Chapter 2—Graphs, Charts, and Tables—Describing Your Data

When applicable, the first few problems in each section will be done following the appropriate step by step procedures outlined in the corresponding sections of the chapter. Following problems will provide key points and the answers to the questions, but all answers can be arrived at using the appropriate steps.

The more difficult problems in this chapter are:

For Section 2.1
2.14, 2.15, 2.16, 2.21, 2.22, 2.23

For Section 2.2
2.39, 2.40, 2.43, 2.45

For Section 2.3
2.54, 2.59, 2.61, 2.62

For End of Chapter
2.74, 2.76, 2.77, 2.78

Section 2.1

2.1   Step 1: List the possible values.
         The possible values for the discrete variable are 0 through 12.
         Step 2: Count the number of occurrences at each value.
         The resulting frequency distribution is shown as follows:

x	Frequency
0	1
1	0
2	2
3	4
4	1
5	2
6	5
7	6
8	1
9	1
10	1
11	0
12	1
Total =	25

2.2 Given n = 2,000, the minimum number of groups for a grouped data frequency distribution determined using the $2^k \geq n$ guideline is:

$$2^k \geq n \quad \text{or} \quad 2^{11} = 2,048 \geq 2,000 \quad \text{Thus, use k} = 11 \text{ groups.}$$

2.3

a. Given n = 1,000, the minimum number of classes for a grouped data frequency distribution determined using the $2^k \geq n$ guideline is:

$$2^k \geq n \quad \text{or} \quad 2^{10} = 1,024 \geq 1,000 \quad \text{Thus, use k} = 10 \text{ classes.}$$

b. Assuming that the number of classes that will be used is 10, the class width is determined as follows:

$$w = \frac{High - Low}{Classes} = \frac{2,900 - 300}{10} = \frac{2,600}{10} = 260$$

Then we round to the nearest 100 points giving a class width of 300.

2.4   Recall that the Ogive is produced by plotting the cumulative relative frequency against the upper limit of each class. Thus, the first class upper limit is 100 and has a relative frequency of 0.2 - 0.0 = 0.2. The second class upper limit is 200 and has a relative frequency of 0.4 - 0.2 = 0.2. Of course, the frequencies are obtained by multiplying the relative frequency by the sample size. As an example, the first class has a frequency of (0.2)50 = 10. The others follow similarly to produce the following distribution

Class	Frequency	Relative Frequency	Cumulative Relative Frequency
0 – < 100	10	0.20	0.20
100 – < 200	10	0.20	0.40
200 – < 300	5	0.10	0.50
300 – < 400	5	0.10	0.60
400 – < 500	20	0.40	1.00
500 – < 600	0	0.00	1.00

2.5

a. There are n = 60 observations in the data set. Using the $2^k \geq n$ guideline, the number of classes, k, would be 6. The maximum and minimum values in the data set are 17 and 0, respectively. The class width is computed to be: w = (17-0)/6 = 2.833, which is rounded to 3. The frequency distribution is

| Class	Frequency
0-2	6
3-5	13
6-8	20
9-11	14
12-14	5
15-17	2
Total =	60

b. To construct the relative frequency distribution divide the number of occurrences (frequency) in each class by the total number of occurrences. The relative frequency distribution is shown below.

Class	Frequency	Relative Frequency
0-2	6	0.100
3-5	13	0.217
6-8	20	0.333
9-11	14	0.233
12-14	5	0.083
15-17	2	0.033
Total =	60	

c. To develop the cumulative frequency distribution, compute a running sum for each class by adding the frequency for that class to the frequencies for all classes above it. The cumulative relative frequencies are computed by dividing the cumulative frequency for each class by the total number of observations. The cumulative frequency and the cumulative relative frequency distributions are shown below.

Class	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
0-2	6	0.100	6	0.100
3-5	13	0.217	19	0.317
6-8	20	0.333	39	0.650
9-11	14	0.233	53	0.883
12-14	5	0.083	58	0.967
15-17	2	0.033	60	1.000
Total =	60			

d. To develop the histogram, first construct a frequency distribution (see part a). The classes form the horizontal axis and the frequency forms the vertical axis. Bars corresponding to the frequency of each class are developed. The histogram based on the frequency distribution from part (a) is shown below.
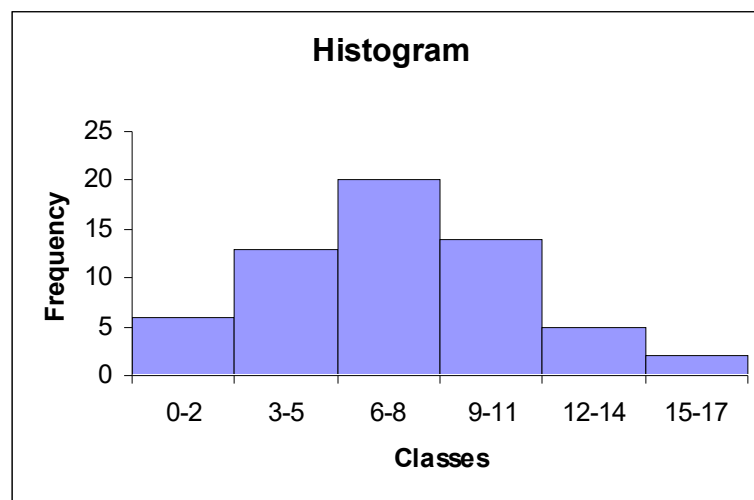
2.6 Note that the first two classes have a total width of 8.05 – 7.85. Thus, the class width equals (8.05 – 7.85)/ 2 = 0.10. The upper class boundary equals the lower class boundary plus the class width. So, as an example, the first class boundary equals 7.85 + 0.10 = 7.95, and so on. The first class has the same relative frequency and cumulative relative frequency. The frequency is obtained by multiplying the sample size times the relative frequency. So the first class frequency equals 50(0.12) = 6. The other class frequencies follow similarly. If we know the frequency we can determine the sample size, and similarly, if we know the sample size we can determine the frequency.

Each cumulative relative frequency is produced by adding the respective class relative frequency to the preceding cumulative relative frequency. So the second class relative frequency is obtained as 0.48 – 0.12 = 0.36. The other calculations follow similarly yielding

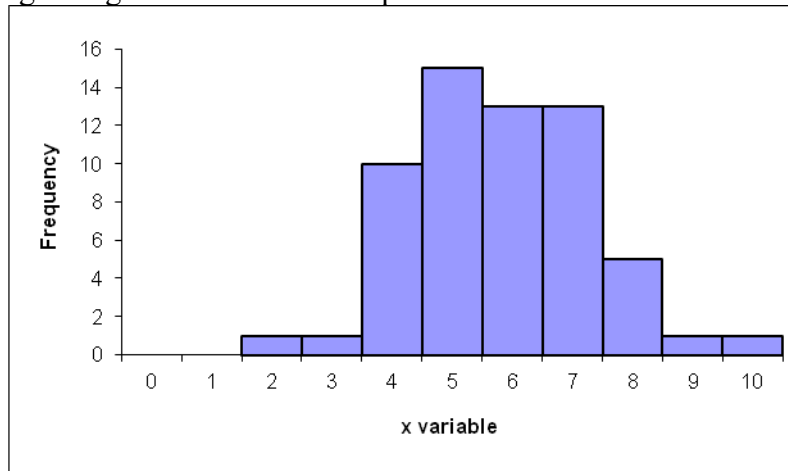| Class	Frequency	Relative Frequency	Cumulative Relative Frequency
7.85 – < 7.95	6	0.12	0.12
7.95 – < 8.05	18	0.36	0.48
8.05 – < 8.15	4	0.08	0.56
8.15 – < 8.25	5	0.10	0.66
8.25 – < 8.35	17	0.34	1.00

2.7

   a. Proportion of days in which no shortages occurred = 1 – proportion of days in which shortages occurred = 1 – 0.24 = 0.76
   b. Less than \$20 off implies that overage was less than \$20 and the shortage was less than \$20 = (proportion of overages less \$20) – (proportion of shortages at most \$20) = 0.56 – 0.08 = 0.48
   c. Proportion of days with less than \$40 over or at most \$20 short = Proportion of days with less than \$40 over – proportion of days with more than \$20 short = 0.96 – 0.08 = 0.86.

2.8

   a. The data do not require grouping. The following frequency distribution is given:

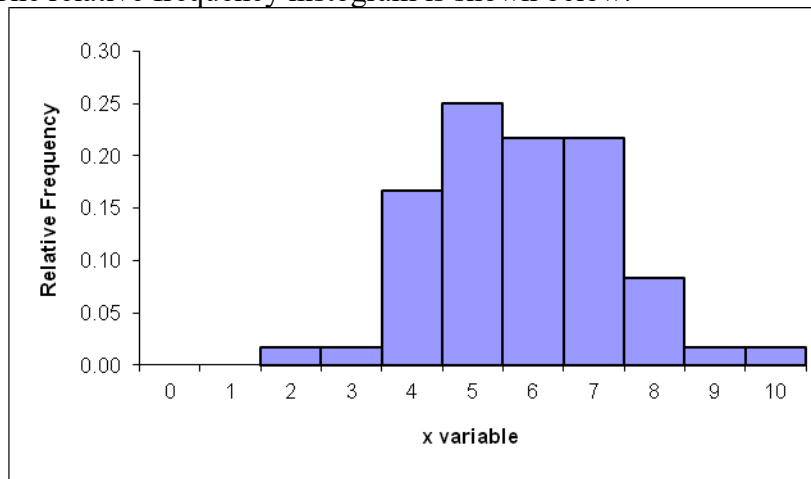| x	Frequency
0	0
1	0
2	1
3	1
4	10
5	15
6	13
7	13
8	5
9	1
10	1

b. The following histogram could be developed.

c. The relative frequency distribution shows the fraction of values falling at each value of x.

x	Frequency	Relative Frequency
0	0	0.00
1	0	0.00
2	1	0.02
3	1	0.02
4	10	0.17
5	15	0.25
6	13	0.22
7	13	0.22
8	5	0.08
9	1	0.02
10	1	0.02
	60	

d. The relative frequency histogram is shown below.

e. The two histograms look exactly alike since the same data are being graphed. The bars represent either the frequency or relative frequency.

2.9
a. Step 1 and Step 2. Group the data into classes and determine the class width:
The problem asks you to group the data. Using the $2^k \geq n$ guideline we get:

$$2^k \geq 60 \text{ so } 2^6 \geq 60$$

Class width is:

$$W = \frac{Maximum - Minumum}{\# Classes} = \frac{10 - 2}{6} = 1.33$$

which we round up to 2.0

Step 3. Define the class boundaries:
Since the data are discrete, the classes are:

Class
2-3
4-5
6-7
8-9
10-11

Step 4. Count the number of values in each class:

Class	Frequency	Relative Frequency
2-3	2	0.0333
4-5	25	0.4167
6-7	26	0.4333
8-9	6	0.1000
10-11	1	0.0167

b. The cumulative frequency distribution is:

Class	Frequency	Cumulative Frequency
2-3	2	2
4-5	25	27
6-7	26	53
8-9	6	59
10-11	1	60

c.

Class	Frequency	Relative Frequency	Cumu. Rel. Freq.
2-3	2	0.0333	0.0333
4-5	25	0.4167	0.4500
6-7	26	0.4333	0.8833
8-9	6	0.1000	0.9833
10-11	1	0.0167	1.000

The relative frequency histogram is:
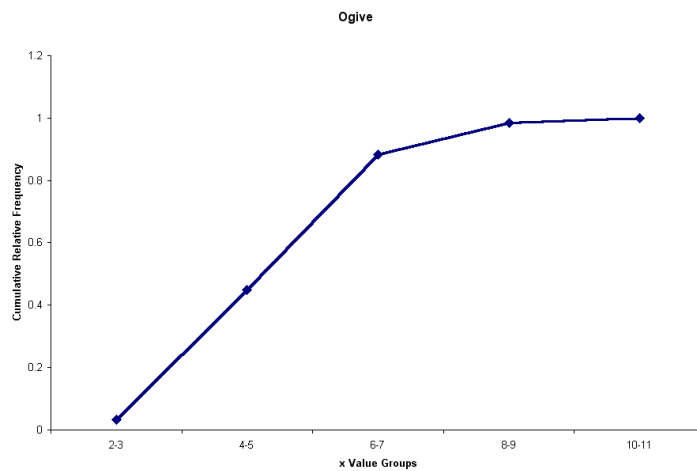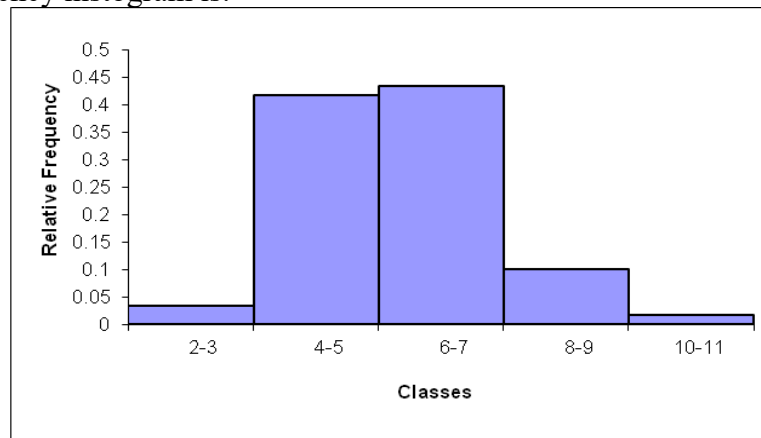
Ogive

d. The ogive is a graph of the cumulative relative frequency distribution.

2.10.
   a. Because the number of possible values for the variable is relatively small, there is no need to group the data into classes. The resulting frequency distribution is:

x	Frequency
0	0
1	2
2	3
3	3
4	10
5	16
6	19
7	7
8	9
9	6
10	2
11	2
12	1
Total =	80

This frequency distribution shows the manager that most customer receipts have 4 to 8 line items.

b. A histogram is a graph of a frequency distribution for a quantitative variable. The resulting histogram is shown as follows.

Line Items on Sales Receipts
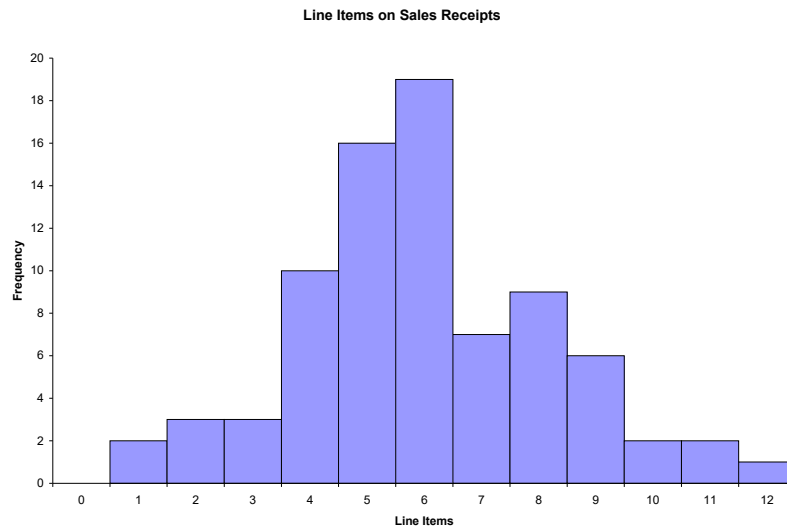
2.11

a.

| | Knowledge Level | | | |
	Savvy	Experienced	Novice	Total
Online Investors	32	220	148	400
Traditional Investors	8	58	134	200
	40	278	282	600

b.

| | Knowledge Level | | |
	Savvy	Experienced	Novice
Online Investors	0.0533	0.3667	0.2467
Traditional Investors	0.0133	0.0967	0.2233

c. The proportion that were both on-line and experienced is 0.3667.
d.   The proportion of on-line investors is 0.6667

2.12

a. The following relative frequency distributions are developed for the two variables:

Rating	Frequency	Rel. Frequency
1	5	0.25
2	8	0.40
3	4	0.20
4	2	0.10
5	1	0.05
Total =	20	

Time Slot	Frequency	Rel. Frequency
1	9	0.45
2	3	0.15
3	5	0.25
4	3	0.15
Total =	20	

b. The joint frequency distribution is a two dimensional table showing responses to the rating on one dimension and time slot on the other dimension. This joint frequency distribution is shown as follows:

Rating ▼	Morning	Afternoon	Evening	Various	Total
Very Good	5				5
Good	4	3		1	8
Fair			3	1	4
Poor			1	1	2
Very Poor			1		1
Total	9	3	5	3	20

c. The joint relative frequency distribution is determined by dividing each frequency by the sample size, 20. This is shown as follows:

Rating ▼	Morning	Afternoon	Evening	Various	Total
Very Good	0.25	0.00	0.00	0.00	0.25
Good	0.20	0.15	0.00	0.05	0.40
Fair	0.00	0.00	0.15	0.05	0.20
Poor	0.00	0.00	0.05	0.05	0.10
Very Poor	0.00	0.00	0.05	0.00	0.05
Total	0.45	0.15	0.25	0.15	1.00

Based on the joint relative frequency distribution, we see that those who advertise in the morning tend to provide higher service ratings. Evening advertisers tend to provide lower ratings. The manager may wish to examine the situation further to see why this occurs.

2.13

a. The weights are sorted from smallest to largest to create the data array.

77	79	80	83	84	85	86
86	86	86	86	86	87	87
87	88	88	88	88	89	89
89	89	89	90	90	91	91
92	92	92	92	93	93	93
94	94	94	94	94	95	95
95	96	97	98	98	99	101

b. Five classes having equal widths are created by subtracting the smallest observed value (77) from the largest value (101) and dividing the difference by 5 to get the width for each class (4.8 rounded to 5). Five classes of width five are then constructed such that the classes are mutually exclusive and all inclusive. Identify the variable of interest. The weight of each crate is the variable of interest. The number of crates in each class is then counted. The frequency table is shown below.

Weight (Classes)	Frequency
77-81	3
82-86	9
87-91	16
92-96	16
97-101	5
Total =	49

c. The histogram can be created from the frequency distribution. The classes are shown on the horizontal axis and the frequency on the vertical axis. The histogram is shown below.
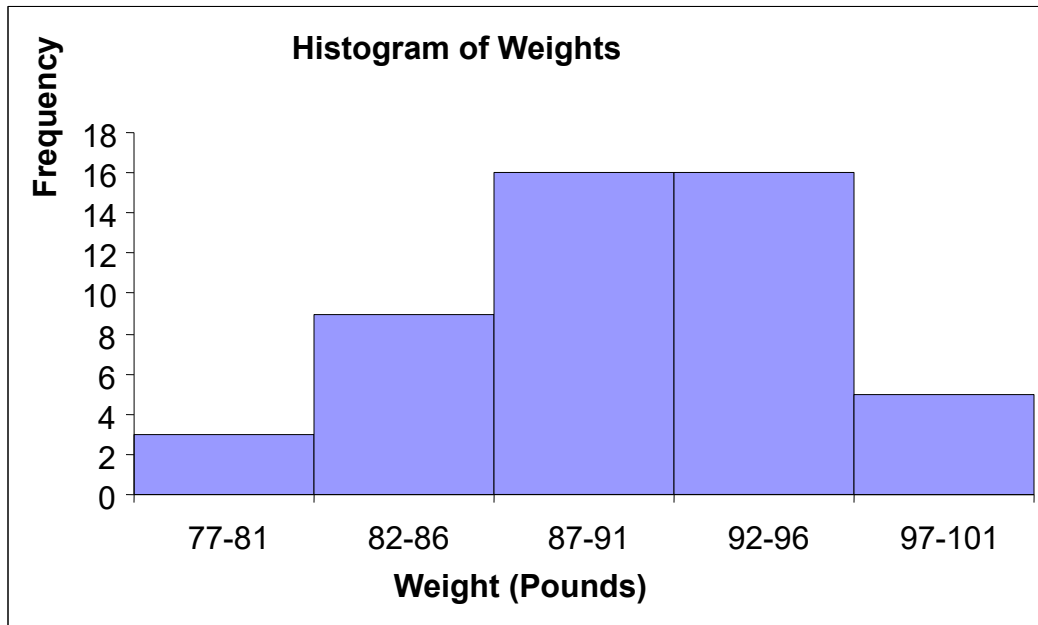
d. **Convert the frequency distribution into relative frequencies and cumulative relative frequencies as shown below.**

Weights (Classes)	Frequency	Relative Frequency	Cumulative Relative Frequency
77-81	3	0.0612	0.0612
82-86	9	0.1837	0.2449
87-91	16	0.3265	0.5714
92-96	16	0.3265	0.8980
97-101	5	0.1020	1.0000
Total =	49		

The percentage of sampled crates with weights greater than 96 pounds is 10.20%.

2.14

a. There are n = 100 values in the data. Then using the $2^k \geq n$ guideline we would need at least k = 7 classes.

b. Using k = 7 classes, the class width is determined as follows:

$$w = \frac{High - Low}{Classes} = \frac{\$376,644 - \$87,429}{7} = \frac{\$289,215}{7} = \$41,316.43$$

Rounding this up to the nearest $1,000, the class width is $42,000.

c. The frequency distribution with seven classes and a class width of $42,000 will depend on the starting point for the first class. This starting value must be at or below the minimum value of $87,429. Student answers will vary depending on the starting point. We have used $85,000. Care should be made to make sure that the classes are mutually exclusive and all-inclusive. The following frequency distribution is developed:

Group	Frequency
$85,000 - $126,999	6
$127,000 - $168,999	12
$169,000 - $210,999	31
$211,000 - $252,999	26
$253,000 - $294,999	14
$295,000 - $336,999	9
$337,000 - $378,999	2
Total =	100

d. The histogram for the frequency distribution in part c. is shown as follows:
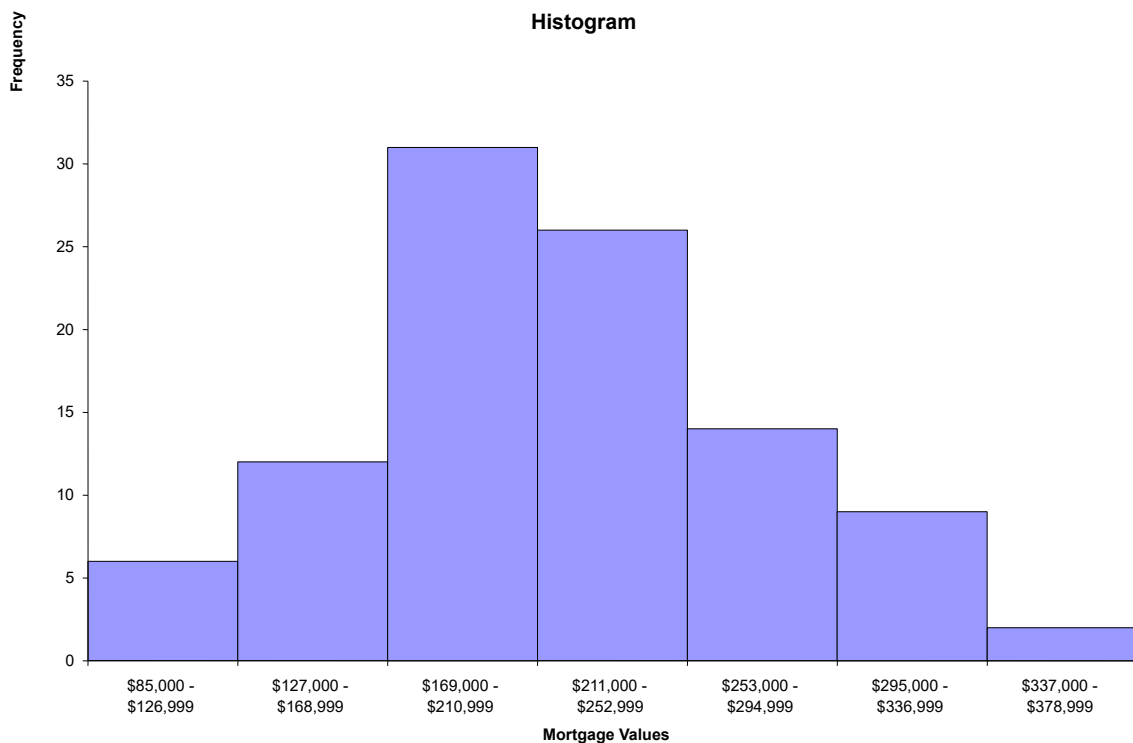
Interpretation should involve a discussion of the range of values with a discussion of where the major classes are located.

2.15

a. $w = \dfrac{L\arg est - smallest}{number\ of\ classes} = \dfrac{214.4 - 105.0}{11} = 9.945 \rightarrow w = 10$. The salaries in the first class are

(105, 105 + 10) = (105, 115). The frequency distribution follows

Classes	Frequency	Relative Frequency	Cumulative Relative Frequency
(105 – <115)	1	0.04	0.04
(115 – <125)	1	0.04	0.08
(125 – <135)	2	0.08	0.16
(135 – <145)	1	0.04	0.20
(145 – <155)	1	0.04	0.24
(155 – <165)	7	0.28	0.52
(165 – <175)	4	0.16	0.68
(175 – <185)	3	0.12	0.80
(185 – <195)	2	0.08	0.88
(195 – <205)	0	0.00	0.88
(205 – <215)	3	0.12	1.00

b. The data shows 8 of the 25, or 0.32 of the salaries are at least 175,000

c. The data shows 18 of the 25, or 0.72 having salaries that are at most $205,000 and a least $135,000.

2.16

a. We are assuming mortgage rates are limited to two decimal places. Students making other assumptions will get a slightly difference histogram. We are also rounding the calculated class width to .15.

Class	Frequency	Relative Frequency
3.46-36.0	3	0.067
3.61-3.75	6	0.133
3.76-3.90	16	0.356
3.91-4.05	14	0.331
4.06-4.20	6	0.133

b. Proportion of rates that are at least 3.76% is the sum of the relative frequencies of the last four classes = 0.356 + 0.311 + 0.133 = 0.800

c.
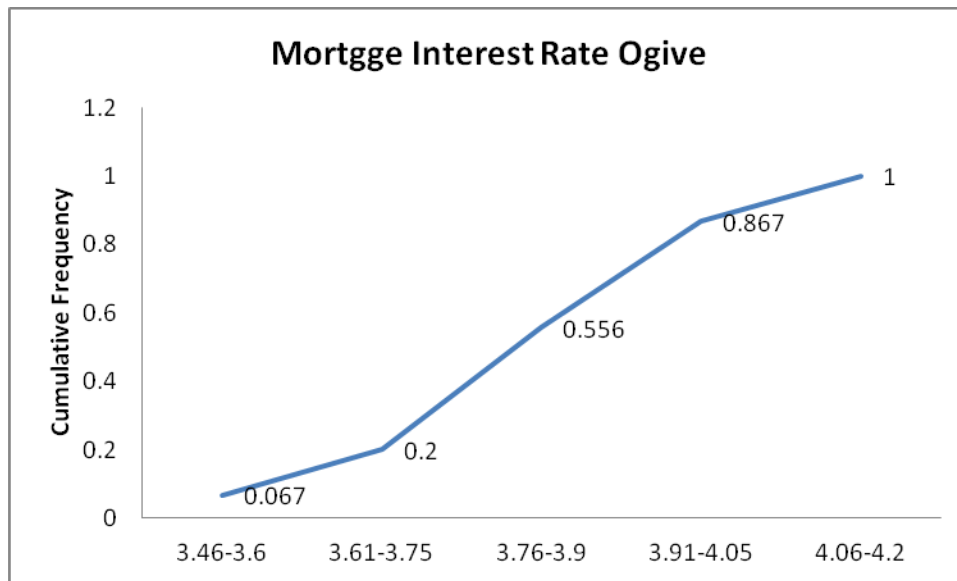
2.17

  a.
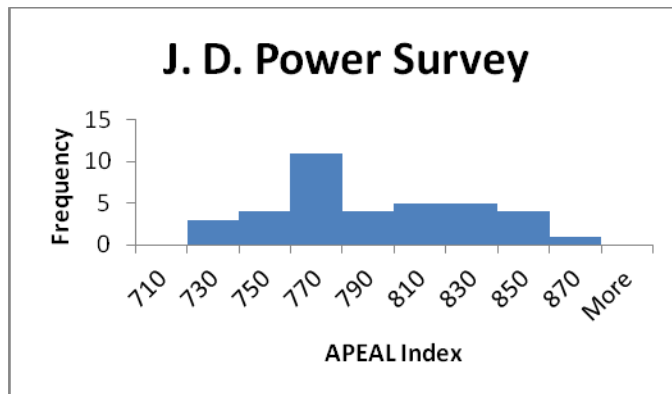
  b. The 2008 average is 782 which is less than the 2005 average of 866. This could indicate that the new models are less appealing to automobile customers, or customers could simply have rising expectations.

2.18

   a.

Classes	Frequency
51 - 53	7
54 - 56	15
57 - 59	28
60 - 62	16
63 - 65	21
66 - 68	9
69 - 71	2
72 - 74	2
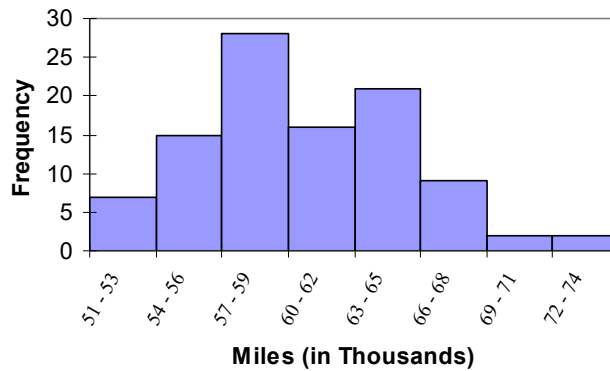
   b. The tread life of at least 50% of the tires is 60,000 or more. The top 10% is greater than 66,000 and the longest tread tire is 74,000. Additional information will vary.

   c.
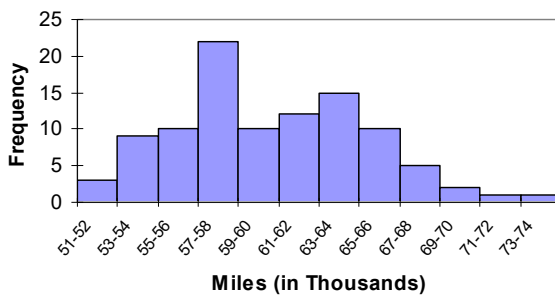
Classes	Frequency
51-52	3
53-54	9
55-56	10
57-58	22
59-60	10
61-62	12
63-64	15
65-66	10
67-68	5
69-70	2
71-72	1
73-74	1

Student will probably say that the 12 classes give better information because it allows you to see more detail about the number of miles the tires can go.

2.19.

   a. There are n = 294 values in the data. Then using the $2^k \geq n$ guideline we would need at least k = 9 classes.

   b. Using k = 9 classes, the class width is determined as follows:

$$w = \frac{High - Low}{Classes} = \frac{32 - 10}{9} = \frac{22}{9} = 2.44$$

Rounding this up to the nearest 1.0, the class width is 3.0.

c. The frequency distribution with nine classes and a class width of 3.0 will depend on the starting point for the first class. This starting value must be at or below the minimum value of 10. Student answers will vary depending on the starting point. We have used 10 as it is nice round number. Care should be made to make sure that the classes are mutually exclusive and all-inclusive. The following frequency distribution is developed:

Rounds	Frequency
10, 11, 12	10
13, 14, 15	31
16, 17, 18	65
19, 20, 21	90
22, 23, 24	64
25, 26, 27	26
28, 29, 30	6
31, 32, 33	2
34, 35, 36	0
Total	294

Students should recognize that by rounding the class width up from 2.44 to 3.0, and by starting the lowest class at the minimum value of 10, the 9th class is actually not needed.

d. Based on the results in part c, the frequency histogram is shown as follows:

Histogram
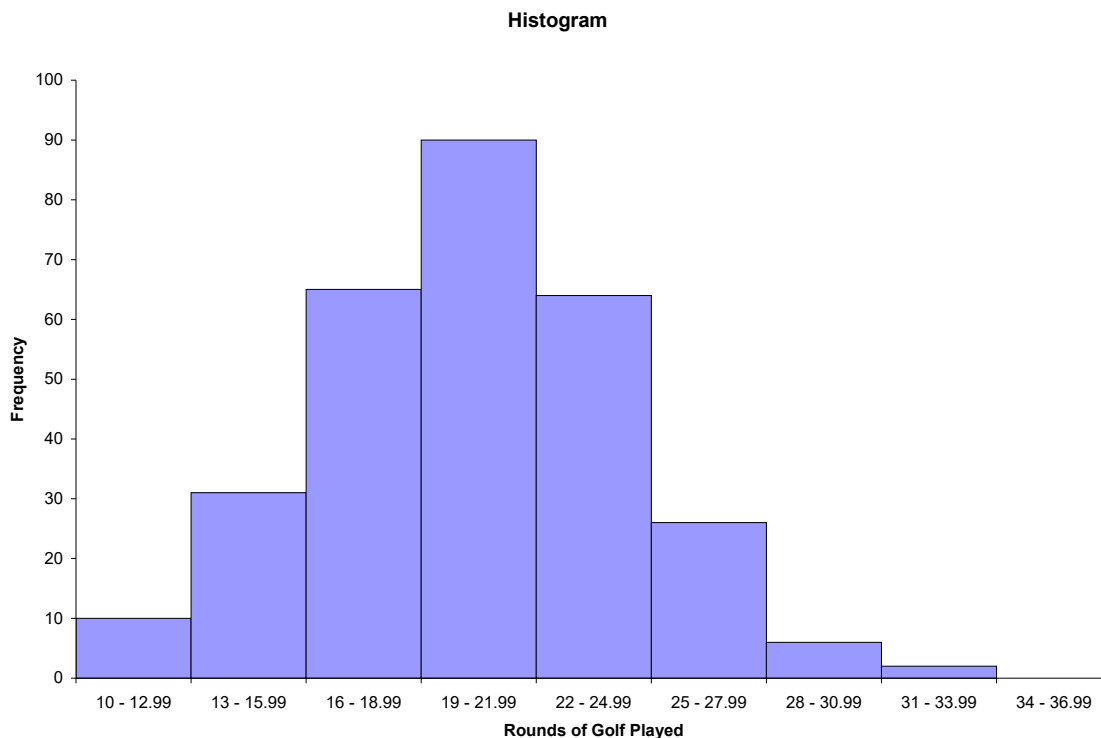
The distribution for rounds of golf played is mound shaped and fairly symmetrical. It appears that the center is between 19 and 22 rounds per year, but the rounds played is quite spread out around the center.

2.20

a.
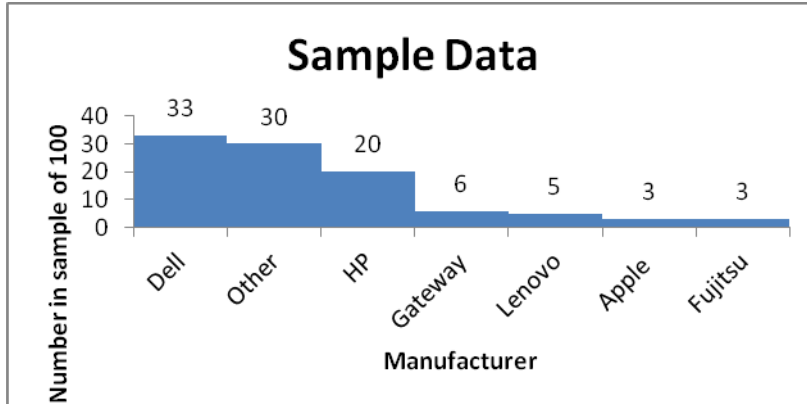
b. Excluding "Other" there are 100 – 30 = 70 percent of the manufacturers. Only Lenovo and Fujitsu have headquarters outside of the United States. There is, therefore, 5 + 3 = 8% for their market share. Therefore, the total market share of US manufacturers excluding "other" = (70 – 8)/70 = 0.89. Therefore, the total market share excluding "other" is 89%

2.21

a. Using the $2^k \geq n$ guideline, the number of classes would be 6. There are 41 airlines. $2^5 = 32$ and $2^6 = 64$. Therefore, 6 classes are chosen.

b. The maximum value is 602,708 and the minimum value is 160 from the Total column. The difference is 602,708 – 160 = 602548. The class width would be 602548/6 = 100424.67. Rounding up to the nearest 1,000 produces a class width of 101,000.

c.

Class	Frequency
0 - < 101,000	33
101,000 - < 202,000	2
202,000 - < 303,000	2
303,000 - < 404,000	1
404,000 - < 505,000	2
505,000 - < 606,000	1

d. Histogram follows:
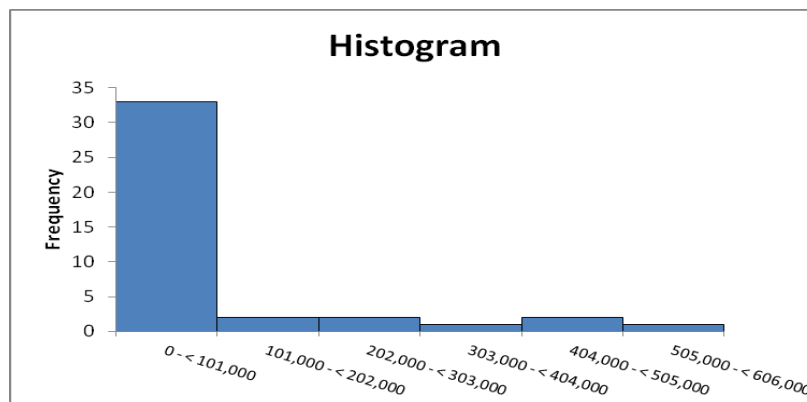
The vast majority of airlines had fewer than 101,000 monthly passengers for December 2011.

2.22
   a. The frequency distribution is:

Satisfaction Level	Frequency
1 = Very Dissatisfied	1
2 = Dissatisfied	82
3 = Nuetral	578
4 = Satisfied	530
5 = Very Satisfied	23
Total =	1214

The frequency distribution shows that over 1,100 people rated the overall service as either neutral or satisfied. While only 83 people expressed dissatisfaction, the manager should be concerned that so many people were in the neutral category. It looks like there is much room for improvement.
   b. The joint relative frequency distribution for "Overall Customer Satisfaction" and "Number of Visits Per Week" is:

Typical Visits Per Week ▼	Very Dissatisfied	Dissatisfied	Nuetral	Satisfied	Very Satisfied	Total
0	0.00	0.00	0.03	0.03	0.00	0.06
1	0.00	0.01	0.09	0.10	0.00	0.21
2	0.00	0.02	0.12	0.08	0.00	0.22
3	0.00	0.01	0.10	0.09	0.00	0.21
4	0.00	0.01	0.07	0.06	0.01	0.14
5	0.00	0.01	0.04	0.04	0.00	0.09
6	0.00	0.00	0.01	0.02	0.00	0.03
7	0.00	0.00	0.01	0.01	0.00	0.02
8	0.00	0.00	0.00	0.00	0.00	0.01
9	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00
Total	0.00	0.07	0.48	0.44	0.02	1.00

The people who expressed dissatisfaction with the service tended to visit 5 or fewer times per week. While 38% of the those surveyed both expressed a neutral rating and visited the club between 1 and 4 times per week.

2.23
   a. Order the observations (coffee consumption) from smallest to largest. The data array is shown below:

3.5	3.8	4.4	4.5	4.6	4.6	4.7	4.7	4.8	4.8	5.0	5.0
5.0	5.0	5.2	5.3	5.3	5.3	5.3	5.3	5.3	5.4	5.4	5.4
5.4	5.5	5.5	5.6	5.6	5.7	5.7	5.7	5.7	5.8	5.8	5.9
5.9	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.1	6.1	6.1
6.1	6.1	6.2	6.2	6.2	6.3	6.3	6.3	6.3	6.3	6.3	6.4
6.4	6.4	6.4	6.4	6.4	6.4	6.5	6.5	6.5	6.5	6.5	6.5
6.5	6.5	6.5	6.6	6.6	6.6	6.6	6.6	6.7	6.7	6.7	6.7
6.7	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.8	6.9	6.9	7.0
7.0	7.0	7.0	7.1	7.1	7.1	7.2	7.2	7.2	7.2	7.2	7.3
7.4	7.4	7.4	7.5	7.5	7.5	7.5	7.5	7.6	7.6	7.6	7.6
7.6	7.6	7.6	7.7	7.7	7.8	7.8	7.8	7.9	7.9	7.9	7.9
8.0	8.0	8.0	8.0	8.0	8.3	8.4	8.4	8.4	8.6	8.9	10.1

b. There are n = 144 observations in the data set. Using the $2^k \geq n$ guideline, the number of classes, k, would be 8. The maximum and minimum values in the data set are 10.1 and 3.5, respectively. The class width is computed to be: w = (10.1-3.5)/8 = 0.825, which is rounded up to 0.9.

Coffee Consumption (kg.)	Frequency
3.5 - 4.3	2
4.4 - 5.2	13
5.3 - 6.1	35
6.2 - 7.0	49
7.1 - 7.9	33
8.0 – 8.8	10
8.9 - 9.7	1
9.8 – 10.6	1

The class with the largest number is 6.2 –7.0 kg of coffee.

c. The histogram can be created from the frequency distribution. The classes are shown on the horizontal axis and the frequency on the vertical axis. The histogram is shown below.
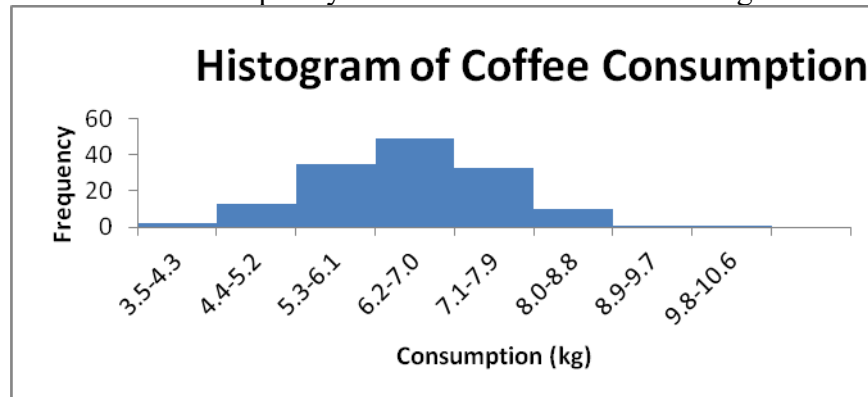
The histogram shows the shape of the distribution. This histogram is showing that fewer people consume small and large quantities and that most individuals consume between 5.3 and 8.0 kg of coffee, with the highest percentage of individuals consuming between 6.2 and 7.0.

d. Convert the frequency distribution into relative frequencies and cumulative relative frequencies as shown below.

Consumption	Frequency	Relative Frequency	Cumulative Relative Frequency
3.5-4.3	2	0.0139	0.0139
4.4-5.2	13	0.0903	0.1042
5.3-6.1	35	0.2431	0.3472
6.2-7.0	49	0.3403	0.6875
7.1-7.9	33	0.2292	0.9167
8.0-8.8	10	0.0694	0.9861
8.9-9.7	1	0.0069	0.9931
9.8-10.6	1	0.0069	1

8.33% (100 - 91.67) of the coffee drinkers sampled consumes 8.0 kg or more annually.

Section 2-2

2.24

    a. The pie chart is shown as follows:

Education Levels
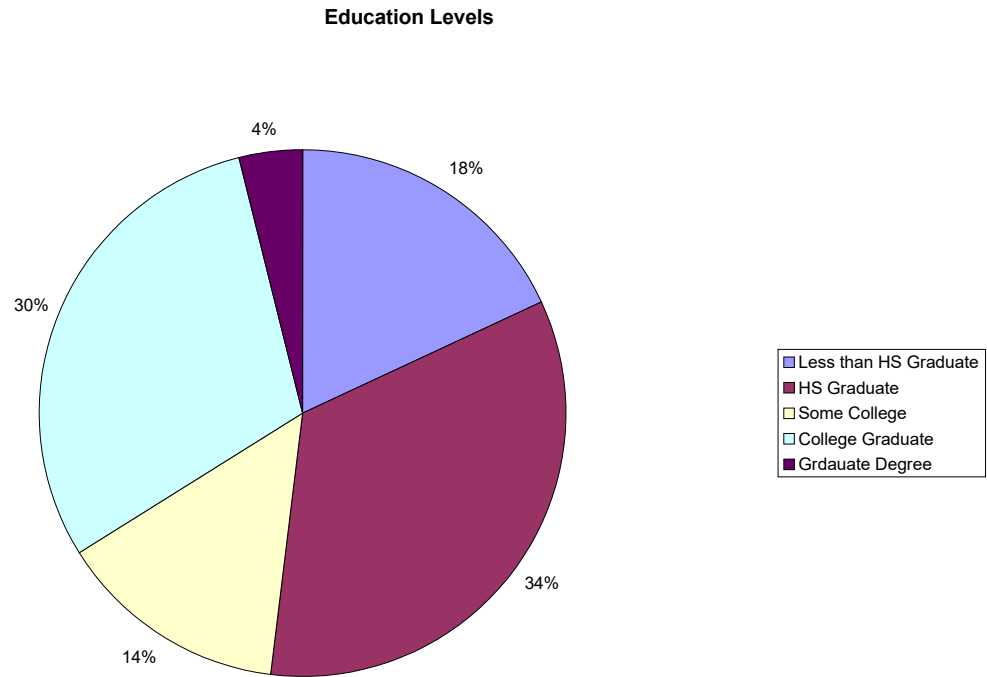
    b. The horizontal bar chart is shown as follows:

Education Levels
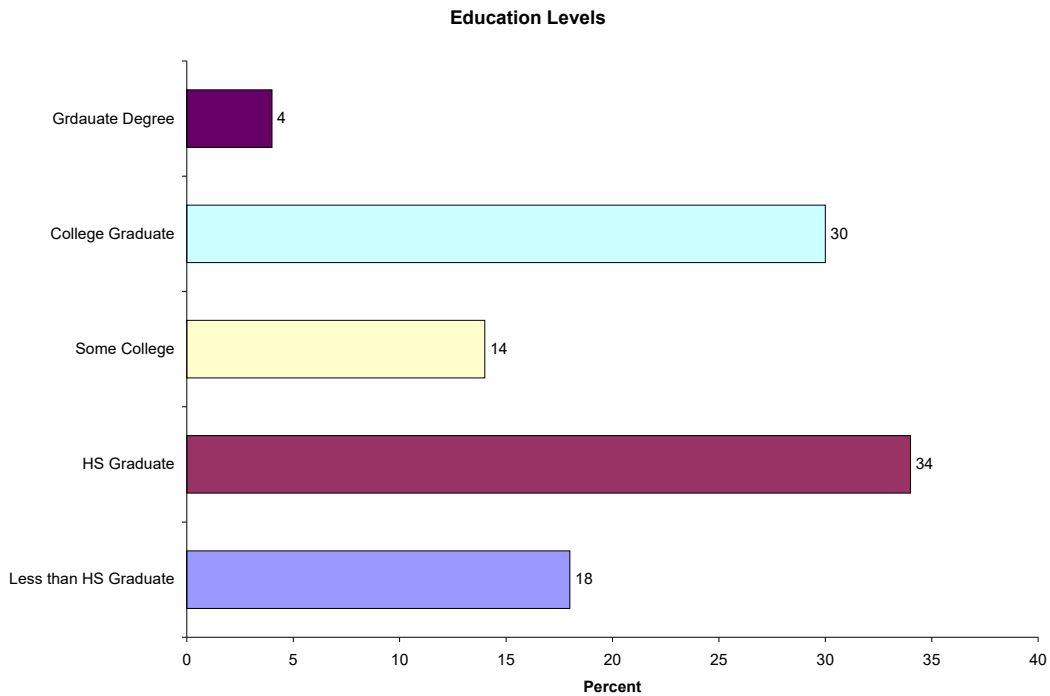
2.25
Step 1. Sort the data from low to high.
This is done in the problem. The lowest value is 0.7 and the highest 6.4.
Step 2. Split the values into a stem and leaf.
Stem = units place   leaf = decimal place
Step 3. List all possible stems from lowest to highest.
Step 4. Itemize the leaves from lowest to highest and place next to the appropriate stems.

Stem-and-Leaf Display		
Stem unit: 1		
0	7 8	
1	0 1 4 7 8	
2	0 0 1 4 8	
3	0 3 8	
4	3 4	
5	3 4 4	
6	3 4	

2.26
a.
Step 1. Define the categories.
The categories are grade level.
Step 2. Determine the appropriate measure.
The measure is the number of students at each grade level.
Step 3. Develop the bar chart.
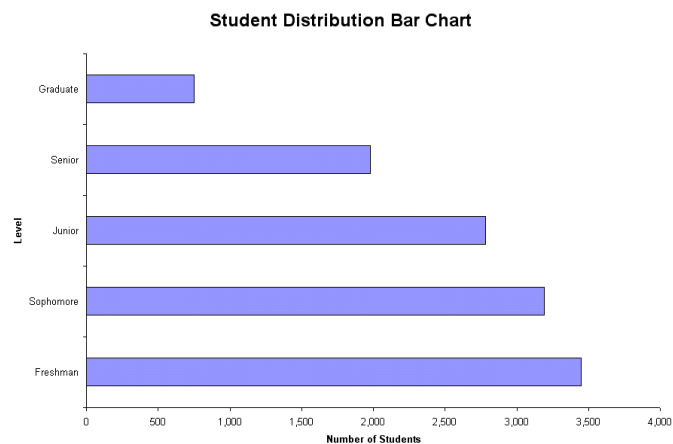
Student Distribution Bar Chart

b.
   Step 1. Define the categories.
   The categories are grade level.
   Step 2. Determine the appropriate measure.
   The measure is the percent of students at each grade level.
   Step 3. Develop the pie chart.

Student Distribution Bar Chart
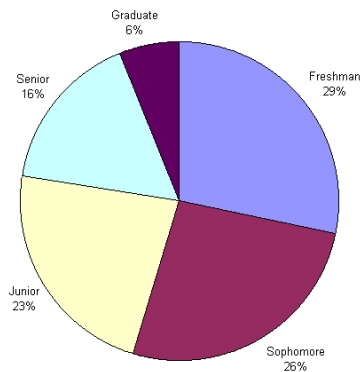
c.   A case can be made for either a bar chart or pie chart. Pie charts are especially good at showing how the total is divided into parts. The bar chart is best to draw attention to specific results. In this case, a discussion might be centered on the posible attrition that takes place in the number of students between Freshman and Senior years.

2.27.   One possible bar chart is shown as follows:

Sales By Product Type and Region
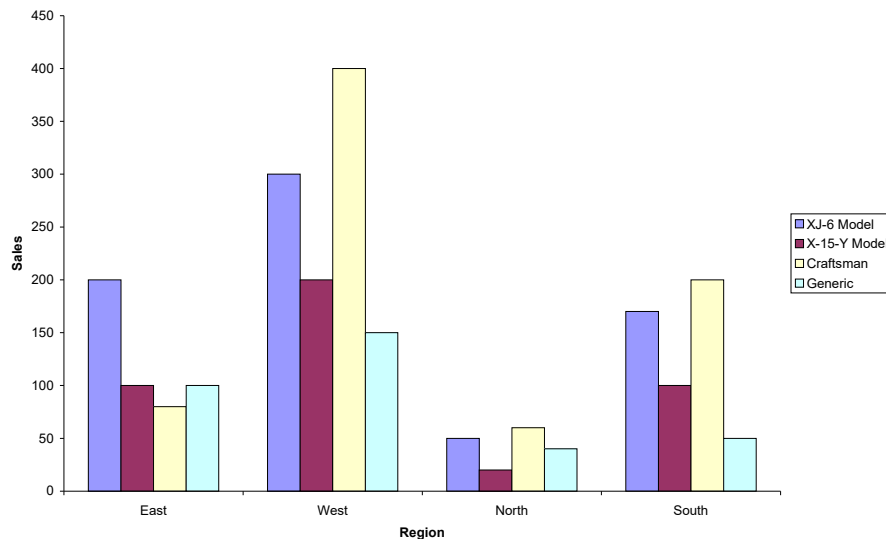
Another way to present the same data is:

Sales By Product Type and Region
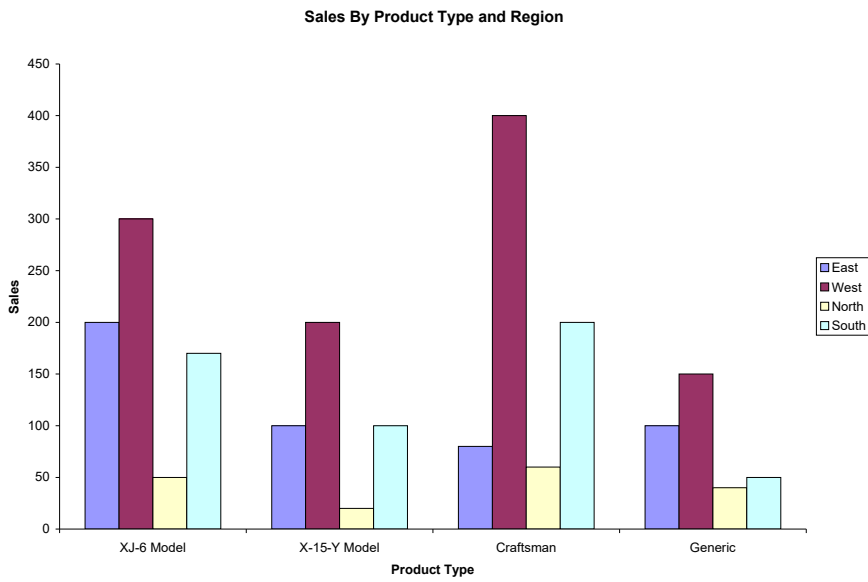
Still another possible way is called a "stacked" bar chart.

Sales By Product Type and Region

2.28

a.

United States Refinery Yields by Product Per Day

- Gasoline — 42%
- Kerosene — 7%
- Diesel and Home Heating Oils — 28%
- Residuals — 12%
- 10%
- 1%

b.

United Kingdom Refinery Yields by Product Per Day

- Gasoline — 25%
- Kerosene — 13%
- Diesel and Home Heating Oils — 31%
- Residuals — 10%
- Asphalt, LPG, and other — 18%
- 3%

c.

United States vs United Kingdom Refinery Yields by Product Category

Categories: Gasoline, Kerosene, Diesel and Home..., Residuals, Asphalt, LPG, and..., Fuel and Loss

- United States
- United Kingdom

2.29

a.

Net Operating Income Distribution by Region

b.

Net Operating Income Distribution by Region

c. The bar chart makes it easier to compare percentages across regions.

2.30

a.

Profit Bar Chart
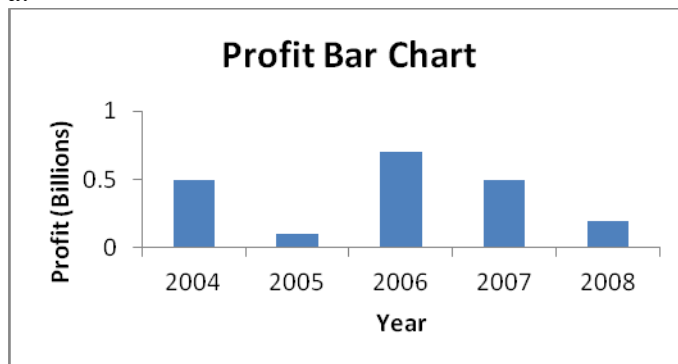
b.
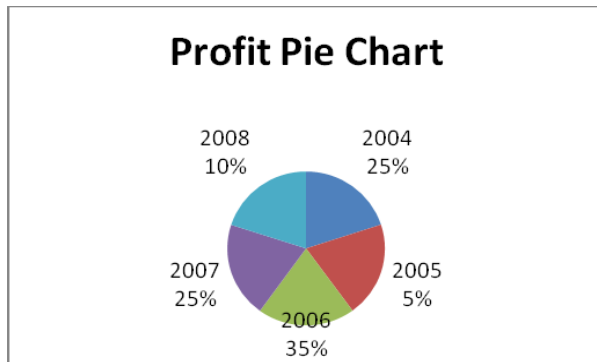
c. Arguments exist for both the pie chart and the bar chart. Pie charts are especially good at showing how the total is divided into parts. The bar chart is best to draw attention to specific results. In this case, it is most likely that the historic change in profits is to be displayed. The bar chart is best at presenting time defined data.

2.31

a.
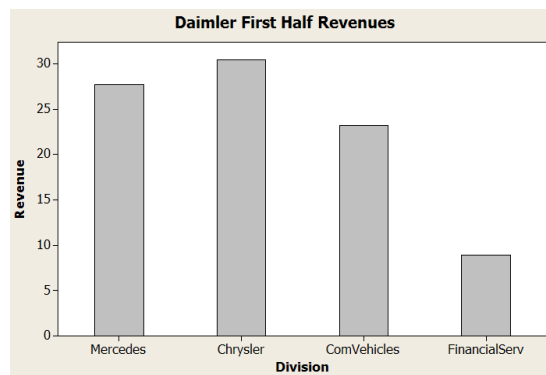
b. Proportion of Vehicle Revenue = 1 – proportion of financial services
$$= 1 - 8.9/90.3$$
$$= 1 - 0.0985 \ = \ 0.9015$$

2.32

a. Pie charts are typically used to show how a total is divided into parts. In this case, the total of the five ratios is not a meaningful value. Thus, a pie chart showing each ratio as a fraction of the total would not be meaningful. Thus a pie chart is not the most appropriate tool. A bar chart would be appropriate.

b

Step 1: Define the categories.
The categories are the five cities where the plants are located
Step 2: Determine the appropriate measure.
The measure of interest is the ratio of manufactured output to the number of employees at the plant
Step 3: Develop the bar chart.
The bar chart is shown as follows:

Step 4:   Interpret the results.
It appears that the number of units manufactured per employee of the plants in the Midwest is larger than in the West.
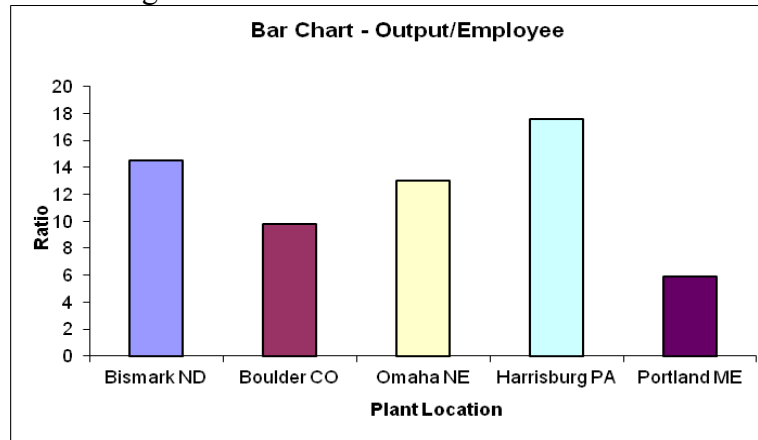
2.33.   Step 1: Define the categories.
The categories are the five years, 2000, 2001….,2004
Step 2: Determine the appropriate measure.
The measure of interest is the number of homes that have a value of $1 million or more.
Step 3: Develop the bar chart.
The horizontal bar chart is shown as follows:
Step 4:   Interpret the results.
The bar chart is skewed below indicating that number of $1 Million houses was growing rapidly. It also appears that that growth is exponential rather than linear.
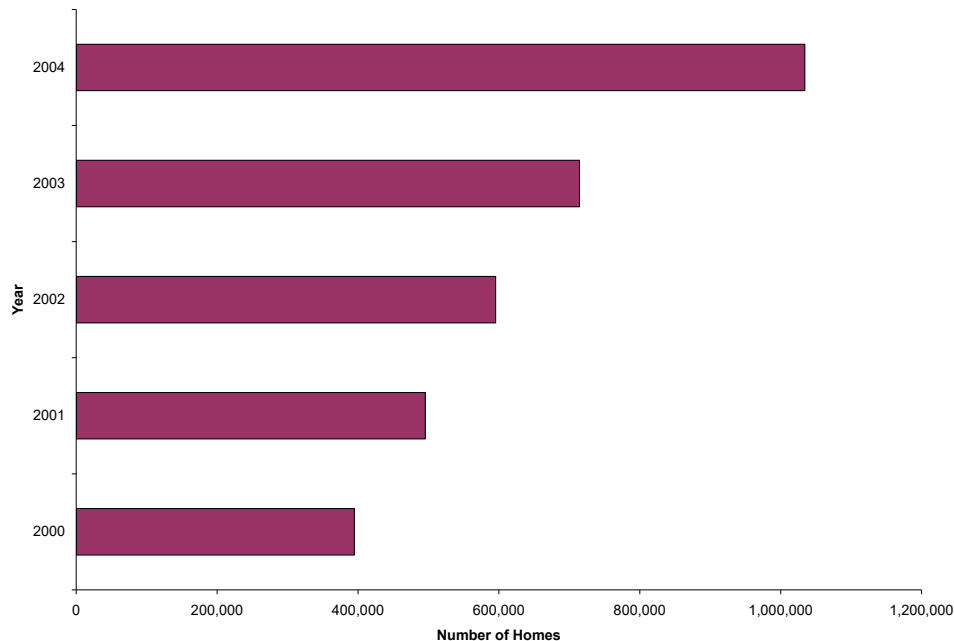
The bar chart is skewed below indicating that number of $1 Million houses is growing rapidly. It also appears that that growth is exponential rather than linear.

2.34 The appropriate graph for these data is a pie chart since the objective is to show the market share for these leading antidepressant products,

Step 1: Define the categories.

The categories are the different drugs

Step 2: Determine the appropriate measure.

The measure of interest is sales in billions of dollars

Step 3: Develop the pie chart.

The pie chart is shown as follows:

Antidepressants - US Market Share
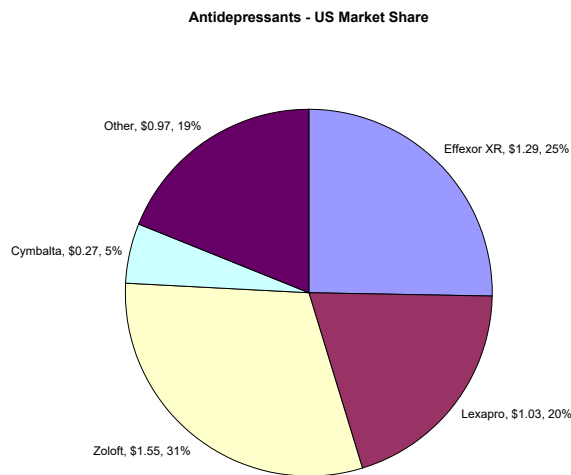
2.35 A bar chart can be used to make the comparison. Shown below are two examples of bar charts which compare North America to the United Kingdom.
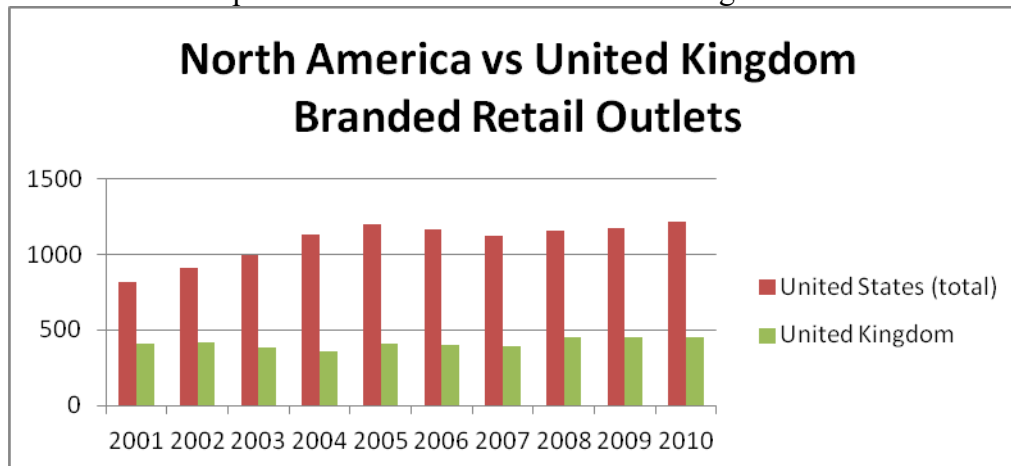
2.36
    a.   The bar chart is shown below. The categories are the Global Segments and the measure for each category is the net sales for the Global Segment.

Proctor & Gamble Net Sales by Global Business Segments

Beauty	20,000
Grooming	8,000
Health Care	12,000
Snacks and Pet Care	3,000
Fabric Care and Home Care	25,000
Baby Care and Family Care	15,500
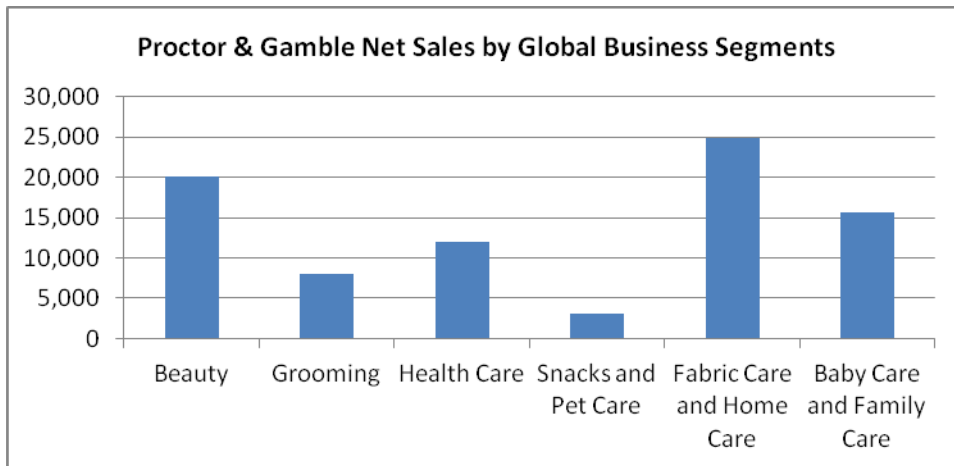
    b.   The pie chart is shown below. The categories are the Global Segments and the measure is the proportion of each segment's total net sales.

2011 Net Sales as a Percentage of Global Segments

- Beauty — 24%
- Grooming — 9%
- Health Care — 14%
- Snacks and Pet Care — 4%
- Fabric Care and Home Care — 30%
- Baby Care and Family Care — 19%
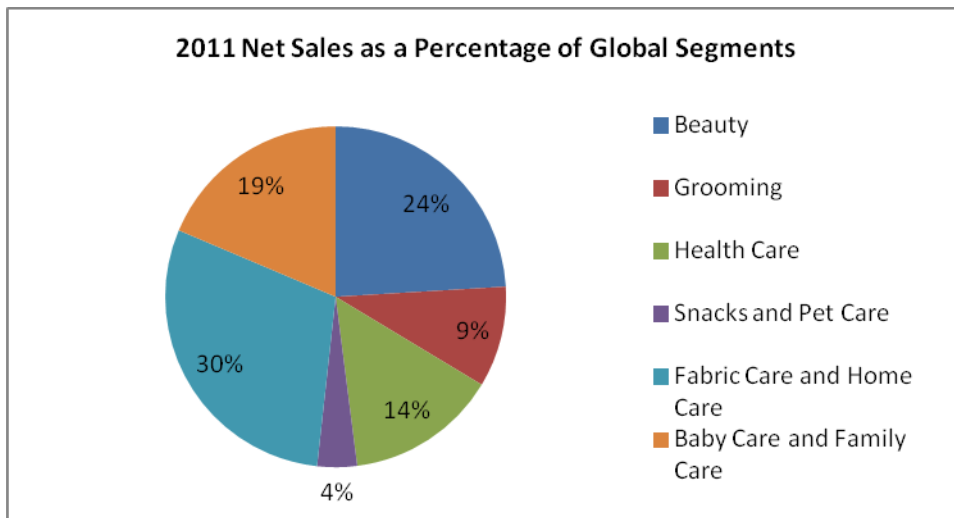
2.37

  a. The following stem and leaf diagram was created using PhStat. The stem unit is 10 and the leaf unit is 1.

Stem-and-Leaf Display for Drive-Thru Service (Seconds)

Stem unit: 10

```
 6 | 8
 7 | 1 3 4 6 9
 8 | 3 5 8
 9 | 0 2 3
10 | 3 5
11 | 0 6 9
12 |
13 | 0 4 8
14 | 5 6 7
15 | 6 6
16 | 2
17 | 8
18 | 1
```

  b. The most frequent speed of service is between 70 and 79 seconds.

2.38

  a. The following stem-and-leaf diagram was developed using PhStat. The stem unit is 10 and the leaf unit is 1.

  Stem-and-Leaf Display for Number of Days to Collect Payment

Stem unit: 10

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

  b. Most payments are collected in the range of 30-39 days.

2.39
a. The bar graph is
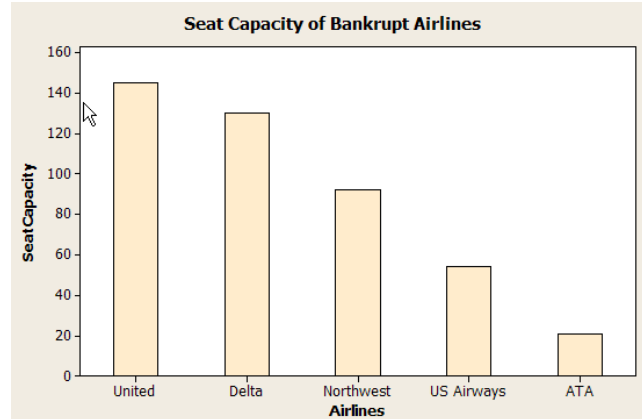
b. The percent equals the individual capacity divided by the total, e.g. United→percent = (145/858)100% = 16.90%, etc. This produces the following pie chart:
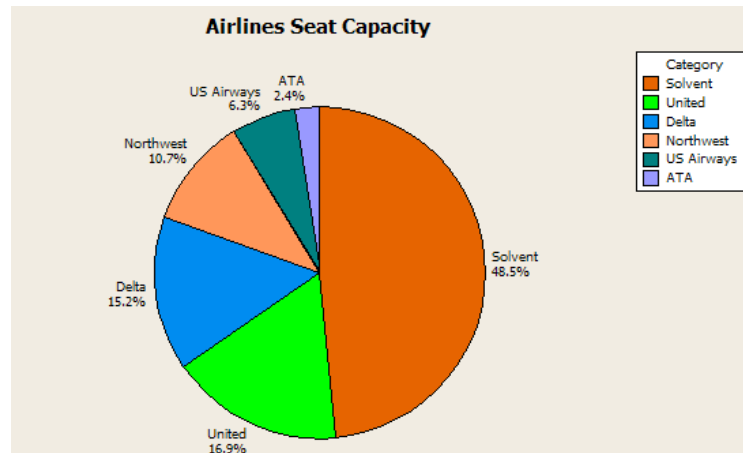
c. The percent of seat capacity of those in bankruptcy = 16.9 + 15.2 + 10.7 + 6.3 + 2.4 = 51.5%. Since this is larger than 50%, their statement was correct.

2.40   a.
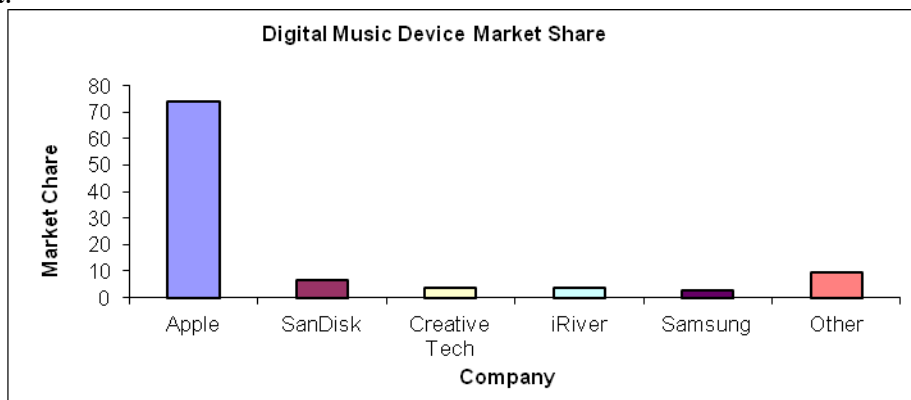
b.
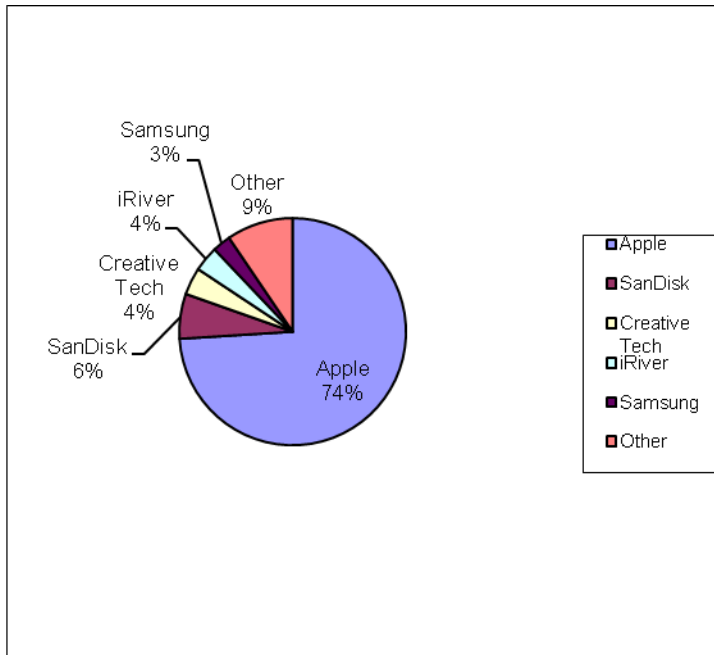
c. A case can be made for either a bar chart or a pie chart. Pie charts are especially good at showing how the total is divided into parts. The bar chart is best to draw attention to specific results. In this case, a discussion might be centered on the relative large percentage attributable to Apple.

2.41
a.

Stem-and-Leaf Display: Days

```
Stem-and-leaf of Days   N  = 50
Leaf Unit = 1.0


  1    0  4
  2    0  7
  6    1  0344
 15    1  566677889
 23    2  00012244
(13)   2  5666777888999
 14    3  000122344
  5    3  5669
  1    4  0
```

b. The shape of the data is slightly skewed to the left. The center of the data appears to be between 24 and 26.

c. $\bar{x} = \dfrac{2428}{50} = 48.56$. This and the data indicates that the mean is larger than indicated by J.D. Power. The difference is that the data set is only a sample of the data. Each sample will produce different results but approximately equal to the population average calculated by J.D. Power.

2.42.
   a. A bar chart is an appropriate graph since there are two categories, males and females. A pie chart could also be used to display the data.
   b. The following steps are used to construct the bar chart:
      Step 1: Define the categories.
           The categories are the two genders, male and female
      Step 2: Determine the appropriate measure.
           The measure of interest is the percentage of credit card holders who are male and female.
      Step 3: Develop the bar chart using computer software such as Excel or Minitab.
           The bar chart is shown as follows:

Bar Chart: Percentage of Credit Card Customers by Gender
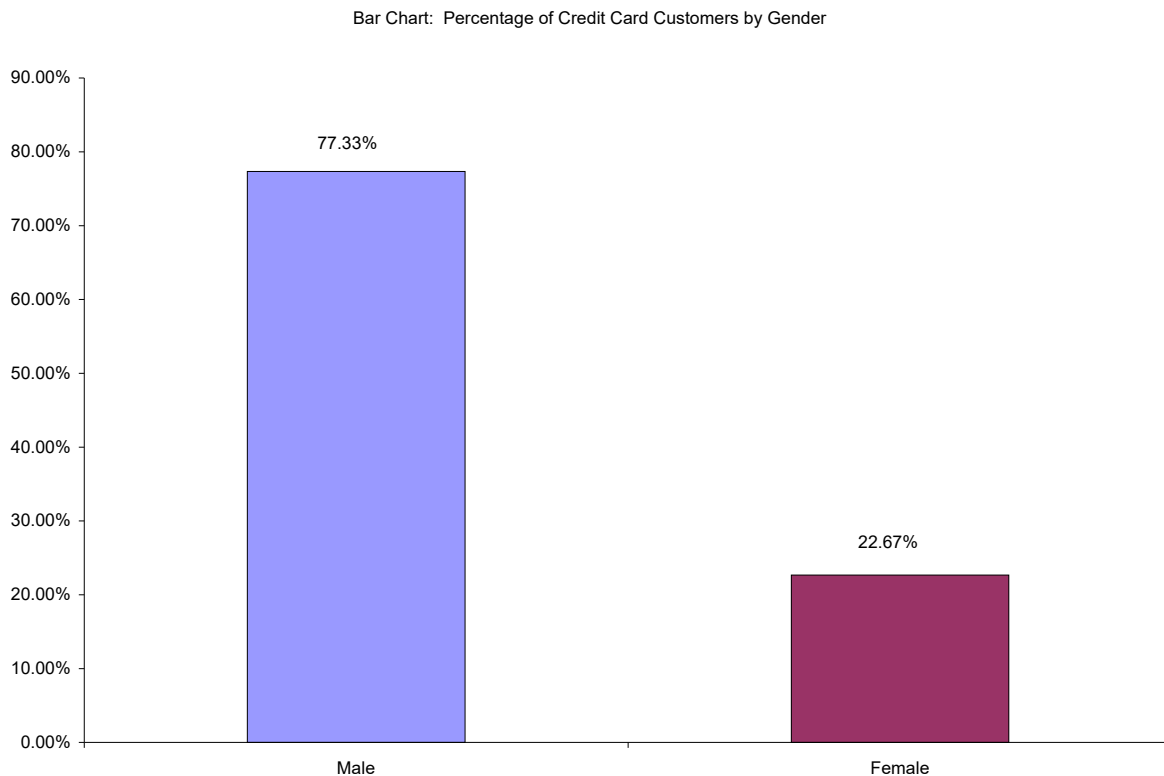
      Step 4 Interpret the results.
      This shows that a clear majority of credit card holders are males (77.33%)

2.43

a. The following are the averages for each hospital computed by summing the charges and dividing by the number of charges:

University Related	Religious Affiliated	Municipally Owned	Privately Held
$6,398	$3,591	$4,613	$5,191

b. The following steps are used to construct the bar chart:

Step 1: Define the categories.

The categories are the four hospital types

Step 2: Determine the appropriate measure.

The measure of interest is the average charge for outpatient gall bladder surgery.

Step 3: Develop the bar chart using computer software such as Excel or Minitab.

The bar chart is shown as follows:

Gall Bladder Charges
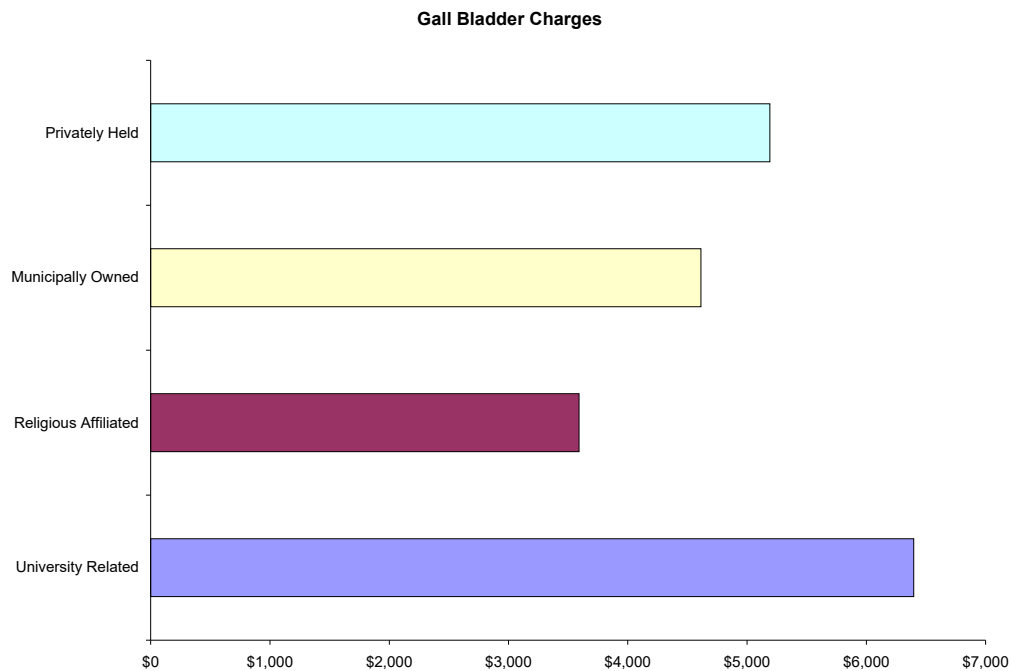
Step 4: Interpret the results.

The largest average charges occurred for gall bladder surgery appears to be in University Related hospitals and the lowest average appears to be in Religious Affiliated hospitals.

c. A pie chart is used to display the parts of a total. In this case the total charges of the four hospital types is not a meaningful number so a pie chart showing how that total is divided among the four hospital types would not be useful or appropriate.

2.44

a.
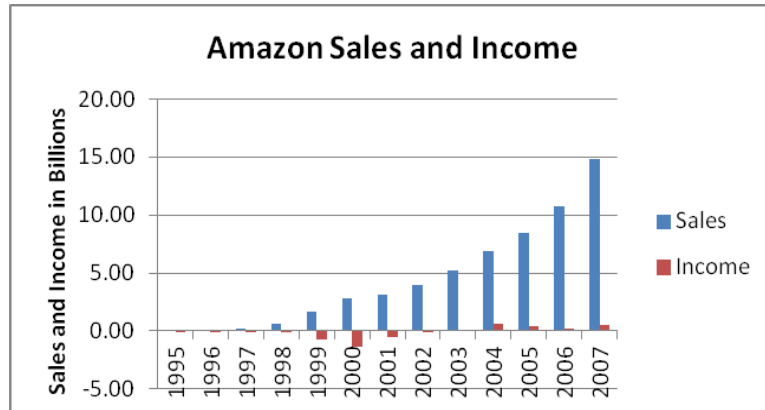
b. There appears to be a linear relationship between sales and years in which the sales were made.

c. In time period between 2000 and 2001, Amazon experienced a decrease in its loses. Prior to this time, each year produced increased loses.

2.45

a.
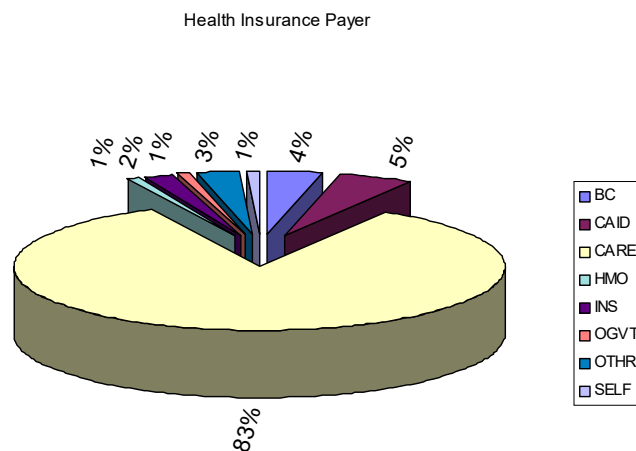
b.
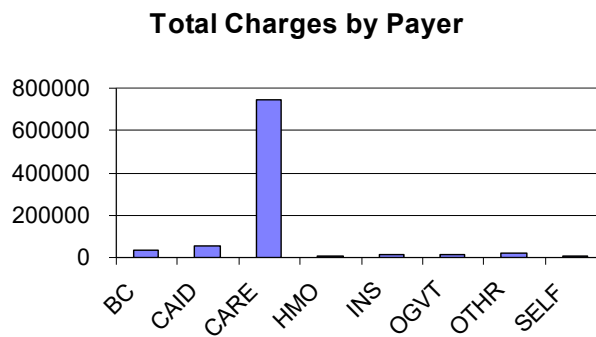
c. Using PHStat the stem & leaf diagram is shown as follows.

```
Stem-and-Leaf Display
for Length of Stay
Stem unit: 1

    1 | 0 0 0 0 0 0 0 0
    2 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    3 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    4 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    5 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    6 | 0 0 0 0 0 0 0 0 0
    7 | 0 0 0 0 0 0 0 0 0 0
    8 | 0 0 0 0 0
    9 | 0 0 0 0
   10 |
   11 | 0 0
   12 | 0
   13 |
   14 |
   15 | 0
   16 | 0
```

d. Excel's pivot table can be used to develop a bar chart. The chart showed is a stacked bar chart.

Section 2-3

2.47
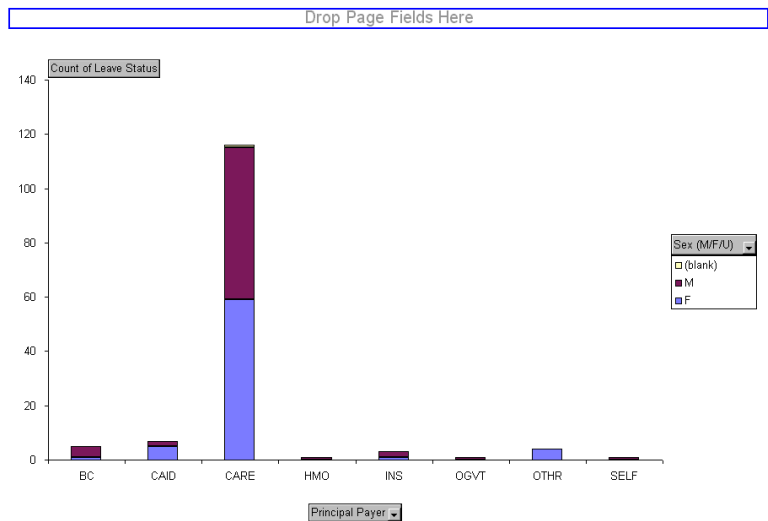
a.
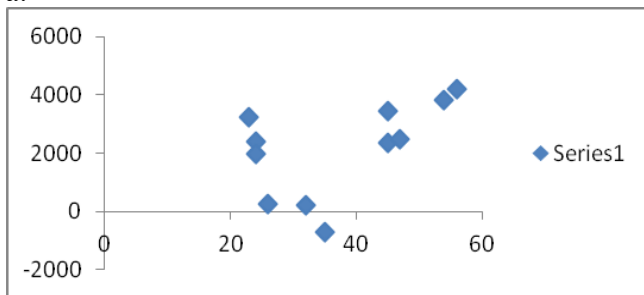
There appears to be a curvilinear relationship between the dependent and independent variables.

b.
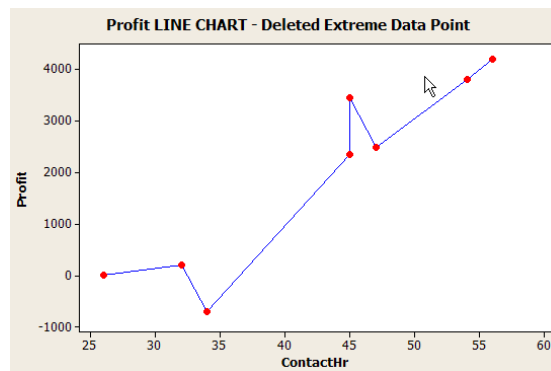
Having removed the extreme data points, the relationship between dependent and independent variables seems to be linear and positive.

2.47   Step 1: Identify the time-series variable
                    The variable of interest is the monthly sales.
        Step 2: Layout the Horizontal and Vertical Axis
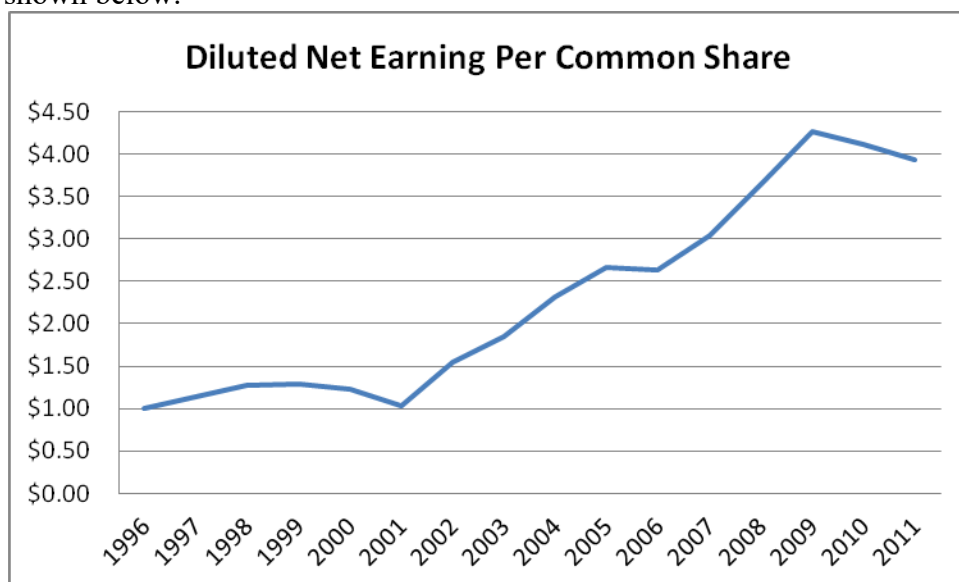   The horizontal axis will be month and the vertical axis is sales .
   Step 3: Plot the values on the graph and connect the points
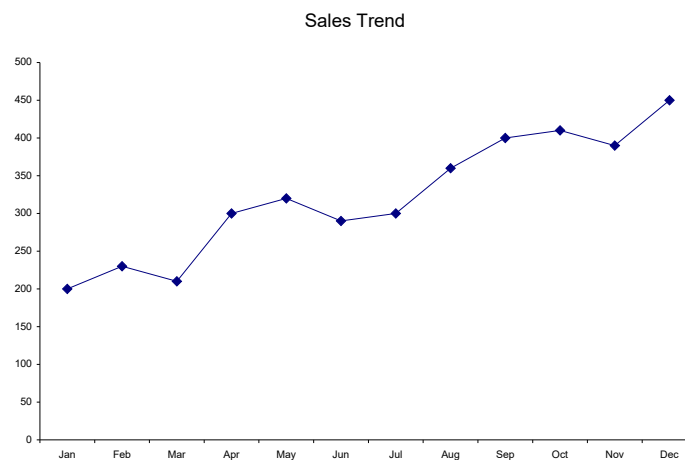
The sales have trended upward over the past 12 months.

2.48   Steps 1 and 2: Identify the two variables of interest
            The variables are y (dependent variable) and x (independent variable)
        Step 3: Establish the scales for the vertical and horizontal axes
            The y variable ranges from 40 to 250 and the x variable ranges from 15.9 to 35.3
        Step 4: Plot the joint values for the two variables by placing a point in the x,y space
            shown as follows:

Scatter Diagram
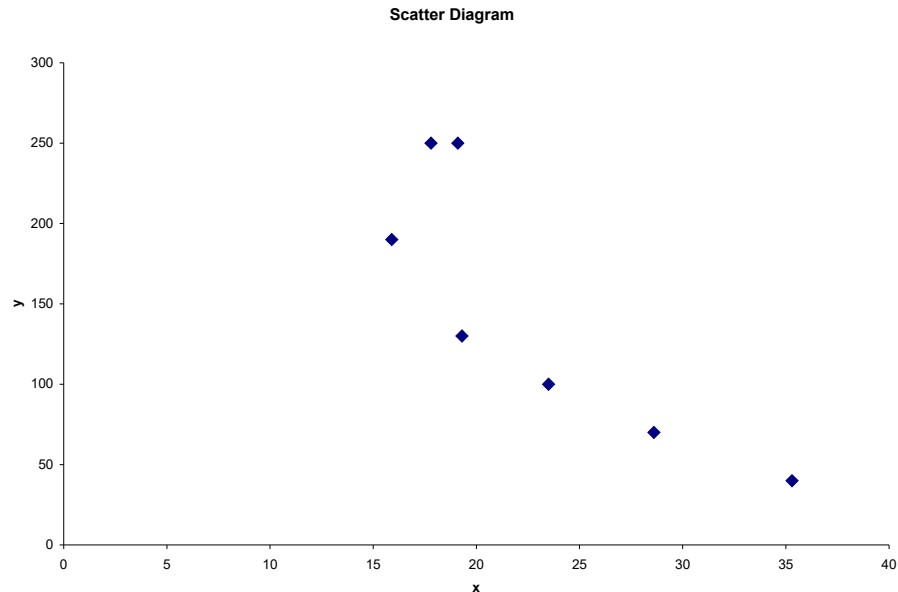
There is negative linear relationship between the two variables.

2.49 The time-series variable is Year-End Dollar Value Deposits ($ in millions) measured over 8 years with a maximum value of 1,380 (million). The horizontal axis will have 8 time periods equally spaced. The vertical axis will start at 0 and go to a value exceeding 2.500. We will use 3,000. The vertical axis will also be divided into 500-unit increments. The line chart of the data is shown below.

**Deposits for Bank of the Ozarks, Inc.**
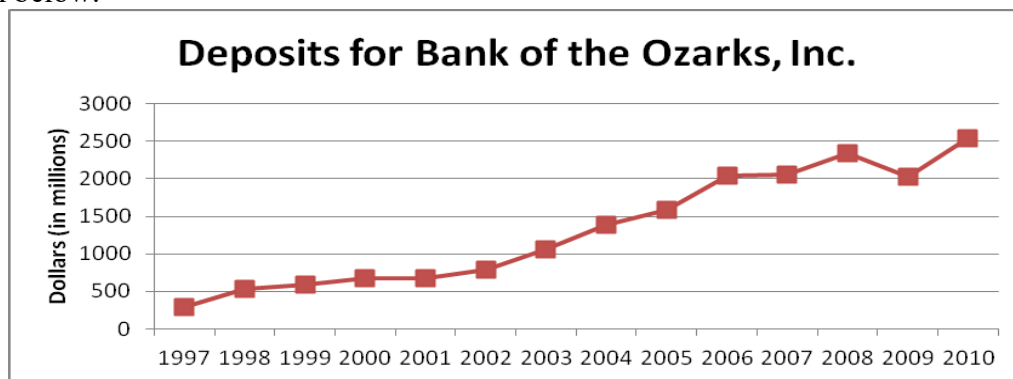
The line chart shows that Year-End Deposits have been increasing since 1997, but have increased more sharply since 2002, and leveled off somewhat between 2006 and 2007 and recently have been more volatile.

2.50

a.
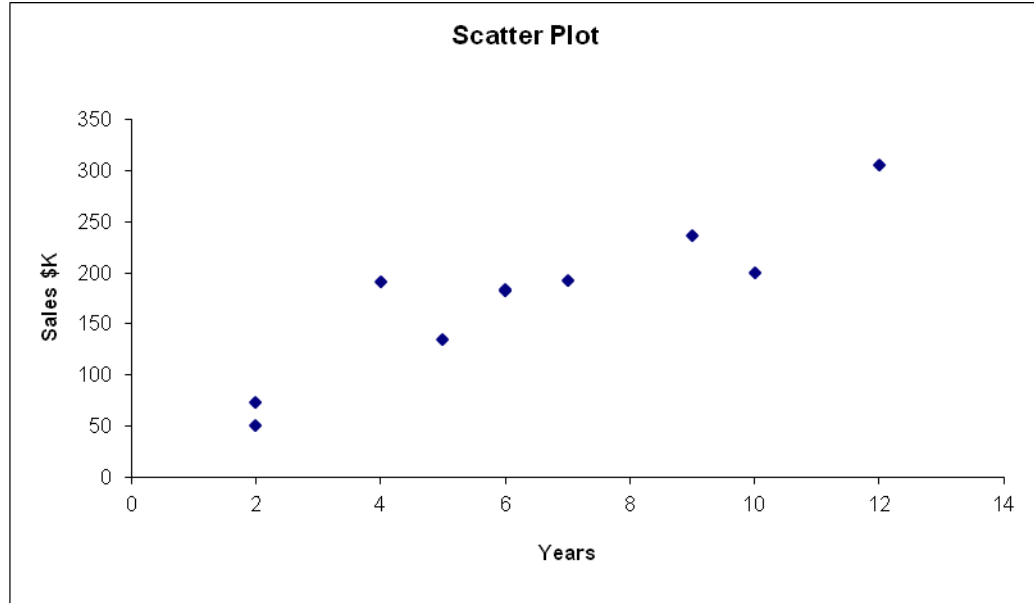
b. The relationship appears to be linear and positive.
c. Note on the line plot that the Years starts at 2 years and stops at 12 years: a range of 10 years. Also the sales increase from 50 to 300 $K: a range of 250. This suggests that 250/10 = 25 $K average increase per one year increase.

2.51

a.
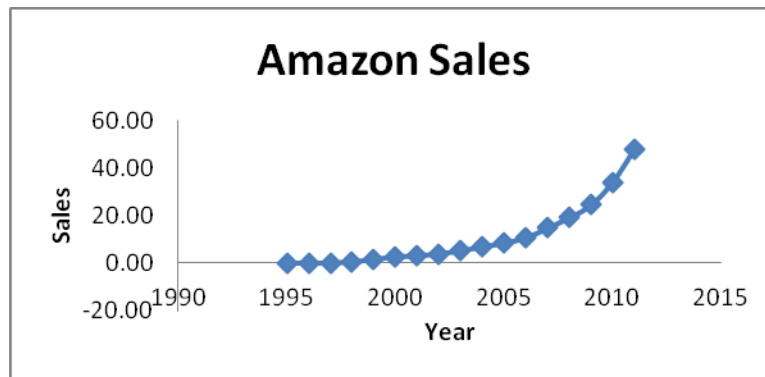
b. The relationship appears to be curvilinear.
c. The largest difference in sales occurred between 2010 and 2011. That difference was almost 14 $billions.

2.52.

      Step 1: Identify the time-series variable
           The variable of interest is annual sales of video games in the U.S.
      Step 2: Layout the Horizontal and Vertical Axis
           The horizontal axis will be the year and the vertical axis is sales.
      Step 3: Plot the values on the graph and connect the points
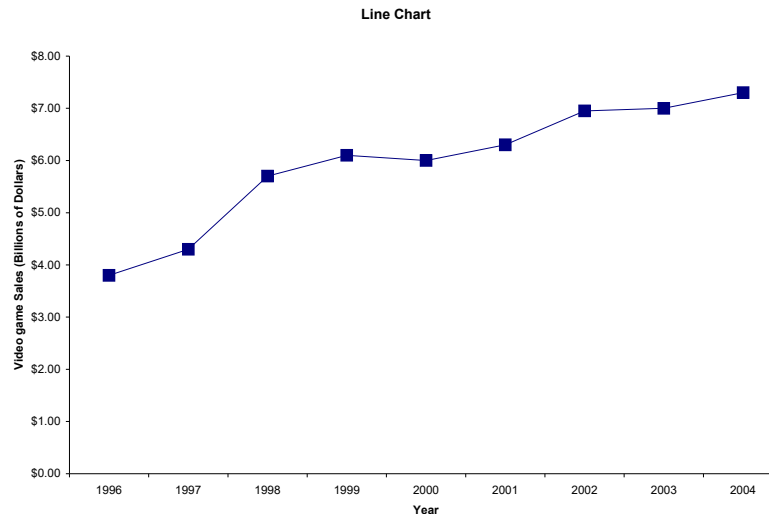
Line Chart

The line chart illustrates that over the nine year period between 1996 and 2004, video game sales in the U.S. have grown quite steadily from a just below $4 billion to over $7 billion.

2.53

    a. The time-series variable is diluted net earnings per common share measured over 16 years with a maximum value of $4.26. The horizontal axis will have 1 time periods equally spaced. The vertical axis will start at 0 and go to a value exceeding $4.26. We will use $4.50. The vertical axis will also be divided into $0.50-unit increments. The line chart of the data is shown below.

Diluted Net Earning Per Common Share

b. The time-series variable is dividends per common share measured over 16 years with a maximum value of $1.03. The horizontal axis will have 16 time periods equally spaced. The vertical axis will start at 0 and go to a value exceeding $1.97. We will use $2.00. The vertical axis will also be divided into $0.20-unit increments. The line chart of the data is shown below.

Dividends Per Common Share
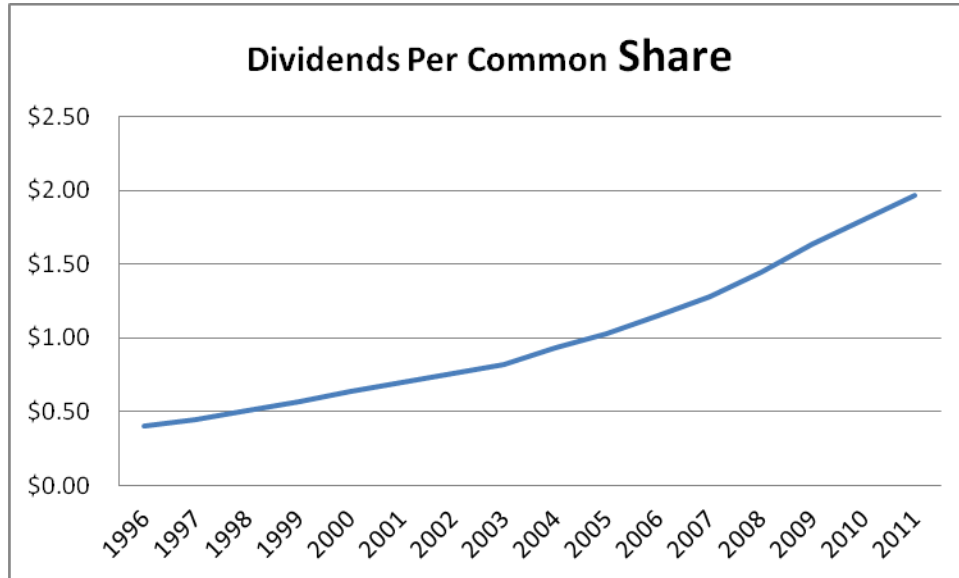
c. One variable is Diluted Net Earnings per Common Share and the other variable is Dividends per Common Share. The variable dividends per common share is the dependent (y) variable. The maximum value for each variable is $4.26 for Diluted Net Earnings and $1.97 for Dividends. The XY Scatter Plot is shown below.

Scatter Plot of Dividends vs Diluted Net Earnings
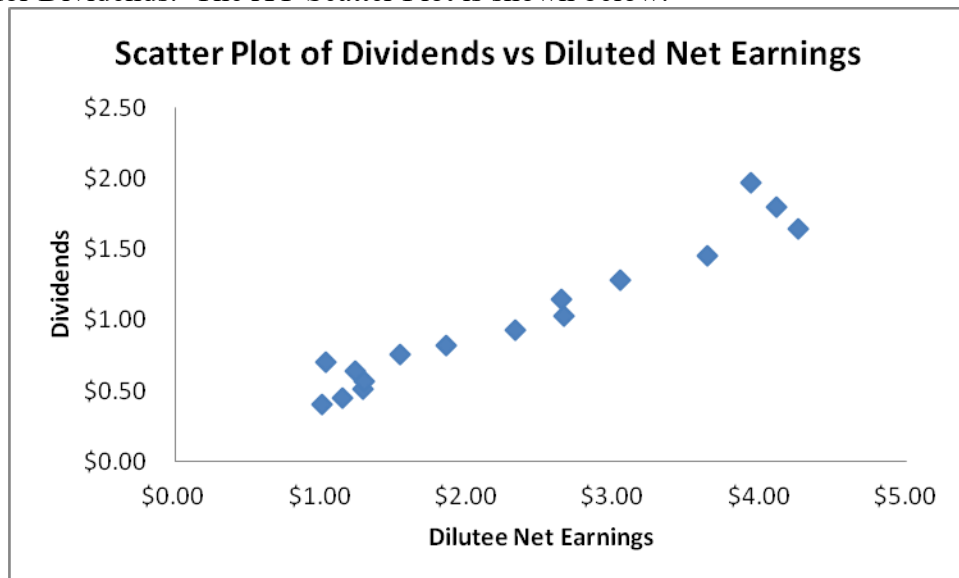
There is a relatively strong positive relationship between the two variables, which is as one would expect. That is, one might expect to see the two variables move in the same direction.

2.54
   a.
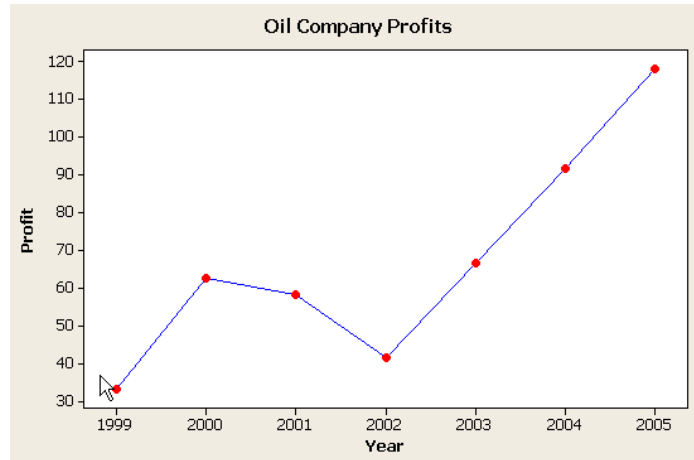
   b. In the interval 1999 – 2002, there appears to be a curvilinear relationship. Whereas in the interval 2002 – 2005, the relationship seems to be linear.

   c. The most recent data indicate that the relationship is linear. Devoid of additional, cooperating data, a linear relationship would be used to project the companies' profits in the year 2006.

2.55.

      Steps 1: Identify the two variables of interest
           In this example, there are two variables of interest, average home attendance and average road game attendance.
      Step 2: Identify the dependent and independent variables.
           Either one of these variables can be selected as the dependent variable. We will select average road game attendance
      Step 3: Establish the scales for the vertical and horizontal axes
           The y variable (average road attendance) ranges from 29,308 to 38,367 and the x variable (average home attendance) ranges from 16,688 to 53,609.
      Step 4: Plot the joint values for the two variables by placing a point in the x,y space shown as follows:
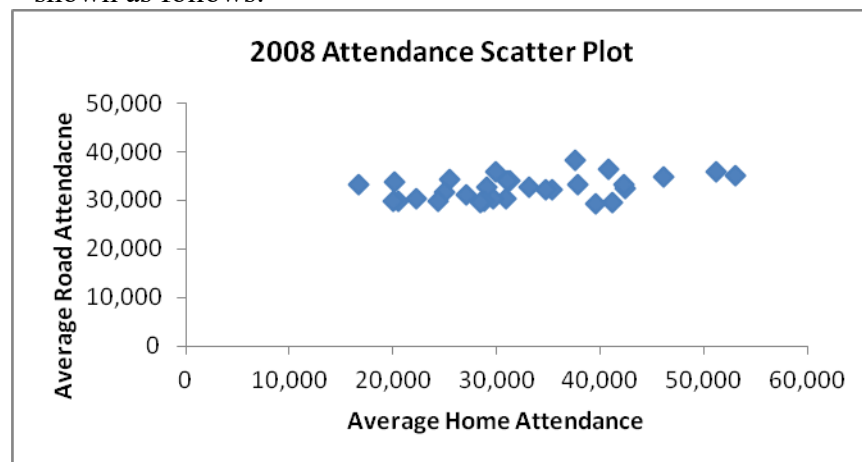
Based on the scatter diagram, it appears that there is a slight positive linear relationship between home and road attendance. However, the relationship is not perfect.

2.56
a. Step 1: Identify the time-series variable
The variable of interest is number of jets owned by two or more owners
Step 2: Layout the Horizontal and Vertical Axis
The horizontal axis will be the year and the vertical axis is the number of jets owned.
Step 3: Plot the values on the graph and connect the points

Fractional Jet Ownership
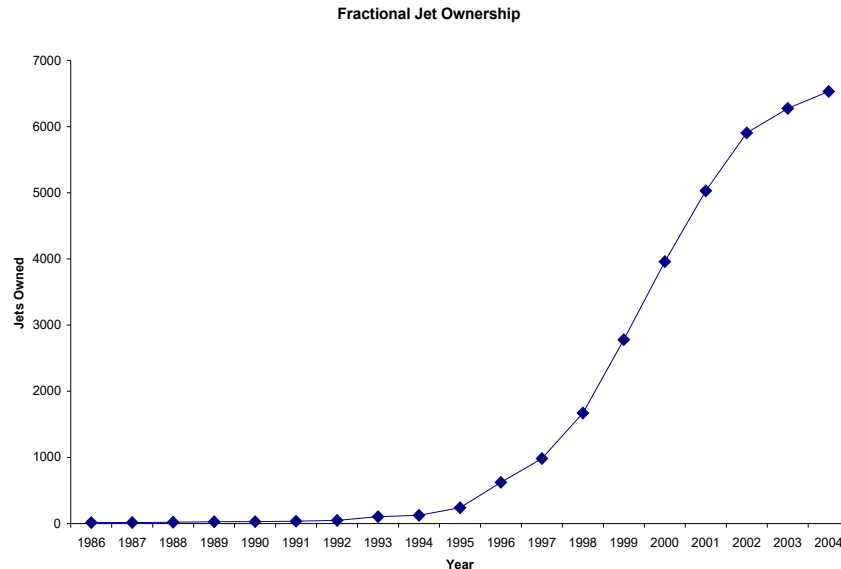
Since 1995, there has been a very steep growth in the number of private jets in fractional ownership status.

2.57. Step 1: Identify the two variables of interest
In this situation, there are two variables, fuel consumption per hour, the dependent variable, and passenger capacity, the independent variable.

Step 2   Identify the dependent and independent variables.

The analyst is attempting to predict passenger capacity using fuel consumption per hour. Therefore, the capacity is the dependent variable and the fuel consumption per hour is the independent variable.
Step 3: Establish the scales for the vertical and horizontal axes
The y variable (fuel consumption) ranges from 631 to 3,529 and the x variable (passenger capacity) ranges from 78 to 405.

Step 4: Plot the joint values for the two variables by placing a point in the x,y space shown as follows:

Scatter Plot
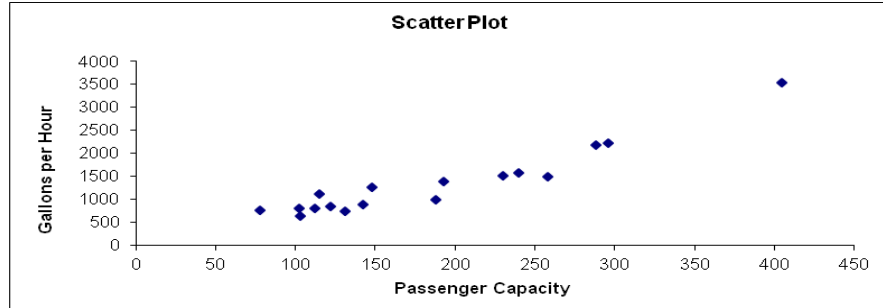
Based on the scatter diagram we see there is a strong positive linear relationship between passenger capacity and fuel consumption per hour.

2.58. Step 1: Identify the time-series variable

In this case, there are seven variables of interest. These are the daily sales for each of the bread types

Step 2: Layout the Horizontal and Vertical Axis

The horizontal axis will be the day and the vertical axis is the number of loaves of bread that were sold.

Step 3: Plot the values on the graph and connect the points.

Japolli Bakery Bread Sales
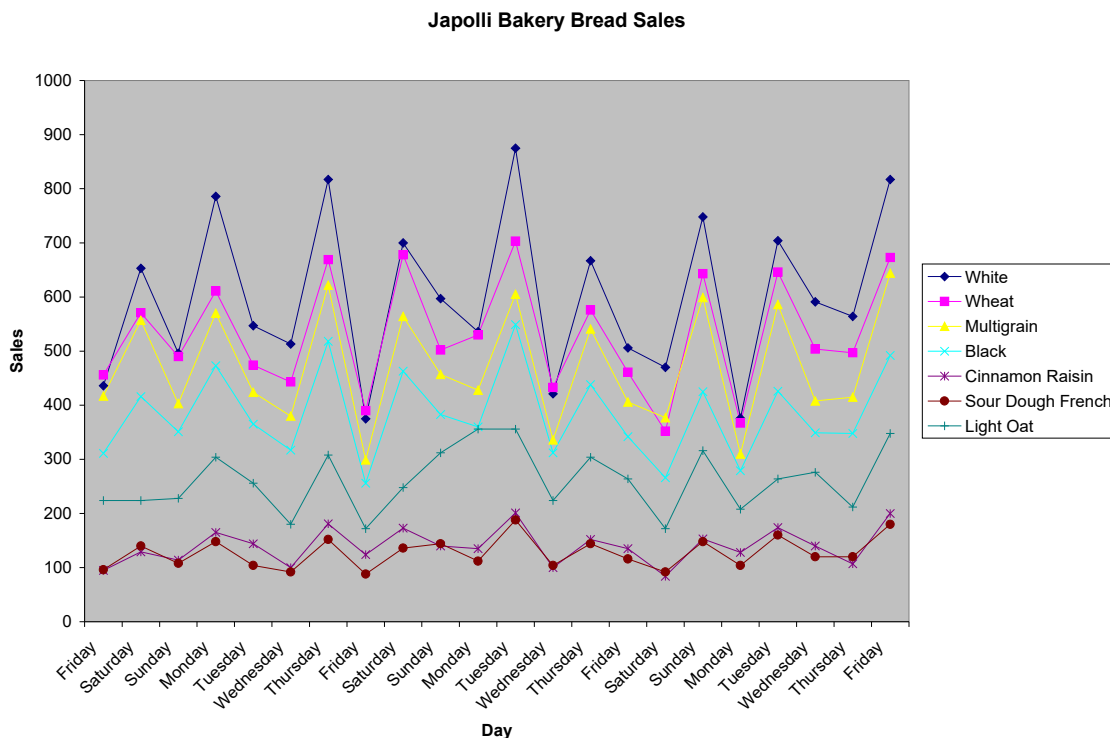
The graph illustrates a general pattern in the bread sales. Higher sales tend to occur for all types of bread on Saturdays, Mondays and Thursdays with Fridays typically the lowest.

2.59

    a. Step 1: Identify the time-series variable

          The variable of interest is annual average price of gasoline in California

        Step 2: Layout the Horizontal and Vertical Axis

          The horizontal axis will be the year and the vertical axis is average price (See Step 3)

        Step 3: Plot the values on the graph and connect the points

California Average Annual Gasoline Prices
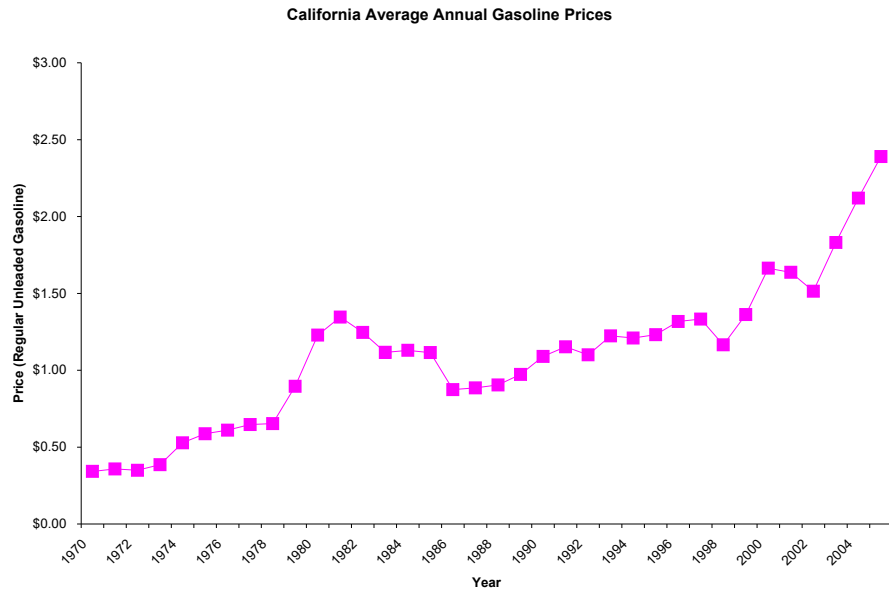
Gasoline prices have trended upward over the 36 year period with some short periods of decline. However, prices rises have been very steep since 1999.

b. Adding the inflation adjusted prices to the graph does not require that we use a different scale. The results of adding the second time-series is shown as follows:

California Average Annual Gasoline Prices
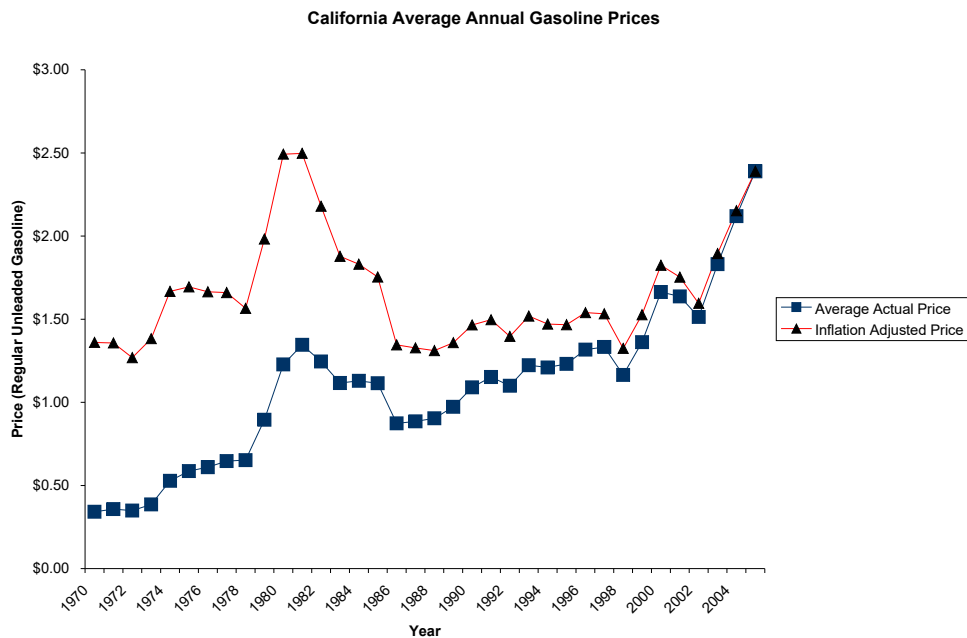
c. The graph in part b. shows an interesting result. That is, although the price of gasoline has risen quite steadily since 1970, when the value of the dollar is taken into account, the overall trend has been more level. In fact, the highest prices (when the inflation index is considered) occurred in 1980 and 1981 at the equivalent of slightly more than $2.50. This exceeds the prices during the 2000-2005 years. Thus, while gasoline prices were high in California in 2005, it is not the worst that has occurred in that state.

2.60

a.

Federal Flood Insurance
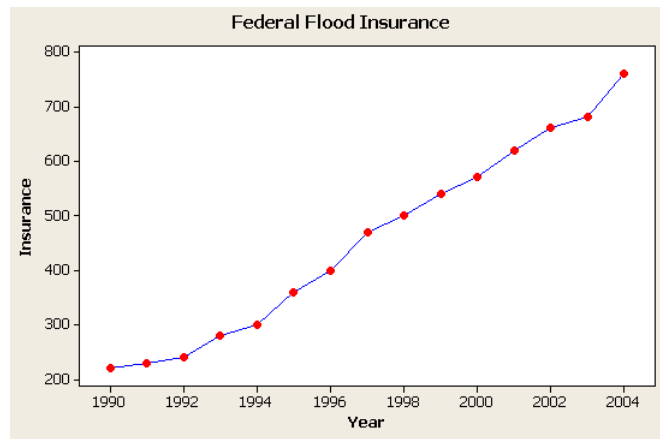
b. The relationship appears to be linear and positive.
c. The average equals the sum divided by the number of data points = 6830/15 = 455.33.

2.61

a.

Patents and Backlog
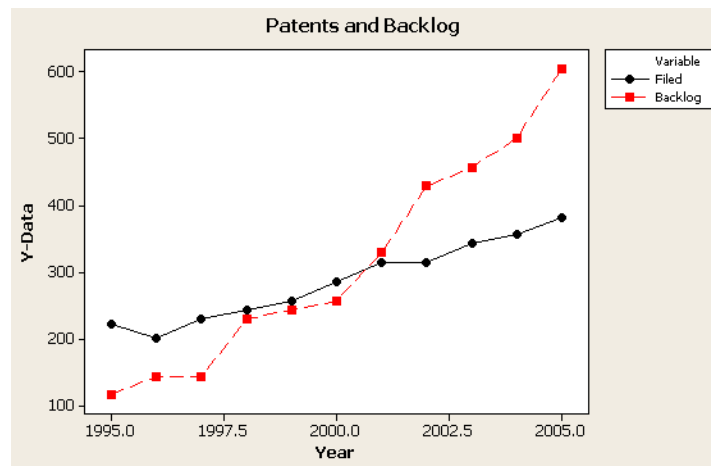
b. Both relationships seem to be linear in nature.
c. This occurred in 1998, 1999, and 2001.

2.62

a.

Time Series Plot of Home Sales

b.

Sales vs Median Price
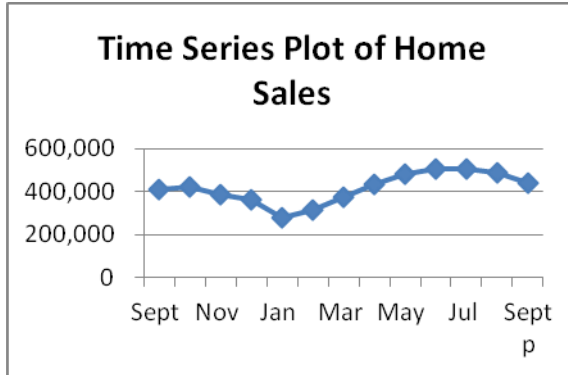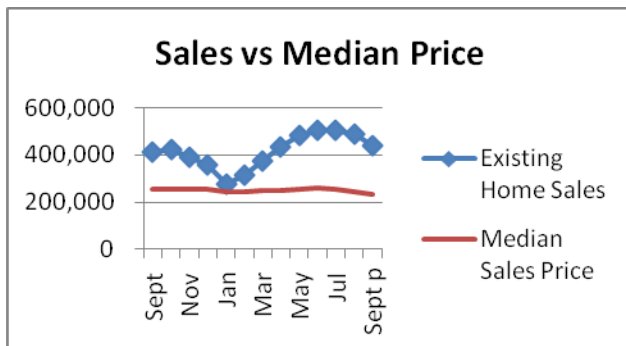
Note: you must convert the sales data into tens of thousands which is obtained by (Sales X 100), e.g., 6.76 million becomes 6.76(100) = 676 (tens of thousands). This is done to produce comparable Y axes for the two line plots.

c. It appears from the line plots that the monthly sales have been fluctuating greatly during this period, dipping in January, heading back up during the summer months and then declining again. Median sales price has shown a steady minor decline during the period.

End of Chapter

2.63 A relative frequency distribution deals with the percentage of the total observations that fall into each class rather than the number that fall into each class. Sometimes decision makers are more interested in percentages than numbers. Politicians, for instance, are often more interested in the percentage of voters that will vote for them (more than 50%) than the total number of votes they will get. Relative frequencies are also valuable when comparing distributions from two populations that have different total numbers.

2.64 Thinking in terms of the types of data discussed in chapter 1, that is nominal, ordinal, interval and ratio, bar charts are visual representations of frequency distributions constructed from nominal or ordinal data.

2.65 Pie charts are effectively used when the data set is made up of parts of a whole, and therefore each part can be converted to a percentage. For instance, if the data involves a budget, a pie chart can represent the percentage of budget each category represents. Or, if the data involves total company sales, a pie chart can be used to represent the percentage contribution to sales for each major product line.

2.66 A line chart is an effective tool to represent the relation between a dependent and an independent variable when values of the independent variable form a natural increasing sequence. In many cases this means the independent variable is a measure of time and the data is time-series data. With a scatter plot the values of the independent variable are determined randomly and not according to a preset sequence.

2.67

    a.
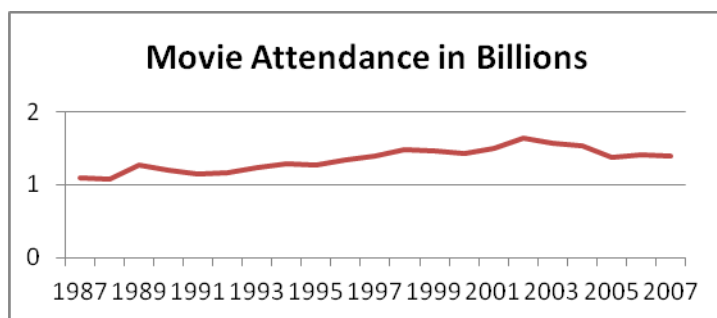
    b. It appears that there is a positive linear relationship between the attendance and the year. However, there does appear to be a sharp decline in the last five years. It could be evidence of a normal cycle since a similar decline occurred in 1990/91 which was followed by a steady climb in attendance for six of the next seven years.

2.68

  a. Using the $2^k \geq n$ guideline:
$$2^k \geq 48 = 2^6 \geq 48$$

To determine the class width, $(17.5 - 0.3)/6 = 2.87$ so round up to 3 to make it easier.
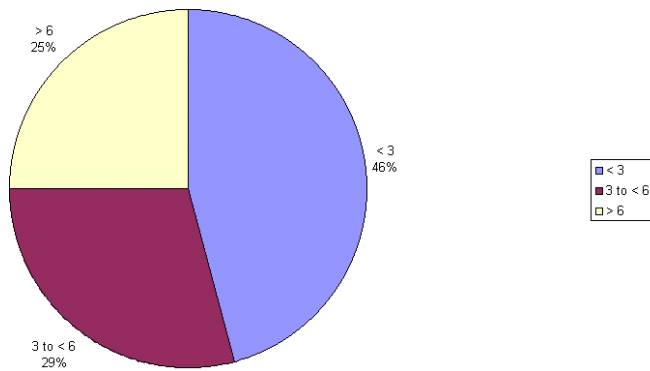
Classes	Frequency
0.1 to 3	27
3.1 to 6.0	9
6.1 to 9.0	6
9.1 to 12	4
12.1 to 15.0	0
15.1 to 18.0	2

b.

Stem-and-Leaf Display		
Stem unit: 1		
0	3 4 5 7	
1	0 0 0 0 4 5 5 9	
2	0 0 0 0 0 4 5 5 5 7	
3	0 0 0 0 0 2 5 5 5 5 6	
4	0 0 0	
5		
6	4 5	
7	5	
8	3	
9	0 0 2	
10		
11	0	
12	0 0	
13		
14		
15		
16	0	
17	5	

c.

Pie Chart - Miles

d.

Bar Chart - Proportion of Employees
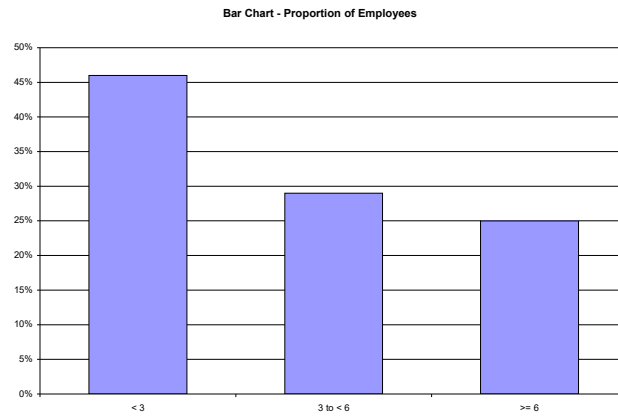
2.69

a.
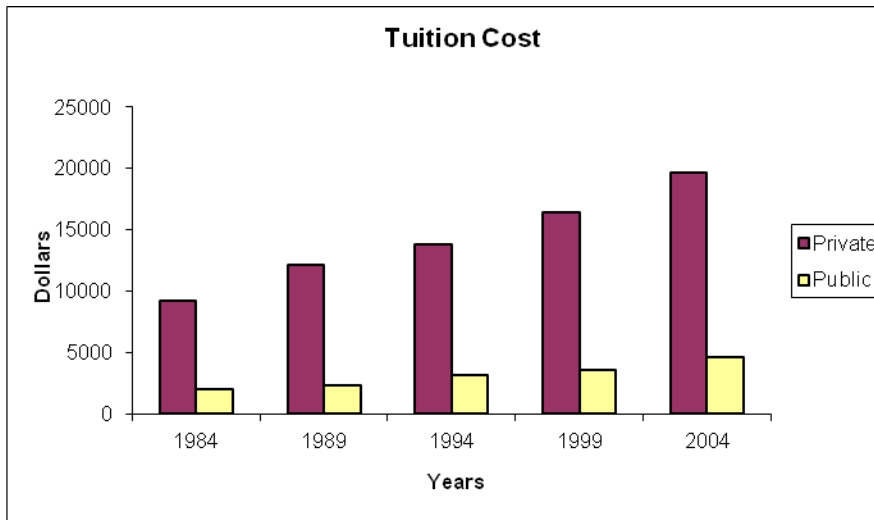
b. Student answers will vary but should include identifying that both private and public college tuition costs have more than doubled in the 20 years of data.

2.70 a.

College Degrees Gender Percents
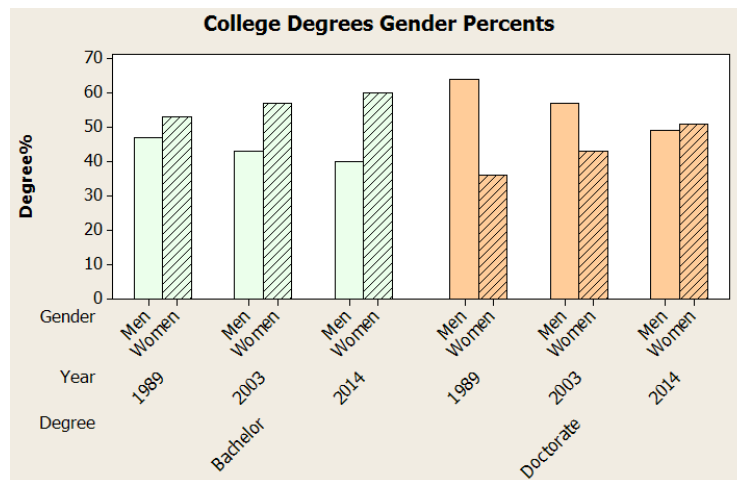
b. In both cases there appears to be a linear decline in the percentage of men receiving the specific degree and an incline in the percentage of women receiving the specific degree.

2.71
a. The frequencies can be calculated by multiplying the relative frequency times the sample size of 1,000.

Class Length (Inches)	Frequency	Relative Frequency
8 < 10	220	0.22
10 < 12	150	0.15
12 < 14	250	0.25
14 < 16	240	0.24
16 < 18	60	0.06
18 < 20	50	0.05
20 < 22	30	0.03

Frequency Distribution of Walleyes
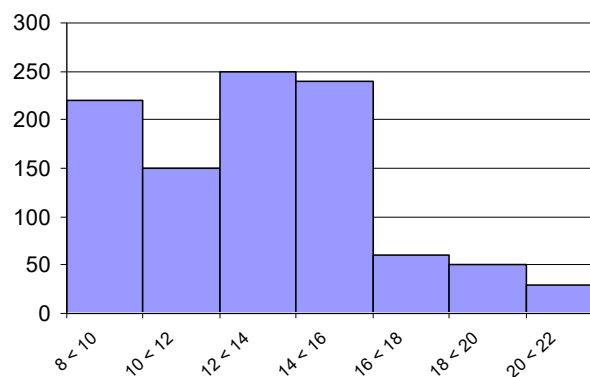
b. The histogram is probably a better representation of the fish length data.

Pie Chart of Walleyes
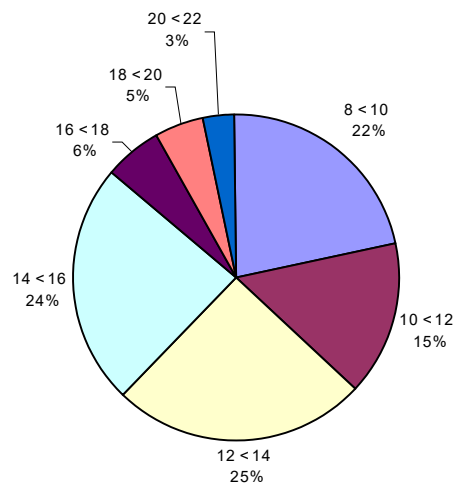
2.72

a. Based upon the following table the percent of class that hold at least 120 seconds (2 minutes) is

$$0.0311 + 0.0244 + 0.0171 + 0.0301 = 0.1029$$

Classes (in seconds)	Number	Rel. Freq.
< 15	456	0.0899
15 < 30	718	0.1415
30 < 45	891	0.1756
45 < 60	823	0.1622
60 < 75	610	0.1202
75 < 90	449	0.0885
90 < 105	385	0.0759
105 < 120	221	0.0435
120 < 150	158	0.0311
150 < 180	124	0.0244
180 < 240	87	0.0171
≥ 240	153	0.0301

Ogive
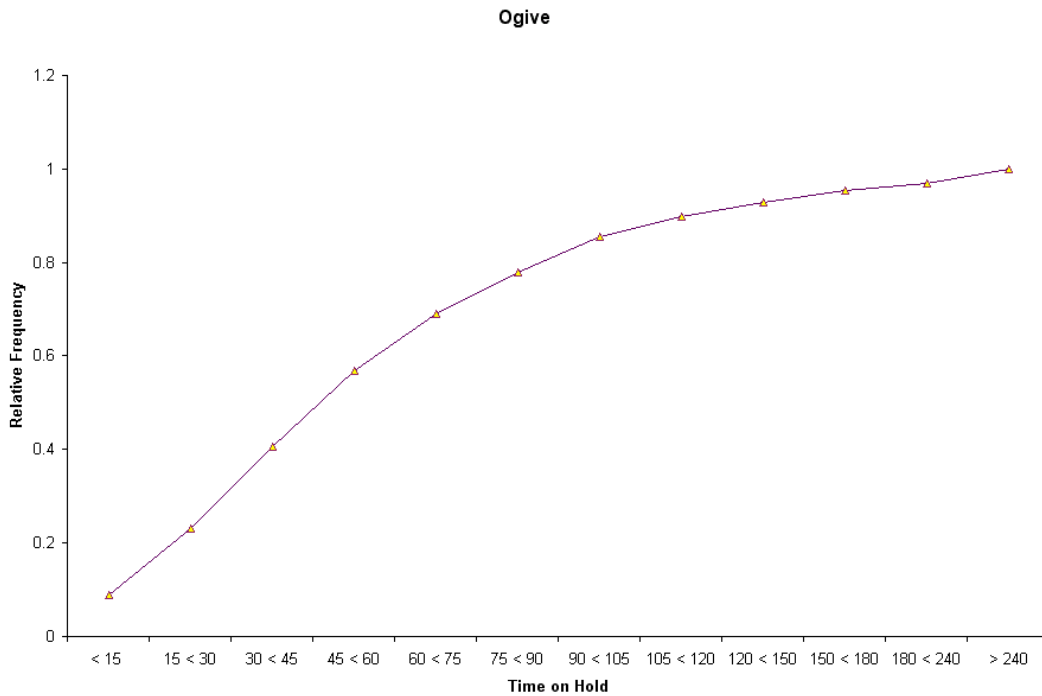
Note: For this problem the class widths are not equal.

b.   The number of people who have to wait 120 seconds (2 minutes) or more is
155 + 124 + 87 + 153 = 522 * $30 = $15,660 month.

2.73
a.   The independent variable is hours and the dependent variable is sales

Scatter Plot of Hours and Sales
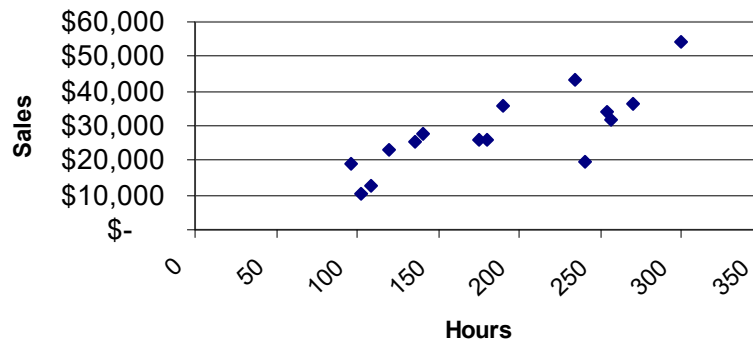
b.   It appears that there is a positive linear relationship between hours worked and weekly sales. It appears that the more hours worked the greater the sales. No stores seem to be substantially different in terms of the general relationship between hours and sales.

2.74

a.
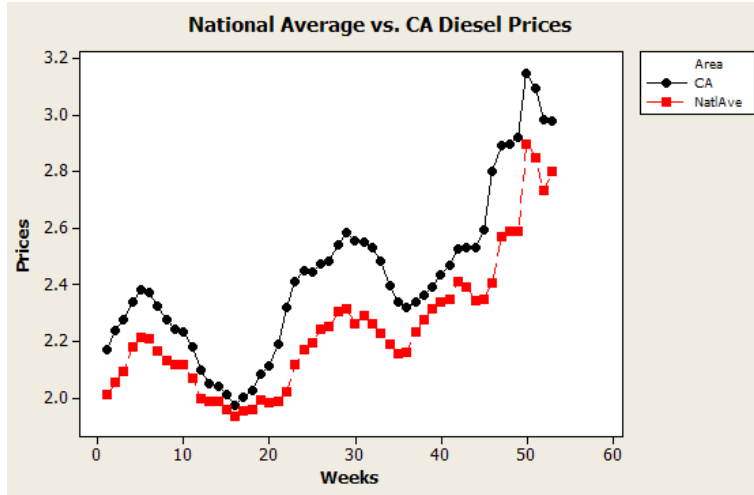
The relationship between diesel prices in CA versus the national average seems to be that CA prices are unanimously higher than the national average and has the structural relationship to the weeks the prices were recorded. That relationship seems to be curvilinear in nature.

b. The time the average diesel price became larger than $3 per gallon ($3.149) was during the 50th week.

c. The smallest price occurred in week 16 and highest price in week 50. So there was 50 - 16 = 34 weeks in which the prices rose 3.149 – 1.973 = 1.176. That is an average rise of 1.176/34 = 0.0346 dollar/week increase. The difference between $4.00 and $3.149 is 0.851. This means that the number of weeks required to reach $4.00 would be 0.851/0.0346 = 24.6 ≈ 25 weeks. Thus, the week in which $4.00 is achieved should be the 75th (50 + 25) week.

2.75
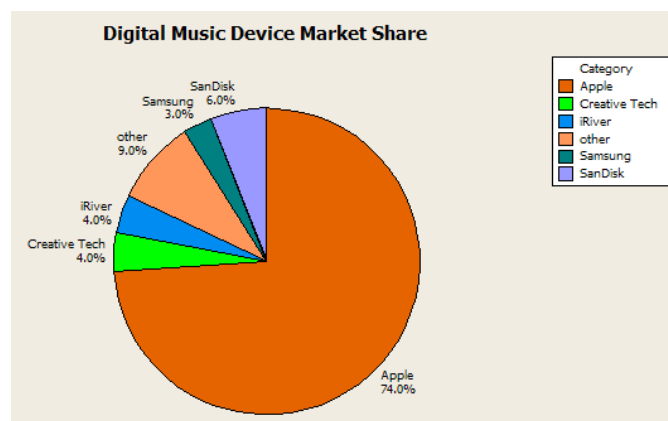
2.76

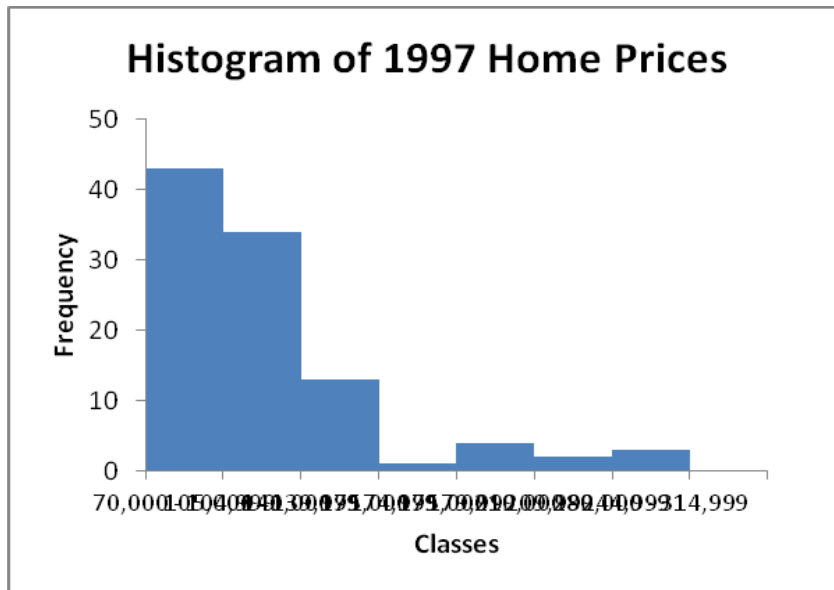 a. Using the $2^k \geq n$

$$2^k \geq 100 \text{ so } 2^7 = 128$$

Determine the width: $(310496 - 70464)/7 = 24{,}290$. Round to 35,000

Classes	Frequency
70,000 - 104,999	43
105,000 - 139,999	34
140,000 - 174,999	13
175,000 - 179,999	1
175,000 - 209,999	4
210,000 - 244,999	2
280,000 - 314,999	3

Histogram of 1997 Home Prices

b.

| Classes	Frequency	Relative Frequency	Cumulative Relative Frequency
70,000 - 104,999	43	0.43	0.43
105,000 - 139,999	34	0.34	0.77
140,000 - 174,999	13	0.13	0.90
175,000 - 179,999	1	0.01	0.91
175,000 - 209,999	4	0.04	0.95
210,000 - 244,999	2	0.02	0.97
280,000 - 314,999	3	0.03	1.00
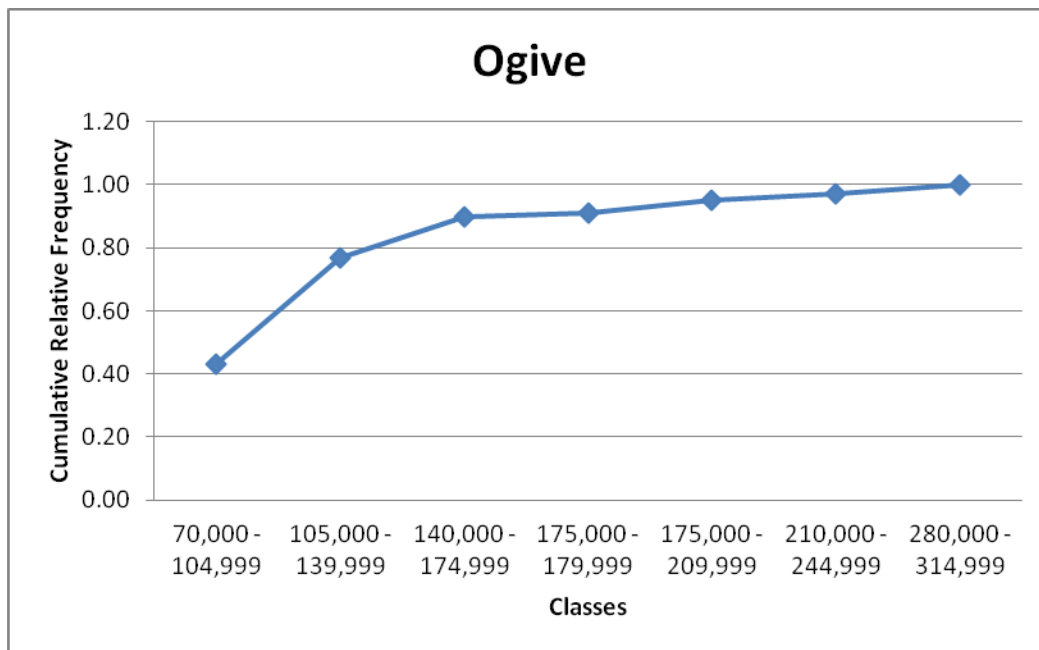
c. Using four classes
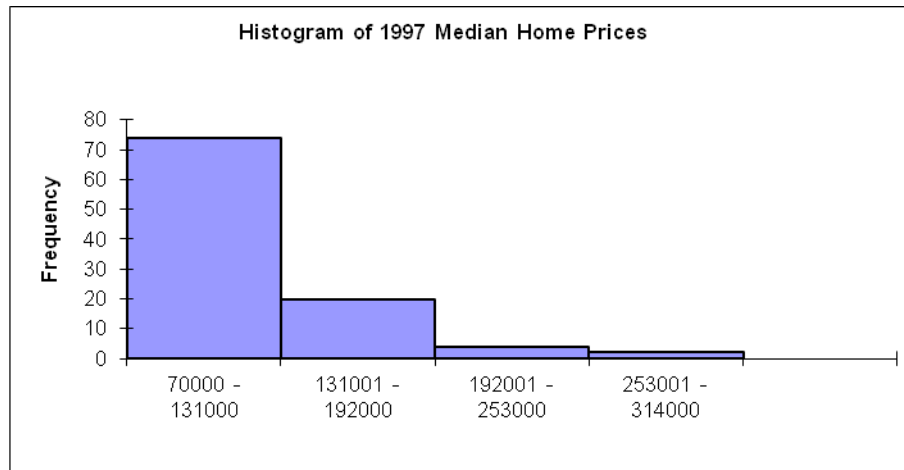
Distribution with 4 classes appears to be more skewed than when 7 classes are used.
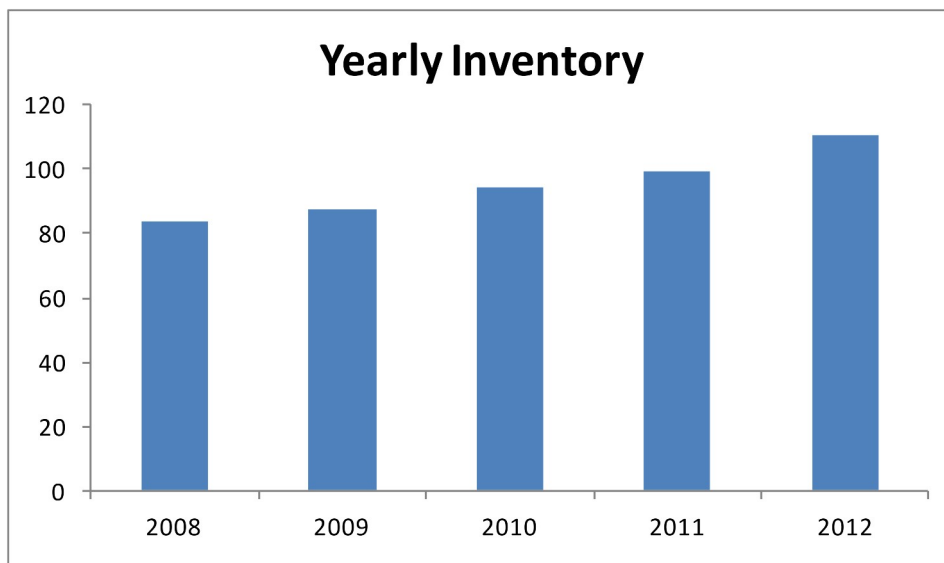Less detail is available.

Cumulative Frequency Ogive
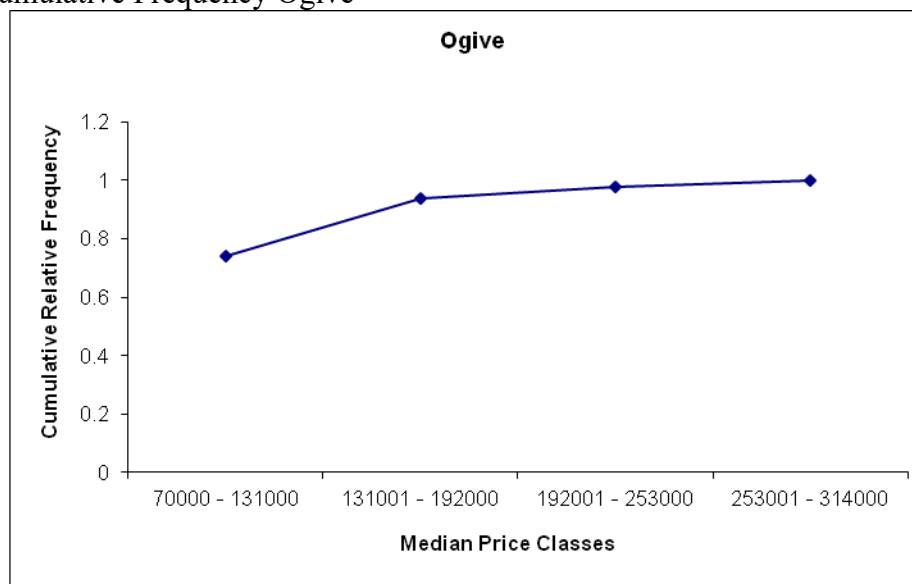
2.77
a.
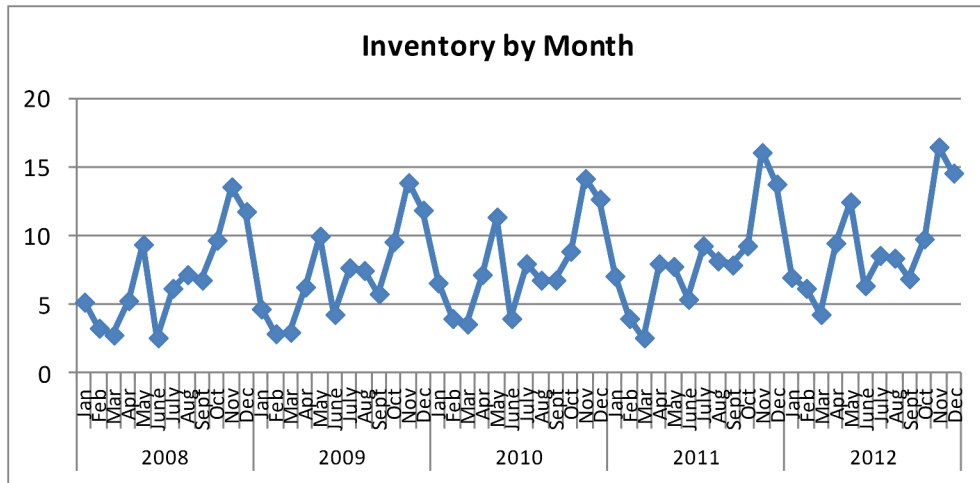
Inventory has been trending slightly up over the five years, but appears to be highly seasonal with predictable highs at certain points each year.
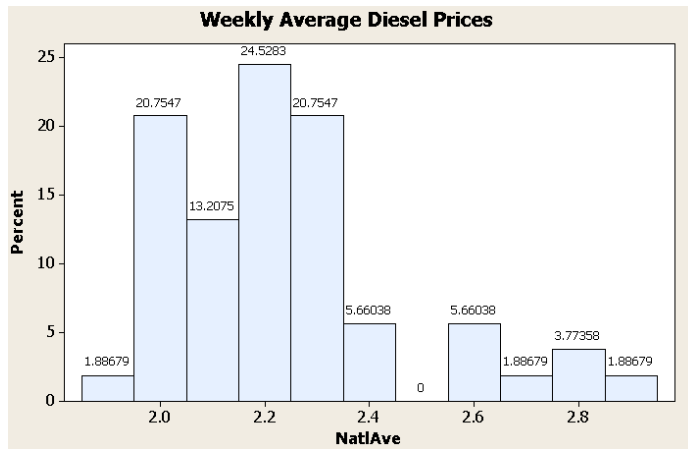
b.

This bar chart is effective for showing the growth in total annual inventory over the five years. However, students should keep in mind that the sum of monthly inventory does not equate to how much inventory the store had on hand at the end of the year. Students might question why the store would graph the total inventory

2.78

a.

Weekly Average Diesel Prices

b. Notice that the class interval with no observations is the 7th class which has boundaries of $2.45 and $2.55. Since the numbers are averages taken across the United States, it is possible that the sampling technique, simply from randomness, didn't select prices in that range. Another possible explanation is that there are retailers in certain locations that generally charge ten cents a gallon more. It is a commonly held belief that California retailers charge more than the national average. This could be the reason. It bears further investigation.