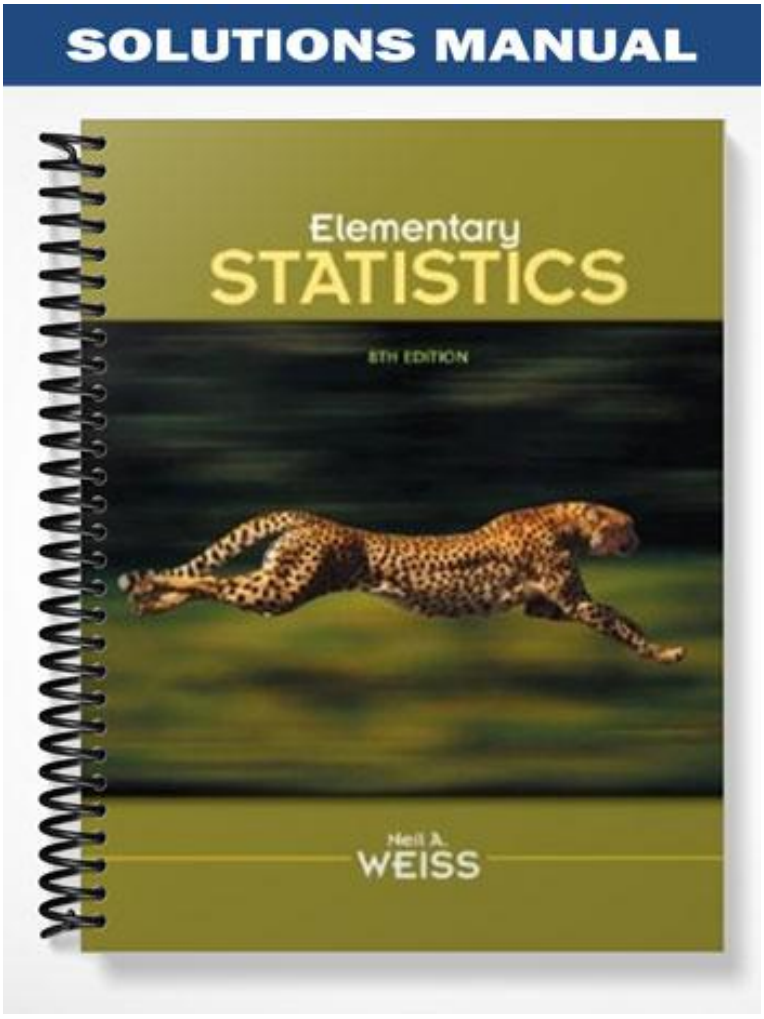
SOLUTIONS MANUAL



CHAPTER 2 ANSWERS

Exercises 2.1

2.1 (a) Hair color, model of car, and brand of popcorn are qualitative variables.

(b) Number of eggs in a nest, number of cases of flu, and number of employees are discrete, quantitative variables.

(c) Temperature, weight, and time are quantitative continuous variables.

2.2 (a) A qualitative variable is a nonnumerically valued variable. Its possible "values" are descriptive (e.g., color, name, gender).

(b) A discrete, quantitative variable is one whose possible values can be listed. It is usually obtained by counting rather than by measuring.

(c) A continuous, quantitative variable is one whose possible values form some interval of numbers. It usually results from measuring.

2.3 (a) Qualitative data result from observing and recording values of a qualitative variable, such as, color or shape.

(b) Discrete, quantitative data are values of a discrete quantitative variable. Values usually result from counting something.

(c) Continuous, quantitative data are values of a continuous variable. Values are usually the result of measuring something such as temperature that can take on any value in a given interval.

2.4 The classification of data is important because it will help you choose the correct statistical method for analyzing the data.

2.5 Of qualitative and quantitative (discrete and continuous) types of data, only qualitative yields nonnumerical data.

2.6 (a) The first column lists states. Thus, it consists of *qualitative* data.

(b) The second column gives the number of serious doctor disciplinary actions in each state in 2005-2007. These data are integers and therefore are *quantitative, discrete* data.

(c) The third column gives ratios of actions per 1,000 doctors for the years 2005-2007. The hint tells us that the possible ratios of positive whole numbers can be listed. For example, 8.33 out of 1,000 could also be listed as 833 out of 100,000. Ratios of whole numbers cannot be irrational. Therefore these data are *quantitative, discrete.*

2.7 (a) The second column consists of *quantitative, discrete* data. This column provides the ranks of the cities with the highest temperatures.

(b) The third column consists of *quantitative, continuous* data since temperatures can take on any value from the interval of numbers found on the temperature scale. This column provides the highest temperature in each of the listed cities.

(c) The information that Phoenix is in Arizona is *qualitative* data since it is nonnumeric.

2.8 (a) The first column consists of *quantitative, discrete* data. This column provides the ranks of the deceased celebrities with the top 5 earnings during the period from October 2004 to October 2005.

(b) The third column consists of *quantitative, discrete* data, the earnings of the celebrities. Since money involves discrete units, such as dollars and cents, the data is discrete, although, for all practical purposes, this data might be considered quantitative continuous data.

2.9 (a) The first column consists of *quantitative, discrete* data. This column

provides the ranks of the top ten countries with the highest number of Wi-Fi locations, as of October 28, 2009. These are whole numbers.

(b) The countries listed in the second column are *qualitative* data since they are nonnumerical.

(c) The third column consists of *quantitative, discrete* data. This column provides the number of Wi-Fi locations in each of the countries. These are whole numbers.

2.10 (a) The first column contains types of products. They are *qualitative* data since they are nonnumerical.

(b) The second column contains number of units shipped in the millions. These are whole numbers and are *quantitative, discrete.*

(c) The third column contains money values. Technically, these are *quantitative, discrete* data since there are gaps between possible values at the cent level. For all practical purposes, however, these are *quantitative, continuous* data.

2.11 The first column contains *quantitative, discrete* data in the form of ranks. These are whole numbers. The second and third columns contain *qualitative* data in the form of names. The last column contains the number of viewers of the programs. Total number of viewers is a whole number and therefore *quantitative, discrete* data.

2.12 Duration is a measure of time and is therefore *quantitative, continuous*. One might argue that workshops are frequently done in whole numbers of weeks, which would be *quantitative, discrete*. The number of students, the number of each gender, and the number of each ethnicity are whole numbers and are therefore *quantitative, discrete*. The genders and ethnicities themselves are nonnumerical and are therefore *qualitative* data. The number of web reports is a whole number and is *quantitative, discrete* data.

2.13 The first column contains *quantitative, discrete* data in the form of ranks. These are whole numbers. The second and fourth columns are nonnumerical and are therefore *qualitative* data. The third and fifth columns are measures of time and weight, both of which are *quantitative, continuous* data.

2.14 Of the eight items presented, only high school class rank involves ordinal data. The rank is ordinal data.

Exercises 2.2

2.15 A frequency distribution of qualitative data is a table that lists the distinct values of data and their frequencies. It is useful to organize the data and make it easier to understand.

2.16 (a) The frequency of a class is the number of observations in the class, whereas, the relative frequency of a class is the ratio of the class frequency to the total number of observations.

(b) The percentage of a class is 100 times the relative frequency of the class. Equivalently, the relative frequency of a class is the percentage of the class expressed as a decimal.

2.17 (a) True. Having identical frequency distributions implies that the total number of observations and the numbers of observations in each class are identical. Thus, the relative frequencies will also be identical.

(b) False. Having identical relative frequency distributions means that the ratio of the count in each class to the total is the same for both frequency distributions. However, one distribution may have twice (or some other multiple) the total number of observations as the other. For example, two distributions with counts of 5, 4, 1 and 10, 8, 2

would be different, but would have the same relative frequency distribution.

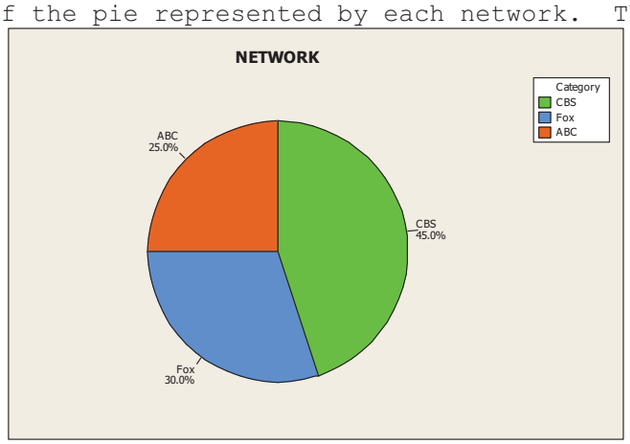
(c) If the two data sets have the same number of observations, either a frequency distribution or a relative-frequency distribution is suitable. If, however, the two data sets have different numbers of observations, using relative-frequency distributions is more appropriate because the total of each set of relative frequencies is 1, putting both distributions on the same basis for comparison.

2.18 (a)-(b)

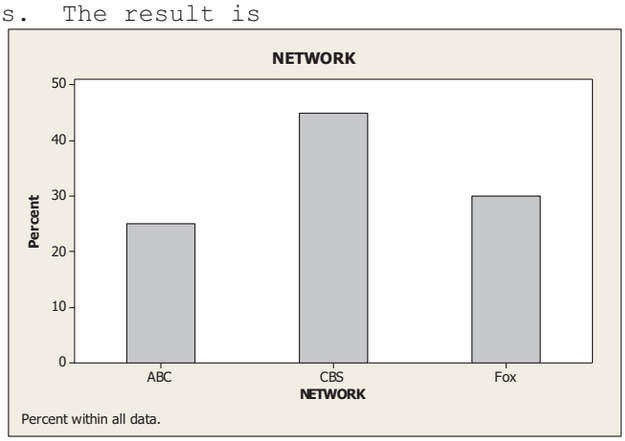
The classes are the days of the week and are presented in column 1. The frequency distribution of the networks is presented in column 2. Dividing each frequency by the total number of shows, which is 20, results in each class's relative frequency. The relative frequency distribution is presented in column 3.

Network	Frequency	Relative Frequency
ABC	5	0.25
CBS	9	0.45
Fox	6	0.30
	20	1.00

(c) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each network. The result is



(d) We use the bar chart to show the relative frequency with which each network occurs. The result is

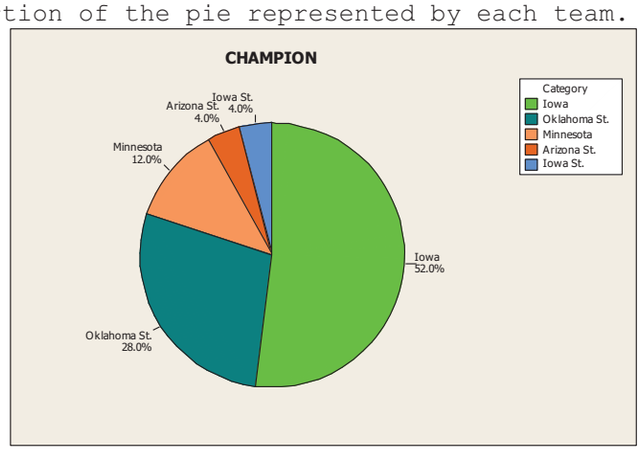


2.19 (a)-(b)

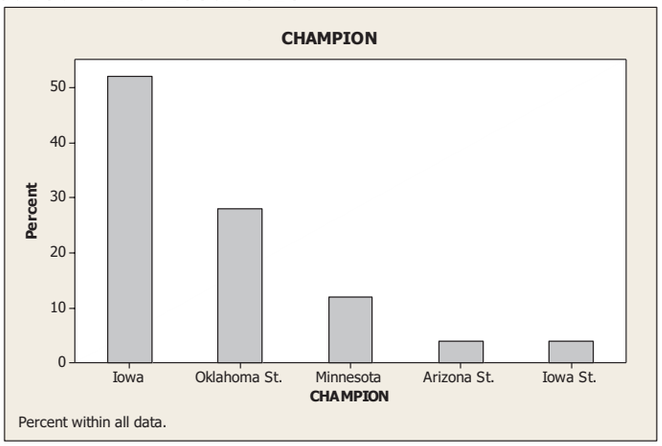
The classes are the NCAA wrestling champions and are presented in column 1. The frequency distribution of the champions is presented in column 2. Dividing each frequency by the total number of champions, which is 25, results in each class's relative frequency. The relative frequency distribution is presented in column 3.

Champion	Frequency	Relative Frequency
Iowa	13	0.52
Iowa St.	1	0.04
Minnesota	3	0.12
Arizona St.	1	0.04
Oklahoma St.	7	0.28
	25	1.00

(c) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each team. The result is



(d) We use the bar chart to show the relative frequency with which each TEAM occurs. The result is

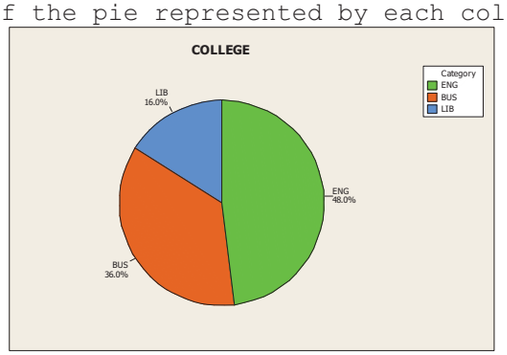


2.20 (a)-(b) The classes are the colleges and are presented in column 1. The frequency distribution of the colleges is presented in column 2. Dividing each frequency by the total number of students in the section of Introduction to Computer Science, which is 25, results in each

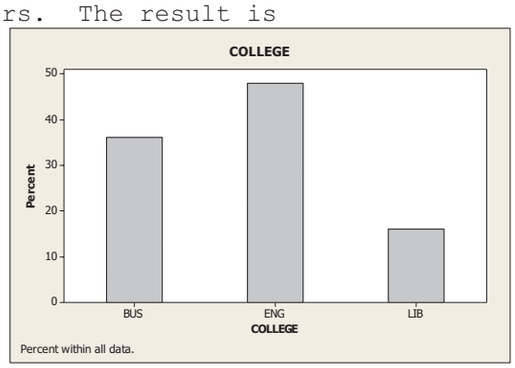
class's relative frequency. The relative frequency distribution is presented in column 3.

College	Frequency	Relative Frequency
BUS	9	0.36
ENG	12	0.48
LIB	4	0.16
	25	1.00

(c) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each college. The result is



(d) We use the bar chart to show the relative frequency with which each COLLEGE occurs. The result is

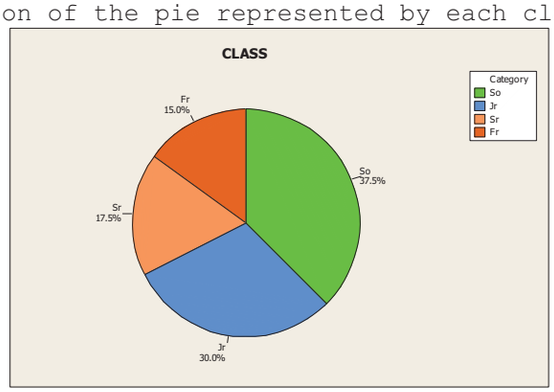


2.21 (a)-(b)

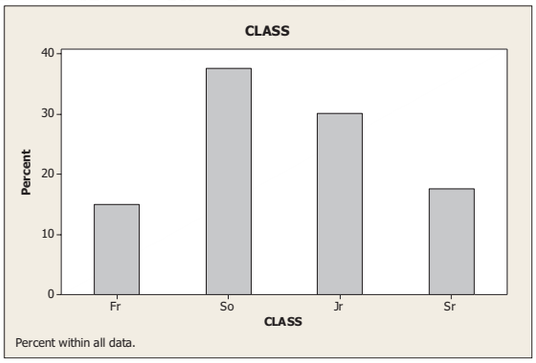
The classes are the class levels and are presented in column 1. The frequency distribution of the class levels is presented in column 2. Dividing each frequency by the total number of students in the introductory statistics class, which is 40, results in each class's relative frequency. The relative frequency distribution is presented in column 3.

Class Level	Frequency	Relative Frequency
Fr	6	0.150
So	15	0.375
Jr	12	0.300
Sr	7	0.175
	40	1.000

(c) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each class level. The result is



(d) We use the bar chart to show the relative frequency with which each CLASS level occurs. The result is

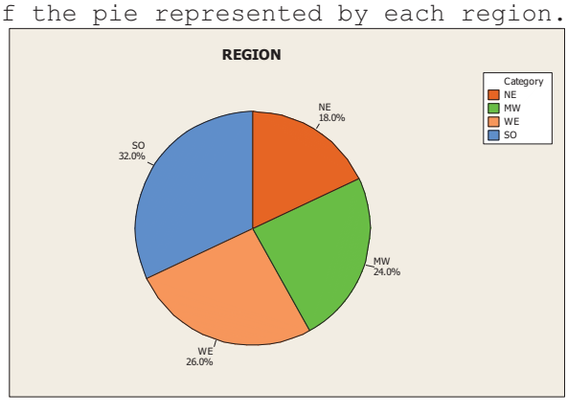


2.22 (a)-(b)

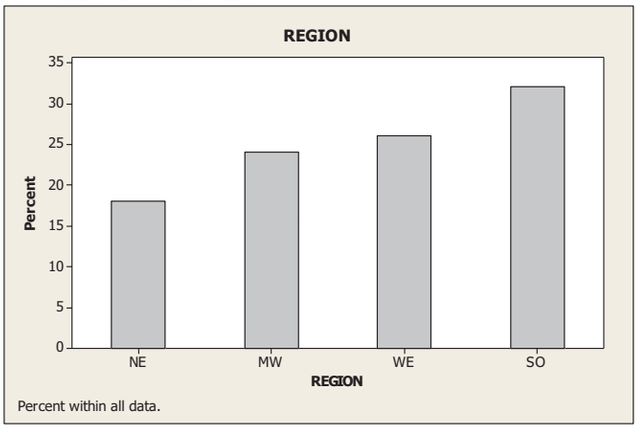
The classes are the regions and are presented in column 1. The frequency distribution of the regions is presented in column 2. Dividing each frequency by the total number of states, which is 50, results in each class's relative frequency. The relative frequency distribution is presented in column 3.

Class Level	Frequency	Relative Frequency
NE	9	0.18
MW	12	0.24
SO	16	0.32
WE	13	0.26
	50	1.00

(c) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each region. The result is



(d) We use the bar chart to show the relative frequency with which each REGION occurs. The result is

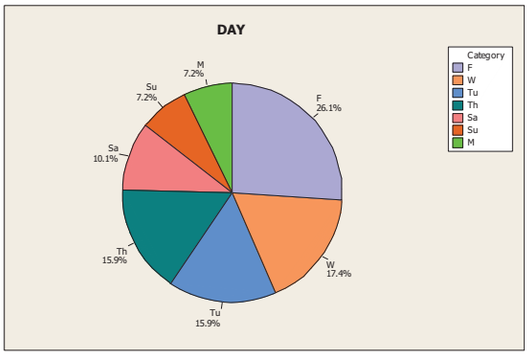


2.23 (a)–(b)

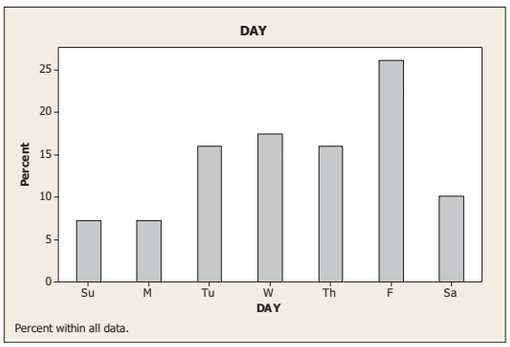
The classes are the days and are presented in column 1. The frequency distribution of the days is presented in column 2. Dividing each frequency by the total number road rage incidents, which is 69, results in each class's relative frequency. The relative frequency distribution is presented in column 3.

Class Level	Frequency	Relative Frequency
Su	5	0.0725
M	5	0.0725
Tu	11	0.1594
W	12	0.1739
Th	11	0.1594
F	18	0.2609
Sa	7	0.1014
	69	1.0000

(c) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each day. The result is



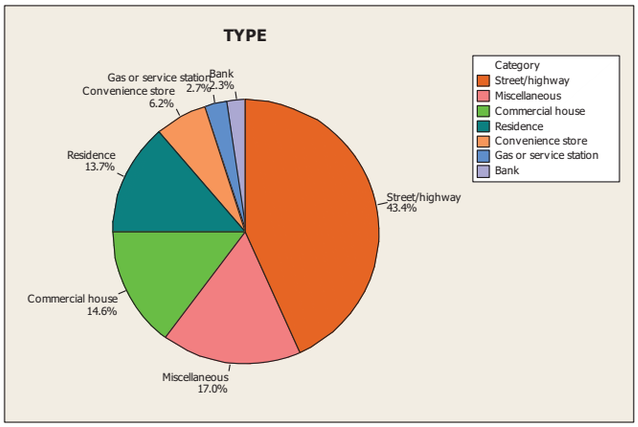
(d) We use the bar chart to show the relative frequency with which each DAY occurs. The result is



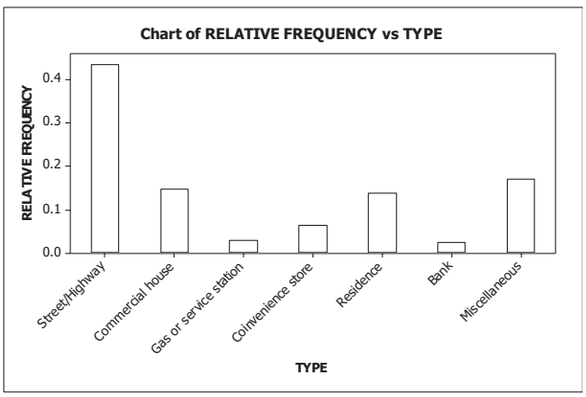
2.24 (a) We first find each of the relative frequencies by dividing each of the frequencies by the total frequency of 413,403

Robbery Type	Frequency	Relative Frequency
Street/highway	179,296	0.4337
Commercial house	60,493	0.1463
Gas or service station	11,362	0.0275
Convenience store	25,774	0.0623
Residence	56,641	0.1370
Bank	9,504	0.0230
Miscellaneous	70,333	0.1701
	413,403	1.0000

(b) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each robbery type. The result is



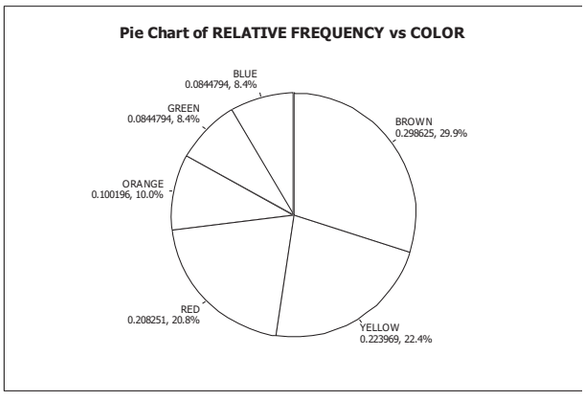
(c) We use the bar chart to show the relative frequency with which each robbery type occurs. The result is



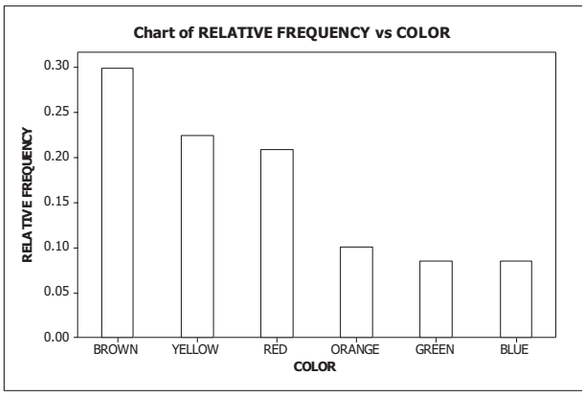
2.25 (a) We first find the relative frequencies by dividing each of the frequencies by the total sample size of 509.

Color	Frequency	Relative Frequency
Brown	152	0.2986
Yellow	114	0.2240
Red	106	0.2083
Orange	51	0.1002
Green	43	0.0845
Blue	43	0.0845
	509	1.0000

(b) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each color of M&M. The result is



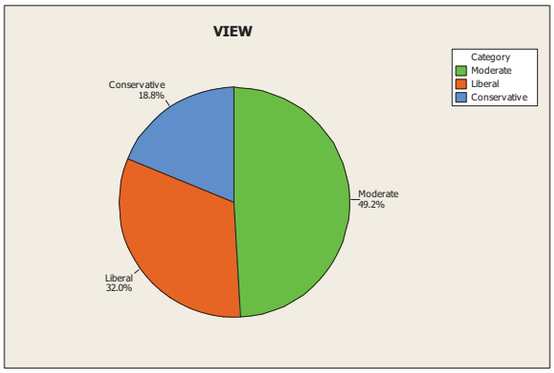
(c) We use the bar chart to show the relative frequency with which each color occurs. The result is



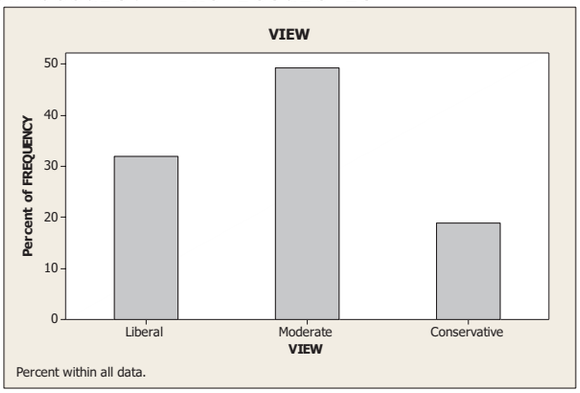
2.26 (a) We first find the relative frequencies by dividing each of the frequencies by the total sample size of 500.

Political View	Frequency	Relative Frequency
Liberal	160	0.320
Moderate	246	0.492
Conservative	94	0.188
	500	1.000

(b) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each political view. The result is



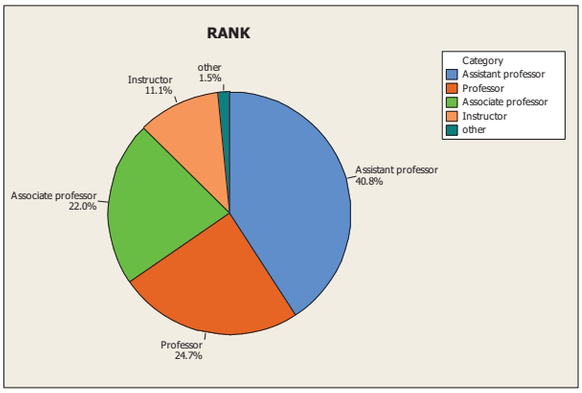
(c) We use the bar chart to show the relative frequency with which each political view occurs. The result is



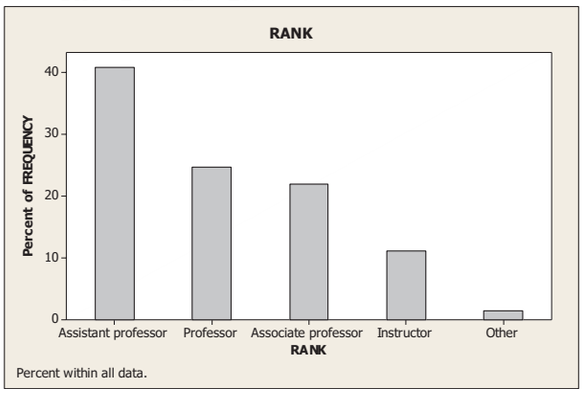
2.27 (a) We first find the relative frequencies by dividing each of the frequencies by the total sample size of 98,993.

Rank	Frequency	Relative Frequency
Professor	24,418	0.2467
Associate professor	21,732	0.2195
Assistant professor	40,379	0.4079
Instructor	10,960	0.1107
Other	1,504	0.0152
	98,993	1.0000

(b) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each rank. The result is



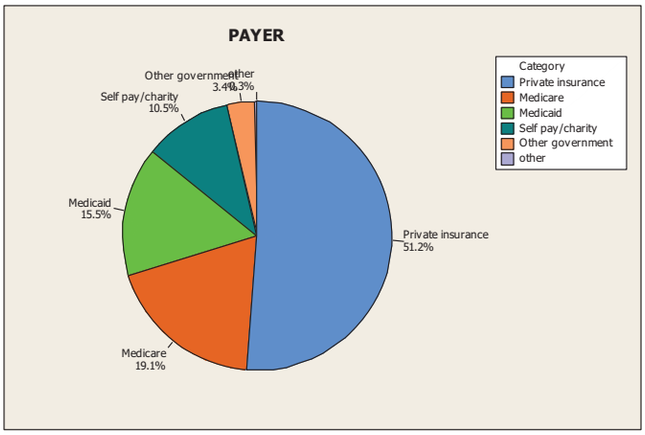
(c) We use the bar chart to show the relative frequency with which each rank occurs. The result is



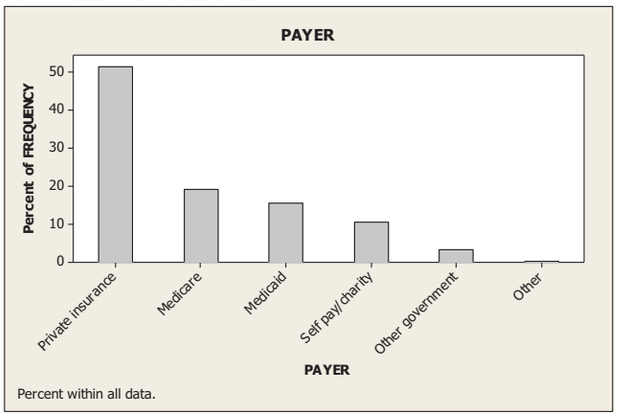
2.28 (a) We first find the relative frequencies by dividing each of the frequencies by the total sample size of 52,389.

Payer	Frequency	Relative Frequency
Medicare	9,983	0.1906
Medicaid	8,142	0.1554
Private insurance	26,825	0.5120
Other government	1,777	0.0339
Self pay/charity	5,512	0.1052
Other	150	0.0029
	52,389	1.0000

(b) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each payer. The result is



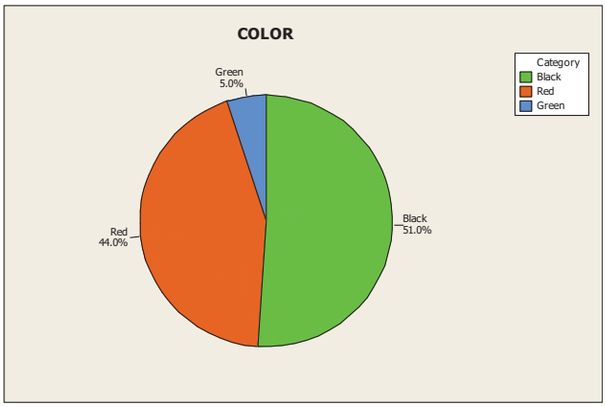
(c) We use the bar chart to show the relative frequency with which each payer occurs. The result is



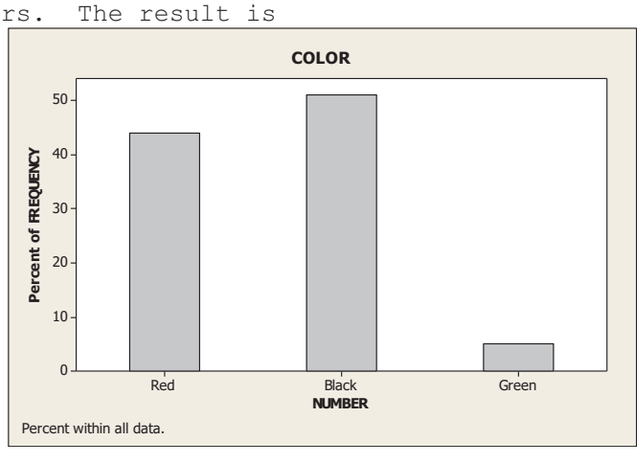
2.29 (a) We first find the relative frequencies by dividing each of the frequencies by the total sample size of 200.

Color	Frequency	Relative Frequency
Red	88	0.44
Black	102	0.51
Green	10	0.05
	200	1.00

(b) We multiply each of the relative frequencies by 360 degrees to obtain the portion of the pie represented by each color. The result is



(c) We use the bar chart to show the relative frequency with which each color occurs. The result is

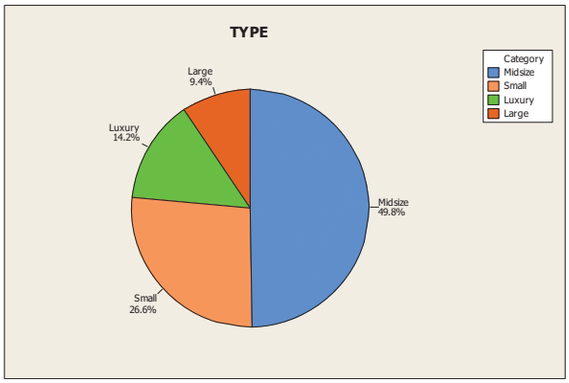


2.30 (a) Using Minitab, retrieve the data from the Weiss-Stats-CD. Column 1 contains the type of the car. From the tool bar, select **Stat ▶ Tables ▶ Tally Individual Variables**, double-click on TYPE in the first box so that TYPE appears in the **Variables** box, put a check mark next to Counts and Percents under Display, and click **OK**. The result is

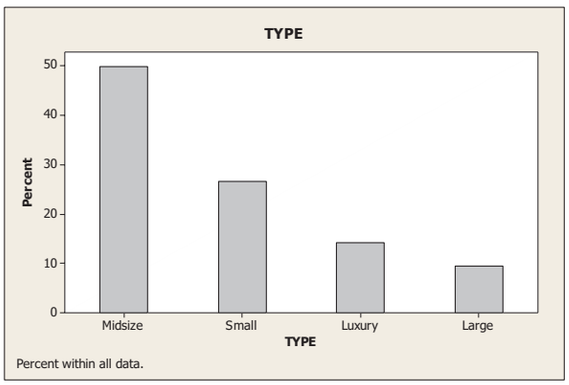
TYPE	Count	Percent
Large	47	9.40
Luxury	71	14.20
Midsize	249	49.80
Small	133	26.60
N=	500	

(b) The relative frequencies were calculated in part(a) by putting a check mark next to Percents. 9.4% of the cars were Large, 14.2% were Luxury, 49.8% were Midsize, and 26.6% were Small.

(c) Using Minitab, select **Graph ▶ Pie Chart**, check Chart counts of unique values, double-click on TYPE in the first box so that TYPE appears in the Categorical Variables box. Click Pie Options, check decreasing volume, click OK. Click Labels, enter TYPE in for the title, click Slice Labels, check Category Name, Percent, and Draw a line from label to slice, Click OK twice. The result is



(d) Using Minitab, select **Graph** ▶ **Bar Chart**, select Counts of unique values, select Simple option, click OK. Double-click on TYPE in the first box so that TYPE appears in the Categorical Variables box. Select Chart Options, check decreasing Y, check show Y as a percent, click OK. Select Labels, enter in TYPE as the title. Click OK twice. The result is

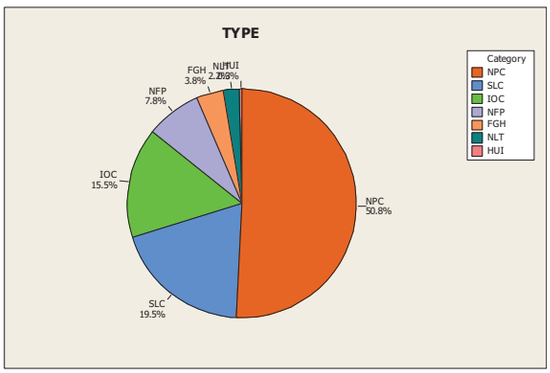


2.31 (a) Using Minitab, retrieve the data from the Weiss-Stats-CD. Column 1 contains the type of the hospital. From the tool bar, select **Stat** ▶ **Tables** ▶ **Tally Individual Variables**, double-click on TYPE in the first box so that TYPE appears in the **Variables** box, put a check mark next to Counts and Percents under Display, and click **OK**. The result is

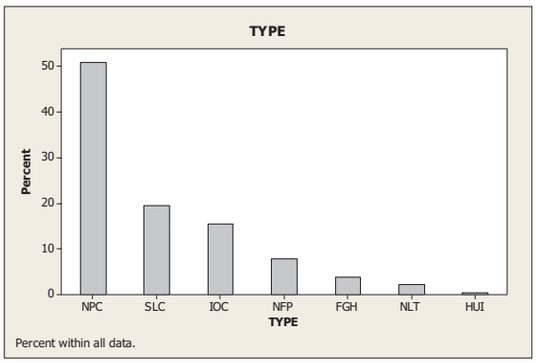
TYPE	Count	Percent
NPC	2919	50.79
IOC	889	15.47
SLC	1119	19.47
FGH	221	3.85
NLT	129	2.24
NFP	451	7.85
HUI	19	0.33
N=	5747	

(b) The relative frequencies were calculated in part(a) by putting a check mark next to Percents. 50.79% of the hospitals were Nongovernment not-for-profit community hospitals, 15.47% were Investor-owned (for-profit) community hospitals, 19.47% were State and local government community hospitals, 3.85% were Federal government hospitals, 2.24% were Nonfederal long term care hospitals, 7.85% were Nonfederal psychiatric hospitals, and 0.33% were Hospital units of institutions.

(c) Using Minitab, select **Graph ▶ Pie Chart**, check Chart counts of unique values, double-click on TYPE in the first box so that TYPE appears in the Categorical Variables box. Click Pie Options, check decreasing volume, click OK. Click Labels, enter TYPE in for the title, click Slice Labels, check Category Name, Percent, and Draw a line from label to slice, Click OK twice. The result is



(d) Using Minitab, select **Graph ▶ Bar Chart**, select Counts of unique values, select Simple option, click OK. Double-click on TYPE in the first box so that TYPE appears in the Categorical Variables box. Select Chart Options, check decreasing Y, check show Y as a percent, click OK. Select Labels, enter in TYPE as the title. Click OK twice. The result is



2.32 (a) Using Minitab, retrieve the data from the Weiss-Stats-CD. Column 2 contains the marital status and column 3 contains the number of drinks per month. From the tool bar, select **Stat ▶ Tables ▶ Tally Individual Variables**, double-click on STATUS and DRINKS in the first box so that both STATUS and DRINKS appear in the **Variables** box, put a check mark next to Counts and Percents under Display, and click **OK**. The results are

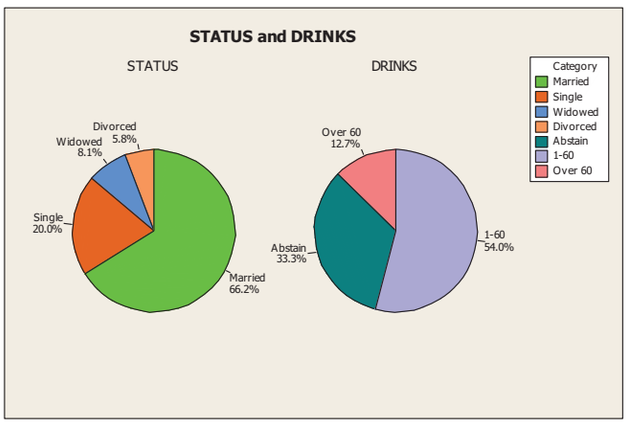
STATUS	Count	Percent
Single	354	19.98
Married	1173	66.20
Widowed	143	8.07
Divorced	102	5.76
N=	1772	

DRINKS	Count	Percent
Abstain	590	33.30
1-60	957	54.01
Over 60	225	12.70
N=	1772	

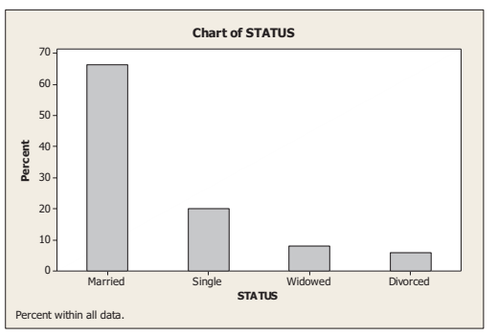
(b) The relative frequencies were calculated in part(a) by putting a check mark next to Percents. For the STATUS variable; 19.98% of the US Adults are single, 66.20% are married, 8.07% are widowed, and 5.76% are

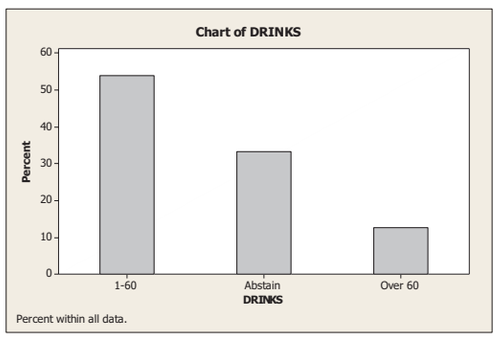
divorced. For the DRINKS variable, 33.30% of US Adults abstain from drinking, 54.01% have 1-60 drinks per month, and 12.70% have over 60 drinks per month.

(c) Using Minitab, select **Graph ▶ Pie Chart**, check Chart counts of unique values, double-click on STATUS and DRINKS in the first box so that STATUS and DRINKS appear in the Categorical Variables box. Click Pie Options, check decreasing volume, click OK. Click Multiple Graphs, check On the Same Graphs, Click OK. Click Labels, click Slice Labels, check Category Name, Percent, and Draw a line from label to slice, Click OK twice. The results are



(d) Using Minitab, select **Graph ▶ Bar Chart**, select Counts of unique values, select Simple option, click OK. Double-click on STATUS and DRINKS in the first box so that STATUS and DRINKS appear in the Categorical Variables box. Select Chart Options, check decreasing Y, check show Y as a percent, click OK. Click OK twice. The results are



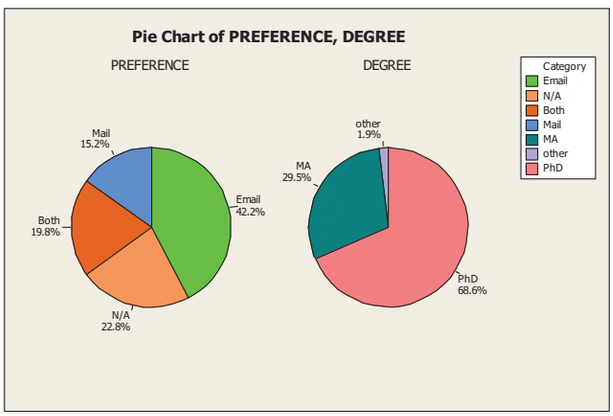


2.33 (a) Using Minitab, retrieve the data from the Weiss-Stats-CD. Column 2 contains the preference for how the members want to receive the ballots and column 3 contains the highest degree obtained by the members. From the tool bar, select **Stat ▶ Tables ▶ Tally Individual Variables**, double-click on PREFERENCE and DEGREE in the first box so that both PREFERENCE and DEGREE appear in the **Variables** box, put a check mark next to Counts and Percents under Display, and click **OK**. The results are

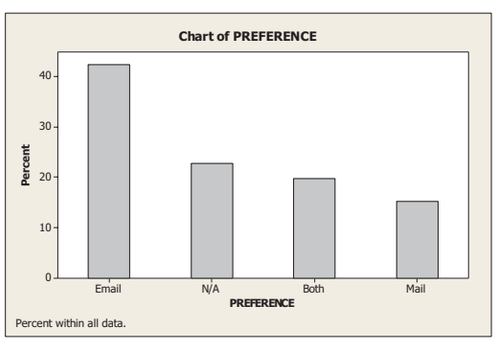
PREFERENCE	Count	Percent	DEGREE	Count	Percent
Both	112	19.79	MA	167	29.51
Email	239	42.23	Other	11	1.94
Mail	86	15.19	PhD	388	68.55
N/A	129	22.79	N=	566	
N=	566				

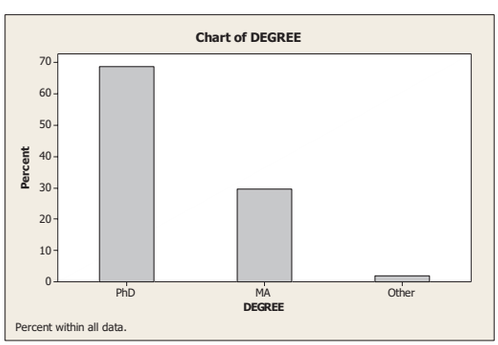
(b) The relative frequencies were calculated in part(a) by putting a check mark next to Percents. For the PREFERENCE variable; 19.79% of the members prefer to receive the ballot by both e-mail and mail, 42.23% prefer e-mail, 15.19% prefer mail, and 22.79% didn't list a preference. For the Degree variable; 29.51% obtained a Master's degree, 68.55% obtained a PhD, and 1.94% received a different degree.

(c) Using Minitab, select **Graph ▶ Pie Chart**, check Chart counts of unique values, double-click on PREFERENCE and DEGREE in the first box so that PREFERENCE and DEGREE appear in the Categorical Variables box. Click Pie Options, check decreasing volume, click OK. Click Multiple Graphs, check On the Same Graphs, Click OK. Click Labels, click Slice Labels, check Category Name, Percent, and Draw a line from label to slice, Click OK twice. The results are



(d) Using Minitab, select **Graph ▶ Bar Chart**, select Counts of unique values, select Simple option, click OK. Double-click on PREFERENCE and DEGREE in the first box so that PREFERENCE and DEGREE appear in the Categorical Variables box. Select Chart Options, check decreasing Y, check show Y as a percent, click OK. Click OK twice. The results are





Exercises 2.3

2.34 One important reason for grouping data is that grouping often makes a large and complicated set of data more compact and easier to understand.

2.35 For class limits, marks, cutpoints and midpoints to make sense, data must be numerical. They do not make sense for qualitative data classes because such data are nonnumerical.

2.36 The most important guidelines in choosing the classes for grouping a data set are: (1) the number of classes should be small enough to provide an effective summary, but large enough to display the relevant characteristics of the data; (2) each observation must belong to one, and only one, class; and (3) whenever feasible, all classes should have the same width.

2.37 In the first method for depicting classes called cutpoint grouping, we used the notation **a – under b** to mean values that are greater than or equal to **a** and up to, but not including **b**, such as 30 – under 40 to mean a range of values greater than or equal to 30, but strictly less than 40. In the alternate method called limit grouping, we used the notation **a-b** to indicate a class that extends from **a** to **b**, including both. For example, 30-39 is a class that includes both 30 and 39. The alternate method is especially appropriate when all of the data values are integers. If the data include values like 39.7 or 39.93, the first method is more advantageous since the cutpoints remain integers; whereas, in the alternate method, the upper limits for each class would have to be expressed in decimal form such as 39.9 or 39.99.

2.38 (a) For continuous data displayed to one or more decimal places, using the cutpoint grouping is best since the description of the classes is simpler, regardless of the number of decimal places displayed.

(b) For discrete data with relatively few distinct observations, the single value grouping is best since either of the other two methods would result in combining some of those distinct values into single classes, resulting in too few classes, possibly less than 5.

2.39 For limit grouping, we find the class mark, which is the average of the lower and upper class limit. For cutpoint grouping, we find the class midpoint, which is the average of the two cutpoints.

2.40 A frequency histogram shows the actual frequencies on the vertical axis; whereas, the relative frequency histogram always shows proportions (between 0 and 1) or percentages (between 0 and 100) on the vertical axis.

2.41 An advantage of the frequency histogram over a frequency distribution is that it is possible to get an overall view of the data more easily. A disadvantage of the frequency histogram is that it may not be possible to determine exact frequencies for the classes when the number of observations is large.

2.42 By showing the lower class limits (or cutpoints) on the horizontal axis, the range of possible data values in each class is immediately known and the class mark (or midpoint) can be quickly determined. This is particularly helpful if it is not convenient to make all classes the same width. The use of the class mark (or midpoint) is appropriate when each class consists of a single value (which is, of course, also the midpoint). Use of the class marks (or midpoints) is not appropriate in other situations since it may be difficult to determine the location of the class limits (or cutpoints) from the values of the class marks (or midpoints), particularly if the class marks (or midpoints) are not evenly spaced. Class Marks (or midpoints) cannot be used if there is an open class.

2.43 If the classes consist of single values, stem-and-leaf diagrams and frequency histograms are equally useful. If only one diagram is needed and the classes consist of more than one value, the stem-and-leaf diagram allows one to retrieve all of the original data values whereas the frequency histogram does not. If two or more sets of data of different sizes are to

be compared, the relative frequency histogram is advantageous because all of the diagrams to be compared will have the same total relative frequency of 1.00. Finally, stem-and-leaf diagrams are not very useful with very large data sets and may present problems with data having many digits in each number.

2.44 The histogram (especially one using relative frequencies) is generally preferable. Data sets with a large number of observations may result in a stem of the stem-and-leaf diagram having more leaves than will fit on the line. In that case, the histogram would be preferable.

2.45 You can reconstruct the stem-and-leaf diagram using two lines per stem. For example, instead of listing all of the values from 10 to 19 on a '1' stem, you can make two '1' stems. On the first, you record the values from 10 to 14 and on the second, the values from 15 to 19. If there are still two few stems, you can reconstruct the diagram using five lines per stem, recording 10 and 11 on the first line, 12 and 13 on the second, and so on.

2.46 For the number of bedrooms per single-family dwelling, single-value grouping is probably the best because the data is discrete with relatively few distinct observations.

2.47 For the ages of householders, given as a whole number, limit grouping is probably the best because the data are given as whole numbers and there are probably too many distinct observations to list them as single-value grouping.

2.48 For additional sleep obtained by a sample of 100 patients by using a particular brand of sleeping pill, cutpoint grouping is probably the best because the data is continuous and the data was recorded to the nearest tenth of an hour.

2.49 For the number of automobiles per family, single-value grouping is probably the best because the data is discrete with relatively few distinct observations.

2.50 For gas mileages, rounded to the nearest number of miles per gallon, limit grouping is probably the best because the data are given as whole numbers and there are probably too many distinct observations to list them as single-value grouping.

2.51 For carapace length for a sample of giant tarantulas, cutpoint grouping is probably the best because the data is continuous and the data was recorded to the nearest hundredth of a millimeter.

2.52 (a) Since the data values range from 0 to 4, we construct a table with classes based on a single value. The resulting table follows.

Number of Siblings	Frequency
0	8
1	17
2	11
3	3
4	1
	40

(b) To get the relative frequencies, divide each frequency by the sample size of 40.

Number of Siblings	Relative Frequency
0	0.200
1	0.425
2	0.275
3	0.075
4	0.025
	1.000

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in part (a) of this exercise. Column 1 demonstrates that the data are grouped using classes based on a single value. These single values in column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical axis units in the frequency histogram are the integers within the range 0 through 17, since these are representative of the magnitude and spread of the frequencies presented in column 2. When classes are based on a single value, the middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

Figure (a)

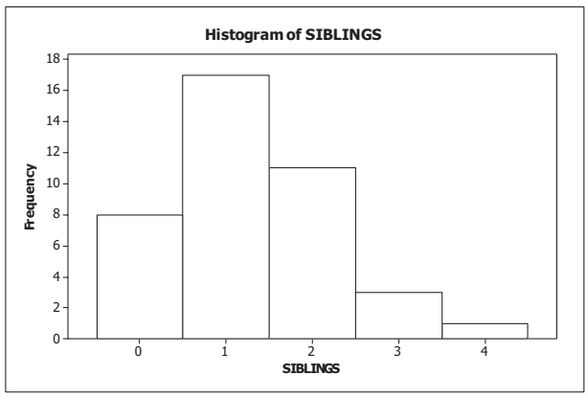
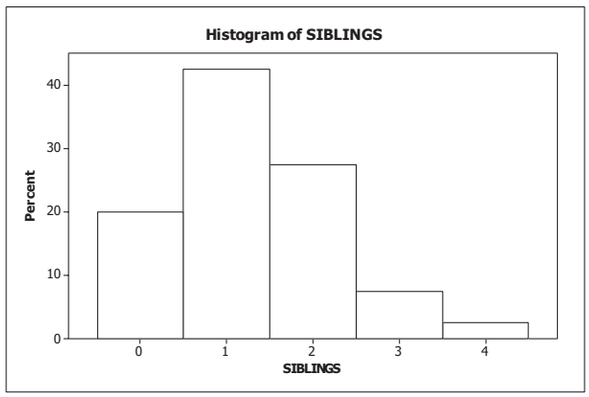


Figure (b)



(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 2 range in size from 0.025 to 0.425. Thus, suitable candidates for vertical-axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.45 (or 45%). The middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

2.53 (a) Since the data values range from 1 to 7, we construct a table with classes based on a single value. The resulting table follows.

Number of Persons	Frequency
1	7
2	13
3	9
4	5
5	4
6	1
7	1
	40

(b) To get the relative frequencies, divide each frequency by the sample size of 40.

Number of Persons	Relative Frequency
1	0.175
2	0.325
3	0.225
4	0.125
5	0.100
6	0.025
7	0.025
	1.000

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in part (a) of this exercise. Column 1 demonstrates that the data are grouped using classes based on a single value. These single values in column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical axis units in the frequency histogram are the integers within the range 0 through 13, since these are representative of the magnitude and spread of the frequencies presented in column 2. When classes are based on a single value, the middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

Figure (a)

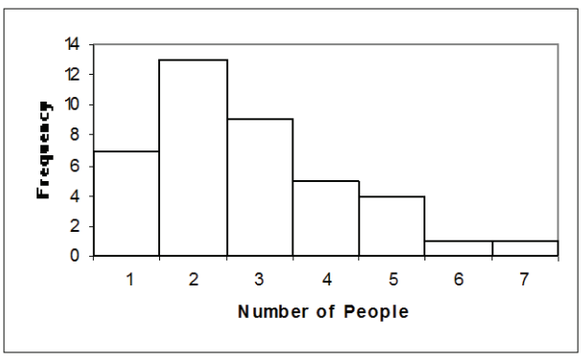
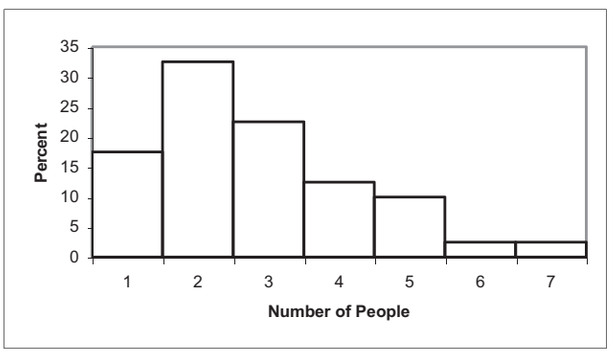


Figure (b)



(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 2 range in size from 0.025 to 0.325. Thus, suitable candidates for vertical-axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.35 (or 35%). The middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

2.54 (a) Since the data values range from 1 to 8, we construct a table with classes based on a single value. The resulting table follows.

Litter Size	Frequency
1	1
2	0
3	1
4	3
5	7
6	6
7	4
8	2
	24

(b) To get the relative frequencies, divide each frequency by the sample size of 24.

Litter Size	Relative Frequency
1	0.042
2	0.000
3	0.042
4	0.125
5	0.292
6	0.250
7	0.167
8	0.083
	1.000

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in part (a) of this exercise. Column 1 demonstrates that the data are grouped using classes based on a single value. These single values in column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical axis units in the frequency histogram are the integers within the range 0 through 7, since these are representative of the magnitude and spread of the frequencies presented in column 2. When classes are based on a single value, the middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

Figure (a)

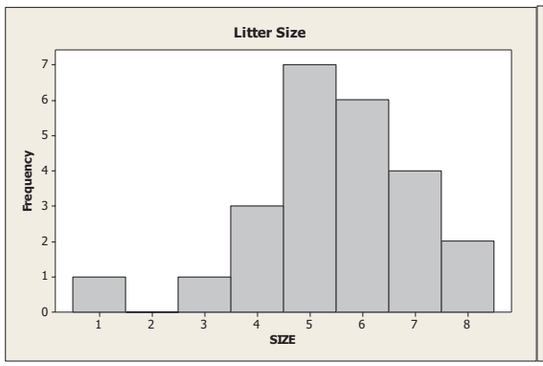
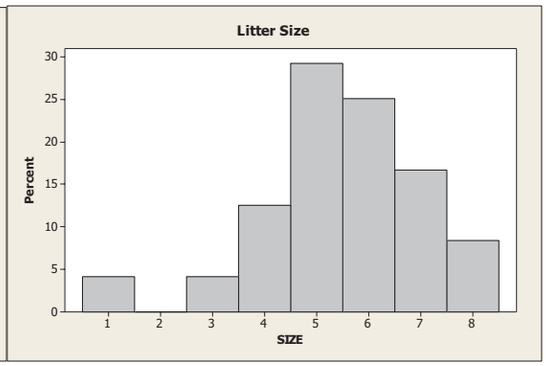


Figure (b)



(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 2 range in size from 0.000 to 0.292. Thus, suitable candidates for vertical-axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.30 (or 30%). The middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

2.55 (a) Since the data values range from 1 to 10, we construct a table with classes based on a single value. The resulting table follows.

Number of Radios	Frequency
1	1
2	1
3	3
4	12
5	6
6	4
7	5
8	4
9	6
10	3
	45

(b) To get the relative frequencies, divide each frequency by the sample size of 45.

Number of Radios	Relative Frequency
1	0.022
2	0.022
3	0.067
4	0.267
5	0.133
6	0.089
7	0.111
8	0.089
9	0.133
10	0.067
	1.000

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in part (a) of this exercise. Column 1 demonstrates that the data are grouped using classes based on a single value. These single values in column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical axis units in the frequency histogram are the integers within the range 1 through 12, since these are representative of the magnitude and spread of the frequencies presented in column 2. When classes are based on a single value, the middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

Figure (a)

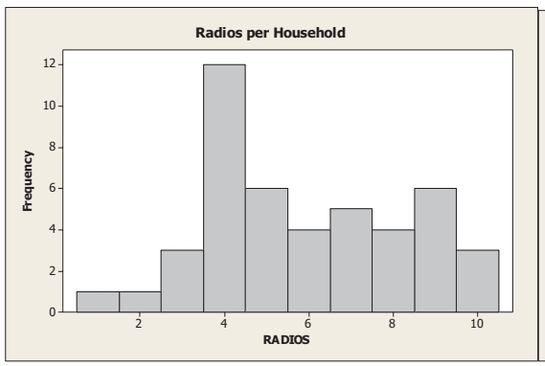
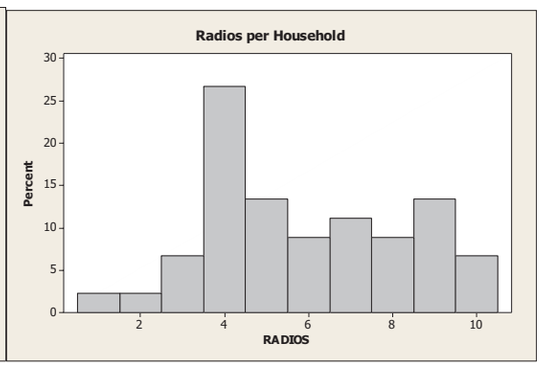


Figure (b)



(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 2 range in size from 0.022 to 0.267. Thus, suitable candidates for vertical-axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.30 (or 30%). The middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

2.56 (a) The first class to construct is 40-49. Since all classes are to be of equal width, and the second class begins with 50, we know that the width of all classes is 50 - 40 = 10. All of the classes are presented

in column 1. The last class to construct is 150-159, since the largest single data value is 155. Having established the classes, we tally the energy consumption figures into their respective classes. These results are presented in column 2, which lists the frequencies.

Consumption (mil. BTU)	Frequency
40-49	1
50-59	7
60-69	7
70-79	3
80-89	6
90-99	10
100-109	5
110-119	4
120-129	2
130-139	3
140-149	0
150-159	2
	50

(b) Dividing each frequency by the total number of observations, which is 50, results in each class's relative frequency. The relative frequencies for all classes are presented in column 2. The resulting table follows.

Consumption (mil. BTU)	Relative Frequency
40-49	0.02
50-59	0.14
60-69	0.14
70-79	0.06
80-89	0.12
90-99	0.20
100-109	0.10
110-119	0.08
120-129	0.04
130-139	0.06
140-149	0.00
150-159	0.04
	1.00

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in part (a) of this exercise. The lower class limits of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical-axis units in the frequency histogram are the even integers 0 through 10, since these are representative of the magnitude and spread of the frequency presented in column 2. The height of each bar in the frequency histogram matches the respective frequency in column 2.

(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 2 vary in size from 0.00 to 0.20. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.20 (or 20%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

Figure (a)

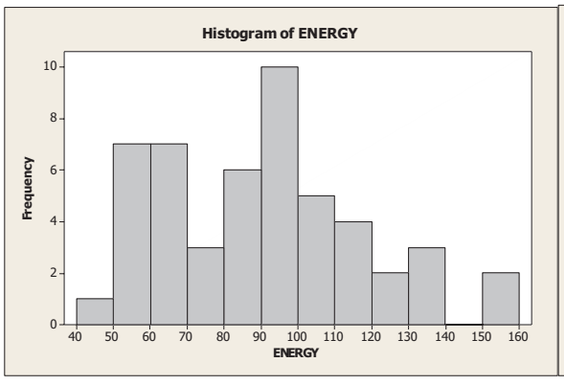
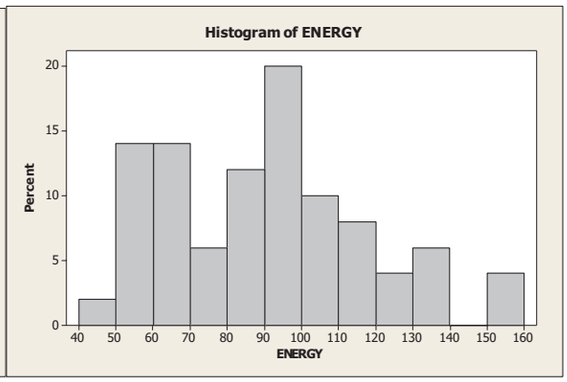


Figure (b)



2.57 (a) The first class to construct is 40-44. Since all classes are to be of equal width, and the second class begins with 45, we know that the width of all classes is 45 - 40 = 5. All of the classes are presented in column 1. The last class to construct is 60-64, since the largest single data value is 61. Having established the classes, we tally the age figures into their respective classes. These results are presented in column 2, which lists the frequencies.

Age	Frequency
40-44	4
45-49	3
50-54	4
55-59	8
60-64	2
	21

(b) Dividing each frequency by the total number of observations, which is 21, results in each class's relative frequency. The relative frequencies for all classes are presented in column 2. The resulting table follows.

Age	Relative Frequency
40-44	0.190
45-49	0.143
50-54	0.190
55-59	0.381
60-64	0.095
	1.000

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in part (a) of this exercise. The lower class limits of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical-axis units in the frequency histogram are the even integers 2 through 8, since these are representative of the magnitude and spread of the frequency presented in column 2. The height of each bar in the frequency histogram matches the respective frequency in column 2.

(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 2 vary in size from 0.095 to 0.381. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.10 (or

10%), from zero to 0.40 (or 40%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

Figure (a)

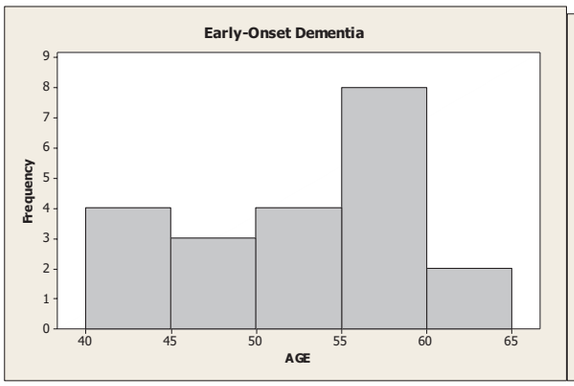
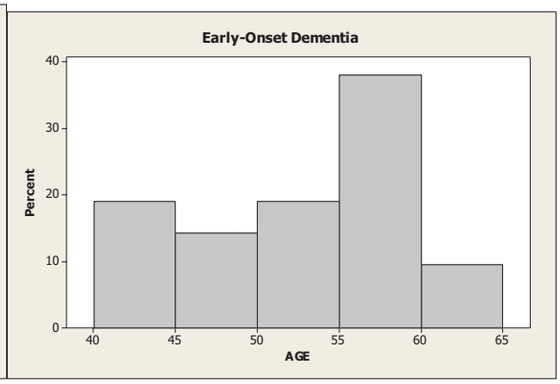


Figure (b)



2.58 (a) The first class to construct is 20-22. Since all classes are to be of equal width, and the second class begins with 23, we know that the width of all classes is 23 - 20 = 3. All of the classes are presented in column 1. The last class to construct is 44-46, since the largest single data value is 45. Having established the classes, we tally the cheese consumption figures into their respective classes. These results are presented in column 2, which lists the frequencies.

Cheese Consumption	Frequency
20-22	2
23-25	3
26-28	4
29-31	7
32-34	6
35-37	5
38-40	3
41-43	3
44-46	2
	35

(b) Dividing each frequency by the total number of observations, which is 35, results in each class's relative frequency. The relative frequencies for all classes are presented in column 2. The resulting table follows.

Cheese Consumption	Relative Frequency
20-22	0.057
23-25	0.086
26-28	0.114
29-31	0.200
32-34	0.171
35-37	0.143
38-40	0.086
41-43	0.086
44-46	0.057
	1.000

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in part (a) of this exercise. The

lower class limits of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical-axis units in the frequency histogram are the even integers 2 through 7, since these are representative of the magnitude and spread of the frequency presented in column 2. The height of each bar in the frequency histogram matches the respective frequency in column 2.

(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 2 vary in size from 0.057 to 0.200. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.20 (or 20%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

Figure (a)

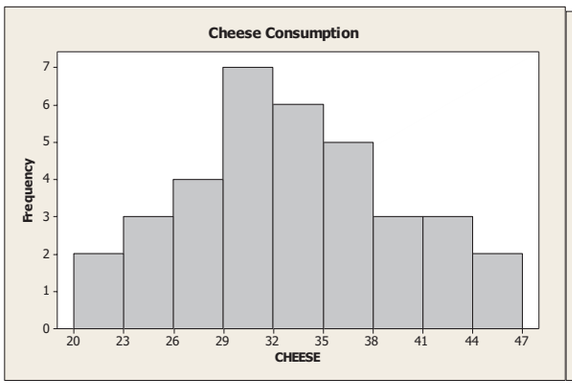
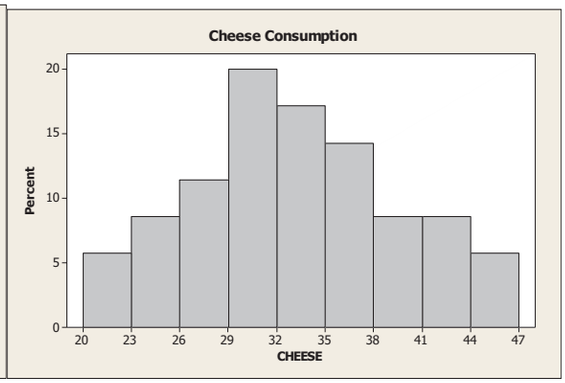


Figure (b)



2.59 (a) The first class to construct is 12-17. Since all classes are to be of equal width, and the second class begins with 18, we know that the width of all classes is 18 - 12 = 6. All of the classes are presented in column 1. The last class to construct is 60-65, since the largest single data value is 61. Having established the classes, we tally the anxiety questionnaire score figures into their respective classes. These results are presented in column 2, which lists the frequencies.

Anxiety	Frequency
12-17	2
18-23	3
24-29	6
30-35	5
36-41	10
42-47	4
48-53	0
54-59	0
60-65	1
	31

(b) Dividing each frequency by the total number of observations, which is 31, results in each class's relative frequency. The relative frequencies for all classes are presented in column 2. The resulting table follows.

Anxiety	Relative Frequency
12-17	0.065
18-23	0.097
24-29	0.194
30-35	0.161
36-41	0.323
42-47	0.129
48-53	0.000
54-59	0.000
60-65	0.032
	1.000

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in part (a) of this exercise. The lower class limits of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical-axis units in the frequency histogram are the even integers 0 through 10, since these are representative of the magnitude and spread of the frequency presented in column 2. The height of each bar in the frequency histogram matches the respective frequency in column 2.

(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 2 vary in size from 0.000 to 0.323. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (or 5%), from zero to 0.35 (or 35%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

Figure (a)

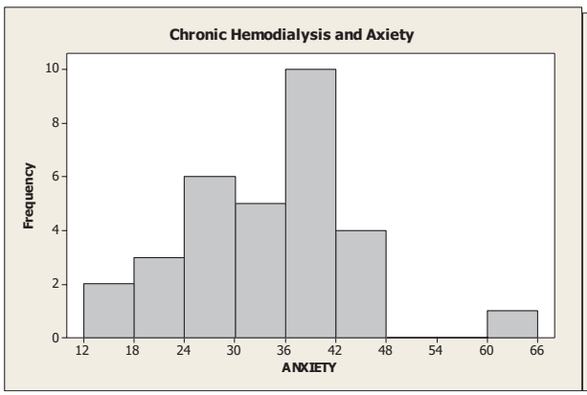
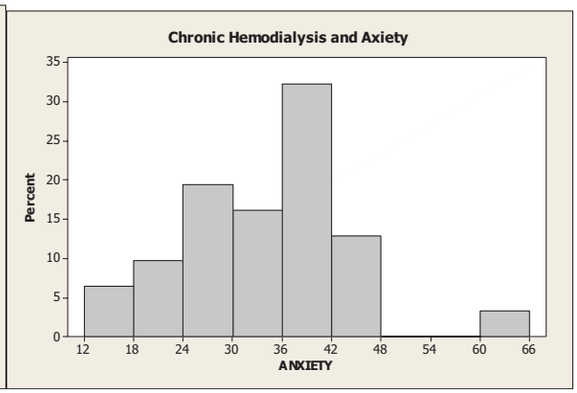


Figure (b)



2.60 (a) The first class to construct is 12 – under 13. Since all classes are to be of equal width 1, the second class is 13 – under 14. All of the classes are presented in column 1. The last class to construct is 19 – under 20, since the largest single data value is 19.492. Having established the classes, we tally the audience sizes into their respective classes. These results are presented in column 2, which lists the frequencies.

Audience (Millions)	Frequency
12 – under 13	3
13 – under 14	5
14 – under 15	4
15 – under 16	4
16 – under 17	1
17 – under 18	1
18 – under 19	1
19 – under 20	1
	20

(b) Dividing each frequency by the total number of observations, which is 20, results in each class's relative frequency. The relative frequencies for all classes are presented in column 2

Audience (Millions)	Relative Frequency
12 – under 13	0.15
13 – under 14	0.25
14 – under 15	0.20
15 – under 16	0.20
16 – under 17	0.05
17 – under 18	0.05
18 – under 19	0.05
19 – under 20	0.05
	1.00

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution obtained in part (a) of this exercise Column 1 demonstrates that the data are grouped using classes with class widths of 1. Suitable candidates for vertical axis units in the frequency histogram are the integers within the range 1 through 5, since these are representative of the magnitude and spread of the frequencies presented in column 2. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution obtained in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.05 to 0.20. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (5%), from zero to 0.20 (20%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

Figure (a)

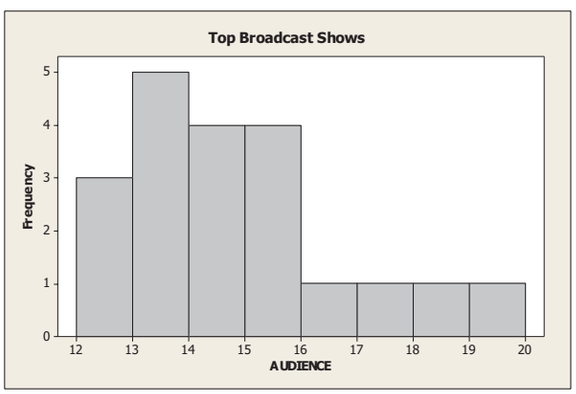
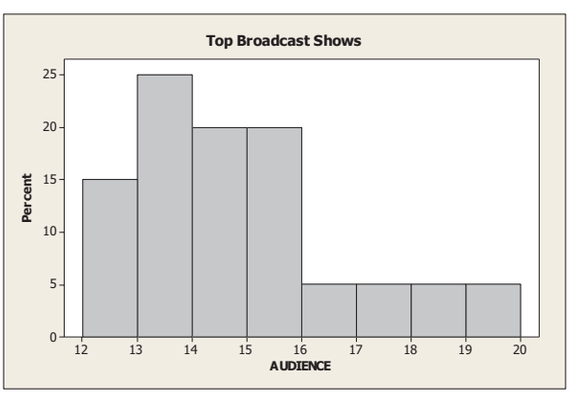


Figure (b)



2.61 (a) The first class to construct is 52 – under 54. Since all classes are to be of equal width 2, the second class is 54 – under 56. All of the classes are presented in column 1. The last class to construct is 74 – under 76, since the largest single data value is 75.3. Having established the classes, we tally the cheetah speeds into their respective classes. These results are presented in column 2, which lists the frequencies.

Speed	Frequency
52 – under 54	2
54 – under 56	5
56 – under 58	6
58 – under 60	8
60 – under 62	7
62 – under 64	3
64 – under 66	2
66 – under 68	1
68 – under 70	0
70 – under 72	0
72 – under 74	0
74 – under 76	1
	35

(b) Dividing each frequency by the total number of observations, which is 35, results in each class's relative frequency. The relative frequencies for all classes are presented in column 2

Speed	Relative Frequency
52 – under 54	0.057
54 – under 56	0.143
56 – under 58	0.171
58 – under 60	0.229
60 – under 62	0.200
62 – under 64	0.086
64 – under 66	0.057
66 – under 68	0.029
68 – under 70	0.000
70 – under 72	0.000
72 – under 74	0.000
74 – under 76	0.029
	1.000

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution obtained in part (a) of this exercise Column 1 demonstrates that the data are grouped using classes with class widths of 2. Suitable candidates for vertical axis units in the frequency histogram are the integers within the range 0 through 8, since these are representative of the magnitude and spread of the frequencies presented in column 2. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution obtained in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.000 to 0.229. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (5%), from zero to 0.25 (25%). The height of each bar in the relative-

frequency histogram matches the respective relative frequency in column 2.

Figure (a)

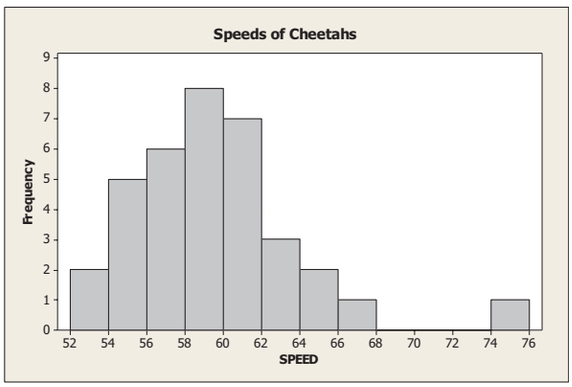
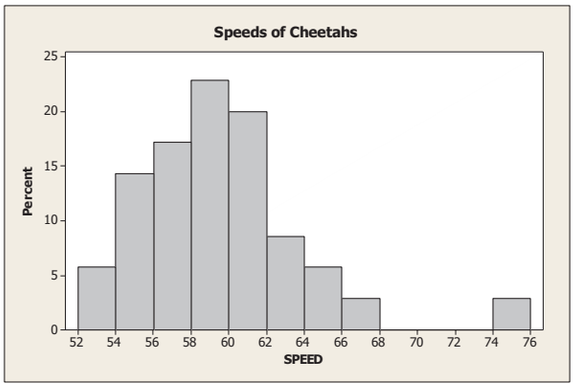


Figure (b)



2.62 (a) The first class to construct is 12 – under 14. Since all classes are to be of equal width 2, the second class is 14 – under 66. All of the classes are presented in column 1. The last class to construct is 26 – under 28, since the largest single data value is 26.4. Having established the classes, we tally the fuel tank capacities into their respective classes. These results are presented in column 2, which lists the frequencies.

Fuel Tank Capacity	Frequency
12 – under 14	2
14 – under 16	6
16 – under 18	7
18 – under 20	6
20 – under 22	6
22 – under 24	3
24 – under 26	3
26 – under 28	2
	35

(b) Dividing each frequency by the total number of observations, which is 35, results in each class's relative frequency. The relative frequencies for all classes are presented in column 2

Fuel Tank Capacity	Relative Frequency
12 – under 14	0.057
14 – under 16	0.171
16 – under 18	0.200
18 – under 20	0.171
20 – under 22	0.171
22 – under 24	0.086
24 – under 26	0.086
26 – under 28	0.057
	1.000

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution obtained in part (a) of this exercise Column 1 demonstrates that the data are grouped using classes with class widths of 2. Suitable candidates for vertical axis units in the frequency histogram are the integers within the range 2 through 7, since these are representative of the magnitude and spread of the frequencies presented in column 2. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

(d) The relative-frequency histogram in Figure (b) is constructed using

the relative-frequency distribution obtained in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.057 to 0.200. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (5%), from zero to 0.20 (20%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

Figure (a)

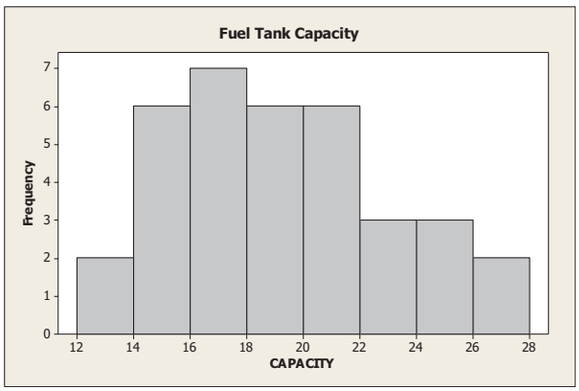
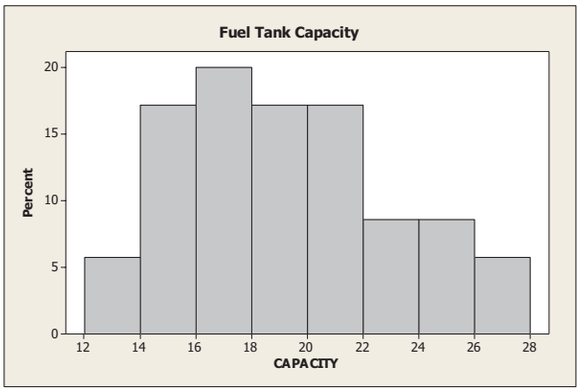


Figure (b)



2.63 (a) The first class to construct is 0 – under 1. Since all classes are to be of equal width 1, the second class is 1 – under 2. All of the classes are presented in column 1. The last class to construct is 7 - under 8, since the largest single data value is 7.6. Having established the classes, we tally the fuel tank capacities into their respective classes. These results are presented in column 2, which lists the frequencies.

Oxygen Distribution	Frequency
0 – under 1	1
1 – under 2	10
2 – under 3	5
3 – under 4	4
4 – under 5	0
5 – under 6	0
6 – under 7	1
7 – under 8	1
	22

(b) Dividing each frequency by the total number of observations, which is 22, results in each class's relative frequency. The relative frequencies for all classes are presented in column 2

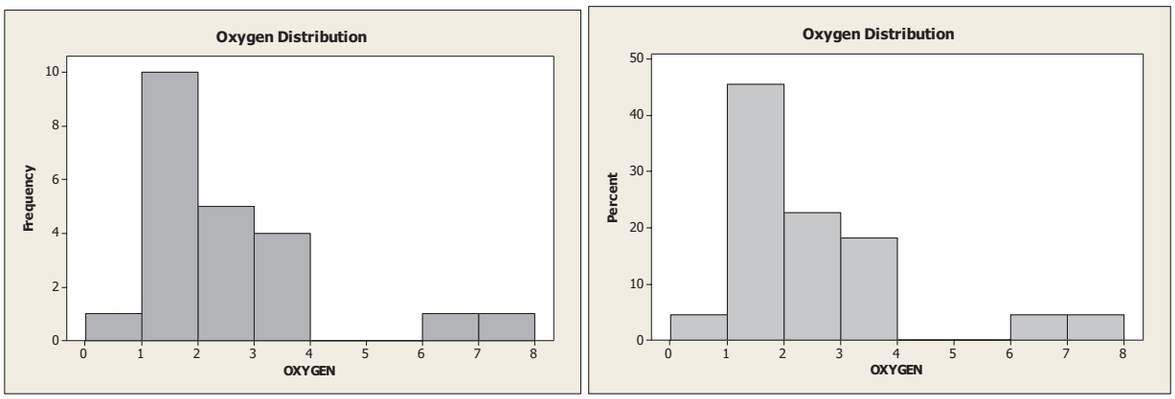
Oxygen Distribution	Relative Frequency
0 – under 1	0.045
1 – under 2	0.455
2 – under 3	0.227
3 – under 4	0.182
4 – under 5	0.000
5 – under 6	0.000
6 – under 7	0.045
7 – under 8	0.045
	1.000

(c) The frequency histogram in Figure (a) is constructed using the frequency distribution obtained in part (a) of this exercise Column 1 demonstrates that the data are grouped using classes with class widths

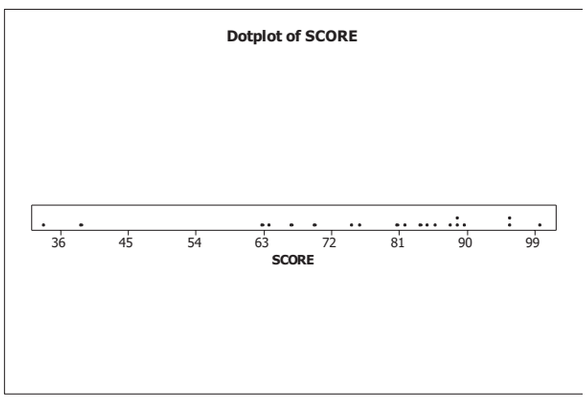
of 2. Suitable candidates for vertical axis units in the frequency histogram are the integers within the range 0 through 10, since these are representative of the magnitude and spread of the frequencies presented in column 2. Also, the height of each bar in the frequency histogram matches the respective frequency in column 2.

(d) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution obtained in part (b) of this exercise. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.000 to 0.455. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05 (5%), from zero to 0.50 (50%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 2.

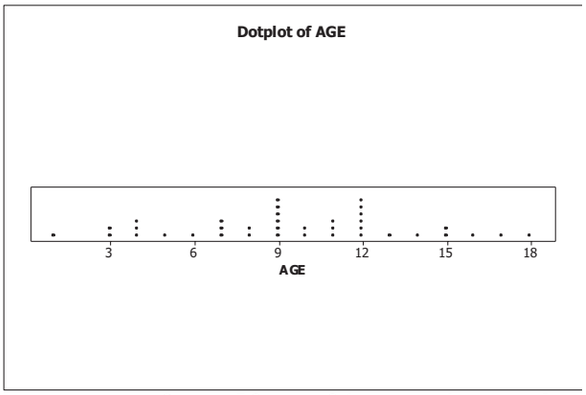
Figure (a) Figure (b)



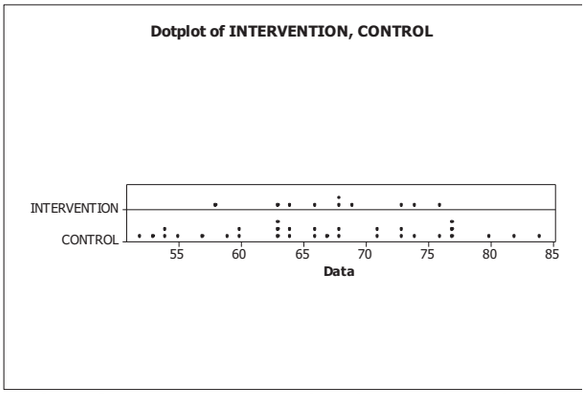
2.64 The horizontal axis of this dotplot displays a range of possible exam scores. To complete the dotplot, we go through the data set and record each exam score by placing a dot over the appropriate value on the horizontal axis.



2.65 The horizontal axis of this dotplot displays a range of possible ages. To complete the dotplot, we go through the data set and record each age by placing a dot over the appropriate value on the horizontal axis.

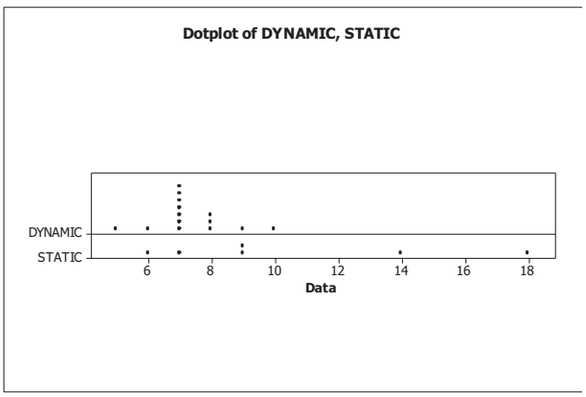


2.66 (a) The data values range from 52 to 84, so the scale must accommodate those values. We stack dots above each value on two different lines using the same scale for each line. The result is



(b) The two sets of pulse rates are both centered near 68, but the Intervention data are more concentrated around the center than are the Control data.

2.67 (a) The data values range from 7 to 18, so the scale must accommodate those values. We stack dots above each value on two different lines using the same scale for each line. The result is



(b) The Dynamic system does seem to reduce acute postoperative days in the hospital on the average. The Dynamic data are centered at about 7

days, whereas the Static data are centered at about 11 days and are much more spread out than the Dynamic data.

2.68 Since each data value consists of 3 or 4 digits ranging from 914 to 1060. The last digit becomes the leaf and the remaining digits are the stems, so we have stems of 91 to 106. The resulting stem-and-leaf diagram is

```
 91|  4
 92|
 93|
 94|  6
 95|  79
 96|  4
 97|  4577
 98|  46789
 99|  015679
100|  1
101|  0478
102|  58
103|  01
104|
105|
106|  0
```

2.69 Since each data value consists of 2 digits, each beginning with 1, 2, 3, or 4, we will construct the stem-and-leaf diagram with these four values as the stems. The result is

```
1|  238
2|  1678899
3|  34459
4|  04
```

2.70 (a) Since each data value consists of a 2 digit number with a one digit decimal, we will make the leaf the decimal digit and the stems the remaining two digit numbers of 28, 29, 30, and 31. The result is

```
28|  8
29|  368
30|  1247
31|  02
```

(b) Splitting into two lines per stem, leafs of 0-4 belong in the first stem and leafs of 5-9 belong in the second stem. The result is

```
28|  8
29|  3
29|  68
30|  124
30|  7
31|  02
```

(c) The stem-and-leaf diagram in part (b) is more useful because by splitting the stems into two lines per stem, you have created more lines. Part (a) had too few lines.

2.71 (a) Since each data value lies between 2 and 93, we will construct the stem-and-leaf diagram with one line per stem. The result is

```
0|  2234799
1|  11145566689
2|  023479
3|  004555
4|  19
5|  5
6|  9
7|  9
8|
9|  3
```

(b) Using two lines per stem, the same data result in the following diagram:

```
0|  2234
0|  799
1|  1114
1|  5566689
2|  0234
2|  79
3|  004
3|  555
4|  1
4|  9
5|
5|  5
6|
6|  9
7|
7|  9
8|
8|
9|  3
```

(c) The stem with one line per stem is more useful. One gets the same impression regarding the shape of the distribution, but the two lines per stem version has numerous lines with no data, making it take up more space than necessary to interpret the data and giving it too many lines.

2.72 (a) Since we have two digit numbers, the last digit becomes the leaf and the first digit becomes the stem. For this data, we have stems of 6 and 7. Splitting the data into five lines per stem, we put the leaves 0-1 in the first stem, 2-3 in the second stem, 4-5 in the third stem, 6-7 in the fourth stem, and 8-9 in the fifth stem. The result is

```
6|  899
7|  0001
7|  22222333333333
7|  44555555
7|  6666677
7|  88
```

(b) Using one or two lines per stem would have given us too few lines.

2.73 (a) Since we have two digit numbers, the last digit becomes the leaf and the first digit becomes the stem. For this data, we have stems of 6, 7, and 8. Splitting the data into five lines per stem, we put the leaves 0-1 in the first stem, 2-3 in the second stem, 4-5 in the third stem, 6-7 in the fourth stem, and 8-9 in the fifth stem. The result is

```
6|  99
7|  11
7|  222233
7|  44444444445555555555
7|  666667
7|  88
8|  1
```

(b) Using one or two lines per stem would have given us too few lines.

2.74 The heights of the bars of the relative-frequency histogram indicate that:

(a) About 27.5% of the returns had an adjusted gross income between $10,000 and $19,999, inclusive.

(b) About 37.5% were between $0 and $9,999; 27.5% were between $10,000 and $19,999; and 19% were between $20,000 and $29,999. Thus, about 84% (i.e., 37.5% + 27.5% + 19%) of the returns had an adjusted gross income less than $30,000.

(c) About 11% were between $30,000 and $39,999; and 5% were between $40,000 and $49,999. Thus, about 16% (i.e., 11% + 5%) of the returns had an adjusted gross income between $30,000 and $49,999. With 89,928,000 returns having an adjusted gross income less than $50,000, the number of returns having an adjusted gross income between $30,000 and $49,999 was 14,388,480 (i.e., 0.16 x 89,928,000).

2.75 The graph indicates that:

(a) 20% of the patients have cholesterol levels between 205 and 209, inclusive.

(b) 20% are between 215 and 219; and 5% are between 220 and 224. Thus, 25% (i.e., 20% + 5%) have cholesterol levels of 215 or higher.

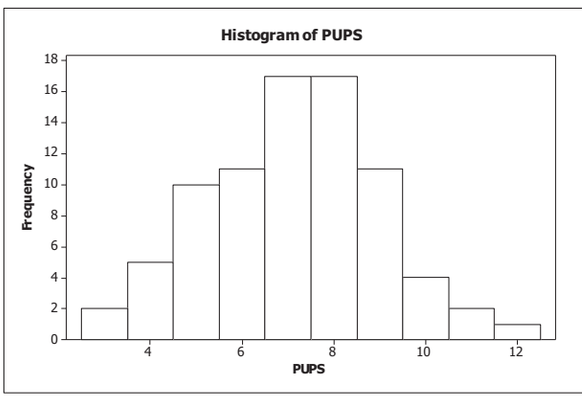
(c) 35% of the patients have cholesterol levels between 210 and 214, inclusive. With 20 patients in total, the number having cholesterol levels between 210 and 214 is 7 (i.e., 0.35 x 20).

2.76 (a) Using Minitab, retrieve the data from the Weiss-Stats-CD. Column 1 contains the numbers of pups borne in a lifetime for each of 80 female Great White Sharks. From the tool bar, select **Stat ▶ Tables ▶ Tally Individual Variables**, double-click on PUPS in the first box so that PUPS appears in the **Variables** box, put a check mark next to Counts and Percents under Display, and click **OK**. The result is

PUPS	Count	Percent
3	2	2.50
4	5	6.25
5	10	12.50
6	11	13.75
7	17	21.25
8	17	21.25
9	11	13.75
10	4	5.00
11	2	2.50
12	1	1.25

(b) After retrieving the data from the WeissStats CD, select **Graph ▶**

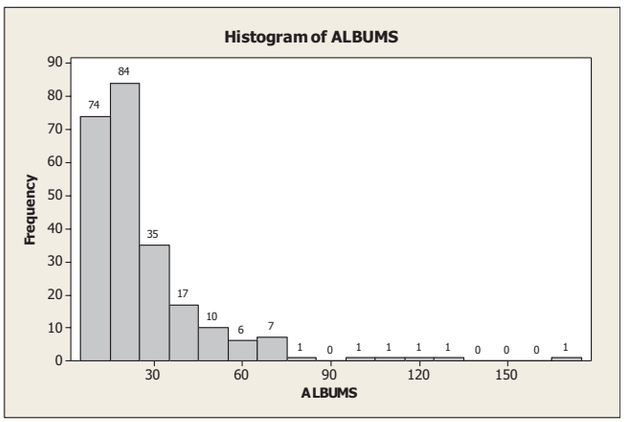
Histogram, choose **Simple** and click **OK**. Double click on PUPS in the first box to enter PUPS in the **Graphs variables** box, and click **OK**. The frequency histogram is



To change to a relative-frequency histogram, before clicking OK the second time, click on the **Scale** button and the **Y-Scale type** tab, and choose **Percent** and click **OK**. The graph will look like the frequency histogram, but will have relative frequencies on the vertical scale instead of counts.

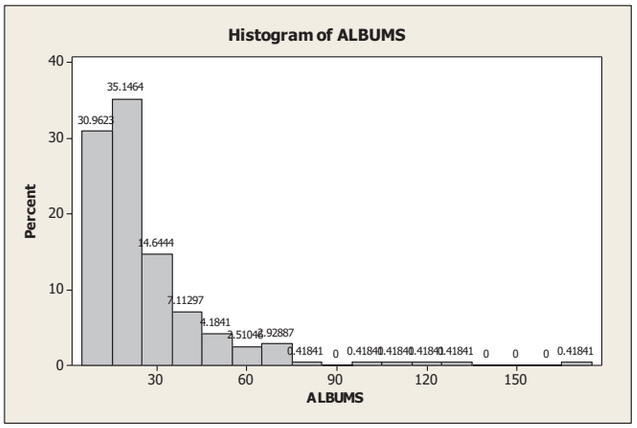
The numbers of pups range from 1 to 12 per female with 7 and 8 pups occurring more frequently than any other values.

2.77 (a) Using Minitab, there is not a direct way to get a grouped frequency distribution. However, you can use an option in creating your histogram that will report the frequencies in each of the classes, essentially creating a grouped frequency distribution. Retrieve the data from the Weiss-Stats-CD. Column 2 contains the number of albums sold, in millions, for the top recording artists. From the tool bar, , select **Graph ▶ Histogram**, choose **Simple** and click **OK**. Double click on ALBUMS to enter ALBUMS in the **Graph variables** box. Click **Labels**, click the **Data Labels** tab, then check **Use y-value labels**. Click **OK** twice. The result is



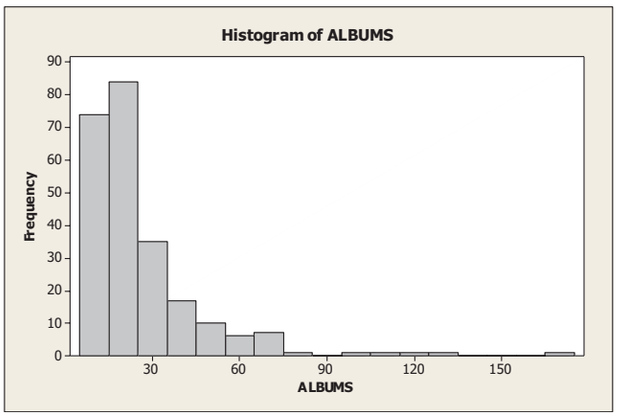
Above each bar is a label for each of the frequencies. Also, the labeling on the horizontal axis is the midpoint for class, where each class has a width of 10. The first class would be 5 – under 15, the second class would be 15 – under 25, etc. To get the relative-

frequency distribution, follow the same steps as above, but also Click the **Scale** button, click the **Y-scale Type**, check **Percent**, then click **OK** twice. The result is



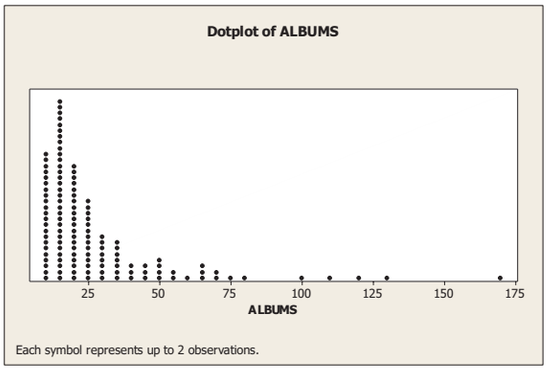
Above each bar is the percentage for that class, essentially creating a relative-frequency distribution. You could also transfer these results into a table.

(b) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Histogram**, choose **Simple** and click **OK**. Double click on ALBUMS to enter ALBUMS in the **Graph variables** box and click **OK**. The result is



The graph shows that there are only a few artists who sell many units and many artists who sell relatively few units.

(c) To obtain the dotplot, select **Graph ▶ Dotplot**, select **Simple** in the **One Y** row, and click **OK**. Double click on ALBUMS to enter ALBUMS in the **Graph variables** box and click **OK**. The result is



(d) The graphs are similar, but not identical. This is because Minitab grouped the data values slightly differently for the two graphs. The overall impression, however, remains the same.

2.78 (a) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Stem-and-Leaf**, double click on PERCENT to enter PERCENT in the **Graph variables** box and enter a 10 in the **Increment** box, and click **OK**. The result is

```
Stem-and-leaf of PERCENT  N  = 51
Leaf Unit = 1.0

 1    7  9
(33)  8  00111222333344456677778888889999
 17   9  00000001111111223
```

(b) Repeat part (a), but this time enter a 5 in the **Increment** box. The result is

```
Stem-and-leaf of PERCENT  N  = 51
Leaf Unit = 1.0

 1    7  9
 16   8  001112223333444
(18)  8  566777778888889999
 17   9  00000001111111223
```

(c) Repeat part (a) again, but this time enter a 2 in the **Increment** box. The result is

```
Stem-and-leaf of PERCENT  N  = 51
Leaf Unit = 1.0

 1    7  9
 6    8  00111
 13   8  2223333
 17   8  4445
 24   8  6677777
(10)  8  8888889999
 17   9  00000001111111
 3    9  223
```

(d) The last graph is the most useful since it gives a better idea of the shape of the distribution. Typically, we like to have five to fifteen classes and this is the only one of the three graphs that satisfies that condition.

2.79 (a) After entering the data from the WeissStats CD, in Minitab, select

Graph ▶ Stem-and-Leaf, double click on RATE to enter RATE in the **Graph variables** box and enter a 10 in the **Increment** box, and click **OK**. The result is

```
Stem-and-leaf of RATE  N  = 51
Leaf Unit = 1.0

 1    1  9
 14   2  0145678888999
(13)  3  0014445667788
 24   4  012223345566677899
 6    5  00124
 1    6  2
```

(b) Repeat part (a), but this time enter a 5 in the **Increment** box. The result is

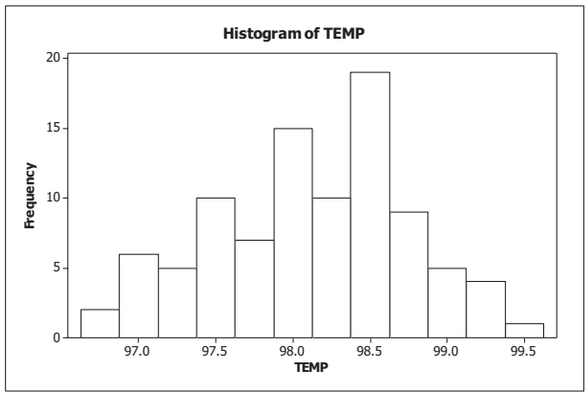
```
Stem-and-leaf of RATE  N  = 51
Leaf Unit = 1.0

 1   1  9
 4   2  014
 14  2  5678888999
 20  3  001444
(7)  3  5667788
 24  4  01222334
 16  4  5566677899
 6   5  00124
 1   5
 1   6  2
```

(c) Repeat part (a) again, but this time enter a 2 in the **Increment** box. The result is

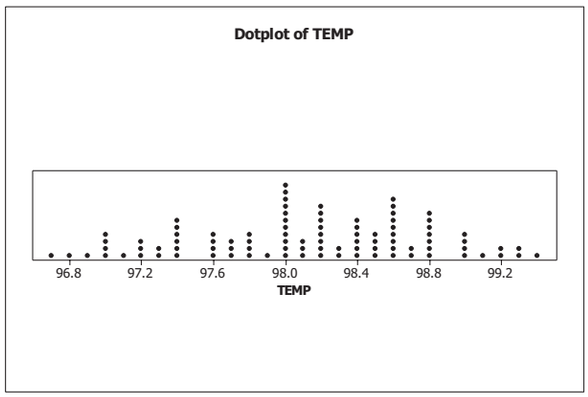
```
Stem-and-leaf of RATE  N  = 51
Leaf Unit = 1.0

 1   1  9
 3   2  01
 3   2
 5   2  45
 7   2  67
 14  2  8888999
 17  3  001
 17  3
 21  3  4445
 25  3  6677
(2)  3  88
 24  4  01
 22  4  22233
 17  4  455
 14  4  66677
 9   4  899
 6   5  001
 3   5  2
 2   5  4
 1   5
 1   5
 1   6
 1   6  2
```

(d) The second graph is the most useful. The third one has more classes than necessary to comprehend the shape of the distribution and has a number of empty stems. Typically, we like to have five to fifteen classes and the first and second diagrams satisfy that condition, but the second one provides a better idea of the shape of the distribution.

2.80 (a) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Histogram**, select **Simple** and click **OK**. double click on TEMP to enter TEMP in the **Graph variables** box and click **OK**. The result is



(b) Now select **Graph ▶ Dotplot**, select **Simple** in the **One Y** row, and click **OK**. Double click on TEMP to enter TEMP in the **Graph variables** box and click **OK**. The result is



(c) Now select **Graph ▶ Stem-and-Leaf**, double click on TEMP to enter TEMP in the **Graph variables** box and click **OK**. Leave the **Increment** box blank to allow Minitab to choose the number of lines per stem. The result is

```
Stem-and-leaf of TEMP  N  = 93
Leaf Unit = 0.10
 1    96  7
 3    96  89
 8    97  00001
 13   97  22233
 19   97  444444
 26   97  6666777
 31   97  88889
 45   98  00000000000111
(10)  98  2222222233
 38   98  4444445555
 28   98  66666666677
 17   98  8888888
 10   99  00001
 5    99  2233
 1    99  4
```

(d) The dotplot shows all of the individual values. The stem-and-leaf diagram used five lines per stem and therefore each line contains leaves with possibly two values. The histogram chose classes of width 0.25. This resulted in, for example, the class with midpoint 97.0 including all of the values 96.9, 97.0, and 97.1, while the class with midpoint 97.25 includes only the two values 97.2 and 97.3. Thus the 'smoothing' effect is not as good in the histogram as it is in the stem-and-leaf diagram. Overall, the dotplot gives the truest picture of the data and allows recovery of all of the data values.

2.81 (a) The classes are presented in column 1. With the classes established, we then tally the exam scores into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of exam scores, which is 20, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3. The class mark of each class is the average of the lower and upper limits. The class marks for all classes are presented in column 4.

Score	Frequency	Relative Frequency	Class Mark
30-39	2	0.10	34.5
40-49	0	0.00	44.5
50-59	0	0.00	54.5
60-69	3	0.15	64.5
70-79	3	0.15	74.5
80-89	8	0.40	84.5
90-100	4	0.20	95.0
	20	1.00	

(b) The first six classes have width 10; the seventh class had width 11.

(c) Answers will vary, but one choice is to keep the first six classes the same and make the next two classes 90-99 and 100-109. Another possibility is 31-40, 41-50, …, 91-100.

2.82 Answers will vary, but by following the steps we first decide on the approximate number of classes. Since there are 40 observations, we should have 7-14 classes. This exercise states we should have approximately seven classes. Step 2 says that we calculate an approximate class width as $(99 - 36)/7 = 9$. A convenient class width close to 9 would be a class width of 10. Step 3 says that we choose a number for the lower class limit which is less than or equal to our minimum observation of 36. Let's choose 35. Beginning with a lower class limit of 35 and width of 10, we have a first class of 35-44, a second class of 45-54, a third class of 55-64, a fourth class of 65-74, a fifth class of 75-84, a sixth class of 85-94, and a seventh class of 95-104. This would be our last class since the largest observation is 99.

2.83 Answers will vary, but by following the steps we first decide on the approximate number of classes. Since there are 37 observations, we should have 7-14 classes. This exercise states we should have approximately eight classes. Step 2 says that we calculate an approximate class width as $(278.8 - 129.2)/8 = 18.7$. A convenient class width close to 18.7 would be a class width of 20. Step 3 says that we choose a number for the lower cutpoint which is less than or equal to our minimum observation of 129.2. Let's choose 120. Beginning with a lower cutpoint of 120 and width of 20, we have a first class of 120 – under 140, a second class of 140 – under 160, a third class of 160 – under 180, a fourth class of 180 - under 200, a fifth

class of 200 – under 220, a sixth class of 220 – under 240, a seventh class of 240 – under 260, and an eighth class of 260 – under 280. This would be our last class since the largest observation is 278.8.

2.84 (a) Tally marks for all 50 students, where each student is categorized by age and gender, are presented in the contingency table given in part (b).

(b) Tally marks in each box appearing in the following chart are counted. These counts, or frequencies, replace the tally marks in the contingency table. For each row and each column, the frequencies are added, and their sums are recorded in the proper "Total" box.

Age (yrs)

Gender	Under 21	21 - 25	Over 25	Total
Male	||||| |||	||||| ||||| ||	||	
Female	||||| ||||| ||	||||| ||||| |||	|||	
Total				

Age (yrs)

Gender	Under 21	21-25	Over 25	Total
Male	8	12	2	22
Female	12	13	3	28
Total	20	25	5	50

(c) The row and column totals represent the total number of students in each of the corresponding categories. For example, the row total of 22 indicates that 22 of the students in the class are male.

(d) The sum of the row totals is 50, and the sum of the column totals is 50. The sums are equal because they both represent the total number of students in the class.

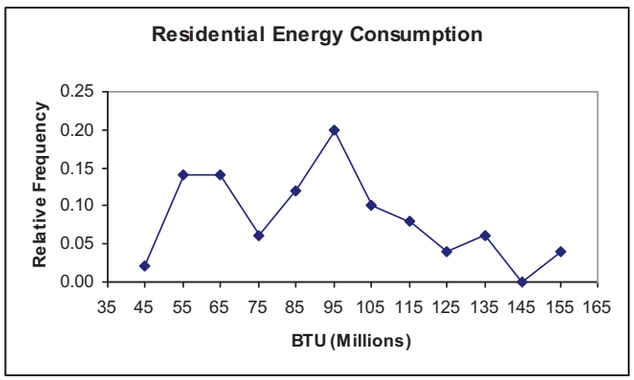
(e) Dividing each frequency reported in part (b) by the grand total of 50 students results in a contingency table that gives relative frequencies.

Age (yrs)

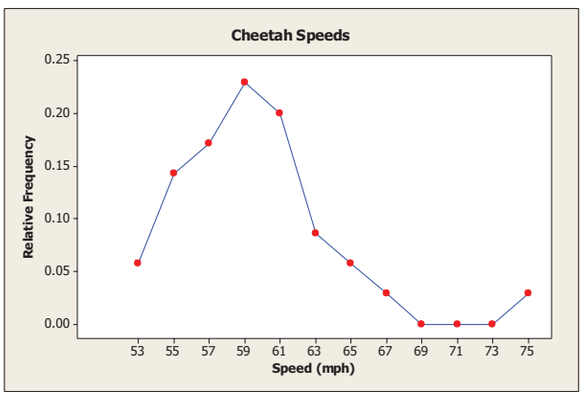
Gender	Under 21	21-25	Over 25	Total
Male	0.16	0.24	0.04	0.44
Female	0.24	0.26	0.06	0.56
Total	0.40	0.50	0.10	1.00

(f) The 0.16 in the upper left-hand cell indicates that 16% of the students in the class are males *and* under 21. The 0.40 in the lower left-hand cell indicates that 40% of the students in the class are under age 21. A similar interpretation holds for the remaining entries.

2.85 Consider columns 1 and 2 of the energy-consumption data given in Exercise 2.56 part (b). Compute the class mark for each class presented in column 1. Pair each class mark with its corresponding relative frequency found in column 2. Construct a horizontal axis, where the units are in terms of class marks and a vertical axis where the units are in terms of relative frequencies. For each class mark on the horizontal axis, plot a point whose height is equal to the relative frequency of the class. Then join the points with connecting lines. The result is a relative-frequency polygon.



2.86 Consider columns 1 and 2 of the Cheetah speed data given in Exercise 2.61 part (b). Compute the midpoint for each class presented in column 1. Pair each midpoint with its corresponding relative frequency found in column 2. Construct a horizontal axis, where the units are in terms of midpoints and a vertical axis where the units are in terms of relative frequencies. For each midpoint on the horizontal axis, plot a point whose height is equal to the relative frequency of the class. Then join the points with connecting lines. The result is a relative-frequency polygon.



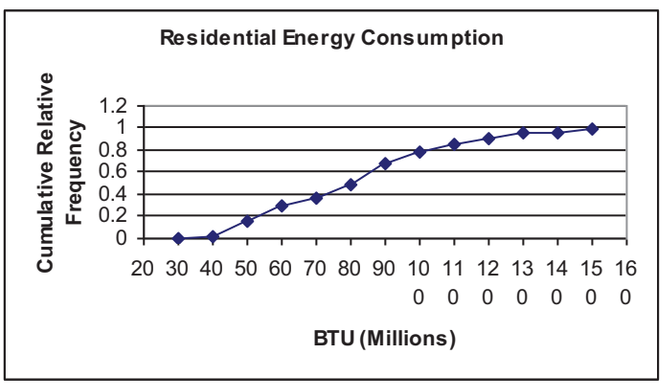
2.87 In single value grouping the horizontal axis would be labeled with the value of each class.

2.88 (a) Consider parts (a) and (b) of the energy-consumption data given in Exercise 2.56. The classes are now reworked to present just the lower class limit of each class. The frequencies are reworked to sum the frequencies of all classes representing values less than the specified lower class limit. These successive sums are the cumulative frequencies. The relative frequencies are reworked to sum the relative frequencies of all classes representing values less than the specified class limits. These successive sums are the cumulative relative frequencies. (Note: The cumulative relative frequencies can

also be found by dividing the each cumulative frequency by the total number of data values.)

Less than	Cumulative Frequency	Cumulative Relative Frequency
40	0	0.00
50	1	0.02
60	8	0.16
70	15	0.30
80	18	0.36
90	24	0.48
100	34	0.68
110	39	0.78
120	43	0.86
130	45	0.90
140	48	0.96
150	48	0.96
160	50	1.00

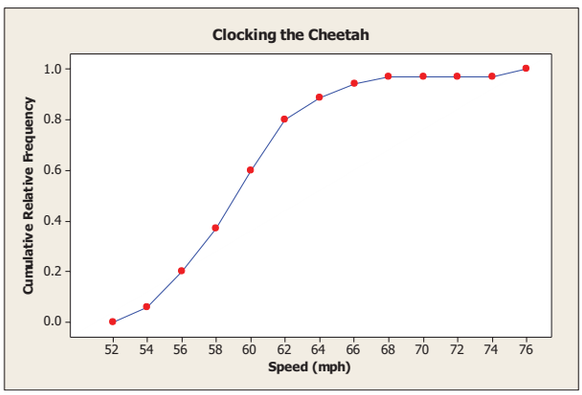
(b) Pair each class limit with its corresponding cumulative relative frequency found in column 3. Construct a horizontal axis, where the units are in terms of the class limits and a vertical axis where the units are in terms of cumulative relative frequencies. For each class limit on the horizontal axis, plot a point whose height is equal to the cumulative relative frequency. Then join the points with connecting lines. The result, presented in the following figure, is an *ogive* using cumulative relative frequencies. (Note: A similar procedure could be followed using cumulative frequencies.)



2.89 (a) Consider parts (a) and (b) of the Cheetah speed data given in Exercise 2.61. The classes are now reworked to present just the lower cutpoint of each class. The frequencies are reworked to sum the frequencies of all classes representing values less than the specified lower cutpoint. These successive sums are the cumulative frequencies. The relative frequencies are reworked to sum the relative frequencies of all classes representing values less than the specified cutpoints. These successive sums are the cumulative relative frequencies. (Note: The cumulative relative frequencies can also be found by dividing the each cumulative frequency by the total number of data values.)

Less than	Cumulative Frequency	Cumulative Relative Frequency
52	0	0.000
54	2	0.057
56	7	0.200
58	13	0.371
60	21	0.600
62	28	0.800
64	31	0.886
66	33	0.943
68	34	0.971
70	34	0.971
72	34	0.971
74	34	0.971
76	35	1.000

(b) Pair each cutpoint with its corresponding cumulative relative frequency found in column 3. Construct a horizontal axis, where the units are in terms of the cutpoints and a vertical axis where the units are in terms of cumulative relative frequencies. For each cutpoint on the horizontal axis, plot a point whose height is equal to the cumulative relative frequency. Then join the points with connecting lines. The result, presented in the following figure, is an *ogive* using cumulative relative frequencies. (Note: A similar procedure could be followed using cumulative frequencies.)



2.90 (a) After rounding each observation to the nearest year, the stem-and-leaf diagram for the rounded ages is

```
5|  334469
6|  6
7|  256689
8|  234678
9|  8
```

(b) After truncating each weight by dropping the decimal part, the stem-and-leaf diagram for the rounded weights is

```
5|  223468
6|  5
7|  245688
8|  123678
9|  7
```

(c) Although there are minor differences between the two diagrams, the overall impression of the distribution of weights is the same for both diagrams.

2.91 (a) After rounding to the nearest 10 and then dropping the final zero, the stem-and-leaf diagram with five lines per stem is

```
 9|  1
 9|
 9|  5
 9|  6667
 9|  8888999999
10|  0000011
10|  223333
10|
10|  6
```

(b) After truncating each observation to the 10s digit, the stem-and-leaf diagram with five lines per stem is

```
 9|  1
 9|
 9|  455
 9|  67777
 9|  88888999999
10|  01111
10|  2233
10|
10|  6
```

(c) The overall impression of the shape of the distribution is the same for the diagrams in parts (a) and (b) although there is a slight shift to lower values in part (b). This is due to truncating instead of rounding.

2.92 Minitab used truncation. Note that there was a data point of 5.8 in the sample. It would have been plotted with a stem of 0 and a leaf of 6 if it had been rounded. Instead Minitab plotted the observation with a stem of 0 and a leaf of 5.

Section 2.4

2.93 (a) The distribution of a data set is a table, graph, or formula that provides the values of the observations and how often they occur.

(b) Sample data are the values of a variable for a sample of the population.

(c) Population data are the values of a variable for the entire population.

(d) Census data are the same as population data, a complete listing of all data values for the entire population.

(e) A sample distribution is the distribution of sample data.

(f) A population distribution is the distribution of population data.

(g) A distribution of a variable is the same as a population distribution.

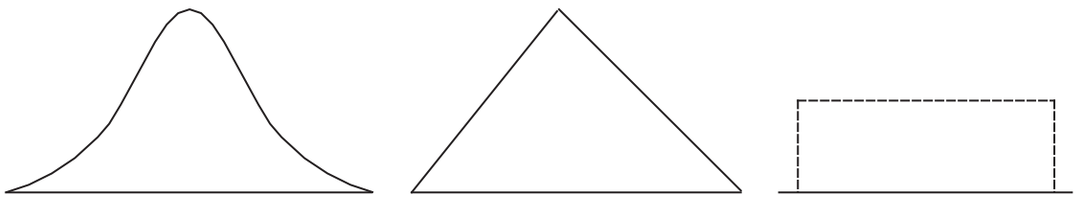
2.94 A smooth curve makes it a little easier to see the shape of a distribution and to concentrate on the overall pattern without being distracted by minor differences in shape.

2.95 A large simple random sample from a bell-shaped distribution would be expected to have roughly a bell-shaped distribution since more sample values should be obtained, on average, from the middle of the distribution.

2.96 (a) Yes. We would expect both simple random samples to have roughly a reverse J-shaped distribution.

(b) Yes. We would expect some variation in shape between the two sample distributions since it is unlikely that the two samples would produce exactly the same frequency table. It should be noted, however, that as the sample size is increased, the difference in shape for the two samples should become less noticeable.

2.97 Three distribution shapes that are symmetric are bell-shaped, triangular, and rectangular, shown in that order below. It should be noted that there are others as well.



Bell-shaped Triangular Uniform (or rectangular)

2.98 (a) The overall shape of the distribution of the number of children of U.S. presidents is right skewed.
(b) The distribution is right skewed.

2.99 (a) Except for the one data value between 74 and 76, this distribution is close to bell-shaped. That one value makes the distribution slightly right skewed.
(b) The distribution is slightly right skewed.

2.100 (a) The distribution is approximately bell-shaped.
(b) The distribution is roughly symmetric.

2.101 (a) The distribution of burrow depths is left skewed.
(b) The distribution is left skewed.

2.102 (a) The distribution of heights is left skewed.
(b) The distribution is left skewed.

2.103 (a) The distribution of shell thickness is approximately bell-shaped.
(b) The distribution is nearly symmetric.

2.104 (a) The distribution of adjusted gross incomes is reverse J-shaped.
(a) The distribution is right skewed.

2.105 (a) The distribution of cholesterol levels appears to be slightly left skewed.
(b) This distribution is nearly symmetric, but is slightly left skewed. Given that the data originated from patients who had high cholesterol levels, one would not expect symmetry. Individuals with low cholesterol levels were not patients and were not included in the testing.

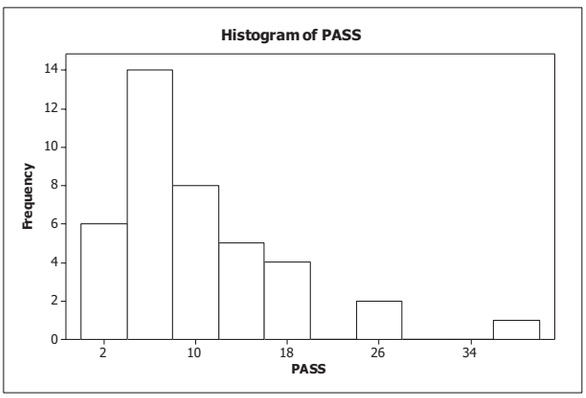
2.106 (a) The distribution of hemoglobin levels for patients with sickle cell disease is approximately uniform.
(b) This distribution is approximately symmetric.

2.107 (a) The distribution of length of stay is nearly reverse J-shaped. It is certainly right skewed.
(b) The distribution is right skewed.

2.108 (a) The frequency distribution for this data is shown in the following table.

Passengers (Millions) Midpoint	Frequency
2	6
6	14
10	8
145	5
18	4
22	0
26	2
30	0
34	0
38	1

(b) The histogram for the distribution is shown below.



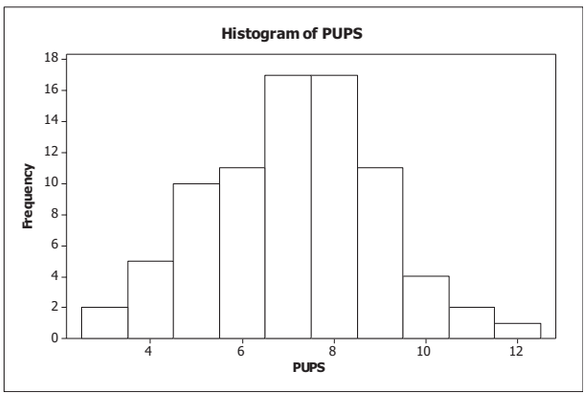
(c) This distribution is very much right skewed.

2.109 (a) The distribution for Year 1 is right skewed and the distribution for Year 2 is reverse J shaped.

(b) Both distributions are right skewed.

(c) Both distributions are rights skewed, however the distribution for Year 1 has a longer right tail indicating more variation than the distribution for Year 2.

2.110 (a) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Histogram**, select **Simple** and click **OK**. Double click on PUPS to enter PUPS in the **Graph variables** box and click **OK**. The result is

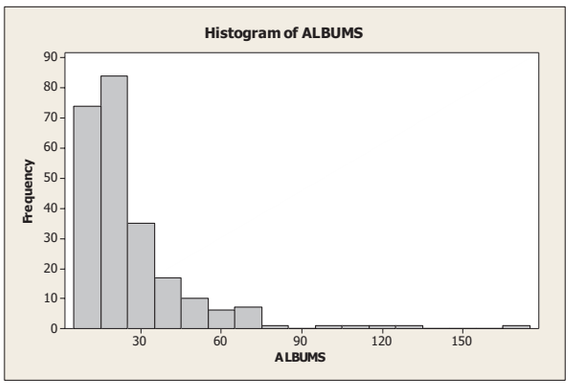


(b) The overall shape of the distribution is bell-shaped.

(c) The distribution is roughly symmetric.

2.111 (a) After entering the data from the WeissStats CD, in Minitab, select

Graph ▶ Histogram, select **Simple** and click **OK**. Double click on ALBUMS to enter ALBUMS in the **Graph variables** box and click **OK**. The result is



(b) The distribution of ALBUMS is definitely right skewed and comes very close to being reverse J-shaped. If the first class had a higher frequency than the second, we would call it reverse J shaped.

(c) The distribution is right skewed.

2.112 (a) In Exercise 2.78, we used Minitab to obtain a stem-and-leaf diagram using 5 lines per stem. That diagram is shown below

```
Stem-and-leaf of PERCENT  N  = 51
Leaf Unit = 1.0

 1    7  9
 6    8  00111
 13   8  2223333
 17   8  4445
 24   8  6677777
(10)  8  8888889999
 17   9  00000001111111
 3    9  223
```

(b) The overall shape of this distribution is left skewed.

(c) We classify the distribution of PERCENT as left skewed.

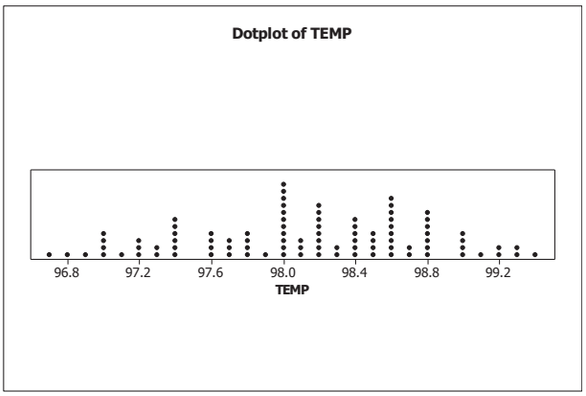
2.113 (a) In Exercise 2.79, we used Minitab to obtain a stem-and-leaf diagram using 2 lines per stem. That diagram is shown below

```
Stem-and-leaf of RATE  N  = 51
Leaf Unit = 1.0

 1   1  9
 4   2  014
 14  2  5678888999
 20  3  001444
(7)  3  5667788
 24  4  01222334
 16  4  5566677899
 6   5  00124
 1   5
 1   6  2
```

(b) The distribution of crime rates is slightly right skewed. Without the largest observation of 62, it would be approximately bell-shaped.

(c) We classify the distribution as right skewed.

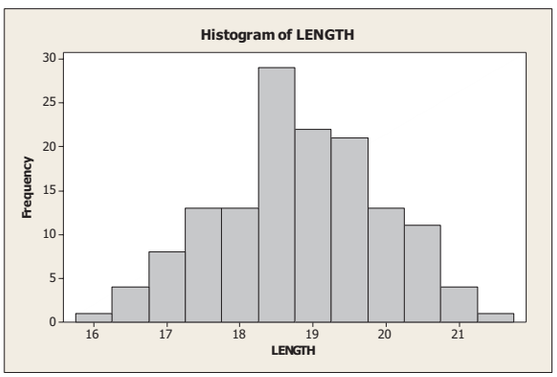
2.114 (a) After entering the data in Minitab, select **Graph ▶ Dotplot**, select **Simple** in the **One Y** row, and click **OK**. Double click on TEMP to enter TEMP in the **Graph variables** box and click **OK**. The result is



(b) The overall distribution of temperatures is roughly triangular or bell shaped.

(c) The distribution is fairly symmetric.

2.115 (a) After entering the data from the WeissStats CD, in Minitab, select **Graph ▶ Histogram**, select **Simple** and click **OK**. Double click on LENGTH to enter LENGTH in the **Graph variables** box and click **OK**. The result is



(b) The distribution of LENGTH is approximately bell shaped.

(c) The distribution is symmetric.

2.116 Class Project. The precise answers to this exercise will vary from class to class.

2.117 The precise answers to this exercise will vary from class to class or individual to individual. Thus your results will likely differ from our results shown below.

(a) We obtained 50 random digits from a table of random numbers. The digits were

4 5 4 6 8 9 9 7 7 2 2 2 9 3 0 3 4 0 0 8 8 4 4 5 3

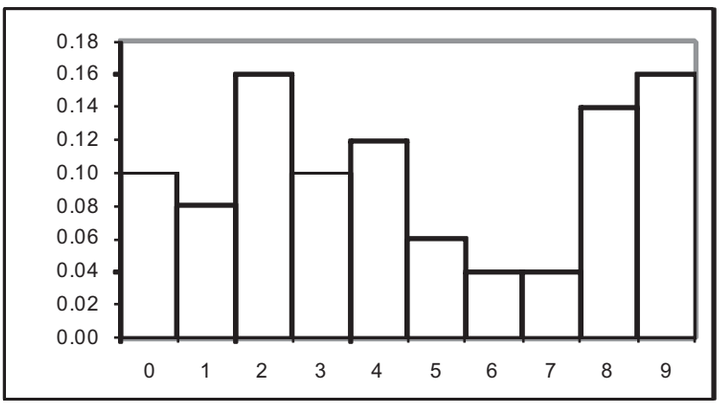
9 2 4 8 9 6 3 0 1 1 0 9 2 8 1 3 9 2 5 8 1 8 9 2 2

(b) Since each digit is equally likely in the random number table, we expect that the distribution would look roughly rectangular.

(c) Using single value classes, the frequency distribution is given by the

following table. The histogram is shown below.

Value	Frequency	Relative-Frequency
0	5	.10
1	4	.08
2	8	.16
3	5	.10
4	6	.12
5	3	.06
6	2	.04
7	2	.04
8	7	.14
9	8	.16



We did not expect to see this much variation.

(d) We would have expected a histogram that was a little more 'even', more like a rectangular distribution, but the relatively small sample size can result in considerable variation from what is expected.

(e) We should be able to get a more evenly distributed set of data if we choose a larger set of data.

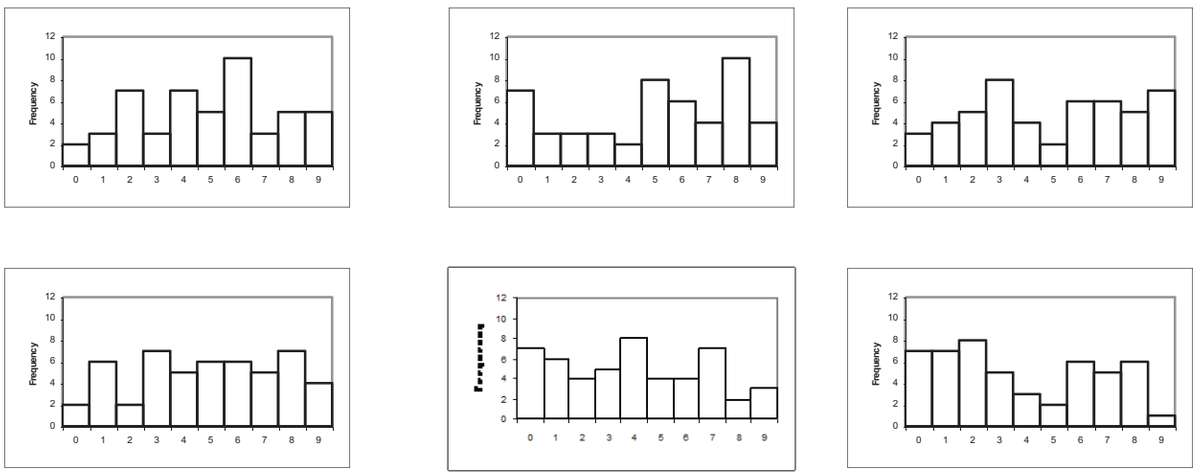
(f) Class project.

2.118 (a-c) Your results will differ from the ones below which were obtained using Excel. Enter a name for the data in cell A1, say RANDNO. Click on cell A2 and enter **=RANDBETWEEN(0,9).** Then copy this cell into cells A3 to A51. There are two ways to produce a histogram of the resulting data in Excel. The easier way is to highlight A!-A51 with the mouse, click on **DDXL** on the toolbar, select **Graphs and Plots**, then choose Histogram in the **Function type** box. Now click on RANDNO in the **Names and Columns** box and drag the name into the **Quantitative Variables** box. Then click OK. A graph and a summary table will be produced. To get five more samples, simply go back to the spreadsheet and press the F9 key. This will generate an entire new sample in Column A and you can repeat the procedure using DDXL. The only disadvantage of this method is that the graphs produced use white lines on a black background.

The second method is a bit more cumbersome and does not provide a summary chart, but yields graphs that are better for reproduction and that can be edited. Generate the data in the same way as was done

above. In cells B1 to B10 enter the integers 0 to 9. These cells are called the **BIN**. Now click on **Tools**, **Data Analysis**, **Histogram**. (If Data Analysis is not in the Tools menu, you will have to add it from the original CD.) Click on the **Input** box and highlight cells A2-A51 with the mouse, then click on the **Bin** box and highlight cells B1-B10. Finally click on the **Output** box and enter **C1**. This will give you a frequency table in columns C and D. Now enter the integers 0 to 9 as text in cells E2 to E11 by entering each digit preceded by a single quote mark, i.e., **'0**, **'1**, etc. In cell F2, enter **=D2**, and copy this cell into F3 through F11. Now highlight the data in columns E and F with the mouse and click the chart icon, pick the **Column** graph type, pick the first sub-type, click on the **Next** button twice, enter any titles desired, remove the legend, and then click on the **Next** button and then the **Finish** button. The graph will appear on the spreadsheet as a bar chart with spaces between the bars. Use the mouse to point to any one of the bars and click with the right mouse button. Choose **Format Data Series**. Click on the **Options** tab and change the **Gap Width** to zero, and click **OK**. Repeat this sequence to produce additional histograms, but use different cells.

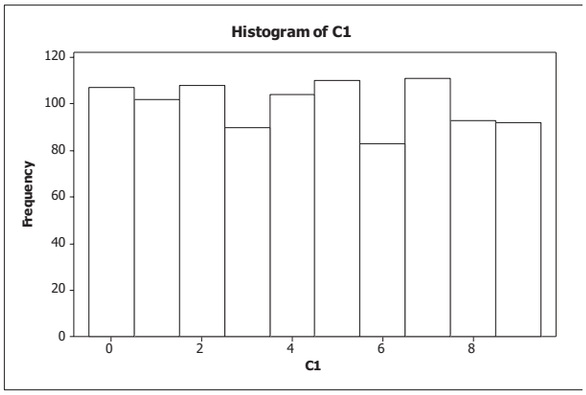
[If you would like to avoid repeating most of the above steps, click near the border of the graph and copy the graph to the Clipboard, then go to Microsoft Word or other word processor, and click on **Edit** on the Toolbar and **Paste Special**. Highlight **Microsoft Excel Chart Object**, and click **OK**. The graphs can be resized in the word processor if necessary. Now go back to Excel and hit the **F9** key. This will produce a completely new set of random numbers. Click on **Tools**, **Data Analysis**, **Histogram**, leave all the boxes as they are and click **OK**. Then click **OK** to overwrite existing data. A new table will be created and the existing histogram will be updated automatically. We used this process for the following histograms.]

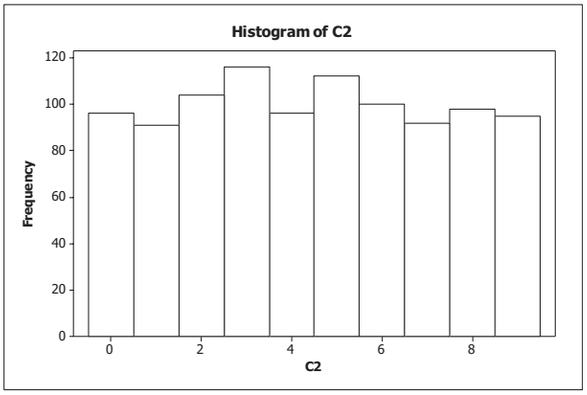


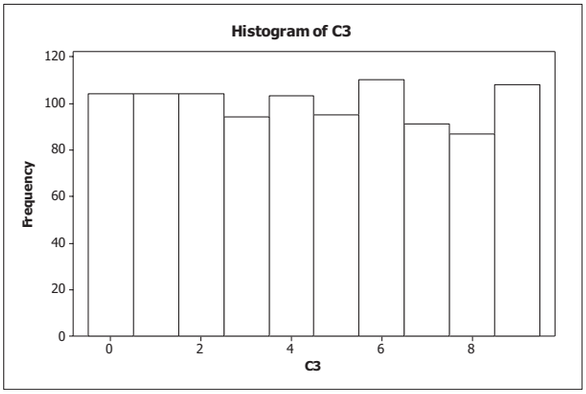
(d) These shapes are about what we expected.

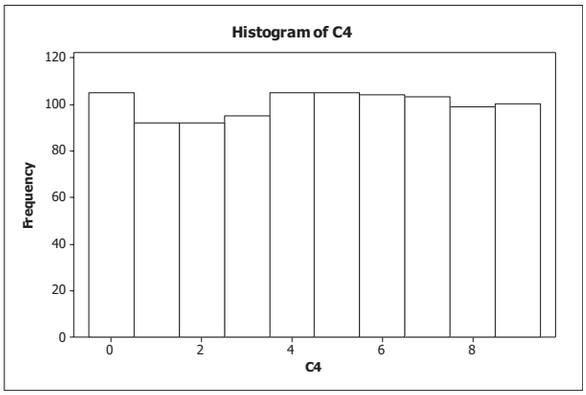
(e) The relative frequency histograms for six samples of digits of size 1000 were obtained using Minitab. Choose **Calc** ▶ **Random Data** ▶ **Integer...**, type 1000 in the **Generate rows of data** test box, click in the **Store in column(s)** text box and type C1 C2 C3 C4 C5 C6, click in the **Minimum value** text box and type 0, type in the **Maximum value** text box and type 9 and click **OK**. Then choose **Graph** ▶ **Histogram**, select the **Simple** version and click **OK**, enter C1 C2 C3 C4 C5 C6 in the **Graph**

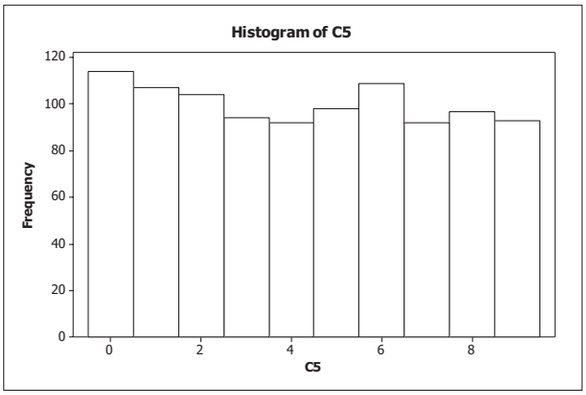
variables text box, C2 in the **Graph 2** text box un **x**, and so on for C3 through C6. Click on the **Multiple Graphs** button. Click on the **On separate graphs** button, and check the boxes for **Same Y** and **Same X**, **including same bins**. Click **OK** and click **OK**. The following graphs resulted.

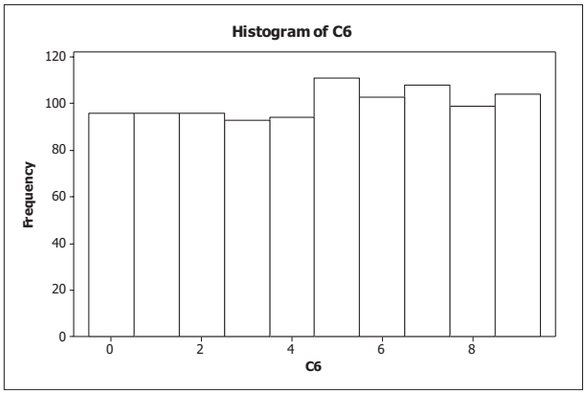










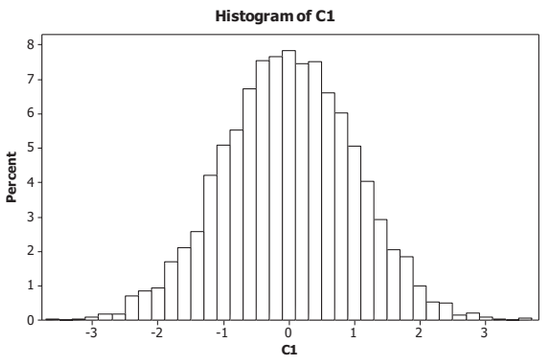


The histograms for samples of size 1000 are much closer to the rectangular distribution we expected than are the ones for samples of size 50.

2.119 (a) Your result will differ from, but be similar to, the one below which was obtained using Minitab. Choose **Calc ▶ Random Data ▶ Normal...**,

type 3000 in the **Generate rows of data** text box, click in the **Store in column(s)** text box and type C1, and click **OK**.

(b) Then choose **Graph ▶ Histogram**, choose the **Simple** version, click **OK**, enter C1 in the **Graph variables** text box, click on the **Scale** button and then on the **Y-Scale Type** tab. Check the **Percent** box and click **OK** twice.



(c) The histogram in part (b) has the shape of a bell symmetric distribution. The sample of 3000 is representative of the population from which the sample was taken.

Section 2.5

2.120 Graphs are sometimes constructed in ways that cause them to be misleading.

2.121 (a) A truncated graph is one for which the vertical axis starts at a value other than its natural starting point, usually zero.

(b) A legitimate motivation for truncating the axis of a graph is to place the emphasis on the ups and downs of the distribution rather than on the actual height of the graph.

(c) To truncate a graph and avoid the possibility of misinterpretation, one should start the axis at zero and put slashes in the axis to indicate that part of the axis is missing.

2.122 Answers will vary.

2.123 (a) A large lower portion of the graph is eliminated. When this is done, differences between district and national averages appear greater than in the original figure.

(b) Even more of the graph is eliminated. Differences between district and national averages appear even greater than in part (a).

(c) The truncated graphs give the misleading impression that, in 2008, the district average is much greater relative to the national average than it actually is.

2.124 (a) A break is shown in the first bar on the left to warn the reader that part of the first bar itself has been removed.

(b) It was necessary to construct the graph with a broken bar to let the reader know that the first bar is actually much taller than it appears. If the true height of the first bar were presented, but without the break, the height would span most of an entire page. This would have used up, perhaps, more room than that desired by the person reporting the graph.

(c) This bar chart is potentially misleading if the reader does not pay attention to the *true* magnitude of the first bar relative to the other three bars. This is precisely the reason for the break in the first bar, however. It is meant to alert the reader that special treatment is to be applied to the first bar. It is actually much taller than it appears. Supplying the numbers for each bar of the graph makes it clear that there was no intention to mislead the reader. This was necessary also because there is no scale on the vertical axis.

2.125 (a) The problem with the bar chart is that it is truncated. That is, the vertical axis, which should start at $0 (in trillions), starts with $7.6 (in trillions) instead. The part of the graph from $0 (in trillions) to $7.6 (in trillions) has been cut off. This truncation causes the bars to be out of correct proportion and hence creates the misleading impression that the money supply is changing more than it actually is.

(b) A version of the bar chart with an untruncated and unmodified vertical axis is presented in Figure (a). Notice that the vertical axis starts at $0.00 (in trillions). Increments are in trillion dollars. In contrast to the original bar chart, this one illustrates that the changes in money supply from week to week are not that different. However, the "ups" and "downs" are not as easy to spot as in the original, truncated bar chart.

(c) A version of the bar chart in which the vertical axis is modified in an acceptable manner is presented in Figure (b). Notice that the special symbol "//" is used near the base of the vertical axis to signify that the vertical axis has been modified. Thus, with this version of the bar chart, not only are the "ups" and "downs" easy to spot but the reader is also aptly warned by the slashes that part of the vertical axis between $0.00 (in trillions) and $7.6 (in trillions) has been removed.

Figure (a)

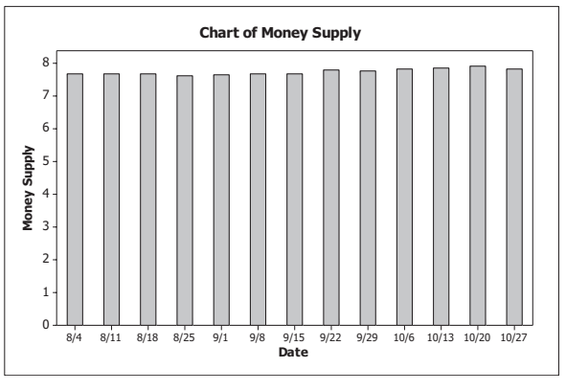
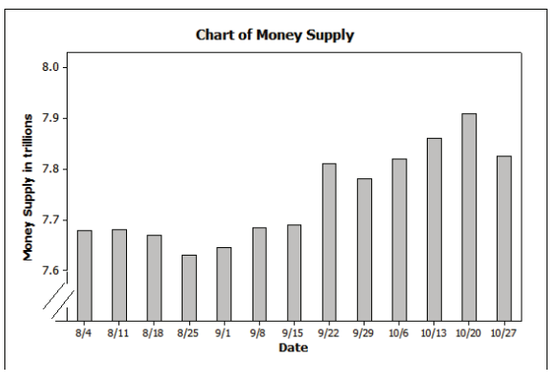


Figure (b)



2.126 (a) 1) The scale on the left begins at 140 million with the graph for the number of licensed drivers beginning at 150.2 million. 2) The scale on the right for the number of Drunk Driving Fatalities begins at 9000 with the graph ending at 13,041. Thus both graphs are truncated, making it look like the number of licensed drivers has increased more dramatically than it really has and making it look like the number of drunk driving fatalities has decreased proportionately more than it really has. 3) The difference between two tic marks on the left-hand scale is 10 million licensed drivers while the difference between two tic marks on the right-hand scale is 2500 fatalities. The graphs don't really cross. If a single scale were used, the fatalities graph would be far below the licensed drivers graph.

(b) All of that being said, to display both sets of data on one graph requires some accommodation due to the relative sizes of the numbers in the two sets of data.

(c) If both graphs are to be presented in one diagram, the axes should both have a broken line between the lowest two tic marks, with the lowest tic mark labeled with a zero. Another option is to produce two separate correctly prepared graphs since the scales are not related anyway.

2.127 (b) Without the vertical scale, it would appear that oil prices dropped about 75% from the first price to the last day shown on the graph.

(c) The actual drop was from about \$82 per barrel to \$58 per barrel, a drop of about \$24 dollars per barrel. This is a drop of 24/82 = 0.29 or about 29%.

(d) The graph is potentially misleading because if the reader doesn't pay attention to the vertical scale, he or she may be led to conclude that the oil price was much more volatile than it actually was.

(e) The graph could be made less potentially misleading by either making the vertical scale range from zero to 85 or by starting at zero and putting a break in the vertical axis to call attention to the fact that part of the vertical axis is missing.

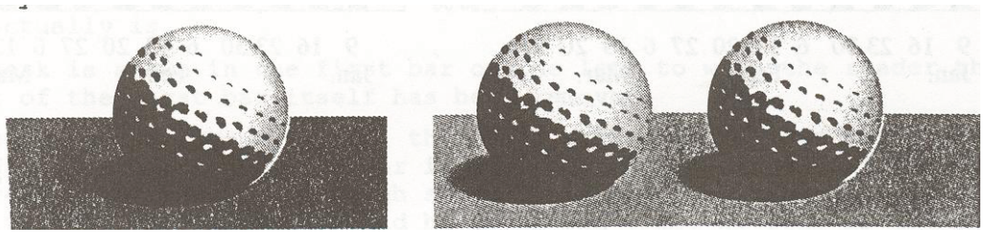
2.128 A correct way in which the developer can illustrate the fact that twice as many homes will be built in the area this year as last year is as follows:

Last Year **This Year**

2.129 (a) The brochure shows a "new" ball with twice the radius of the "old" ball. The intent is to give the impression that the "new" ball lasts roughly twice as long as the "old" ball. However, if the "new" ball has twice the radius of the "old" ball, the "new" ball will have eight times the volume of the "old" ball (since the volume of a sphere is proportional to the cube of its radius, or the radius $2^3 = 8$). Thus, the scaling is improper because it gives the impression that the "new" ball lasts eight times as long as the "old" ball rather than merely two times as long.

Old Ball New Ball

(b) One possible way for the manufacturer to illustrate the fact that the "new" ball lasts twice as long as the "old" ball is to present pictures of two balls, side by side, each of the same magnitude as the picture of the "old" ball and to label this set of two balls "new ball". This will illustrate the point that a purchaser will be getting twice as much for the money.



Old Ball New Ball

Review Problems For Chapter 2

1. (a) A variable is a characteristic that varies from one person or thing to another.

(b) Variables are quantitative or qualitative.

(c) Quantitative variables can be discrete or continuous.

(d) Data are values of a variable.

(e) The data type is determined by the type of variable being observed.

2. A frequency distribution of qualitative data is a table that lists the distinct values of data and their frequencies. It is useful to organize the data and make it easier to understand. A relative-frequency distribution of qualitative data is a table that lists the distinct values of data and a ratio of the class frequency to the total number of observations, which is called the relative frequency.

3. For both quantitative and qualitative data, the frequency and relative-frequency distributions list the values of the distinct classes and their frequencies and relative frequencies. For single value grouping of quantitative data, it is the same as the distinct classes for qualitative data. For class limit and cutpoint grouping in quantitative data, we create groups that form distinct classes similar to qualitative data.

4. The two main types of graphical displays for qualitative data are the bar chart and the pie chart.

5. The bars do not abut in a bar chart because there is not any continuity between the categories. Also, this differentiates them from histograms.

6. Answers will vary. One advantage of pie charts is that it shows the proportion of each class to the total. One advantage of bar charts is that it emphasizes each individual class in relation to each other. One disadvantage of pie charts is that if there are too many classes, the chart becomes confusing. Also, if a class is really small relative to the total, it is hard to see the class in a pie chart.

7. Single value grouping is appropriate when the data are discrete with relatively few distinct observations.

8. (a) The second class would have lower and upper limits of 9 and 14. The class mark of this class would be the average of these limits and equal 11.5.

(b) The third class would have lower and upper limits of 15 and 20.

(c) The fourth class would have lower and upper limits of 21 and 26. This class would contain an observation of 23.

9. (a) If the width of the class is 5, then the class limits will be four whole numbers apart. Also, the average of the two class limits is the class mark of 8. Therefore, the upper and lower class limits must be

two whole numbers above and below the number 8. The lower and upper class limits of the first class are 6 and 10.

(b) The second class will have lower and upper class limits of 11 and 15. Therefore, the class mark will be the average of these two limits and equal 13.

(c) The third class will have lower and upper class limits of 16 and 20.

(d) The fourth class has limits of 21 and 25, the fifth class has limits of 26 and 30. Therefore, the fifth class would contain an observation of 28.

10. (a) The common class width is the distance between consecutive cutpoints, which is 15 - 5 = 10.

(b) The midpoint of the second class is halfway between the cutpoints 15 and 25, and is therefore 20.

(c) The sequence of the lower cutpoints is 5, 15, 25, 35, 45, ... Therefore, the lower and upper cutpoints of the third class are 25 and 35.

(d) Since the third class has lower and upper cutpoints of 25 and 35, an observation of 32.4 would belong to this class.

11. (a) The midpoint is halfway between the cutpoints. Since the class width is 8, 10 is halfway between 6 and 14.

(b) The class width is also the distance between consecutive midpoints. Therefore, the second midpoint is at 10 + 8 = 18.

(c) The sequence of cutpoints is 6, 14, 22, 30, 38, ... Therefore the lower and upper cutpoints of the third class are 22 and 30.

(d) An observation of 22 would go into the third class since that class contains data greater than or equal to 22 and strictly less than 30.

12. (a) If lower class limits are used to label the horizontal axis, the bars are placed between the lower class limit of one class and the lower class limit of the next class.

(b) If lower cutpoints are used to label the horizontal axis, the bars are placed between the lower cutpoint of one class and the lower cutpoint of the next class.

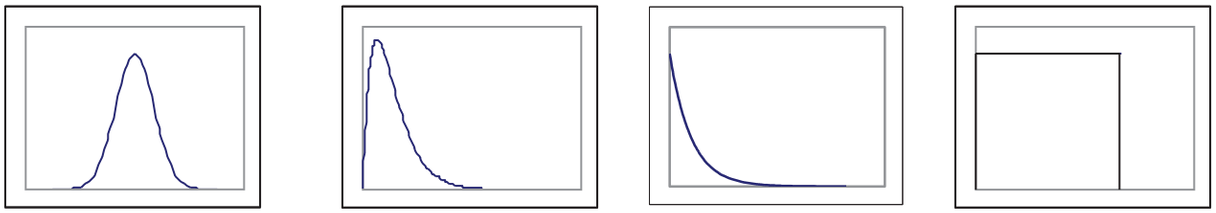
(c) If class marks are used to label the horizontal axis, the bars are placed directly above and centered over the class mark for that class.

(d) If midtpoints are used to label the horizontal axis, the bars are placed directly above and centered over the midpoint for that class.

13. (a) A single value frequency histogram for the prices of DVD players would be identical to the dotplot in example 2.16 because the classes would be 197 through 224 and the height for each bar in the frequency histogram would reflect the number of dots above each observation in the dotplot.

(b) No. The frequency histogram with cutpoint or class limit grouping would combine some of the single values together which would change the frequencies and the heights of the bars corresponding to those classes.

14. Bell-shaped Right skewed Reverse J shape Uniform



15. (a) Slightly skewed to the right. Assuming that the most typical heights are around 5'10", most heights below that figure would still be above 5'4", whereas heights above 5'10" extend to around 7'.

(b) Skewed to the right. High incomes extend much further above the mean income than low incomes extend below the mean.

(c) Skewed to the right. While most full-time college students are in the 17-22 age range, there are very few below 17 while there are many above 22.

(d) Skewed to the right. The main reason for the skewness to the right is that those students with GPAs below fixed cutoff points have been suspended by the time they would have been seniors.

16. (a) The distribution of the large simple random sample will reflect the distribution of the population, so it would be left-skewed as well.

(b) No. The randomness in the samples will almost certainly produce different sets of observations resulting in shapes that are not identical.

(c) Yes. We would expect both of the simple random samples to reflect the shape of the population and be left-skewed.

17. (a) The first column ranks the hydroelectric plants. Thus, it consists of *quantitative, discrete* data.

(b) The fourth column provides measurements of capacity. Thus, it consists of *quantitative, continuous* data.

(c) The third column provides nonnumerical information. Thus, it consists of *qualitative* data.

18. (a) The first class to construct is 40-44. Since all classes are to be of equal width, and the second class begins with 45, we know that the width of all classes is 45 - 40 = 5. All of the classes are presented in column 1 of the grouped-data table in the figure below. The last class to construct does not go beyond 65-69, since the largest single data value is 69.

Age at Inauguration	Frequency	Relative Frequency	Class Mark
40–44	2	0.045	42
45–49	7	0.159	47
50–54	13	0.295	52
55–59	12	0.273	57
60–64	7	0.159	62
65–69	3	0.068	67
	44	1.000	

(b) By averaging the lower and upper limits for each class, we arrive at the class mark for each class. The class marks for all classes are presented in column 4.

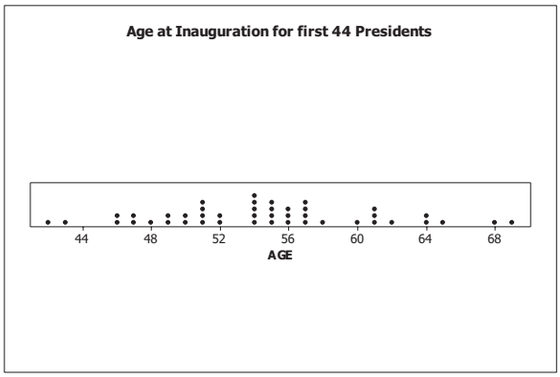
(c) Having established the classes, we tally the ages into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 44, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3.

(d) The frequency histogram presented below is constructed using the frequency distribution presented above; i.e., columns 1 and 2. Notice that the lower cutpoints of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical-axis units in the frequency histogram are the even integers within the range 2 through 14, since these are representative of the magnitude and spread of the frequencies presented in column 2. The height of each bar in the frequency histogram matches the respective frequency in column 2.

(e) The overall shape of the inauguration ages is somewhere between triangular and bell-shaped.

(f) The distribution is roughly symmetric.

19. The horizontal axis of this dotplot displays a range of possible ages for the 44 Presidents of the United States. To complete the dotplot, we go through the data set and record each age by placing a dot over the appropriate value on the horizontal axis.



20. (a) Using *one* line per stem in constructing the ordered stem-and-leaf diagram means vertically listing the numbers comprising the stems *once*. The leaves are then placed with their respective stems in order. The ordered stem-and-leaf diagram using one line per stem is presented in the following figure.

```
4|  236677899
5|  00111122444445555666777778
6|  0111244589
```

(b) Using *two* lines per stem in constructing the ordered stem-and-leaf diagram means vertically listing the numbers comprising the stems *twice*. In turn, if the leaf is one of the digits 0 through 4, it is ordered and placed with the first of the two stem lines. If the leaf is one of the digits 5 through 9, it is ordered and placed with the second of the two stem lines. The ordered stem-and-leaf diagram using two lines per stem is presented in the following figure.

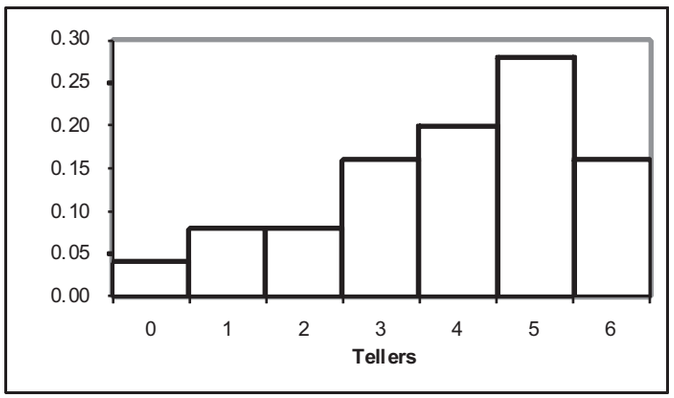
```
4|  23
4|  6677899
5|  0011112244444
5|  555566677778
6|  0111244
6|  589
```

(c) The stem-and-leaf diagram in part (b) corresponds to the frequency distribution of Problem 18.

21. (a) The frequency and relative-frequency distribution presented below is constructed using classes based on a single value. Since each data value is one of the integers 0 through 6, inclusive, the classes will be 0 through 6, inclusive. These are presented in column 1. Having established the classes, we tally the number of busy tellers into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 25, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3.

Number Busy	Frequency	Relative Frequency
0	1	0.04
1	2	0.08
2	2	0.08
3	4	0.16
4	5	0.20
5	7	0.28
6	4	0.16
	25	1.00

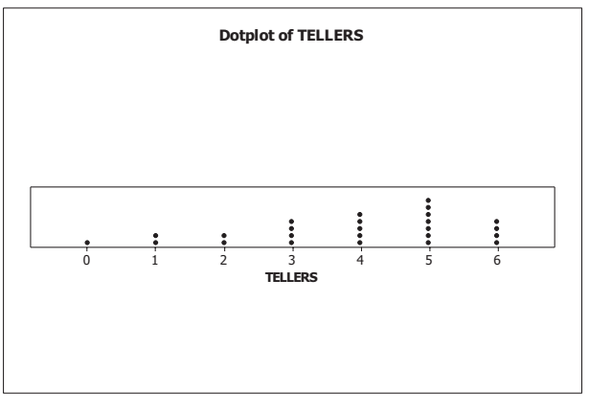
(b) The following relative-frequency histogram is constructed using the relative-frequency distribution presented in part (a); i.e., columns 1 and 3. Column 1 demonstrates that the data are grouped using classes based on a single value. These single values in column 1 are used to label the horizontal axis of the relative-frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.04 to 0.28 (4% to 28%). Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 0.05, starting with zero and ending at 0.30. The middle of each histogram bar is placed directly over the single numerical value represented by the class. Also, the height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.



(c) The overall shape of this distribution is left skewed.

(d) The distribution is left skewed.

(e)



(f) Since both the histogram and the dotplot are based on single value grouping, they both convey exactly the same information.

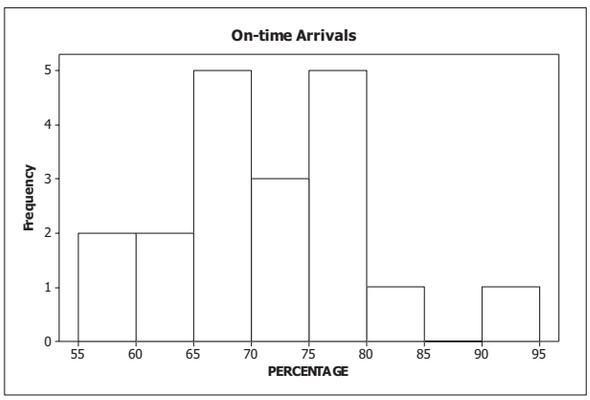
22. (a) The classes will begin with the class 55 – under 60. The second class will be 60 – under 65. The classes will continue like this until the last class, which will be 90 – under 95, since the largest observation is 92.2. The classes can be found in column 1 of the frequency distribution in part (c).

(b) The midpoints of the classes are the averages of the lower and upper cutpoint for each class. For example, the first midpoint will be 57.5, the second midpoint will be 62.5. The midpoints will continue like this until the last midpoint, which will be 92.5. The midpoints can be found in column 4 of the frequency distribution in part (c).

(c) The first and second columns of the following table provide the frequency distribution. The first and third columns of the following table provide the relative-frequency distribution for percentages of on-time arrivals for the airlines.

Percent On-Time	Frequency	Relative Frequency	Midpoint
55 – under 60	2	0.105	57.5
60 – under 65	2	0.105	62.5
65 – under 70	5	0.263	67.5
70 – under 75	3	0.158	72.5
75 – under 80	5	0.263	77.5
80 – under 85	1	0.053	82.5
85 – under 90	0	0.000	87.5
90 – under 95	1	0.053	92.5
	19	1.000	

(d) The frequency histogram is constructed using columns 1 and 2 from the frequency distribution from part (c). The cutpoints are used to label the horizontal axis. The vertical axis is labeled with the frequencies that range from 0 to 5.



(e) After rounding each observation to the nearest whole number, the stem-and-leaf diagram with two lines per stem for the rounded percentages is

```
5|  99
6|  3
6|  567789
7|  34
7|  566788
8|  1
8|
9|  2
```

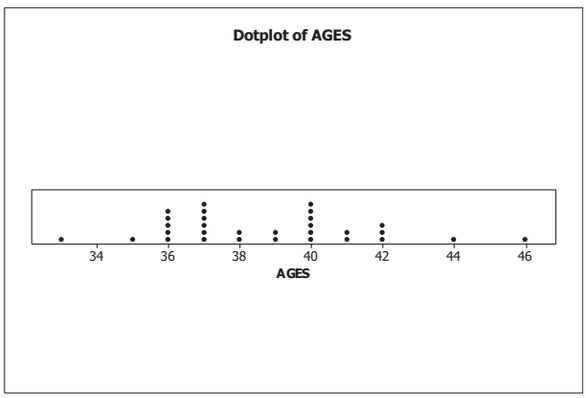
(f) After obtaining the greatest integer in each observation, the stem-and-

leaf diagram with two lines per stem is

```
 5|  89
 6|  34
 6|  57778
 7|  244
 7|  66777
 8|  0
 8|
 9|  2
```

(g) The stem-and-leaf diagram in part (f) corresponds to the frequency distribution in part (d) because the observations weren't rounded in constructing the frequency distribution.

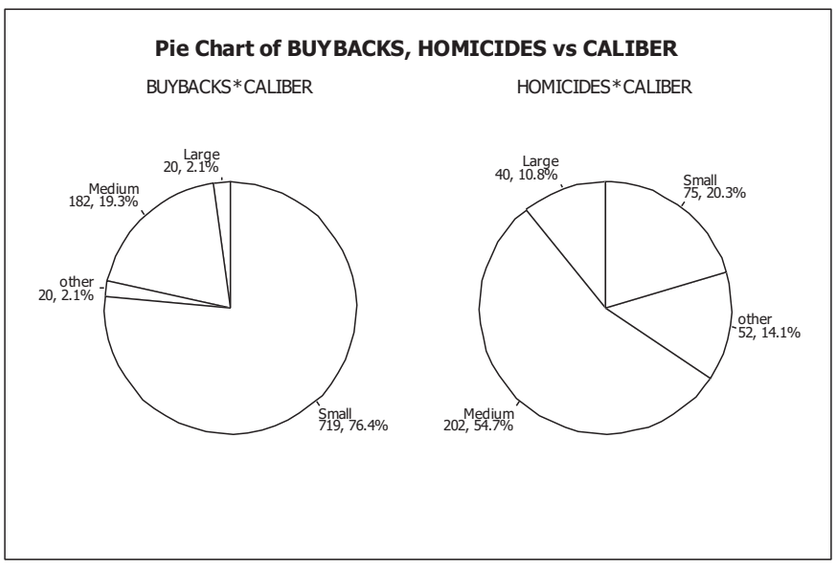
23. (a) The dotplot of the ages of the oldest player on each major league baseball team is



(b) The overall shape of the distribution of ages is bimodal.

(c) The distribution is roughly symmetric.

24. (a)-(b) The two pie-charts are shown below. In each case, the proportion of the circle for each category is found by multiplying 360 degrees by the category frequency and dividing by the total frequency (941 for Buybacks and 369 for Homicides).



(c) The most striking characteristic of the two charts is that most of the

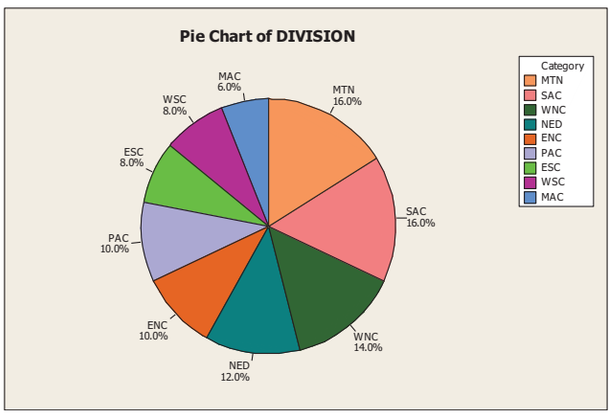
buybacks are of small caliber guns while most of the homicides are committed with medium caliber guns. It should also be noted that small and medium caliber guns are the two largest categories of both buybacks and homicides, accounting for 95.7% of the buybacks and 75.0% of the homicides.

25. (a) The population consists of the states in the United States. The variable is the division.

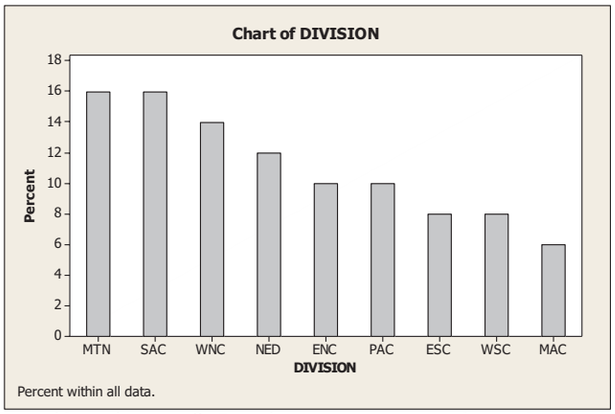
(b) The frequency and relative frequency distribution for Region is shown below.

Region	Frequency	Relative Frequency
East North Central	5	0.10
East South Central	4	0.08
Middle Atlantic	3	0.06
Mountain	8	0.16
New England	6	0.12
Pacific	5	0.10
South Atlantic	8	0.16
West North Central	7	0.14
West South Central	4	0.08
	50	1.000

(c) The pie chart for Region is shown below.



(d) The bar chart for Region is shown below.

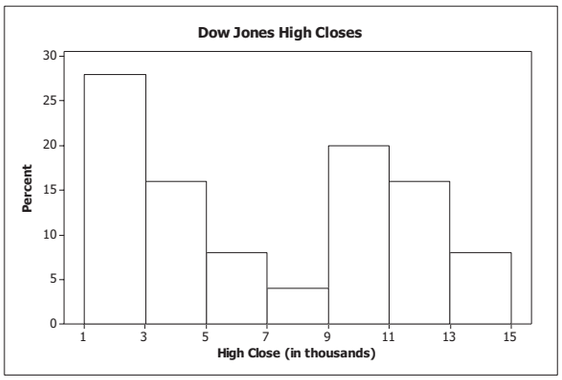


(e) The distribution of Region seems to be almost uniformly distributed between the categories. The smallest region is the Middle Atlantic at a frequency of 3, the largest region is the Mountain and South Atlantic at a frequency of 8.

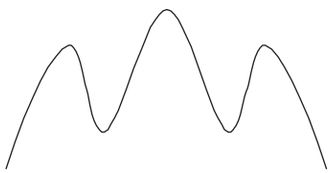
26. (a) The first class to construct is 1 – under 3. Since all classes are to be of equal width, we know that the width of all classes is 3 - 1 = 2. All of the classes are presented in column 1 of Figure (a) below. The last class to construct is 13 – under 15, since the largest single data value is 14164.53. Having established the classes, we tally the highs into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 25, results in each class's relative frequency. The relative frequencies for all classes are presented in column 3.

High close (in thousands)	Freq.	Relative Frequency
1 – under 3	7	0.28
3 – under 5	4	0.16
5 – under 7	2	0.08
7 – under 9	1	0.04
9 – under 11	5	0.20
11 – under 13	4	0.16
13 – under 15	2	0.08
	25	1.000

(b) The following relative-frequency histogram is constructed using the relative-frequency distribution presented above; i.e., columns 1 and 3. The lower cutpoints of column 1 are at the left edges of each rectangle in the relative-frequency histogram. We notice that the relative frequencies presented in column 3 range in size from 0.00 to 0.28. Thus, suitable candidates for vertical axis units in the relative-frequency histogram are increments of 5%, from 0.00 (0%) to 0.30 (30%). The height of each bar in the relative-frequency histogram matches the respective relative frequency in column 3.



27. Answers will vary, but here is one possibility:



28. (a) The break in the third bar is to emphasize that the bar as shown is not as tall as it should be.

(b) The bar for the space available in coal mines is at a height of about 30 billion tonnes. To accurately represent the space available in saline aquifers (10,000 billion tonnes), the third bar would have to be over 300 times as high as the first bar and more than 10 times as high as the second bar. If the first two bars were kept at the sizes shown, there wouldn't be enough room on the page for the third bar to

be shown at its correct height. If a reasonable height were chosen for the third bar, the first bar wouldn't be visible. The only apparent solution is to present the third bar as a broken bar.

29. (a) Covering up the numbers on the vertical axis totally obscures the percentages.

(b) Having followed the directions in part (a), we might conclude that the percentage of women in the labor force for 2000 is about three and one-half times that for 1960.

(c) Not covering up the vertical axis, we find that the percentage of women in the labor force for 2000 is about 1.8 times that for 1960.

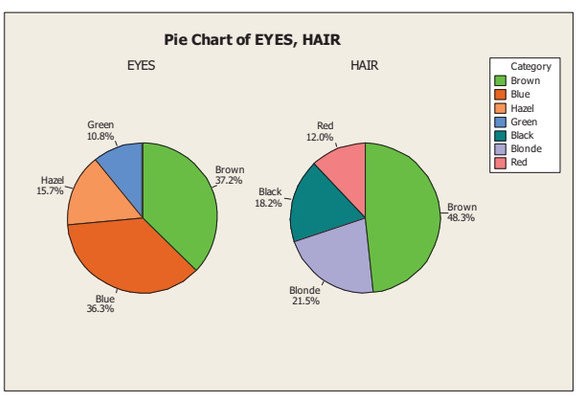
(d) The graph is potentially misleading because it is truncated. Notice that vertical axis units begin at 30 rather than at zero.

(e) To make the graph less potentially misleading, we can start it at zero instead of 30.

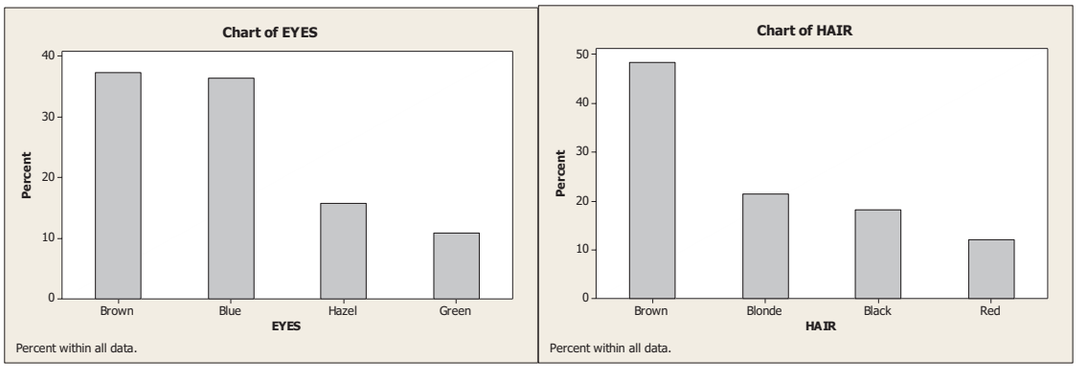
30. (a) Using Minitab, retrieve the data from the Weiss-Stats-CD. Column 2 contains the eye color and column 3 contains the hair color for the students. From the tool bar, select **Stat ▶ Tables ▶ Tally Individual Variables**, double-click on EYES and HAIR in the first box so that both EYES and HAIR appear in the **Variables** box, put a check mark next to Counts and Percents under Display, and click **OK**. The results are

EYES	Count	Percent	HAIR	Count	Percent
Blue	215	36.32	Black	108	18.24
Brown	220	37.16	Blonde	127	21.45
Green	64	10.81	Brown	286	48.31
Hazel	93	15.71	Red	71	11.99
N=	592		N=	592	

(b) Using Minitab, select **Graph ▶ Pie Chart**, check Chart counts of unique values, double-click on EYES and HAIR in the first box so that EYES and HAIR appear in the Categorical Variables box. Click Pie Options, check decreasing volume, click OK. Click Multiple Graphs, check On the Same Graphs, Click OK. Click Labels, click Slice Labels, check Category Name, Percent, and Draw a line from label to slice, Click OK twice. The results are

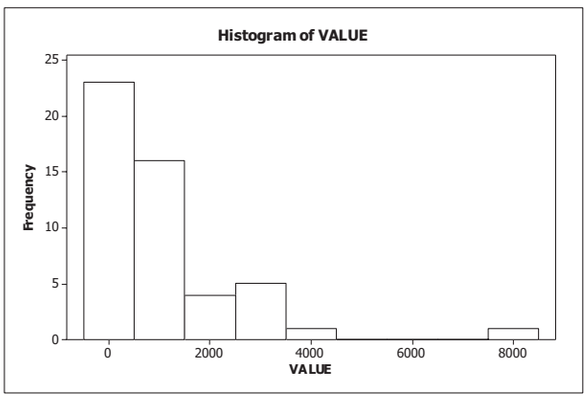


(c) Using Minitab, select **Graph ▶ Bar Chart**, select Counts of unique values, select Simple option, click OK. Double-click on EYES and HAIR in the first box so that EYES and HAIR appear in the Categorical Variables box. Select Chart Options, check decreasing Y, check show Y as a percent, click OK. Click OK twice. The results are

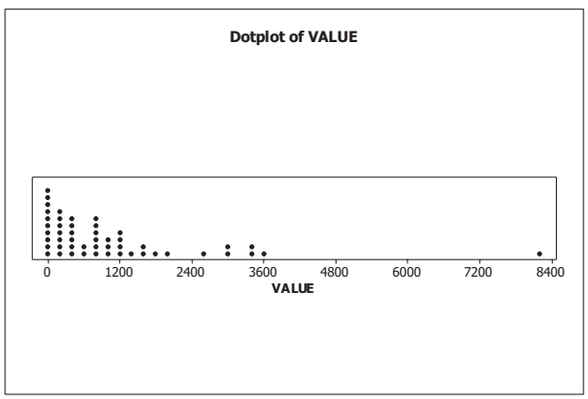


31. (a) The population consists of the states of the U.S. and the variable under consideration is the value of the exports of each state.

(b) Using Minitab, we enter the data from the WeissStats CD, choose **Graph** ▶ **Histogram**, click on **Simple** and click **OK**. Then double click on VALUE to enter it in the **Graph variables** box and click **OK**. The result is



(c) For the dotplot, we choose **Graph** ▶ **Dotplot**, click on **Simple** from the **One Y** row and click **OK**. Then double click on VALUE to enter it in the **Graph variables** box and click **OK**. The result is



(d) For the stem-and-leaf plot, we choose **Graph ▶ Stem-and-Leaf**, double click on VALUE to enter it in the **Graph variables** box and click **OK.** The result is

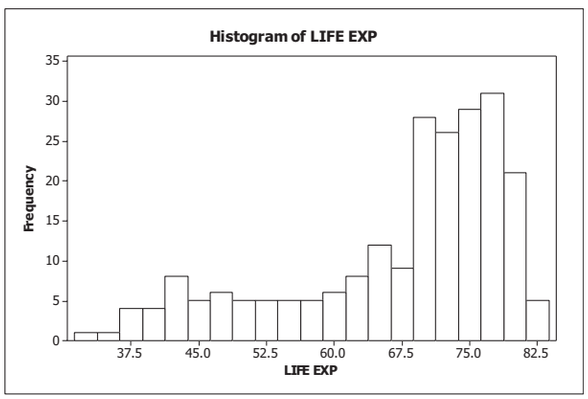
```
Stem-and-leaf of VALUE  N  = 50
Leaf Unit = 100

 23   0  00000000001112222344444
(10)  0  5677888899
 17   1  012224
 11   1  5679
 7    2
 7    2  69
 5    3  034
 2    3  6
 1    4
 1    4
 1    5
 1    5
 1    6
 1    6
 1    7
 1    7
 1    8  2
```

(e) The overall shape of the distribution is reverse J shaped.

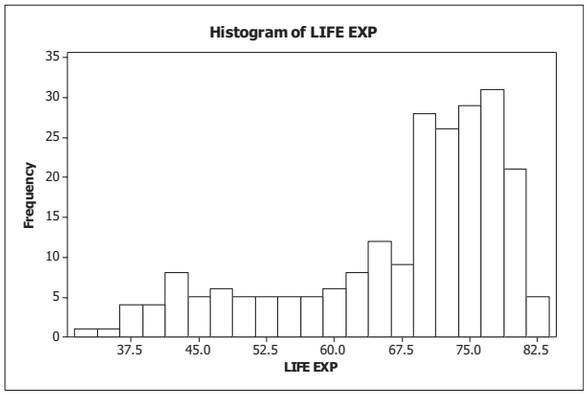
(f) The distribution is right skewed.

32. (a) The population consists of countries of the world, and the variable under consideration is the expected life in years for people in those countries.

(b) Using Minitab, we enter the data from the WeissStats CD, choose **Graph ▶ Histogram**, click on **Simple** and click **OK.** Then double click on LIFE EXP to enter it in the **Graph variables** box and click **OK.** The result is



(c) For the dotplot, we choose **Graph ▶ Dotplot**, click on **Simple** from the **One Y** row and click **OK.** Then double click on LIFE EXP to enter it in the **Graph variables** box and click **OK.** The result is



(d) For the stem-and-leaf plot, we choose **Graph ▶ Stem-and-Leaf**, double click on LIFE EXP to enter it in the **Graph variables** box and click **OK**. The result is

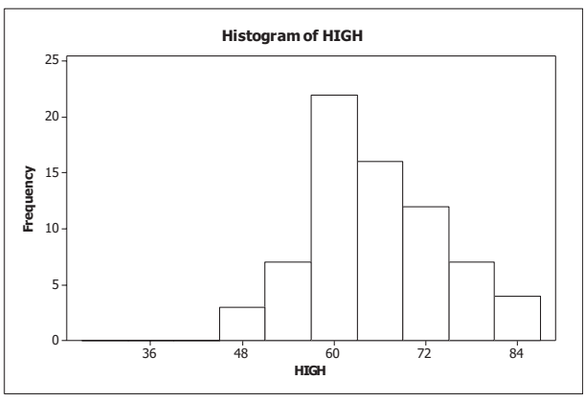
```
Stem-and-leaf of LIFE EXP  N  = 224
Leaf Unit = 1.0

 1    3  2
 9    3  56788999
 21   4  112222233444
 31   4  5567788889
 42   5  01111333444
 50   5  56667779
 70   6  00111112233333444444
 99   6  55556666777788888899999999999
(52)  7  00000000011111111111112222222223333333333444444444
 73   7  55555555555556666666666777777777778888888888888888889999999999
 11   8  00000011113
```

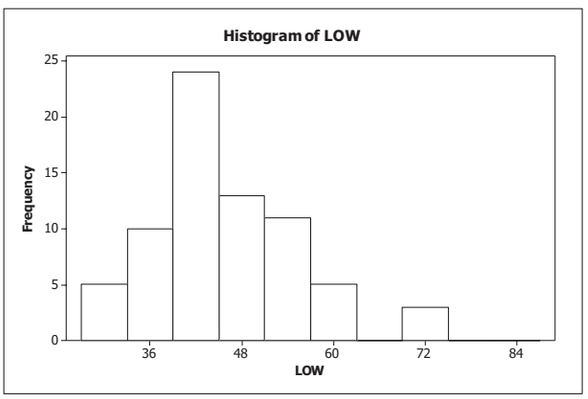
(e) The overall shape of the distribution is left skewed.

(f) This distribution is classified as left skewed.

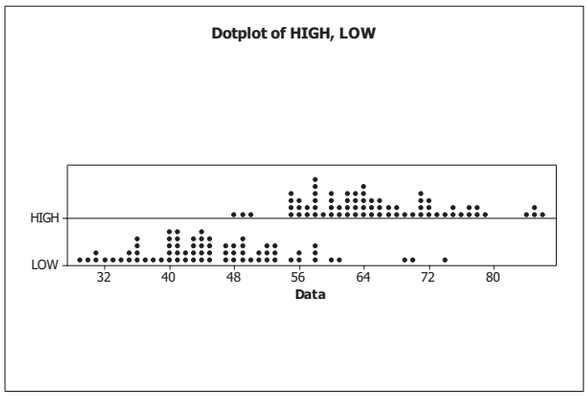
33. (a) The population consists of cities in the U.S., and the variables under consideration are their annual average maximum and minimum temperatures.

(b) Using Minitab, we enter the data from the WeissStats CD, choose **Graph ▶ Histogram**, click on **Simple** and click **OK**. Double click on HIGH to enter it in the **Graph variables** box, and double click on LOW to enter it in the **Graph variables** box. Now click on the **Multiple graphs** button and click to **Show Graph Variables on separate graphs** and also check both boxes under **Same Scales for Graphs**, and click **OK** twice. The result is





(c) For the dotplot, we choose **Graph ▶ Dotplot**, click on **Simple** from the **Multiple Y's** row and click **OK.** Then double click on HIGH and then LOW to enter them in the **Graph variables** box and click **OK.** The result is



(d) For the stem-and-leaf diagram, we choose **Graph ▶ Stem-and-Leaf**, double click on HIGH and then on LOW to enter then in the **Graph variables** box and click **OK.** The result is

```
Stem-and-leaf of HIGH  N  = 71
Leaf Unit = 1.0

 3    4  789
 6    5  444
 21   5  555677777888999
(17)  6  00001122222333444
 33   6  55556677779
 22   7  00001122234
 11   7  5577889
 4    8  444
 1    8  5

Stem-and-leaf of LOW  N  = 71
Leaf Unit = 1.0


 1    2  9
 7    3  001234
 19   3  555556779999
(20)  4  00011111223333344444
 32   4  5567777888899
 19   5  11122223
 11   5  5667889
 4    6  1
 3    6  9
 2    7  04
```

(e) Both variables have distributions that are slightly right skewed.
(f) Both distributions are close to symmetric, but are slightly right skewed. It would take only a few changes in the data to make the distributions approximately symmetric.

Using the FOCUS Database: Chapter 2

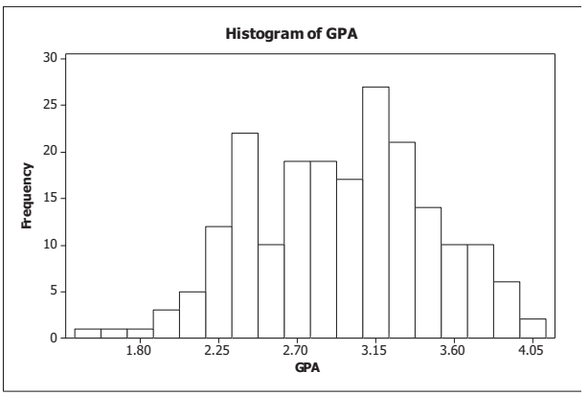
We use the *Menu commands* in Minitab to complete parts (a)-(e). The data sets in the Focus database and their names have already been stored in the file FOCUS.MTW on the WeissStats CD supplied with the text, and, assuming that that

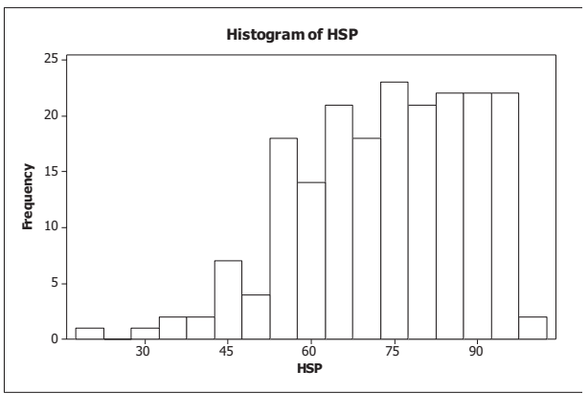
disk is in Drive D, all information can be recovered if you

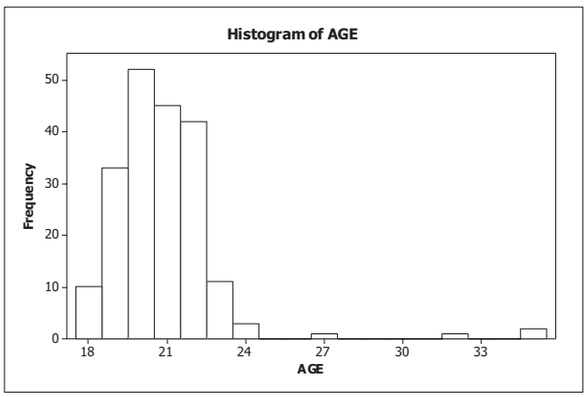
- Choose **File ▶ Open Worksheet...** and **Look in: d:\FocusSample** or **d:\Focus** (depending on which part of this exercise you are working on) and Click **OK**

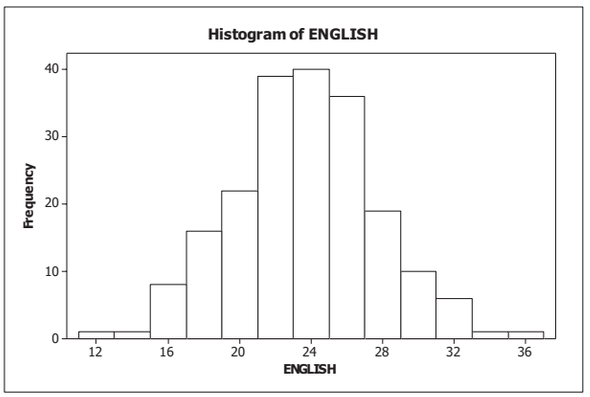
(a) UWEC is a school that attracts good students. HSP reflects pre-college experience and will tend to be left-skewed since fewer students with lower high school percentile scores will have been admitted, but exceptions are made for older students whose high school experience is no longer relevant. GPA will probably show left skewness tendencies since many, but not all, students with lower cumulative GPAs (below 2.0 on a 4-point scale) will likely have been suspended and will not appear in the database, but there are also upper limits on these scores, so the scores will tend to bunch up nearer to the high end than to the low end. AGE will be right skewed because there are few students below the typical 17-22 ages, but many above that range. ENGLISH, MATH, and COMP will be closer to bell-shaped. The ACT typically is taken only by high school students intending to go to college, and the scores are designed to roughly follow a bell-shaped curve. Individual colleges may, however, have a different profile that reflects their admission policies.

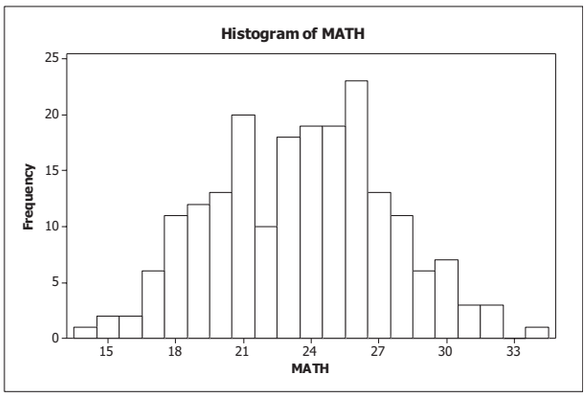
(b) Using Minitab with FocusSample, choose **Graph ▶ Histogram...**, select the **Simple** version, and Click **OK**. Then specify HSP GPA AGE ENGLISH MATH COMP in the **Graph variables** text box and click on the button for **Multiple Graphs**. Click on the button for **On separate graphs** and click **OK** twice. The results are

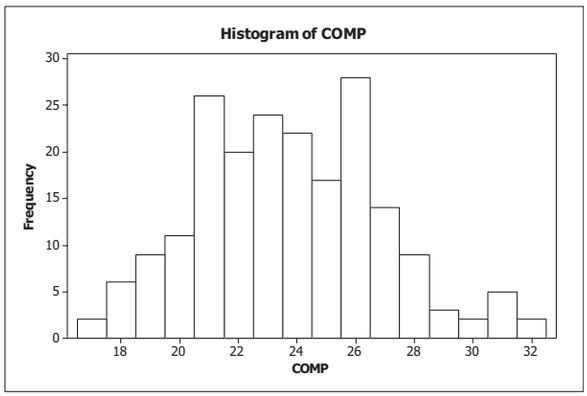






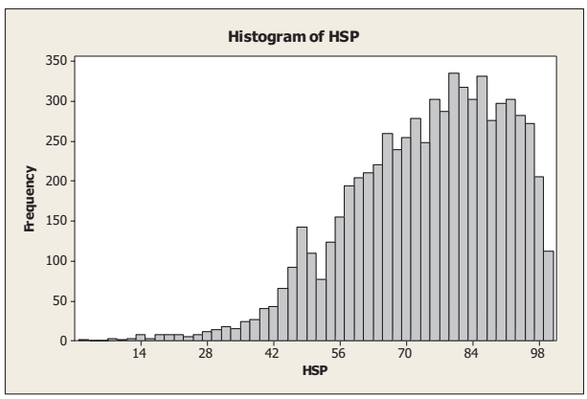


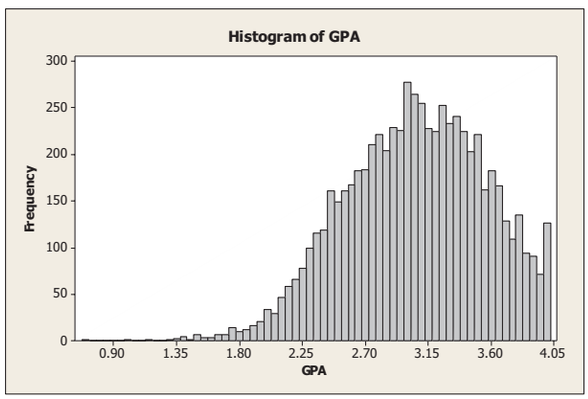


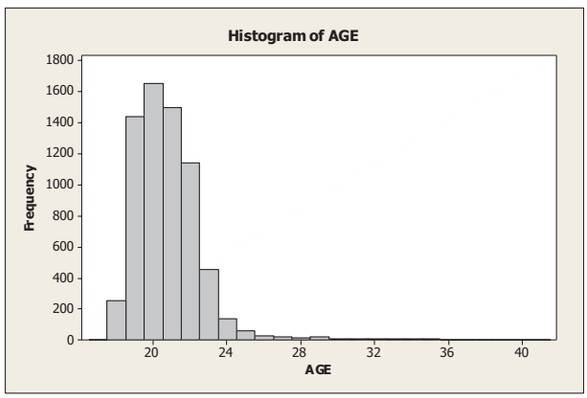


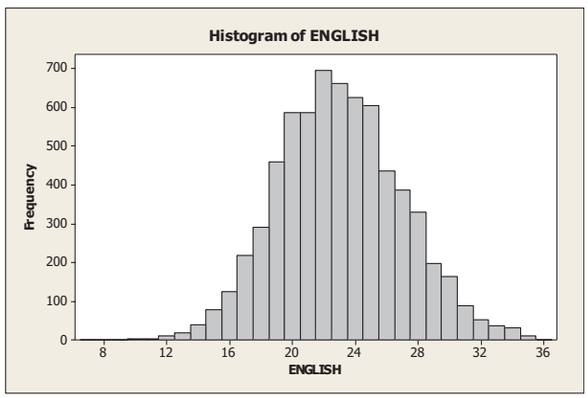
The graphs compare quite well with the educated guesses for all six variables.

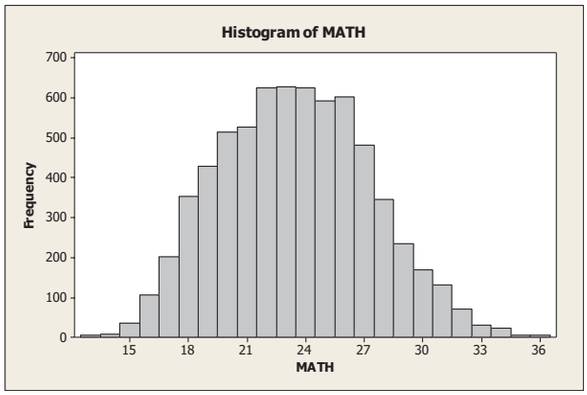
(c) Using Minitab with Focus, choose **Graph ▶ Histogram...**, select the **Simple** version, and Click **OK**. Then specify HSP GPA AGE ENGLISH MATH COMP in the **Graph variables** text box and click on the button for **Multiple Graphs**. Click on the button for **On separate graphs** and click **OK** twice. The results are

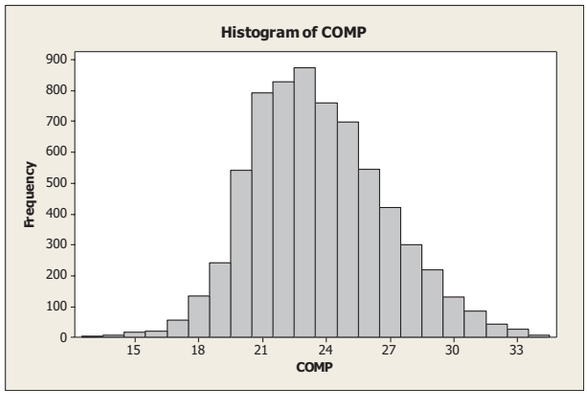








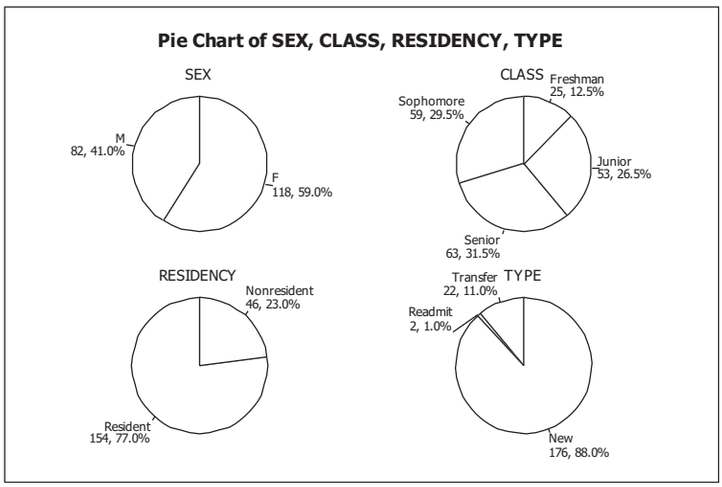




We were correct on the first five variables: HSP and GPA are left skewed, AGE is right skewed, ENGLISH and MATH are fairly symmetric. COMP is close to symmetric, but is slightly right skewed. Comparing the graphs for the sample with those for the entire population, we see similarities between each pair of graphs, but the outline of the histogram for the entire population is much smoother than that of the histogram for the sample.

(d) Using Minitab and the FocusSample file, choose **Graph ▶ Piechart**, click on the **Chart raw data** button, specify SEX CLASS RESIDENCY TYPE in the **Graph variables** text box and click on the **Labels** button. Now click on the tab for **Slice Labels**, check all four boxes and click **OK**, click on the button for **Multiple Graphs** and ensure that the button for **On the same graph** is checked, and click **OK** twice.

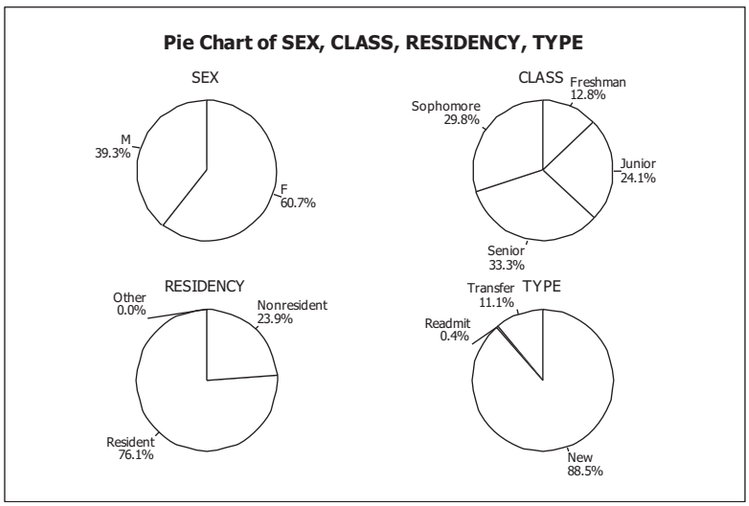
Once the graphs are displayed, we right clicked on the legend that was shown and selected **Delete** since we already had provided for each slice of the graphs to be labeled.



From the graph of SEX, we see that about 59% of the students are females. From the graph of CLASS, we see that the student sample is about 12.5% Freshmen, 29.5% Sophomores, 26.5% Juniors, and 31.5% Seniors. From the graph of RESIDENCY, we see that about 77% of the students are Wisconsin residents and 23$ are nonresidents. From the graph of TYPE, we see that 88.0% of the students were admitted initially as new students, 11.0% were admitted initially as transfer students, and 1.0% are readmits, that is, students who were initially new or transfer students, left the university,

and were later readmitted.

(e) Now repeat part (d) using the entire Focus file. The results are



From the graph of SEX, we see that about 61% of the students are females. From the graph of CLASS, we see that the student population is about 13% Freshmen, 30% Sophomores, 24% Juniors, and 33% Seniors. From the graph of RESIDENCY, we see that about 76% of the students are Wisconsin residents and 24$ are nonresidents. From the graph of TYPE, we see that 85.5% of the students were admitted initially as new students, 11.1% were admitted initially as transfer students, and 0.4% are readmits, that is, students who were initially new or transfer students, left the university, and were later readmitted. We would expect that the two sets of graphs would be approximately the same, but not identical since the sample contains only 200 students out of a population of 6738. This is, in fact, the case. The percentages in each sample graph are very close to the percentages in the corresponding population graph.

Case Study: 25 Highest Paid Women

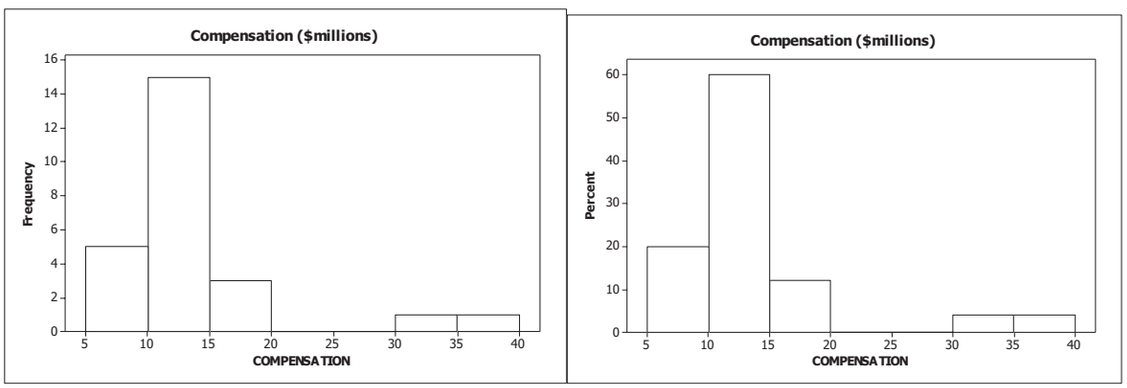
(a) The first column variable is ranks and is qualitative (If discussed, the variable can also be considered ordindal). The second column variable is Name and is qualitative. The third column variable is Company and is qualitative. The fourth column variable is Compensation in millions and is quantitative discrete.

(b) The first class to construct is "5-under 10." Since all classes are to be of equal width, and the second class begins with 10, we know that the width of all classes is 10-5 = 5. All of these classes are presented in column 1. The last class to construct is "35-under 40," since the largest data value is 38.6. Having established the classes, we tally the compensation data into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 25, results in the relative frequencies for each class which are presented in column 3. By averaging the lower and upper cutpoints for each class, we arrive at the class midpoint for each class. The midpoints for all classes are presented in column 4.

Compensation ($million)	Frequency	Relative Frequency	Midpoints
5 – under 10	5	0.20	7.5
10 – under 15	15	0.60	12.5
15 – under 20	3	0.12	17.5
20 – under 25	0	0.00	22.5
25 – under 30	0	0.00	27.5
30 – under 35	1	0.04	32.5
35 – under 40	1	0.04	37.5
	25	1.00	

(c) The frequency and relative-frequency histograms for compensation are constructed using the frequency and relative-frequency distribution presented in part (b); i.e. The lower class cutpoints of column 1 are used to label the horizontal axis of the histograms. Suitable candidates for vertical axis units in the frequency histogram are the integers 0 through 15, since these are representative of the magnitude and spread of the frequencies presented in column 2. The height of each bar matches the respective frequency in column 2. Candidates for the vertical axis units in the relative-frequency histogram are the values 0.00 to 0.60 in increments of 0.10 (10%). Figure (a) is the frequency histogram and Figure (b) is the relative-frequency histogram.

Figure (a) Figure (b)



(d) The shape of the distribution in part (c) is right skewed. Most of the women made between 5 and 20 million. But there were two women who made over 30 million.

(e) After truncating each observation, the stem-and-leaf diagram of compensation using two lines per stem is

```
0|  88999
1|  000011111122444
1|  566
2|
2|
3|  4
3|  8
```

(f) After rounding each observation, the stem-and-leaf diagram of compensation using two lines per stem is

```
0|  9999
1|  0000111222223
1|  555666
2|
2|
3|  4
3|  9
```

(g) The stem-and-leaf diagram in part (e) corresponds to the frequency histogram in part (c) because the observations were not rounded when forming the frequency histogram.

(h) After rounding each observation to the nearest whole number, the dotplot for compensation is