$ credit hours

c. Find the median and IQR for the number
   of credit hours.

The median is 16.0 credit hours.
IQR = Q3 – Q1 = 20 – 14.5 = 5.5 credit
hours

d. Is it more appropriate to use the mean
   and standard deviation or the median
   and IQR to summarize theses data?
   Explain.

It is more appropriate to use the median
and IQR to summarize these data, because
these data are not unimodal and
symmetric.

2. Suppose that the student taking 22 credit hours in the data set in the previous question was
    actually taking 28 credit hours instead of 22 (so we would replace the 22 in the data set with
    28). Indicate whether changing the number of credit hours for that student would make each
    of the following summary statistics increase, decrease, or stay about the same:

    a. mean   increase

    b. median   stay about the same

    c. range   increase

    d. IQR   stay about the same

    e. standard deviation   increase

Statistics Quiz C—Chapter 2   *Name* _____

1. The students in a biology class kept a record of the height
    (in centimeters) of plants for a class experiment.

49	67	38	55	62
54	36	41	56	43
48	75	44	60	48
52	48	53	59	32

   a. Sketch a histogram for these data.   b. Find the mean and standard deviation of
                                                     the plant heights.

   c. Is it appropriate to use the mean and   d. Describe the distribution of plant heights.
      standard deviation to summarize these
      data? Explain.

2. All students in a physical education class completed a basketball free-throw shooting event
    and the highest number of shots made was 32. The next day a student who had just
    transferred into the school completed the event, making 35 shots. Indicate whether adding the
    new student's score to the rest of the data made each of these summary statistics increase,
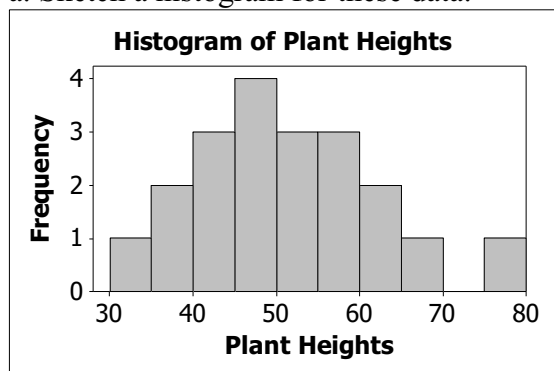   decrease, or stay about the same:

    a. mean   _____

    b. median   _____

    c. range   _____

    d. IQR   _____

    e. standard deviation   _____

Statistics Quiz C—Chapter 2—KEY

1. The students in a biology class kept a record of the height
    (in centimeters) of plants for a class experiment.

49	67	38	55	62
54	36	41	56	43
48	75	44	60	48
52	48	53	59	32

a. Sketch a histogram for these data.

b. Find the mean and standard deviation of the plant heights.

$$\bar{x} = 51.0 \text{ cm}; \quad s = 10.6 \text{ cm}$$

c. Is it appropriate to use the mean and standard deviation to summarize these data? Explain.

Yes, the data are roughly unimodal and symmetric with no outliers.

d. Describe the distribution of plant heights.

The data are roughly symmetric with no outliers; however there is a small gap from 70 to 75 cm. The average plant height is 51.0 centimeters, with a standard deviation of 10.6 centimeters. The range of plant heights is 43 centimeters. The distribution of plant heights has a mode between 45 and 49 centimeters.

2. All students in a physical education class completed a basketball free-throw shooting event and the highest number of shots made was 32. The next day a student who had just transferred into the school completed the event, making 35 shots. Indicate whether adding the new student's score to the rest of the data made each of these summary statistics increase, decrease, or stay about the same:

a. mean   increase

b. median   stay about the same

c. range   increase

d. IQR   stay about the same

e. standard deviation   increase

Statistics Quiz D—Chapter 2   *Name* _____

1. A brake and muffler shop reported the repair bills, in
    dollars, for their customers yesterday.

88	283	312	290	172
154	400	381	346	181
203	118	143	252	227
56	192	292	213	422

    a. Sketch a histogram for these data.

    b. Find the mean and standard deviation of
       the repair costs.

    c. Is it appropriate to use the mean and
       standard deviation to summarize these
       data? Explain.

    d. Describe the distribution of repair costs.

2. In a survey of 1,500 millennials (ages 18-
    34), *Business Insider* asked which
    services were used to consume video
    content. Use the data on the right to
   construct an appropriate display **and**
    describe your graph.

Service	Percent
YouTube	81%
Netflix	79%
Hulu	37%
Amazon Prime	34%
HBO	21%
Crackle	12%
None	3%
Other	2%

Statistics Quiz D—Chapter 2—KEY

1. A brake and muffler shop reported the repair bills, in dollars, for their customers yesterday.

88	283	312	290	172
154	400	381	346	181
203	118	143	252	227
56	192	292	213	422

a. Sketch a histogram for these data.
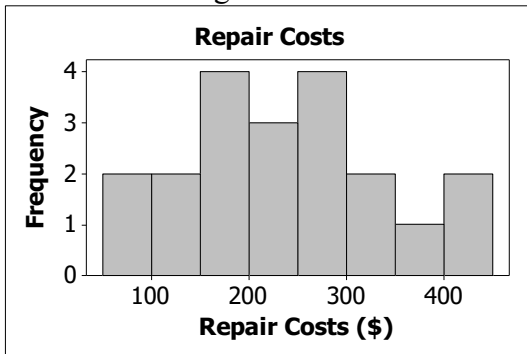
b. Find the mean and standard deviation of the repair costs.

$$\bar{x} = \$236.25; \quad s = \$103.43$$

c. Is it appropriate to use the mean and standard deviation to summarize these data? Explain.

Yes, the data are roughly unimodal and symmetric with no outliers.

d. Describe the distribution of repair costs.

The repair costs averaged $236.25, ranging from $56 to $422 with a standard deviation of $103.43. The distribution was approximately symmetric, with typical repair costs clustered between $150 and $300.

2. In a survey of 1,500 millennials (ages 18-34), *Business Insider* asked which services were used to consume video content. Use the data on the right to construct an appropriate display **and** describe your graph.

Service	Percent
YouTube	81%
Netflix	79%
Hulu	37%
Amazon Prime	34%
HBO	21%
Crackle	12%
None	3%
Other	2%
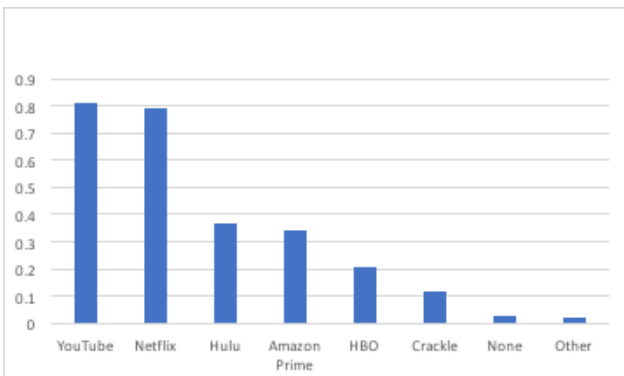
YouTube and Netflix are clearly the dominant choices. And as these percentages add to more than 100%, it is clear that most millennials use more than one service, making the pie chart an inappropriate choice. Hulu and Amazon are used by about one-third of the millennials, while Crackle and HBO are clearly the least popular.

Dollars for Students

In 2017 the U.S. Census Bureau published *Public Education Finances*, reporting the average amount (dollars per student) spent by public schools in each state during the 2015 school year. (The table seen here divides states according to whether they lie east or west of the Mississippi River).

Write a report describing the amounts states spent to educate their children.

Eastern State	$ per Student	Western State	$ per Student
Alabama	9,125	Alaska	20,172
Connecticut	18,377	Arizona	7,489
Delaware	14,120	Arkansas	9,694
D.C.	19,396	California	10,467
Florida	8,881	Colorado	9,245
Georgia	9,427	Hawaii	12,855
Illinois	13,755	Idaho	6,923
Indiana	9,687	Iowa	10,944
Kentucky	9,630	Kansas	10,040
Maine	13,257	Louisiana	11,010
Maryland	14,192	Minnesota	11,949
Massachusetts	15,592	Missouri	10,147
Michigan	11,482	Montana	11,028
Mississippi	8,456	Nebraska	11,946
New Hampshire	14,697	Nevada	8,615
New Jersey	18,235	New Mexico	9,752
New York	21,206	North Dakota	13,320
North Carolina	8,687	Oklahoma	8,082
Ohio	11,637	Oregon	10,442
Pennsylvania	14,717	South Dakota	8,937
Rhode Island	15,179	Texas	8,861
South Carolina	9,953	Utah	6,575
Tennessee	8,726	Washington	10,735
Vermont	18,039	Wyoming	16,055
Virginia	11,237		
West Virginia	11,359		
Wisconsin	11,375		

A complete report will include a visual display (stem-and-leaf plot), appropriate statistics, and a well-written description of the expenditures (in context, of course).

Statistics - Investigative Task   *Chapter 2*

	Components	Comments
Think	Demonstrates clear understanding of statistical concepts, vocabulary, and procedures in analyzing and describing these data.	

Show	**Visual/Numerical**:
	  o stem-and-leaf plot
	  o plot well-labeled
	  o plot correctly constructed
	  o summary statistics correct	

Tell	**Verbal**: Describes the distribution of expenditures *in context*, including…
	  o shape (skewed right by higher spending in the East)
	  o center (median)
	  o spread (IQR)
	The written analysis…
	  o also interprets at least one quartile, or the max or min *in context*
	  o identifies the W's
	  o uses statistical vocabulary correctly
	  o avoids speculation	

Components are scored as **E**ssentially correct, **P**artially correct, or **I**ncorrect

1: **Visual/Numerical**

      E—Has all 4 features

      P—Has only 3 of the 4 features, but attempts an appropriate graph (ex: histogram)

      I—Graph is not appropriate (ex: bar chart), has many errors, or is missing

2: **Shape**

      E—Identifies skewness to the right, attributed to generally higher spending in the East.

      P—Mis-identifies skewness OR overlooks East/West spending difference

      I—Description of shape is missing or incorrect

3: Center and Spread

      E—Correctly interprets median and IQR in context

      P—Correctly interprets only one of median/IQR OR lacks context OR uses mean and SD

      I— Has more than one of the three shortcomings described in "P"

4: Written Analysis

      E—Has all 4 listed properties.

      P—Has only 2 or 3 of the listed properties.

      I—Has fewer than of the properties.

Scoring

- E's count 1 point, P's are 1/2

- Grade: A = 4, B = 3, etc., with +/- based on rounding (ex: 3.5 rounded to 3 is a B+)

Name _____ *Grade* ____

NOTE: We present a model solution with some trepidation. <u>This is not a scoring key</u>, just an example. Many other approaches could fully satisfy the requirements outlined in the scoring rubric. That (not this) is the standard by which student responses should be evaluated.

Model Solution—Investigative Task—Dollars for Students

Dollars Spent per Student

The U.S. Census published *Public Education Finances*, a collection of educational information in 2017. Among the data reported was information on average educational spending by state during the 2015 school year, in dollars per student.

Spending by many states was near the nationwide median of $11010 per student. The middle 50% of states spent between $9245 and $14120 per student, an interquartile range of $4875. The distribution of the average number of education dollars spent per student was skewed to the right. Thirteen Eastern States spent in excess of $13,000 per student, while only one western state spent more than $13000. The maximum of $21206 was spent per student in New York. Utah spent the least, $6575 per student.

```
 6 | 69
 7 | 5
 8 | 15677999
 9 | 1246778
10 | 0014579
11 | 002445699
12 | 9
13 | 338
14 | 1277
15 | 26
16 | 1
17 |
18 | 024
19 | 4
20 | 2
21 | 2
```

Key: 6|7 = 6650 − 6749
Dollars spent per student

Statistics Quiz E—Chapter 2   *Name _____*

1. A automobile marketing firm conducts a study to see what types of cars people owned before buying an American car. The results are shown below.

Previous Ownership	Frequency
American	760
Japanese	375
Korean	72
German	37
Other	24
Total	**1268**

The relative frequency of those who owned Japanese cars previously who now bought American cars is

A. 59.9 %
B. 29.6%
C. 5.7%
D. 14.9%
E. 2.9%

2. A restaurant uses comment cards to get feedback from its customers about newly added items to the menu. It recently introduced homemade organic veggie burgers. Customers who tried the new burger were asked if they would order it again. The data are summarized in the table below. What percentage of customers would definitely order the veggie burger again?

Response	*Frequency*
Definitely would.	10
Most likely would.	40
Maybe	12
Definitely would not.	3

A. 10%
B. 15%
C. 20%
D. 40%
E. 77%

3. A restaurant uses comment cards to get feedback from its customers about newly added items to the menu. It recently introduced homemade organic veggie burgers. Customers who tried the new burger were asked if they would order it again. The data are summarized in the table below. What percentage of customers would most likely or definitely order the veggie burger again?

Response	Frequency
Definitely would.	10
Most likely would.	40
Maybe	12
Definitely would not.	3

A. 10%
B. 15%
C. 40%
D. 50%
E. 77%

4. A restaurant uses comment cards to get feedback from its customers about newly added items to the menu. It recently introduced homemade organic veggie burgers. Customers who tried the new burger were asked if they would order it again. Which of the following would be an appropriate method for displaying the data shown in the table?

Response	Frequency
Definitely would.	10
Most likely would.	40
Maybe	12
Definitely would not.	3

A. Histogram.
B. Dotplot.
C. Pie chart.
D. Stem-and-leaf display.
E. Both C and D.

5. Below is a histogram of salaries (in $) for a sample of U.S. marketing managers.

Histogram of Mktg Manager Salaries
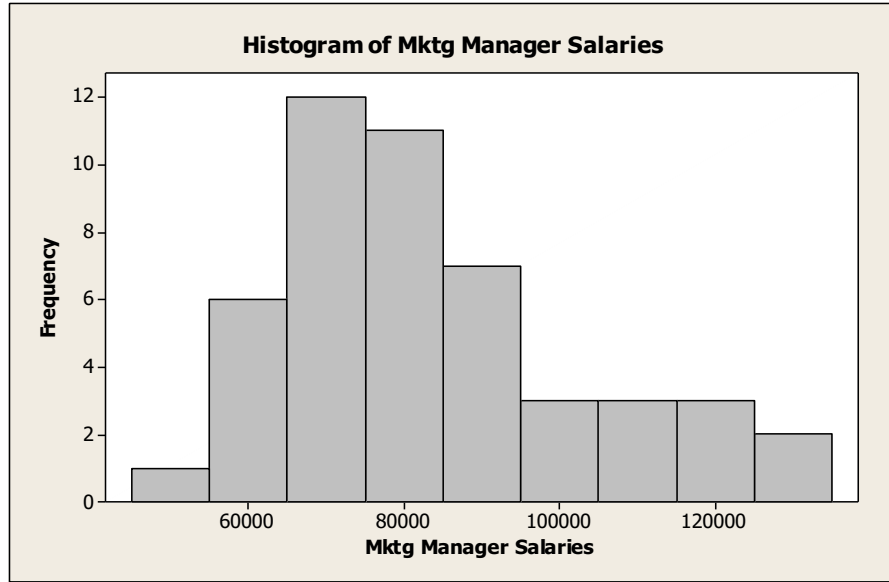
The shape of this distribution is
A. symmetric.
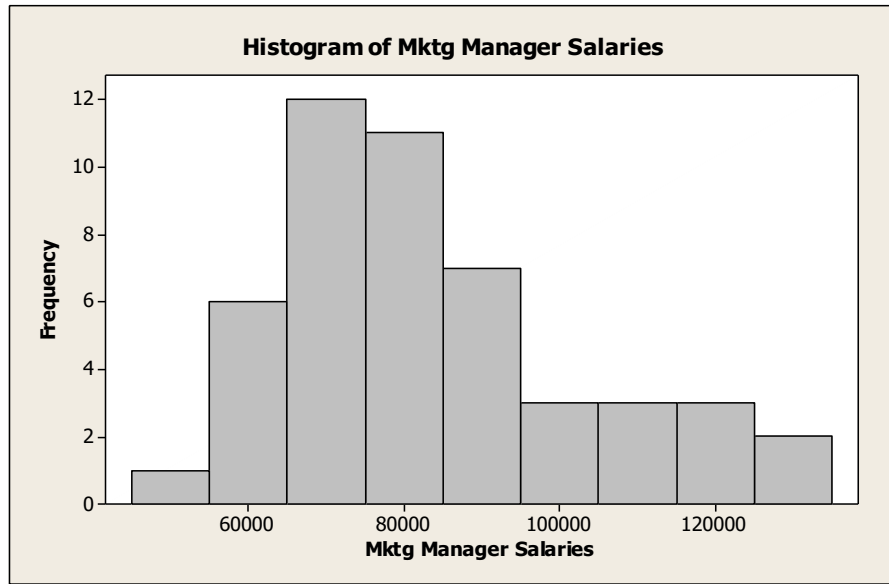B. bimodal.
C. right skewed.
D. left skewed.
E. normal.

6. Here is the five number summary for salaries of U.S. marketing managers.

Min	Q1	Median	Q3	Max
46360	69693	77020	91750	129420

The IQR is
A. $83,060.
B. $22.057.
C. $69,693.
D. $77.020.
E. $14,566.

7. Below is a histogram of salaries (in $) for a sample of U.S. marketing managers.

The most appropriate measure of central tendency for these data is the
A. median.
B. mean.
C. mode.
D. range.
E. standard deviation.

8. Consider the five number summary for salaries of U.S. marketing managers.

Min	Q1	Median	Q3	Max
46360	69693	77020	91750	129420

Suppose the marketing manager who was earning $129,420 got a raise and is now earning $140,000. Which of the following statement is true?
A. The mean would increase.
B. The median would increase.
C. The range would increase.
D. Both A and C.
E. All of the above.

9. The following table shows data for total assets ($ billion) for a small sample of U.S. banks (4th quarter 2018).

BANK	ASSETS ($ billion)
State Street Bank and Trust	231.0
Discover Bank	104.3
Bank of the West	87.6
Citizens Bank	126.9
Northern Trust	131.9
The Huntington National Bank	105.8
KeyBank	136.9
People's United	43.9

The mean for the total assets data ($ billion) is
A. $98.7.
B. $76.3.
C. $135.6.
D. $121.0.
E. $62.4.

10. The following table shows representative recent closing share prices for a small sample of companies based in India in January 2019.

COMPANY	CLOSING SHARE PRICE
20 Microns	42.00
Kuantum Papers	467.95
Bank of MA	14.65
Photoquip	35.65
Saksoft	284.35
Marg LTD	15.10
Galaxy ENT	29.15
Sonatasoft	305.00
EDynamics	0.67
DB Corp.	179.00

The standard deviation in closing share prices is
A. $163.5.
B. $25.8.
C. $136.6.
D. $62.7.
E. $267.6.

Statistics Quiz E—Chapter 2—KEY

1. B

2. B

3. E

4. C

5. C

6. B

7. A

8. D

9. D

10. A

Statistics Quiz F—Chapter 2   *Name* _____

1. A clothing store uses comment cards to get feedback from its customers about newly added items. It recently introduced plus size fashion wear. Customers who purchased the items were asked to fill out an online comment survey giving 10% off the next purchase. The data are summarized in the table below. What percentage of customers were at least satisfied with the item(s) purchased (Satisfied or Very satisfied)?

Response	Frequency
Very satisfied.	15
Satisfied.	30
Less than fully satisfied.	12
Not satisfied.	4

A. 49.2%
B. 73.8%
C. 24.6%
D. 26.2%
E. 68.9%

2. A clothing store uses comment cards to get feedback from its customers about newly added items. It recently introduced plus size fashion wear. Customers who purchased the items were asked to fill out an online comment survey giving 10% off the next purchase. The data are summarized in the table below. What percentage of customers were less than fully satisfied?

Response	Frequency
Very satisfied.	15
Satisfied.	30
Less than fully satisfied.	12
Not satisfied.	4

A. 26.2%
B. 6.5%
C. 75.4%
D. 49.2%
E. 19.6%

3. A clothing store uses comment cards to get feedback from its customers about newly added items. It recently introduced plus size fashion wear. Customers who purchased the items were asked to fill out an online comment survey giving 10% off the next purchase. The data are summarized in the table below. Which of the following would be an appropriate method for displaying the data shown in the table?

Response	Frequency
Very satisfied.	15
Satisfied.	30
Less than fully satisfied.	12
Not satisfied.	4

A. Histogram.
B. Dotplot.
C. Pie chart.
D. Stem-and-leaf display.
E. Both C and D.

4. The following table shows total assets ($ billion) for a small sample of U.S. banks in the 4th quarter of 2018.

BANK	ASSETS ($ billion)
State Street Bank and Trust	231.0
Discover Bank	104.3
Bank of the West	87.6
Citizens Bank	126.9
Northern Trust	131.9
The Huntington National Bank	105.8
KeyBank	136.9
People's United	43.9

The standard deviation for these data is
A. $ 53.55 billion.
B. $ 100.35 billion.
C. $ 75.68 billion.
D. $ 84.21 billion.
E. $ 89.79 billion.

5. The following table shows representative recent closing share prices for a small sample of companies based in India in January 2019.

COMPANY	*CLOSING SHARE PRICE*
20 Microns	42.00
Kuantum Papers	467.95
Bank of MA	14.65
Photoquip	35.65
Saksoft	284.35
Marg LTD	15.10
Galaxy ENT	29.15
Sonatasoft	305.00
EDynamics	0.67
DB Corp.	179.00

The mean closing share prices is
A. $12.78

B. $ 137.35
C. $ 38.83
D. $ 284.35
E. $ 163.50

6. Consider the five number summary of hourly wages ($) for a sample of sales managers.

Min	Q1	Median	Q3	Max
20.94	37.64	44.77	49.34	67.11

The range for these data is
A. $11.70
B. $46.17
C. $67.11
D. $20.94
E. $44.77

7. Consider the five number summary of hourly wages ($) for a sample of sales managers.

Min	Q1	Median	Q3	Max
20.94	37.64	44.77	49.34	67.11

The IQR for these data is
A. $11.70
B. $46.17
C. $67.11
D. $20.94
E. $44.77

8. Consider the five number summary of hourly wages ($) for a sample of sales managers. Suppose the mean hourly wage is $38.50. What can we say about the shape of the distribution?

Min	Q1	Median	Q3	Max
20.94	37.64	44.77	49.34	67.11

A. The distribution of hourly wages for sales managers is symmetric.
B. The distribution of hourly wages for sales managers is skewed right.
C. The distribution of hourly wages for sales managers is skewed left.
D. The distribution of hourly wages for sales managers is bimodal.
E. None of the above.

9. Consider the five number summary of hourly wages ($) for a sample of advertising / promotion managers.

Min	Q1	Median	Q3	Max
19.64	29.36	34.18	40.86	57.26

Suppose there had been an error and that the lowest hourly wage was $15.50 instead of $19.64. This would result in
A. an increase in the median.
B. an increase in the standard deviation.
C. a decrease in the range.
D. a decrease in the IQR.
E. an increase in the mean.

10. Of the following stem-and-leave plots of four data sets each containing 11 observations, which represents the set of data that has the greatest standard deviation?

Variable: Set A	**Variable: Set B**	**Variable: Set C**	**Variable: Set D**
Decimal point is at the colon. Leaf unit = 0.1	Decimal point is at the colon. Leaf unit = 0.1	Decimal point is at the colon. Leaf unit = 0.1	Decimal point is at the colon. Leaf unit = 0.1
0 : 0	0 : 012345	0 : 0123	0 : 012
1 : 0	1 :	1 :	1 :
2 : 0	2 :	2 :	2 :
3 : 0	3 :	3 :	3 :
4 : 0	4 :	4 : 9	4 : 89
5 : 0	5 :	5 : 01	5 : 012
6 : 0	6 :	6 :	6 :
7 : 0	7 :	7 :	7 :
8 : 0	8 :	8 :	8 :
9 : 0	9 : 6789	9 : 789	9 : 89
10 : 0	10 : 0	10 : 0	10 : 0

A. Set A.
B. Set B.
C. Set C.
D. Set D.
E. both Set C and Set D.

Statistics Quiz F—Chapter 2—KEY

1. B

2. E

3. C

4. A

5. B

6. B

7. A

8. C

9. B

10. B