Chapter 2

Summarizing and Graphing Data

Section 2-2, Basic Skills and Concepts

1. No. For each class, the frequency tells us how many values fall within the given range of values, but there is no way to determine the exact IQ scores represented in the class.

2. If percentages are used, the sum should be 100%. If proportions are used, the sum should be 1.

3. No. The sum of the percentages is 199% not 100%, so each respondent could answer "yes" to more than one category. The table does not show the distribution of a data set among all of several different categories. Instead, it shows responses to five separate questions.

4. The gap in the frequencies suggests that the table includes heights of two different populations: students and faculty/staff.

5. Class width: 10.

    Class midpoints: 24.5, 34.5, 44.5, 54.5, 64.5, 74.5, 84.5.

    Class boundaries: 19.5, 29.5, 39.5, 49.5, 59.5, 69.5, 79.5, 89.5.

6. Class width: 10.

    Class midpoints: 24.5, 34.5, 44.5, 54.5, 64.5, 74.5.

    Class boundaries: 19.5, 29.5, 39.5, 49.5, 59.5, 69.5, 79.5.

7. Class width: 10.

    Class midpoints: 54.5, 64.5, 74.5, 84.5, 94.5, 104.5, 114.5, 124.5.

    Class boundaries: 49.5, 59.5, 69.5, 79.5, 89.5, 99.5, 109.5, 119.5, 129.5.

8. Class width: 5.

    Class midpoints: 2, 7, 12, 17, 22, 27, 32, 37.

    Class boundaries: -0.5, 4.5, 9.5, 14.5, 19.5, 24.5, 29.5, 34.5, 39.5.

9. Class width: 2.

    Class midpoints: 3.95, 5.95, 7.95, 9.95, 11.95.

    Class boundaries: 2.95, 4.95, 6.95, 8.95, 10.95, 12.95.

10. Class width: 2.

    Class midpoints: 3.95, 5.95, 7.95, 9.95, 11.95.

    Class boundaries: 2.95, 4.95, 6.95, 8.95, 10.95, 12.95, 14.95.

11. No. The frequencies do not satisfy the requirement of being roughly symmetric about the maximum frequency of 34.

12. Yes. The frequencies start low, increase to the maximum frequency of 43, and then decrease. Also, the frequencies are approximately symmetric about the maximum frequency of 43.

13. 18, 7, 4.

14. 12, 12, 6, 2.

15. On average, the actresses appear to be younger than the actors.

Age When Oscar Was Won	Relative Frequency (Actresses)	Relative Frequency (Actors)
20 – 29	32.9%	1.2%
30 – 39	41.5%	31.7%
40 – 49	15.9%	42.7%
50 – 59	2.4%	15.9%
60 – 69	4.9%	7.3%
70 – 79	1.2%	1.2%
80 – 89	1.2%	0.0%

16. The differences are not substantial. Based on the given data, males and females appear to have about the same distribution of white blood cell counts.

White Blood Cell Counts	Relative Frequency (Males)	Relative Frequency (Females)
3.0 – 4.9	20.0%	15.0%
5.0 – 6.9	37.5%	40.0%
7.0 – 8.9	27.5%	22.5%
9.0 – 10.9	12.5%	17.5%
11.0 – 12.9	2.5%	0.0%
13.0 – 14.9	0.0%	5.0%

17. The cumulative frequency table is

Age (years) of Best Actress When Oscar Was Won	Cumulative Frequency
Less than 30	27
Less than 40	61
Less than 50	74
Less than 60	76
Less than 70	80
Less than 80	81
Less than 90	82

18. The cumulative frequency table is

Age (years) of Best Actor When Oscar Was Won	Cumulative Frequency
Less than 30	1
Less than 40	27
Less than 50	62
Less than 60	75
Less than 70	81
Less than 80	82

19. Enter the label **Last Digit** in cell A1 of an Excel worksheet followed by the last digit data. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the data range including the label (**A1:A38**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **Number** of intervals and enter **10**. Click **OK**. The first three columns of the XLSTAT output are displayed at the top of the next page. The values in the lower bound columns are the last digits.

Because there are disproportionately more 0s and 5s, it appears that the heights were reported instead of measured. Consequently, it is likely that the results are not very accurate.

Lower bound	Upper bound	Frequency
0	1	9
1	2	2
2	3	1
3	4	3
4	5	1
5	6	15
6	7	2
7	8	0
8	9	3
9	10	1

20. Enter the label **Last Digit** in cell A1 of an Excel worksheet followed by the last digit data. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the data range including the label (**A1:A51**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **Number** of intervals and enter **10**. Click **OK**. The first three columns of the XLSTAT output are displayed on the right. The values in the lower bound columns are the last digits.

Because there are disproportionately more 0s and 5s, it appears that the heights were reported instead of measured. Consequently, it is likely that the results are not very accurate.

Lower bound	Upper bound	Frequency
0	1	26
1	2	1
2	3	1
3	4	2
4	5	2
5	6	12
6	7	1
7	8	0
8	9	4
9	10	1

21. Open the **MBODY** data file on your data disk. Enter the label **Lower** in cell P1 followed by the lower class limits of 40, 50, 60, 70, 80, 90, and 100, as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the PULSE data including the label (**B1:B41**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**P1:P8**). Click **OK**. The first three columns of the XLSTAT output are displayed below. Yes, the distribution appears to be a normal distribution.

P
Lower
40.0
50.0
60.0
70.0
80.0
90.0
100.0

Lower bound	Upper bound	Frequency
40	50	1
50	60	7
60	70	17
70	80	9
80	90	5
90	100	1

22. Open the **FBODY** data file on your data disk. Enter the label **Lower** in cell P1 followed by the lower class limits of 50, 60, 70, 80, 90, 100, and 110, as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the PULSE data including the label (**B1:B41**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**P1:P8**). Click **OK**. The first three columns of the XLSTAT output are displayed below. Yes, the distribution appears to be a normal distribution.

P
Lower
50.0
60.0
70.0
80.0
90.0
100.0
110.0

Lower bound	Upper bound	Frequency
50	60	1
60	70	8
70	80	18
80	90	5
90	100	6
100	110	2

23. Open the **QUAKE** data file on your data disk. Enter the label **Lower** in cell D1 followed by the lower class limits of 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 as displayed at the top of the next page on the left. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the MAG data including

the label (**A1:A51**). Select **Continuous**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**D1:D8**). Click **OK**. The first three columns of the XLSTAT output are displayed below on the right. No, the distribution does not appear to be a normal distribution.

D
Lower
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Lower bound	Upper bound	Frequency
0	0.5	5
0.5	1	15
1	1.5	19
1.5	2	7
2	2.5	2
2.5	3	2

24. Open the **QUAKE** data file on your data disk. Enter the label **Lower** in cell D1 followed by the lower class limits of 1, 5, 9, 13, 17, and 21 as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the DEPTH data including the label (**B1:B51**). Select **Continuous**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**D1:D7**). Click **OK**. The first three columns of the XLSTAT output are displayed below. No, the distribution does not appear to be a normal distribution.

D
Lower
1
5
9
13
17
21

Lower bound	Upper bound	Frequency
1	5	7
5	9	21
9	13	4
13	17	12
17	21	6

25. Open the **MBODY** data file on your data disk. Enter the label **Lower** in cell P1 followed by the lower class limits of 4.0, 4.4, 4.8, 5.2, 5.6, and 6.0 as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the RED data including the label (**H1:H41**). Select **Continuous**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**P1:P7**). Click **OK**. The first three columns of the XLSTAT output are displayed below. Yes, the distribution appears to be roughly a normal distribution.

P
Lower
4.0
4.4
4.8
5.2
5.6
6.0

Lower bound	Upper bound	Frequency
4	4.4	2
4.4	4.8	7
4.8	5.2	15
5.2	5.6	13
5.6	6	3

26. Open the **FBODY** data file on your data disk. Enter the label **Lower** in cell P1 followed by the lower class limits of 3.6, 4.0, 4.4, 4.8, 5.2, 5.6, and 6.0 as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the RED data including the label (**H1:H41**). Select **Continuous**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**P1:P8**). Click **OK**. The first three columns of the XLSTAT output are displayed below. Yes, the distribution appears to be roughly a normal distribution.

P
Lower
3.6
4.0
4.4
4.8
5.2
5.6
6.0

Lower bound	Upper bound	Frequency
3.6	4	2
4	4.4	13
4.4	4.8	15
4.8	5.2	7
5.2	5.6	2
5.6	6	1

27. Open the **FLIGHTS** data file on your data disk. Enter the label **Lower** in cell H1 followed by the lower class limits as displayed at the top of the next page on the left. Select the **XLSTAT** add-in. Select **Describing data**.

Select **Histograms**. Enter the range of the Arr Delay data including the label (**F1:H49**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**H1:H8**). Click **OK**. The first three columns of the XLSTAT output are displayed below on the right. Yes. Among the 48 flights, 36 arrived on time or early, and 45 of the flights arrived no more than 30 minutes late.

H
Lower
-60
-30
0
30
60
90
120

Lower bound	Upper bound	Frequency
-60	-30	11
-30	0	25
0	30	9
30	60	1
60	90	0
90	120	2

28. Open the **FLIGHTS** data file on your data disk. Enter the label **Lower** in cell H1 followed by the lower class limits as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the Taxi Out data including the label (**D1:D49**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**H1:H10**). Click **OK**. The first three columns of the XLSTAT output are displayed below. No. The times vary from a low of 12 minutes to a high of 49 minutes. It appears that many flights taxi out quickly, but many other flights require much longer times, so it would be difficult to predict the taxi-out time with reasonable accuracy.

H
Lower
10
15
20
25
30
35
40
45
50

Lower bound	Upper bound	Frequency
10	15	10
15	20	20
20	25	9
25	30	1
30	35	2
35	40	2
40	45	2
45	50	2

29. Enter the categories and frequencies in an Excel worksheet. Then use formulas to compute the relative frequencies. The completed table is displayed on the right.

Category	Frequency	Relative Frequency
Male Survivors	361	16.2%
Males Who Died	1395	62.8%
Female Survivors	345	15.5%
Females Who Died	122	5.5%
Total	2223	100.0%

30. Enter the label **Cause** in cell A1 of an Excel worksheet followed by the data given in Exercise 30. Click **INSERT** at the top of the screen. Select **Pivot Table**. Select **Select a table or range** and enter the data range including the label (**A1:A51**). Select **New Worksheet**. Click **OK**. In the section to the right of the worksheet, click and drag the Cause field button to ROWS. Click and drag the Cause field button to Σ VALUES. To change the counts to percentages, click directly on **Count of Cause** to the right of the worksheet and click the down arrow. Click on **Value Field Settings**. Click on **Show Values As**. Click the arrow to the right of No Calculation. Select **% of Column Total**. Click **OK**. The completed table is displayed below.

Row Labels	Count of Cause
E	18.00%
H	24.00%
O	12.00%
T	46.00%
Grand Total	100.00%

31. Pilot error is the most serious threat to aviation safety. Better training and stricter pilot requirements can improve aviation safety.

Cause	Relative Frequency
Pilot Error	50.5%
Other Human Error	6.1%
Weather	12.1%
Mechanical	22.2%
Sabotage	9.1%

32. The digit 0 appears to have occurred with a higher frequency than expected, but in general the differences are not very substantial, so the selection process appears to be functioning correctly. The digits are qualitative data because they do not represent measures or counts of anything. The digits could be replaced by the first 10 letters of the alphabet, and the lottery would be essentially the same.

Digit	Relative Frequency
0	16.7%
1	8.3%
2	10.0%
3	10.0%
4	6.7%
5	9.2%
6	7.5%
7	8.3%
8	7.5%
9	15.8%

Section 2-2, Beyond the Basics

33. Open the **CANS** data file on your data disk. Enter the label **Lower** in cell D1 followed by the lower class limits displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the CANS111 data including the label (**B1:B176**). Select **Continuous**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**D1:D18**). Click **OK**. The first three columns of the XLSTAT output including the outlier are displayed below on the left. Next construct the distribution excluding the outlier. Make the necessary adjustments to the input ranges. The first three columns of the XLSTAT excluding the outlier are displayed below on the right. An outlier can dramatically affect the frequency table.

D
Lower
200
220
240
260
280
300
320
340
360
380
400
420
440
460
480
500
520

Lower bound	Upper bound	Frequency
200	220	6
220	240	5
240	260	12
260	280	36
280	300	87
300	320	28
320	340	0
340	360	0
360	380	0
380	400	0
400	420	0
420	440	0
440	460	0
460	480	0
480	500	0
500	520	1

Lower bound	Upper bound	Frequency
200	220	6
220	240	5
240	260	12
260	280	36
280	300	87
300	320	28

34.

| Number of Data Values	Ideal Number of Classes
16 – 22	5
23 – 45	6
46 – 90	7
91 – 181	8
182 – 362	9
363 – 724	10
725 – 1448	11
1449 – 2896	12

Section 2-3, Basic Skills and Concepts

1. It is easier to see the distribution of the data by examining the graph of the histogram than by the numbers in the frequency distribution.

2. Not necessarily. Because those with special interests are more likely to respond, and the voluntary response sample is likely to consist of a group having characteristics that are fundamentally different than those of the population.

3. With a data set that is so small, the true nature of the distribution cannot be seen with a histogram. The data set has an outlier of 1 minute. That duration time corresponds to the last flight, which ended in an explosion that killed seven crew members.

4. When referring to a normal distribution, the term normal has a meaning that is different from its meaning in ordinary language. A normal distribution is characterized by a histogram that is approximately bell-shaped. Determination of whether a histogram is approximately bell-shaped does require subjective judgment.

5. Identifying the exact value is not easy, but answers not too far from 200 are good answers.

6. Class width of 2 inches. Approximate lower limit of first class of 43 inches. Approximate upper limit of first class of 45 inches.

7. The tallest person is about 108 inches, or about 9 feet tall. That tallest height is depicted in the bar that is farthest to the right in the histogram. That height is an outlier because it is very far from all of the other heights. The height of 9 feet must be an error, because the height of the tallest human ever recorded was 8 feet 11 inches.

8. The first group appears to be adults. Knowing that the people entered a museum on a Friday morning, we can reasonably assume that there were many school children on a field trip and that they were accompanied by a smaller group of teachers and adult chaperones and other adults visiting the museum by themselves.

9. Enter the label **Last Digit** in cell A1 of an Excel worksheet followed by the last digit data. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the data range including the label (**A1:A38**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **Number** of intervals and enter **10**. Click the **Charts** tab. Select **Histogram** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**.

    The digits 0 and 5 seem to occur much more than the other digits, so it appears that the heights were reported and not actually measured. This suggests that the results might not be very accurate.
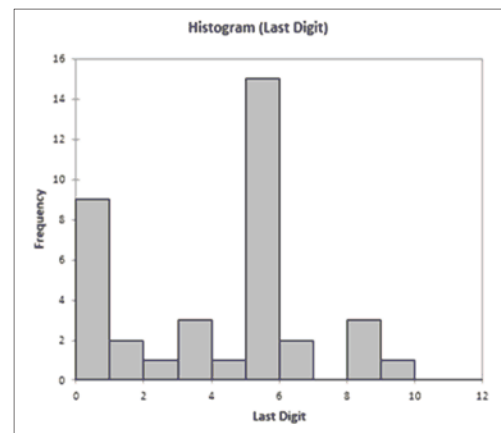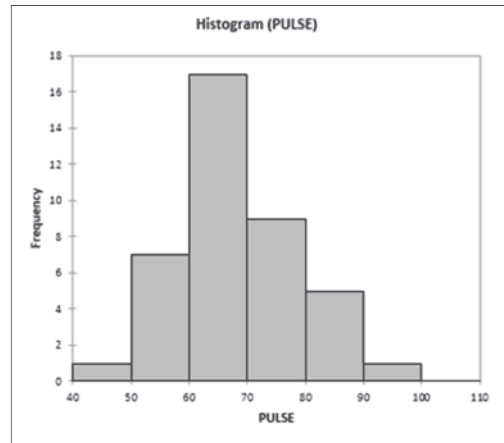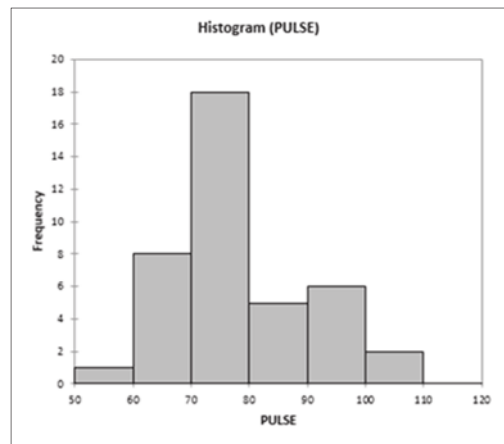
10. Enter the label **Last Digit** in cell A1 of an Excel worksheet followed by the last digit data. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the data range including the label (**A1:A51**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **Number** of intervals and enter **10**. Click the **Charts** tab. Select **Histogram** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**.

    The digits 0 and 5 seem to occur much more often than the other digits, so it appears that the heights were reported and not measured. This suggests that the results might not be very accurate.
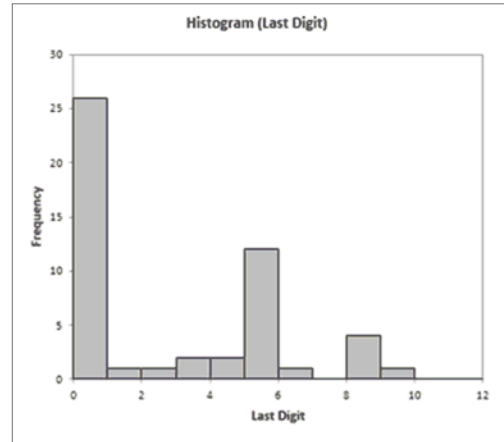
11. Open the **MBODY** data file on your data disk. Enter the label **Lower** in cell P1 followed by the lower class limits as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the PULSE data including the label (**B1:B41**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**P1:P8**). Click the **Charts** tab. Select **Histogram** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**. The histogram does appear to depict a normal distribution. The frequencies increase to a maximum and then tend to decrease, and the histogram is symmetric with the left half being roughly a mirror image of the right half.
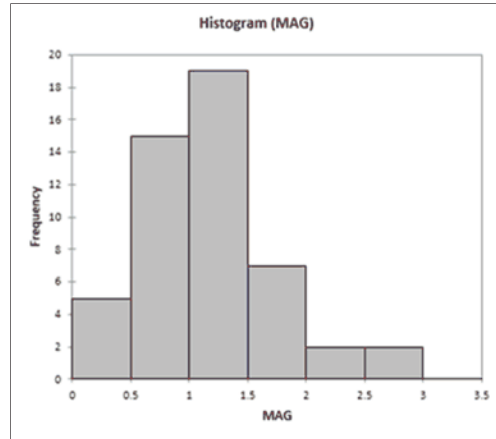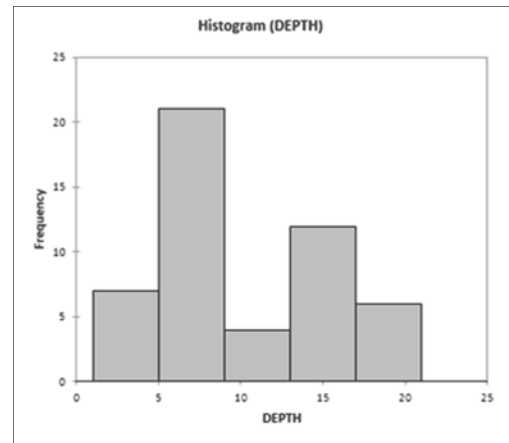
P
Lower
40.0
50.0
60.0
70.0
80.0
90.0
100.0

12. Open the **FBODY** data file on your data disk. Enter the label **Lower** in cell P1 followed by the lower class limits displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the PULSE data including the label (**B1:B41**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**P1:P8**). Click the **Charts** tab. Select **Histogram** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**.   The histogram appears to roughly approximate a normal distribution. The frequencies generally increase to a maximum and then tend to decrease, and the histogram is symmetric with the left half being roughly a mirror image of the right half.

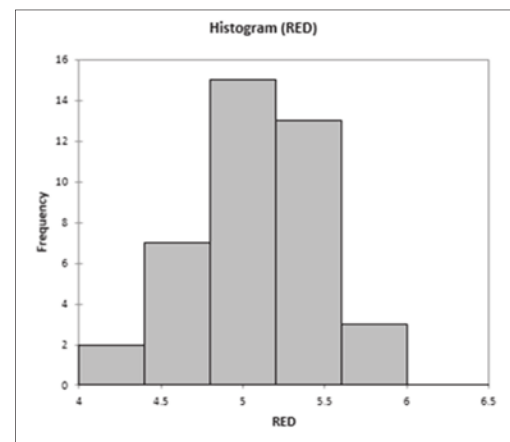| P |
|---|
| **Lower** |
| 50.0 |
| 60.0 |
| 70.0 |
| 80.0 |
| 90.0 |
| 100.0 |
| 110.0 |

13. Open the **QUAKE** data file on your data disk. Enter the label **Lower** in cell D1 followed by the lower class limits displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the MAG data including the label (**A1:A51**). Select **Continuous**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**D1: D8**). Click the **Charts** tab. Select **Histogram** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**. The histogram is displayed at the top of the next page. The histogram appears to roughly approximate a normal distribution. The frequencies increase to a maximum and then tend to decrease, and the histogram is symmetric with the left half being roughly a mirror image of the right half.

D
Lower
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Histogram (MAG)

14. Open the **QUAKE** data file on your data disk. Enter the label **Lower** in cell D1 followed by the lower class limits displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the DEPTH data including the label (**B1:B51**). Select **Continuous**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**D1:D7**). Click the **Charts** tab. Select **Histogram** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**.   No, the histogram does not appear to approximate a normal distribution. The frequencies do not increase to a maximum and then decrease, and the histogram is not symmetric with the left half being a mirror image of the right half.

D
Lower
1
5
9
13
17
21

Histogram (DEPTH)

15. Open the **MBODY** data file on your data disk. Enter the label **Lower** in cell P1 followed by the lower class limits of 4.0, 4.4, 4.8, 5.2, 5.6, and 6.0 as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the RED data including the label (**H1:H41**). Select **Continuous**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**P1:P7**). Click the **Charts** tab. Select **Histogram** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**. The histogram appears to roughly approximate a normal distribution. The frequencies increase to a maximum and then tend to decrease, and the histogram is symmetric with the left half being roughly a mirror image of the right half.

P
Lower
4.0
4.4
4.8
5.2
5.6
6.0

Histogram (RED)

16. Open the **FBODY** data file on your data disk. Enter the label **Lower** in cell P1 followed by the lower class limits of 3.6, 4.0, 4.4, 4.8, 5.2, 5.6, and 6.0 as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the RED data including the label (**H1:H41**). Select **Continuous**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**P1:P8**). Click the **Charts** tab. Select **Histogram** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**. The histogram is displayed at the top of the next page. The histogram appears to roughly approximate a normal distribution. The frequencies increase to a maximum and then tend to decrease, and the histogram is symmetric with the left half being roughly a mirror image of the right half.

| P |
|---|
| Lower |
| 3.6 |
| 4.0 |
| 4.4 |
| 4.8 |
| 5.2 |
| 5.6 |
| 6.0 |
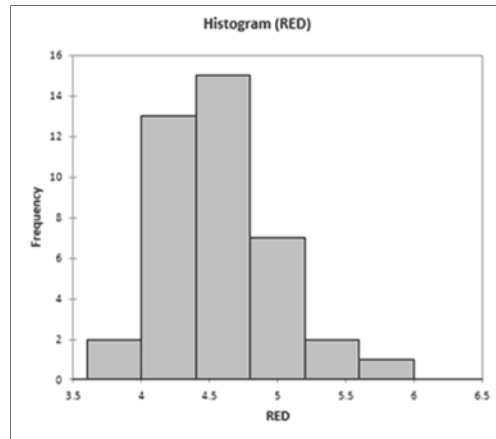
Histogram (RED)

17. Open the **FLIGHTS** data file on your data disk. Enter the label **Lower** in cell H1 followed by the lower class limits as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the Arr Delay data including the label (**F1:H49**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**H1:H8**). Click the **Charts** tab. Select **Histogram** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**. The two leftmost bars depict flights that arrived early, and the other bars to the right depict flights that arrived late.

H
Lower
-60
-30
0
30
60
90
120

Histogram (Arr Delay)
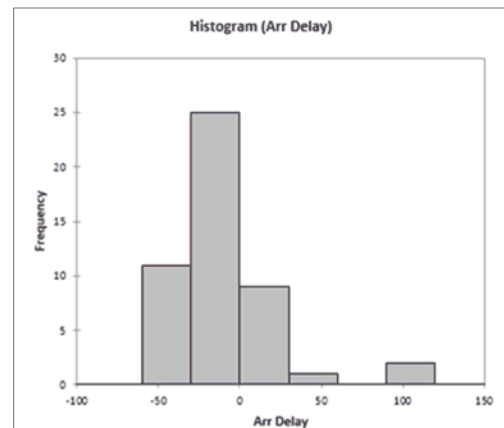
18. Open the **FLIGHTS** data file on your data disk. Enter the label **Lower** in cell H1 followed by the lower class limits as displayed on the right. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the Taxi Out data including the label (**D1:D49**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter range of the user defined lower class limits (**H1:H10**). Click the **Charts** tab. Select **Histogram** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**. Yes, the entire distribution would be more concentrated with less spread.

H
Lower
10
15
20
25
30
35
40
45
50

Histogram (Taxi Out)
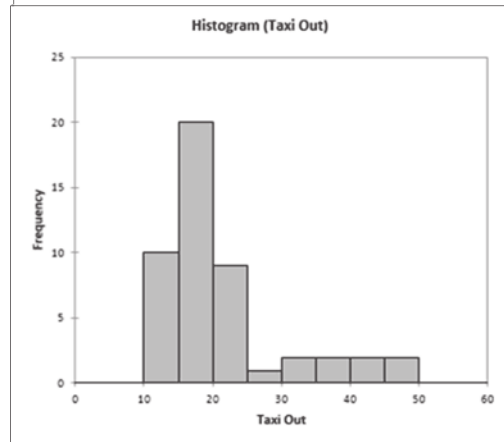
Section 2-3, Beyond the Basics

19. The ages of actresses are lower than those of actors.
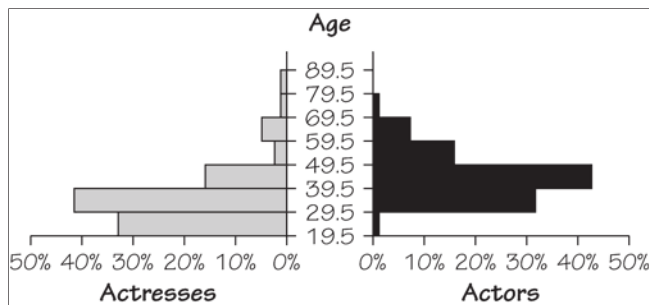
20. a.   107 inches to 109 inches; 8 feet 11 inches to 9 feet 1 inch.

    b.   The heights of the bars represent numbers of people, not heights. Because there are many more people between 43 inches tall and 55 inches tall, they have the tallest bars in the histogram, but they have the lowest actual heights. They have the tallest bars because there are more of them.

Section 2-4, Basic Skills and Concepts

1.   In a Pareto chart, the bars are arranged in descending order according to frequencies. The Pareto chart helps us understand data by drawing attention to the more important categories, which have the highest frequencies.

2.   A scatter plot is a plot of paired quantitative data, and each pair of data is plotted as a single point. The scatterplot requires paired quantitative data. The configuration of the plotted points can help us determine whether there is some relationship between two variables.

3.   The data set is too small for a graph to reveal important characteristics of the data. With such a small data set, it would be better to simply list the data or place them in a table.

4.   The sample is a voluntary response sample since the students report their scores to the website. Because the sample is a voluntary response sample , it is very possible that it is not representative of the population, even if the sample is very large. Any graph based on the voluntary response sample would have a high chance of showing characteristics that are not actual characteristics of the population.

5.   Open the **POTUS** data file on your data disk. Select the **XLSTAT** add-in. Select **Visualizing data**. Select **Scatter plots**. X: Enter the range of the presidents' heights including the label (E1:E39). Y: Enter the range of the opponents' heights including the label (F1:F39). Select **Variable labels**. Click **OK**. Because the points are scattered throughout with no obvious pattern, there does not appear to be a correlation.
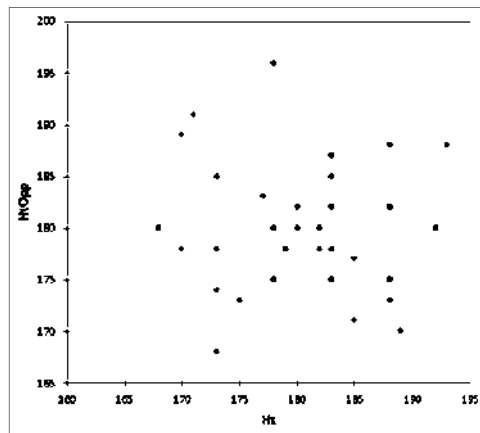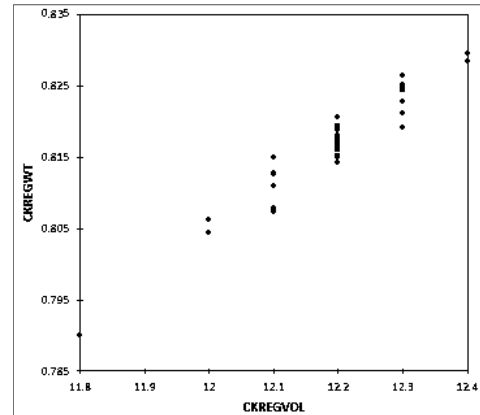

6.   Open the **IQBRAIN** data file on your data disk. Select the **XLSTAT** add-in. Select **Visualizing data**. Select **Scatter plots**. X: Enter the range of the brain volume data including the label (E1:E21). Y: Enter the range of the IQ scores including the label (D1:D21). Select **Variable labels**. Click **OK**. The scatter plot is displayed at the top of the next page. The configuration of the points does not support the hypothesis that people with larger brains have larger IQ scores.
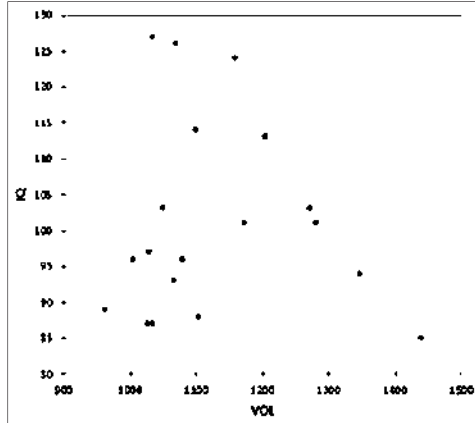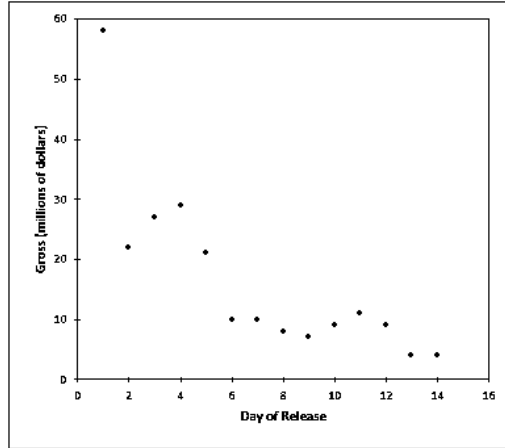
7. Open the **BEARS** data file on your data disk. Select the
    XLSTAT add-in. Select **Visualizing data**. Select **Scatter
    plots**. X: Enter the range of the chest size data including the
    label (H1:H55). Y: Enter the range of the weight data including
    the label (I1:I55). Select **Variable labels**.   Click **OK**. Yes.
    There is a very distinct pattern showing that bears with larger
   chest sizes tend to weigh more.
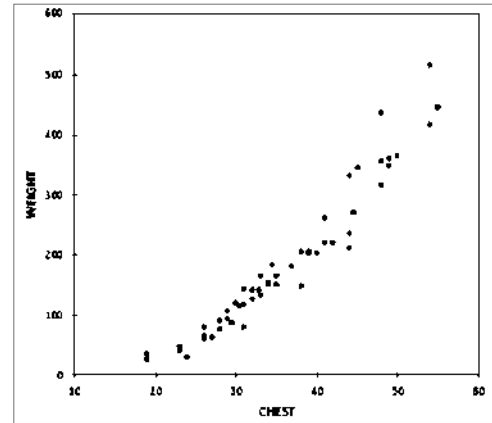
8. Open the **COLA** data file on your data disk. Select the
    XLSTAT add-in. Select **Visualizing data**. Select **Scatter plots**.
    X: Enter the range of the CKREGVOL data including the label
    (B1:B37). Y: Enter the range of the CKREGWT including the
    label (A1:A37). Select **Variable labels**.   Click **OK**. Yes. There
    is a very distinct pattern showing that cans of Coke with larger
    volumes tend to weigh more. Another notable feature of the
    scatterplot is that there are five groups of points that are stacked
    above each other. This is due to the fact that the measured
    volumes were rounded to one decimal place, so the different
    volume amounts are often duplicated, with the result that points
   are stacked vertically.

9. Enter the label **Day of Release** in cell A1 of an Excel worksheet followed by the numbers 1 through 14. Enter
    the label **Gross (millions of dollars)** in cell B1 followed by the amounts given in Exercise 9. Select the
    XLSTAT add-in. Select **Visualizing data**. Select **Scatter plots**. X: Enter the range of the Day of Release data
    including the label (A1:A15). Y: Enter the range of the Gross (millions of dollars) data including the label
    (B1:B15). Select **Variable labels**. Click **OK**. The scatter plot is displayed at the top of the next page. The
    first amount is highest for the opening day, when many Harry Potter fans are most eager to see the movie; the
    third and fourth values are from the first Friday and the first Saturday, which are the popular weekend days
   when movie attendance tends to spike.

10. Enter the label **Year** in cell A1 of an Excel worksheet followed by the years 1990 through 2009. Enter the label **Home Runs** in cell B1 followed by the amounts given in Exercise 10. Select the **XLSTAT** add-in. Select **Visualizing data**. Select **Scatter plots**. X: Enter the range of the years including the label (A1:A21). Y: Enter the range of the home run data including the label (B1:B21). Select **Variable labels**. Click **OK**. The numbers of home runs rose from 1990 to 2000, but after 2000 there was a very gradual decline.
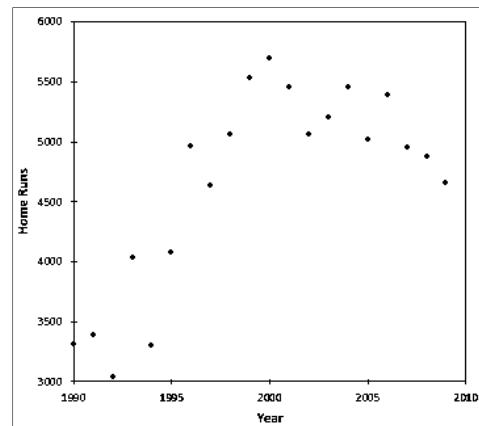
11. Yes, because the configuration of the points is roughly a bell shape, the volumes appear to be from a normally distributed population. The volume of 11.8 oz. appears to be an outlier.
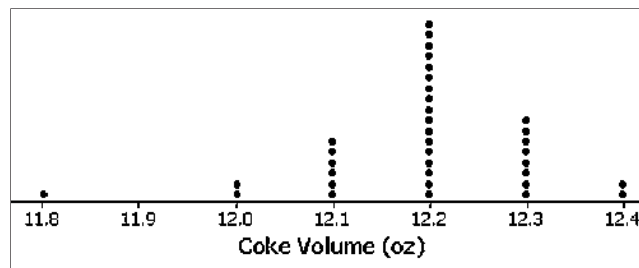
12. No, because the configuration of points is not at all a bell shape, the amounts do not appear to be from a normally distributed population.
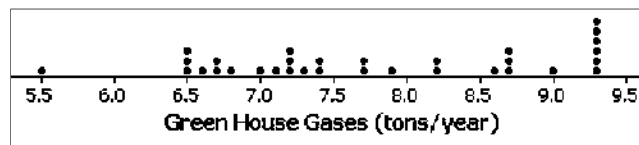
13. Open the **CRASH** data file on your data disk. Select the **XLSTAT** add-in. Select **Describing data**. Select **Descriptive statistics**. Enter the range of the PLVS data including the label (H1:H22). Select **Sample labels**. Click the **Options** tab. Select **Charts**. Click the **Charts (1)** tab. Select **Stem-and-leaf plots**. Click **OK**. The stem-and-leaf plot is displayed at the top of the next page. No. The distribution is not dramatically far from being a normal distribution with a bell shape, so there is not strong evidence against a normal distribution.

Stem-and-leaf plot (PLVS):

Unit:	10

```
4 | 5
5 | 3 3 3 5 5 7 9
6 | 1 1 1 6 7
7 | 1 1 1 5 5 6 8
8 | 4
```

14. There are no outliers. The distribution is not dramatically far from being a normally distribution with a bell shape, so there is not strong evidence against a normal distribution.

```
12 | 6 8
13 | 1 2 3 4 5 5 6 6 6 7 7 8 9 4
14 | 0 0 0 3 3 5
```

15. Enter the four frustrating sound responses in column A of an Excel worksheet (A1:A4). Enter the frequencies associated with each of the responses in column B (B1:B4). Sort the two columns by frequency arranged in descending order as displayed below on the left. Click and drag over the range A1:B4 to select these cells for the pareto chart. Click **INSERT** at the top of the screen. Select Column Chart in the Charts group by clicking on the column chart figure. Select the leftmost figure in the 2-D Column. The pareto chart is displayed below on the right.

| A	B
Response	Frequency
Car Alarm	388
Jackhammer	279
Crying Babies	209
Barking Dogs	128
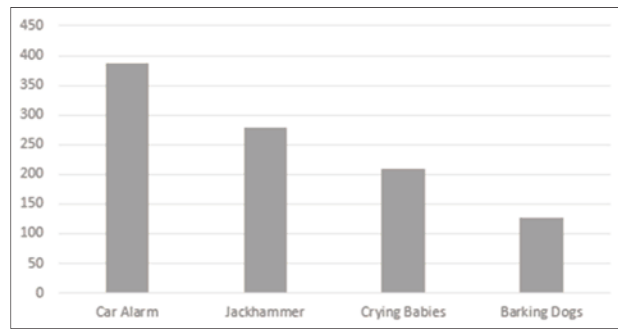
16. Enter the names of the five countries in column A of an Excel worksheet (A1:A5). Enter the instruction time associated with each country in column B (B1:B5). Sort the two columns by instruction time arranged in descending order as displayed below on the left. Click and drag over the range A1:B5 to select these cells for the pareto chart. Click **INSERT** at the top of the screen. Select Column Chart in the Charts group by clicking on the column chart figure. Select the leftmost figure in the 2-D Column row. The pareto chart is displayed below on the right. To remain competitive in the world, the United States should require more weekly instruction time.

| A	B
China	26.9
France	24.8
U.K.	24.6
Japan	23.8
U.S.	22.2
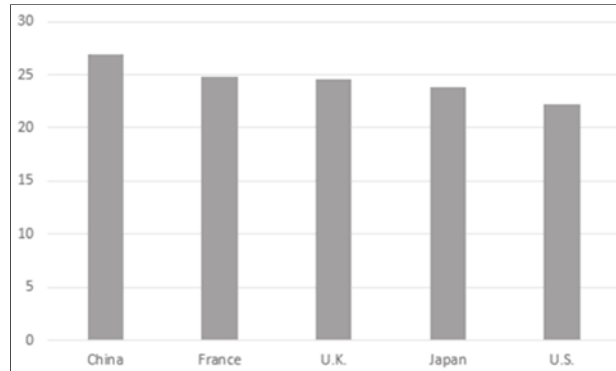
17. Click and drag over the Exercise 15 data range to select these cells for the pie chart. Click **INSERT** at the top of the screen. Select Pie Chart in the Charts group by clicking on the pie chart figure. Select the leftmost figure in the 2-D Pie row.
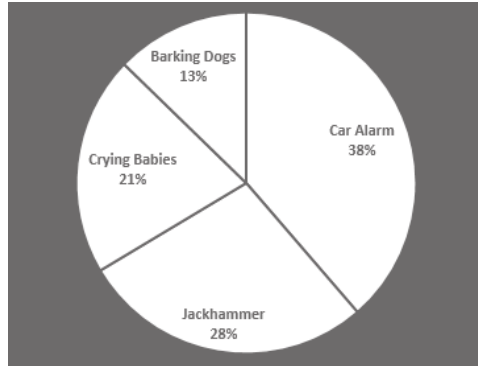

18. Click and drag over the Exercise 16 data range to select these cells for the pie chart. Click **INSERT** at the top of the screen. Select Pie Chart in the Charts group by clicking on the pie chart figure. Select the leftmost figure in the 2-D Pie row. Because there is not a single total number of hours of instruction time that is partitioned among the five countries, it does not make sense to use a pie chart for the given data.
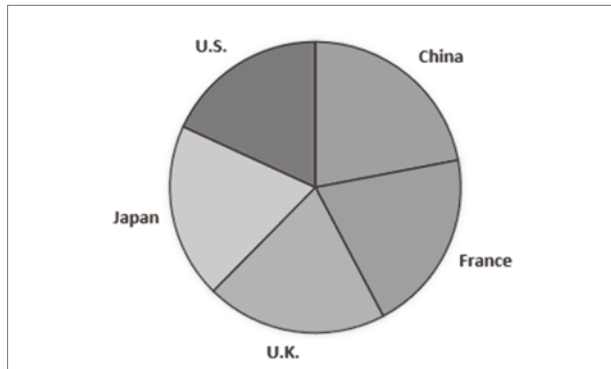
19. Follow the instructions in Exercise 23 in Section 2-2 to obtain the frequency distribution output. Cut and paste the lower bound values and the frequency values so that they are in adjacent columns of an Excel worksheet. Click and drag over the range of the lower bound and frequency cells to select these cells for the frequency polygon. Click **INSERT** at the top of the screen. Select **Scatter** in the Charts group. Select the rightmost figure in the second row, **Scatter with Straight Lines and Markers**. The frequency polygon is displayed below. The frequency polygon appears to roughly approximate a normal distribution. The frequencies increase to a maximum and then tend to decease, and the graph is symmetric with the left half being roughly a mirror image of the right half.
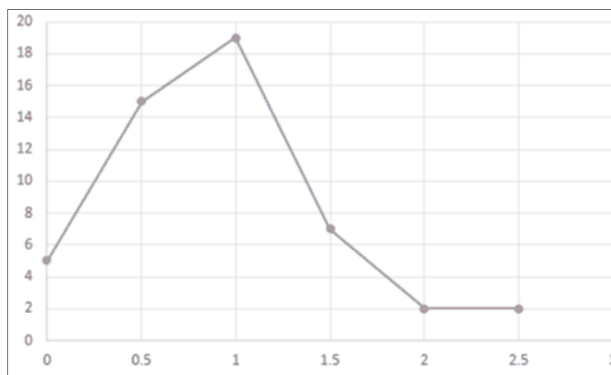
20. Follow the instructions in Exercise 24 in Section 2-2 to obtain the frequency distribution output. Cut and paste the lower bound values and the frequency values so that they are in adjacent columns of an Excel worksheet. Click and drag over the range of the lower bound and frequency cells to select these cells for the frequency polygon. Click **INSERT** at the top of the screen. Select **Scatter** in the Charts group. Select the rightmost figure in the second row, **Scatter with Straight Lines and Markers**. The frequency polygon is displayed below. No, the frequency polygon does not appear to approximate a normal distribution. The frequencies do not increase to a maximum and then decrease, and the graph is not symmetric with the left half being a mirror image of the right half.
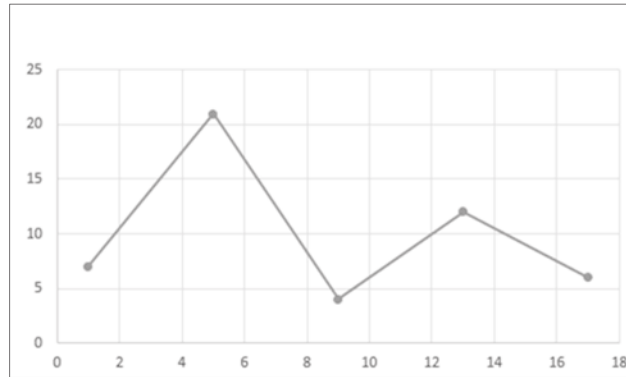
21. The vertical scale does not start at 0, so the difference is exaggerated. The graphs make it appear that Obama got about twice as many votes as McCain, but Obama actually got about 69 million votes compared to 60 million to McCain.

22. The fare doubled from $1 to $2, but when the $2 bill is shown with twice the width and twice the height of the $1 bill, the $2 bill has an area that is four times that of the $1 bill, so the illustration greatly exaggerates the increase in fare.

23. China's oil consumption is 2.7 times (or roughly 3 times) that of the United States, but by using a larger barrel that is three times as wide and three times as tall (and also three times as deep) as the smaller barrel, the illustration has made it appear that the larger barrel has a volume that is 27 times that of the smaller barrel. The actual ratio of US consumption to China's consumption is roughly 3 to 1, but the illustration makes it appear to be 27 to 1.

24. The actual braking distances are 133 ft., 136 ft., and 143 ft., so the differences are relatively small, but the illustration has a scale that begins at 130 ft., so the differences are grossly exaggerated.

Section 2-4, Beyond the Basics

25. The ages of actresses are lower than those of actors.

```
                     Actresses            Actors
     999999988887777666655554421 | 2 | 9
 99888877655555555444333333222111000 | 3 | 00122244455666777788888999
             9655322111110 | 4 | 000111111122222233334445555677788999
                        40 | 5 | 0112222346677
                      3110 | 6 | 000222
                         4 | 7 | 6
                         0 | 8 |
```

26. a.   96 | 5 9

97 | 0 0 0 1 1 1 2 3 3 3 4 4 4

97 | 5 5 6 6 6 6 6 7 8 8 8 8 9 9 9

98 | 0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 2 2 2 3 3 4 4 4 4 4 4 4 4 4 4 4 4

98 | 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 8 8 8 8 8 8 8 9 9

99 | 0 0 1 2 4

99 | 5 6

b.   The condensed stemplot reduces the number of rows so that the stemplot is not too large to be understandable.

6 – 7 | 79 * 778

8 – 9 | 45678 * 049

10 – 11 | 348 * 234477

12 – 13 | 01234 * 5

14 – 15 | 05 * 4569

16 – 17 | * 049

18 – 19 | * 6

20 – 21 | 1 * 3

Chapter Quick Quiz

1.   The class width is 1.00

2.   The class boundaries are –0.005 and 0.995

3.   No

4.   61 min., 62 min., 62 min., 62 min., 62 min., 67 min., and 69 min.

5.   No

6.   Bar graph

7.   Scatterplot

8.   Pareto Chart

9.   The distribution of the data set

10.   The bars of the histogram start relatively low, increase to a maximum value and then decrease. Also, the histogram is symmetric with the left half being roughly a mirror image of the right half.
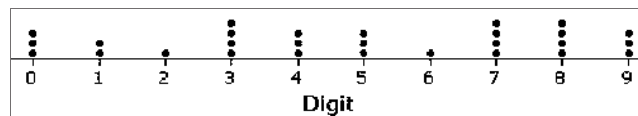
Review Exercises

1.   Enter the label **cm^3** in cell A1 of an Excel worksheet followed by the brain volume data given in Review Exercise 1. Enter the label **Lower** in cell B1followed by the lower bounds 900, 1000, 1100, 1200, 1300, 1400, and 1500. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the data range including the label (**A1:A21**). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **User defined** and enter the range of the lower bounds, B1:B6. Click the **Charts** tab. Select **Histograms** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**. The first three columns of the XLSTAT output are displayed at the top of the next page.

Lower bound	Upper bound	Frequency
900	1000	1
1000	1200	14
1200	1300	3
1300	1400	1
1400	1500	1

2. The histogram was included in the Review Exercise 1 output. The histogram is displayed below. No, the distribution does not appear to be normal because the graph is not symmetric.
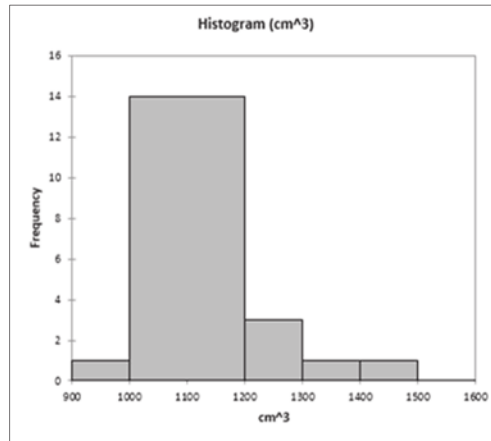
3. Although there are differences among the frequencies of the digits, the differences are not too extreme given the relatively small sample size, so the lottery appears to be fair.


4. The sample size is not large enough to reveal the true nature of the distribution of IQ scores for the population from which the sample is obtained.

8 | 7 7 9

9 | 6 6

10 | 1 3 3

5. A time-series graph is best. It suggests that the amounts of carbon monoxide emissions in the United States are increasing. To construct a time-series graph, begin by entering the label **Year** in cell A1 of an Excel worksheet followed by the numbers 1 through 10. Enter the label **Carbon Monoxide Emissions** followed by the data given in Review Exercise 5. Click and drag over the data range including labels (A1:B11). Click **INSERT** at the top of the screen. Select **Scatter** in the **Charts** group. Select Scatter with Straight Lines and Markers. The time-series graph is displayed at the top of the next page.
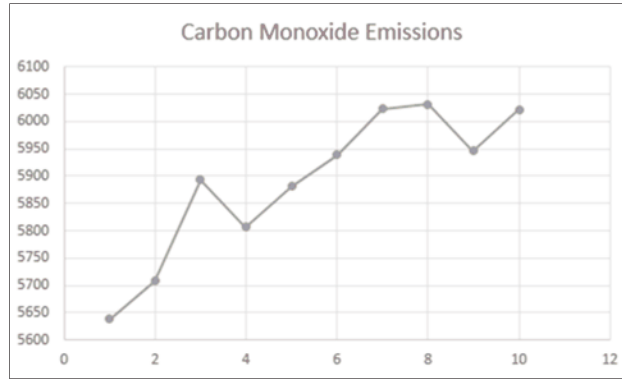
6.   A scatterplot is best. The scatterplot does not suggest that there is a relationship. To construct a scatterplot, continue using the Excel worksheet that you constructed for Review Exercise 5. Add a column with the nitrious oxide emissions data as shown below on the left. Click and drag over the range of the carbon monoxide emissions and nitrious oxide emissions to select these data for the graph. Click **INSERT** at the top of the screen. Select Scatter in the Charts group. If you would like, add horizontal and vertical axis titles.

Year	Carbon Monoxide Emissions	Nitrious Oxide Emissions
1	5638	351
2	5708	349
3	5893	345
4	5807	339
5	5881	335
6	5939	335
7	6024	362
8	6032	371
9	5946	376
10	6022	384
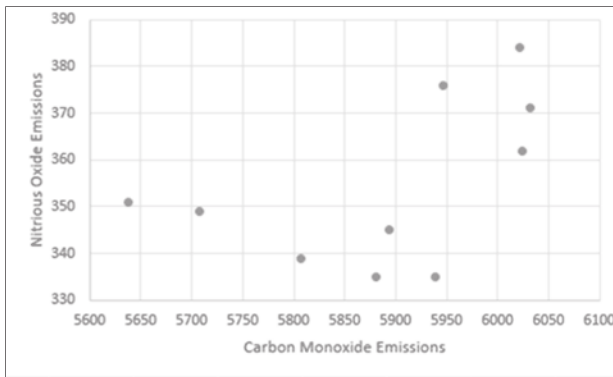
7.   A Pareto chart is best. To construct a Pareto chart, enter the data in an Excel worksheet and sort the data by sales in descending order. Click and drag over the category and sales data to select the cells. Click **INSERT** at the top of the screen. Select Column chart in the Charts group. Select the leftmost figure in the 2-D Column row. If you would like, add horizontal and vertical axis titles.
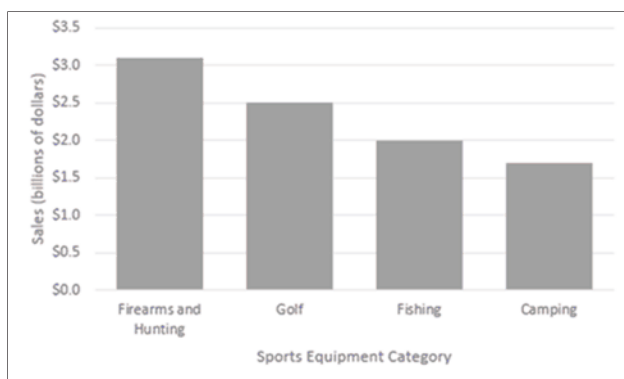
Cumulative Review Exercises

1. Pareto chart.

2. Nominal, because the responses consist of names only. The responses do not measure or count anything, and they cannot be arranged in order according to some quantitative scale.

3. Voluntary response sample. The voluntary response sample is not likely to be representative of the population, because those with special interests or strong feelings about the topic are more likely than others to respond and their views might be very different from those of the general population.

4. By using a vertical scale that does not begin at 0, the graph exaggerates the differences in the numbers of responses. The graph could be modified by starting the vertical scale at 0 instead of 50.

5. The percentage is $\dfrac{241}{641} = 0.376 = 37.6\%$. Because the percentage is based on a sample and not a population that percentage is a statistic.

6. Enter the label **Minutes** in cell A1 of an Excel worksheet followed by the times data given in Exercise 6. Select the **XLSTAT** add-in. Select **Describing data**. Select **Histograms**. Enter the range of the times data including the label (A1:A21). Select **Discrete**. Select **Sample labels**. Click the **Options** tab. Select **Range** and enter **10**. Click the **Charts** tab. Select **Histograms** and **Bars**. Select **Frequency** for the ordinate of the histogram. Click **OK**. The first three columns of the frequency distribution output are shown below.

Lower bound	Upper bound	Frequency
0	10	2
10	20	3
20	30	9
30	40	4
40	50	2
50	60	0

7. The histogram was included in the output for Exercise 6.   Because the frequencies increase to a maximum and then decrease and the left half of the histogram is roughly a mirror image of the right half, the data appear to be from a population with a normal distribution.

Histogram (Minutes)
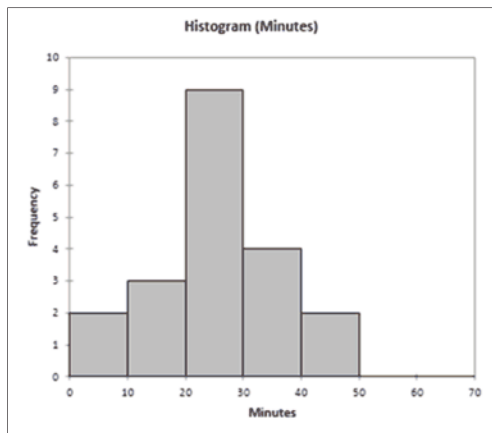
8. Stemplot

    0 | 0 5

    1 | 2 5 5

    2 | 0 2 4 5 5 5 7 7 8

    3 | 0 0 5 5

    4 | 0 5