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.

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:

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

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:

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

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 $0.4 - 0.2 = 0.2$. Ofcourse, 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 $2^k > n$ 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: $w = (17-0)/66 = 2.833$, 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.
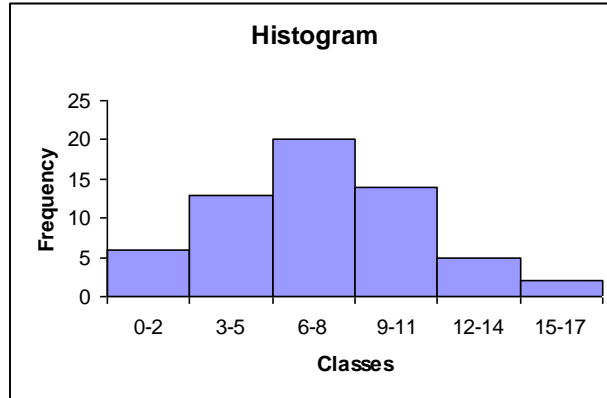
Class	Frequency	Relative Frequency
0–2	6	0.100
3–5	13	0.217
6–8	20	0.333
9–11	14	0.233
12–14	5	0.083
15–17	2	0.033
	Total = 60	

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.   Equal class width $= \dfrac{12-4}{2} = 4$

So the first class upper boundary $= 4 + 4 = 8$, and we can complete the rest of the missing figures for the classes.

Then, use the following steps in the distribution to complete the table:

Class	Frequency	Cumulative Frequency	Relative Frequency
$4 - < 8$	**Step 1:** Same as CF = 5	5	**Step 8:** $5 \div 50 = 0.1$
$8 - < 12$	**Step 2:** $50(0.34) = 17$	**Step 5:** $5 + 17 = 22$	0.34
$12 - < 16$	7	**Step 6:** $22 + 7 = 29$	**Step 9:** $7 \div 50 = 0.14$
$16 - < 20$	**Step 3:** $50(0.18) = 9$	**Step 7:** $29 + 9 = 38$	0.18
$20 - < 24$	**Step 4:** $50 - (5+17+7+9)= 12$	50	**Step 10:** $12 \div 50 = 0.24$

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

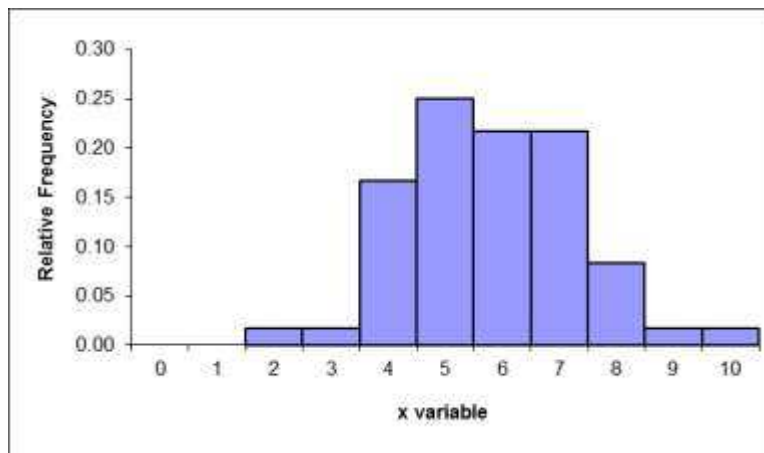| x	Frequency
0	0
1	0
2	1
3	1
4	10
5	15
6	13
7	13
8	5
9	1
10	1

b. The following histogram could be developed.
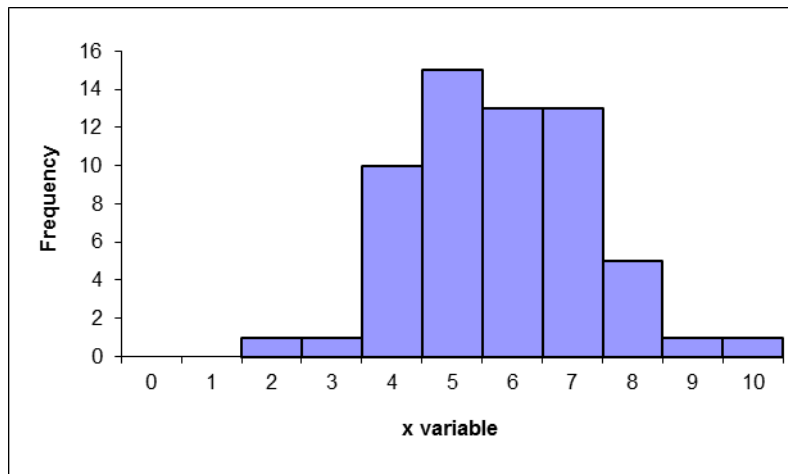
c. The relative frequency distribution shows the fraction of values falling at each value of *x*.

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.

2.8. 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:
    $2^k \geq 60$ so $2^6 \geq 60$

Class width is:

$$W = \frac{\text{Maximum} - \text{Minumum}}{\text{Number of Classes}} = \frac{10-2}{6} = 1.33$$

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	Cumulative Relative Frequency
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:
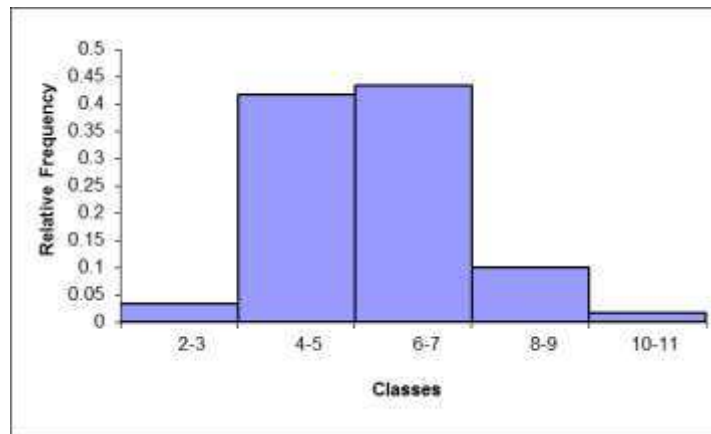
d. The ogive is a graph of the cumulative relative frequency distribution.
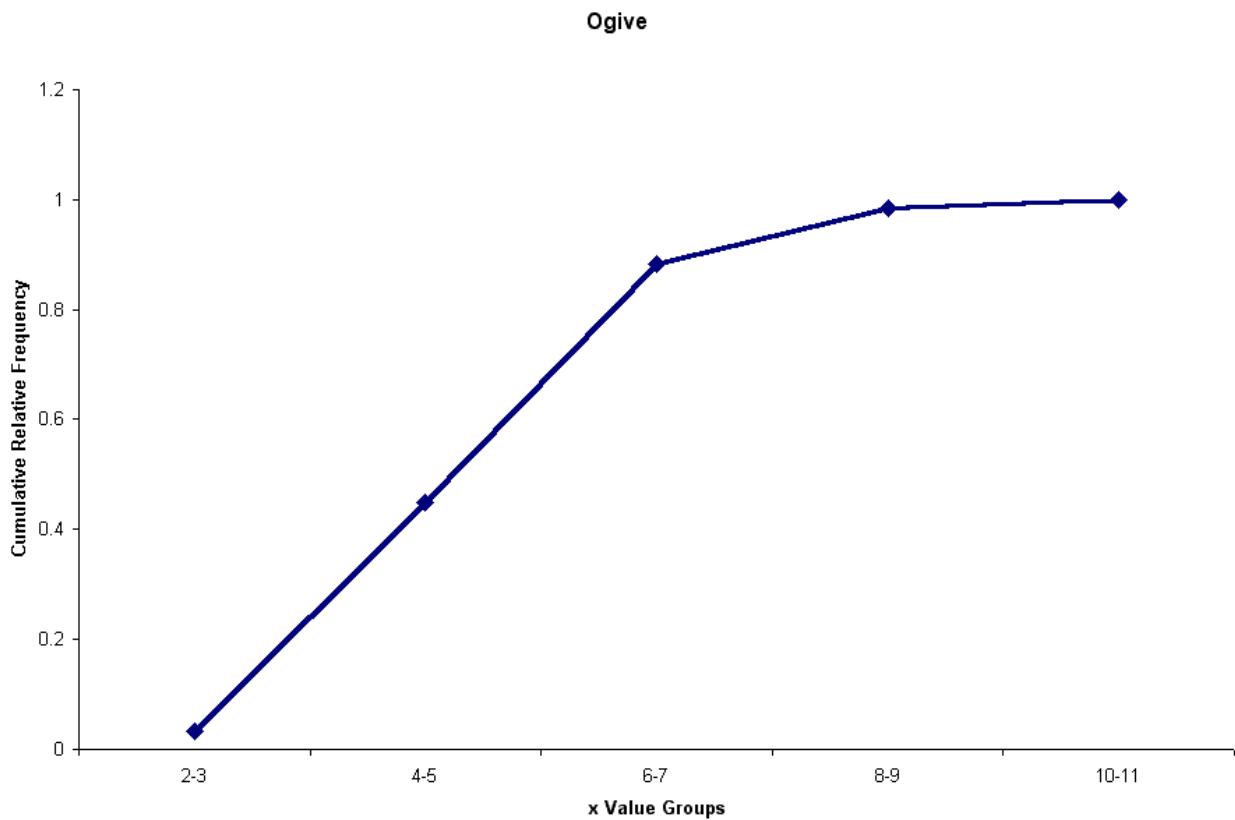
2.9. a. Step 1: Determine the number of classes, $2^k \geq n$

$2^6 = 64$, which is greater than $n = 61$; so $k = 6$

Step 2: Determine the class width,

$$W = \frac{max - min}{k} = \frac{29 - 4}{6} = 4.1667 \text{ (Round up to 5)}$$
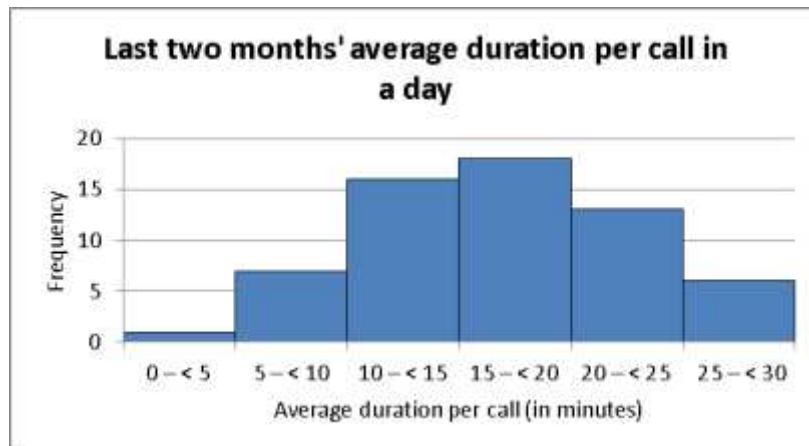
Step 3: Define the class boundaries:

Average duration per call (in minutes)
$0 - < 5$
$5 - < 10$
$10 - < 15$
$15 - < 20$
$20 - < 25$
$25 - < 30$

Step 4: Determine the class frequency for each class.

Average duration per call (in minutes)	Frequency
$0 - < 5$	1
$5 - < 10$	7
$10 - < 15$	16
$15 - < 20$	18
$20 - < 25$	13
$25 - < 30$	6

The most concentration of the distribution is from class $10 - < 15$ to $20 - < 25$.

b.

2.10. a. Joint frequency distribution:

Type of toys	1st Week	2nd Week	Total
Soft toys	9	21	30
Educational toys	17	13	30
Physical activity toys	8	17	25
Total	34	51	85

b. Joint relative frequency distribution:

Type of toys	1st Week	2nd Week	Total
Soft toys	0.11	0.25	0.36
Educational toys	0.20	0.15	0.35
Physical activity toys	0.09	0.20	0.29
Total	0.40	0.60	1.00

c. In the first week, the best-selling toys are the educational toys with a total sales of 17. In the second week, the best-selling toys are the soft toys with a total sales of 21.

d. $0.15 + 0.2 = 0.35 = 35\%$; thus, the educational toys and physical activity toys sold in the second week constitute 35% of the total toy sales.

2.11. a. The two distributions are:

Housing Code	Frequency	Relative Frequency	Satisfaction	Frequency	Relative Frequency
A	8	0.4	1	4	0.2
B	12	0.6	2	8	0.4
Total	20	1.0	3	8	0.4
			Total	20	1.0

b. The joint frequency distribution is:

Housing Code	Satisfaction 1	Satisfaction 2	Satisfaction 3	Total
A	1	3	4	8
B	3	5	4	12
Total	4	8	8	20

c. The joint relative frequency distribution is:

Housing Code	Satisfaction 1	Satisfaction 2	Satisfaction 3	Total
A	0.05	0.15	0.2	0.4
B	0.15	0.25	0.2	0.6
Total	0.2	0.4	0.4	1.0

The distribution shows that most of the customers who favored the apartment were satisfied with the newly launching apartment. Only a minority of customers who selected the Tower Block as their preferred housing showed low satisfaction

2.12. 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.
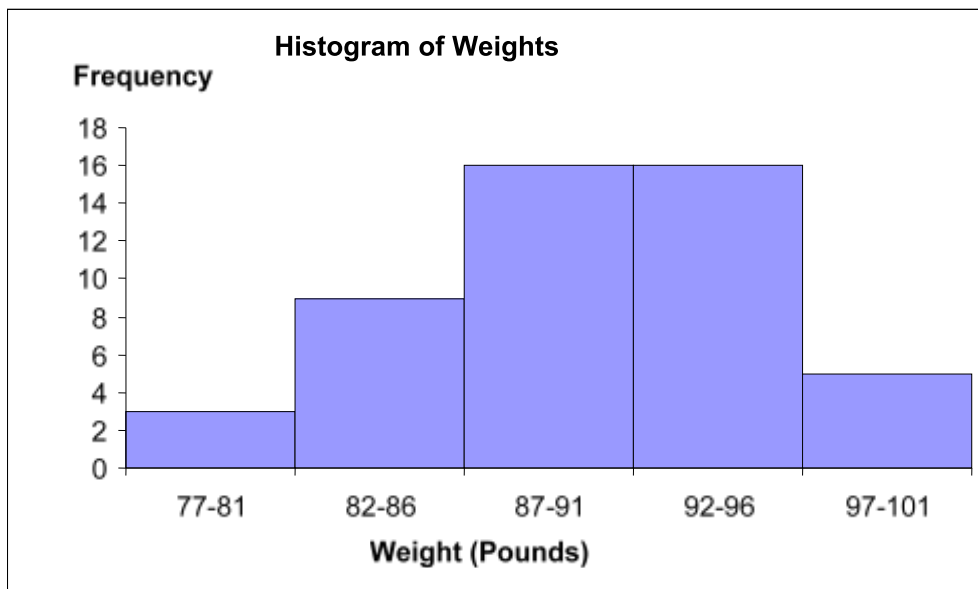
d. Convert the frequency distribution into relative frequencies and cumulative relative frequencies as shown below.

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%.

2.13. a. Use $2k \geq n$ to determine the number of classes.
   $27 = 128$, which is greater than $n = 50$; so the number of classes, $k = 7$

b. Determine the class width,

$$W = \frac{\text{max} - \text{min}}{k} = \frac{4,200 - 2,145}{7} = 293.57 \text{ and round up to } 300$$

c. The cumulative relative frequency distribution:

Required Salary	Frequency	Relative Frequency	Cumulative Relative Frequency
2,100 – < 2,400	2	0.04	0.04
2,400 – < 2,700	10	0.2	0.24
2,700 – < 3,000	12	0.24	0.48
3,000 – < 3,300	11	0.22	0.7
3,300 – < 3,600	5	0.1	0.8
3,600 – < 3,900	7	0.14	0.94
3,900 – < 4,200	3	0.06	1
Total	50	1	

d.
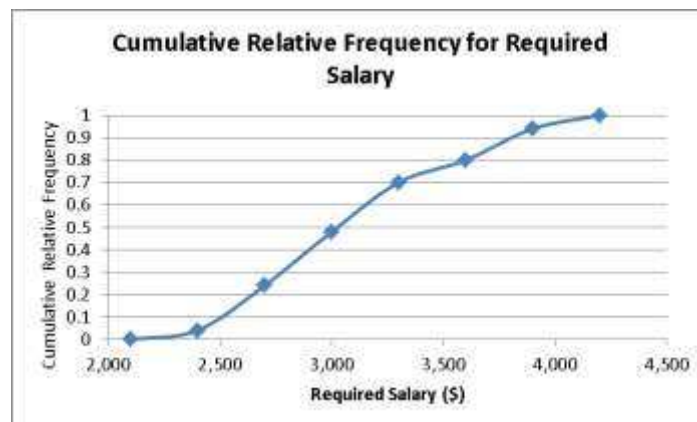
Based on either the distribution in part c. or the above graph, the manager will interview those whose salary requirement is not more than $3,300.

2.14. a. Sort the customers' age:

13	17	18	19	20	21	22	23	23	26
27	27	28	29	33	36	36	36	37	37
37	39	41	44	45	45	48	51	52	52

b. Given k = 5, the class width, $W = \dfrac{max - min}{k} = \dfrac{52 - 13}{5} = 7.8$, and round up to 8.

The frequency distribution:

Age	Frequency
13 – < 21	5
21 – < 29	8
29 – < 37	5
37 – < 45	6
45 – < 53	6

c.
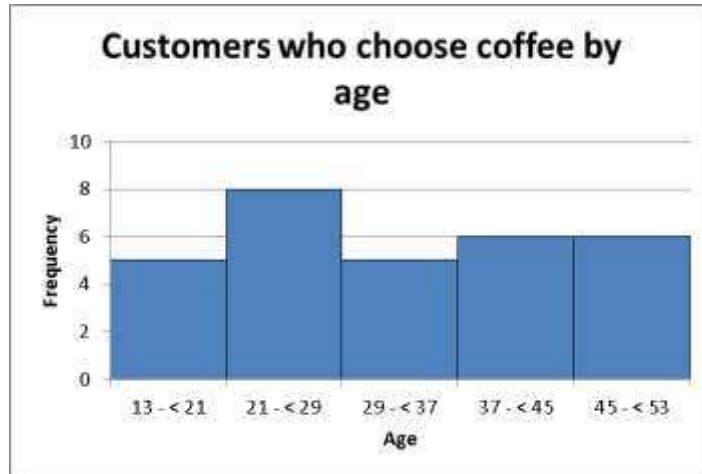
d.

Age	Frequency	Relative frequency	Cumulative Relative Frequency
13 – < 21	5	0.1667	0.1667
21 – < 29	8	0.2667	0.4333
29 – < 37	5	0.1667	0.6
37 – < 45	6	0.2	0.8
45 – < 53	6	0.2	1
Total	30	1.0000	

Thus, 43.33% of the café's customers who chose to drink coffee are less than 29 years old.

2.15.   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
3.46–3.60	3	0.067
3.61–3.75	6	0.133
3.76–3.90	16	0.356
3.91–4.05	14	0.311
4.06–4.20	6	0.133
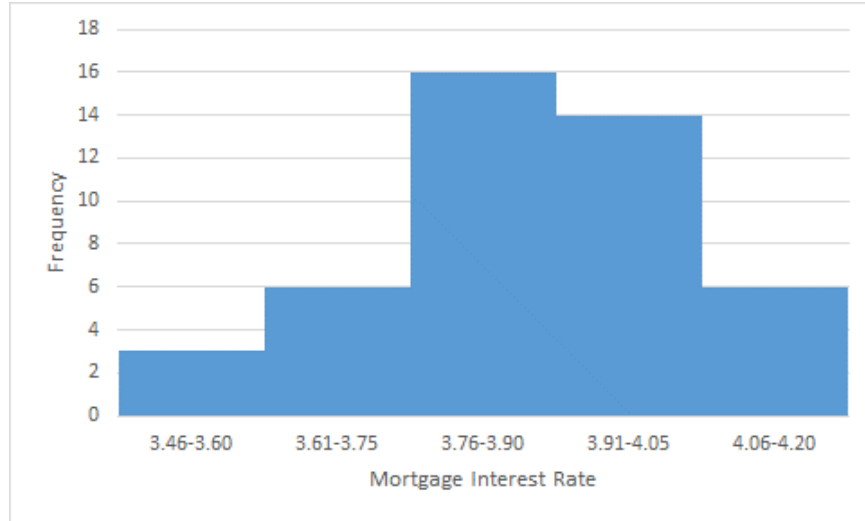
   b.   Proportion of rates that are at least 3.76% is the sum of the relative frequencies of the last four classes $= 0.356 + 0.311 + 0.133 = 0.800$

   c.
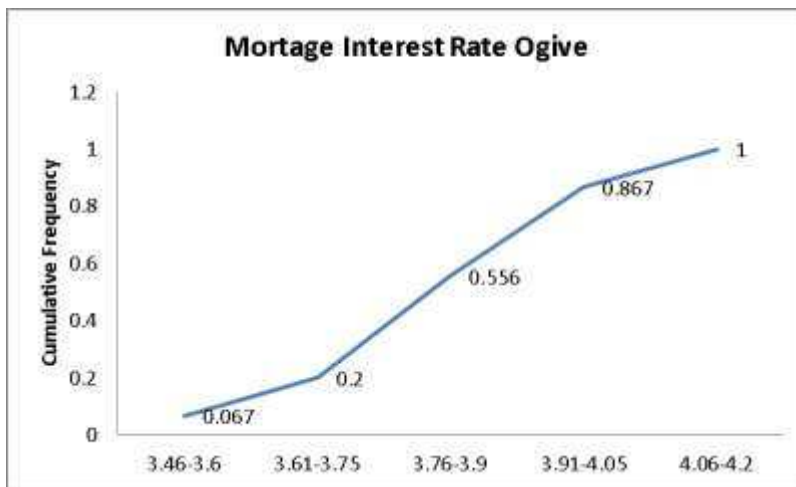
2.16. a.
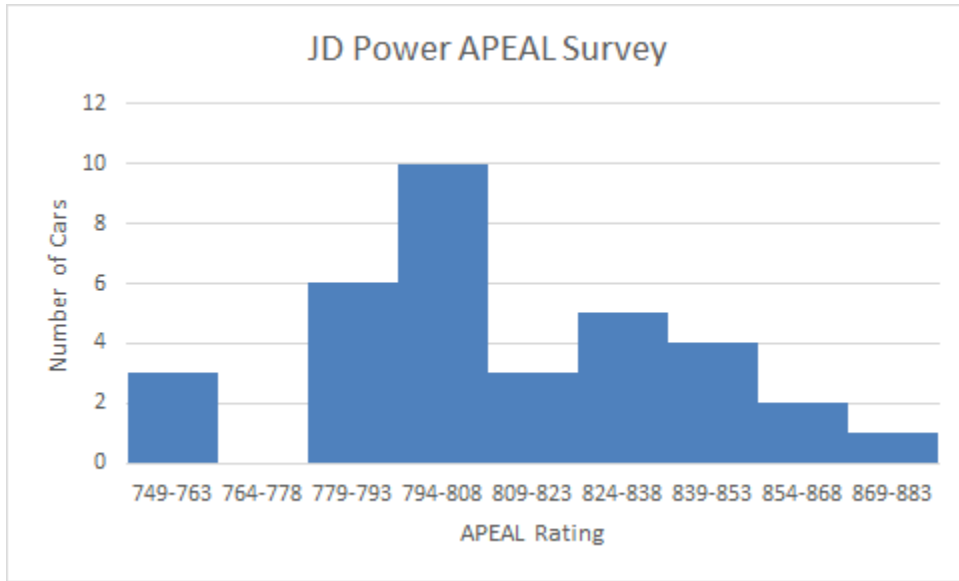
b.   The 2015 average is 798 which exceeds the 2008 average of 790. This could indicate that the new models are slightly more appealing on average to automobile customers, or customers could simply have reduced expectations.

2.17. 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
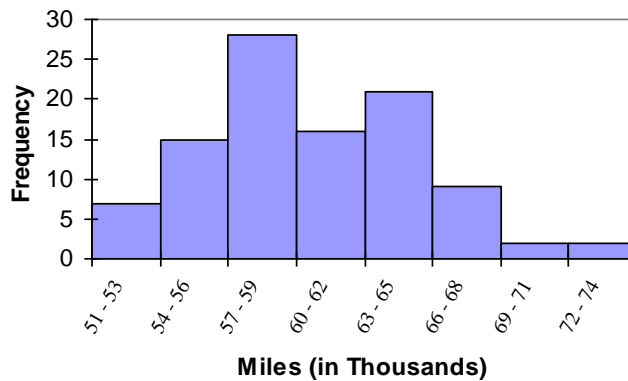
b.   The tread life of at least 50% of the tires is 60,000 or more. The top 10% is greater than 66,000 and the longest tread tire is 74,000. Additional information will vary.

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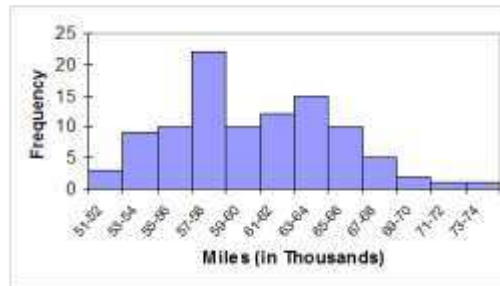
Students will probably say that the 12 classes give better information because it allows you to see more detail about the number of miles the tires can go.

2.18.   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:

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

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.

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

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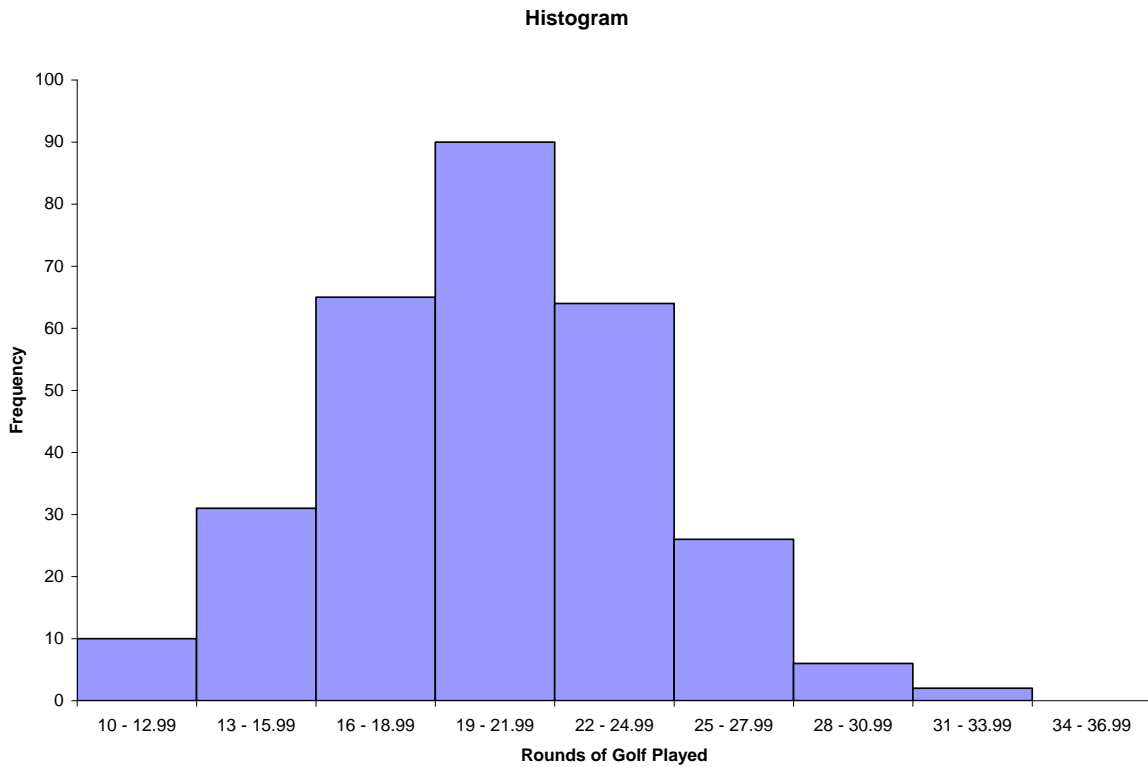
The distribution for rounds of golf played is mound shaped and fairly symmetrical. It appears that the center is between 19 and 22 rounds per year, but the rounds played is quite spread out around the center.

2.19.
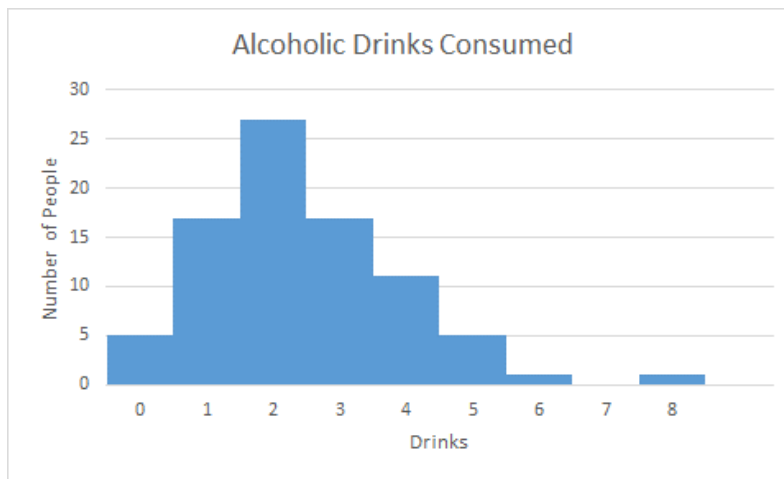
2.20.   a.   Using the $2^k \geq n$ guideline, the number of classes would be 6. There are 41 airlines. $2^5 = 32$ and $2^6 = 64$. Therefore, 6 classes are chosen.

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

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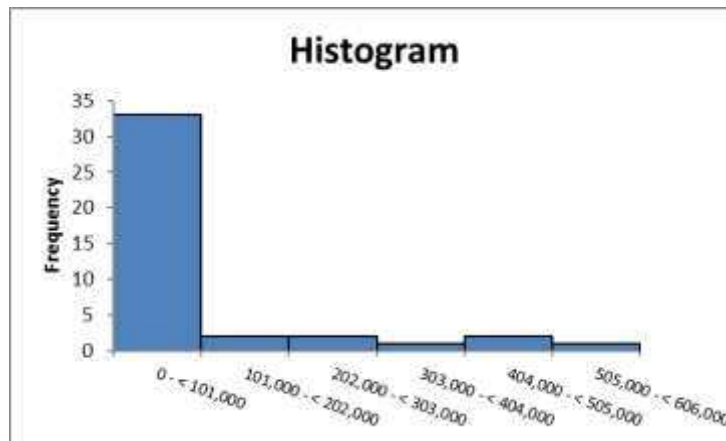
The vast majority of airlines had fewer than 101,000 monthly passengers for December 2011.

2.21. a. The frequency distribution is:

Satisfaction Level	Frequency
1 = Very Dissatisfied	1
2 = Dissatisfied	82
3 = Neutral	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 Service 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.

2.22. a. 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.
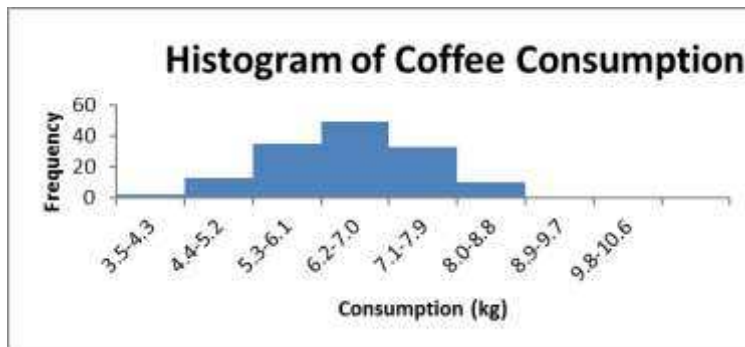
The histogram shows the shape of the distribution. This histogram is showing that fewer people consume small and large quantities and that most individuals consume between 5.3 and 8.0 kg of coffee, with the highest percentage of individuals consuming between 6.2 and 7.0.

b. 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

2.23. a. The pie chart is shown as follows:

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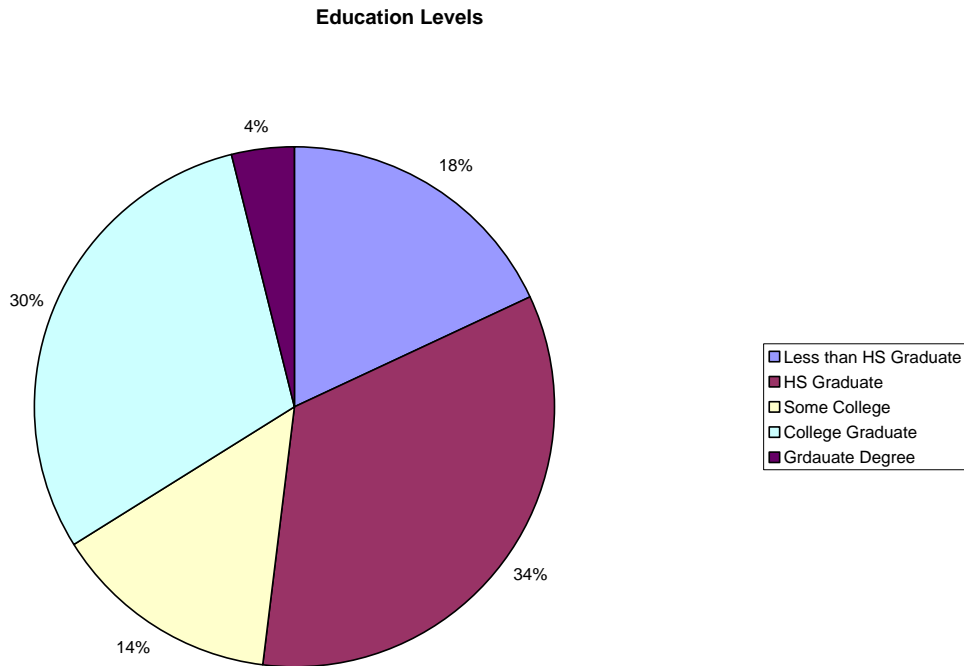
b. The horizontal bar chart is shown as follows:

Education Levels
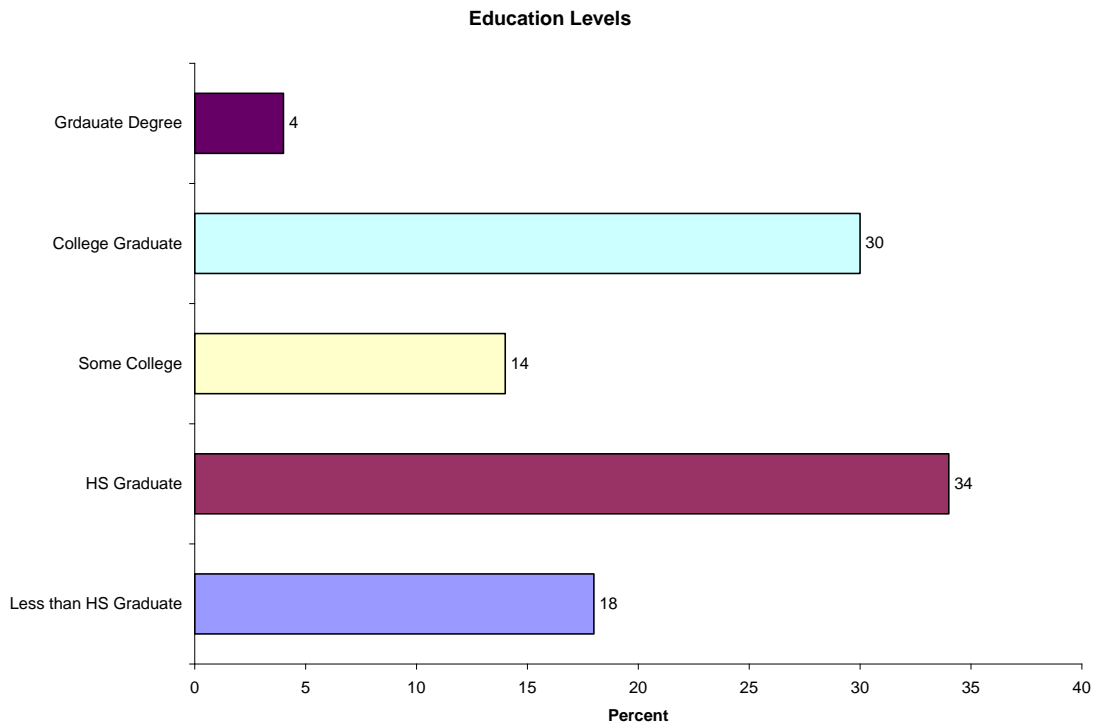
2.24.

Stem	Leaves
0	1, 2, 3, 3, 4, 5
1	0, 0, 0, 7, 9
2	6, 6, 7, 8
3	0, 1, 3, 3, 3, 6
4	3, 8, 9, 9, 9
5	0, 1, 3, 8

Key: 5/1 = 51 years old

2.25. 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
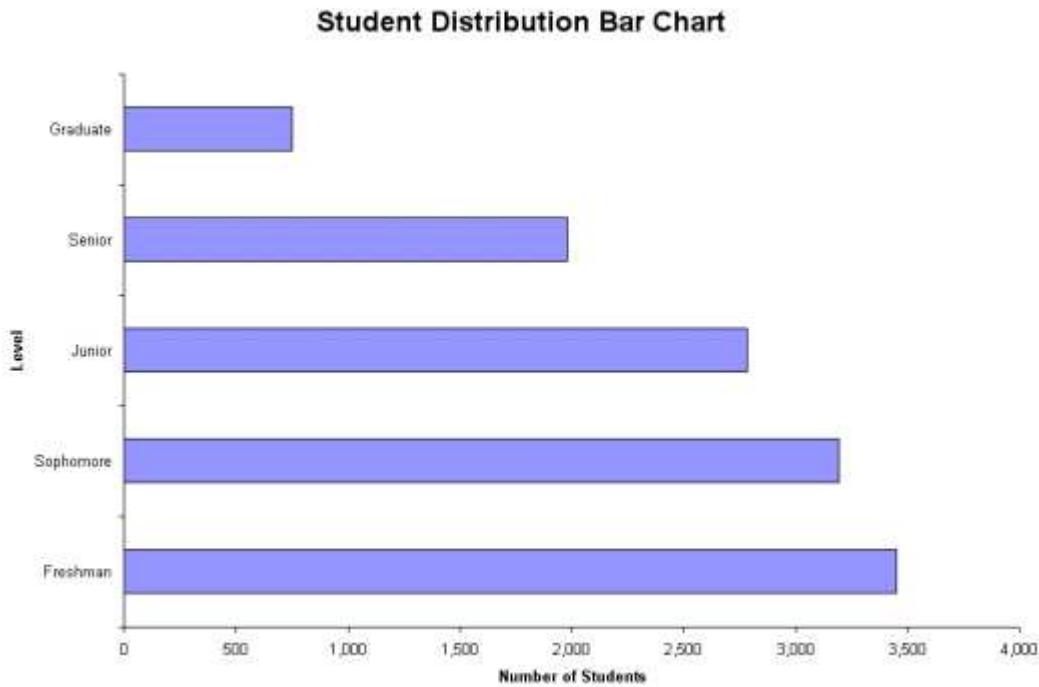
b. Step 1: Define the categories.

The categories are grade level.

Step 2: Determine the appropriate measure.

The measure is the number of students at each grade level.

Step 3: Develop the pie chart.

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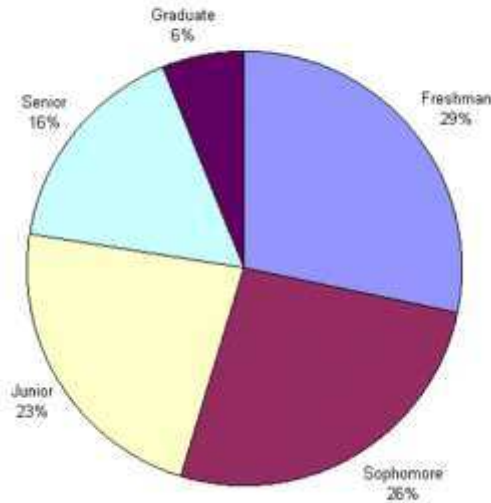
c.   A case can be made for either a bar chart or pie chart. Pie charts are especially good at showing how the total is divided into parts. The bar chart is best to draw attention to specific results. In this case, a discussion might be centered on the possible attrition that takes place in the number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

2.26.   One possible bar chart is shown as follows:
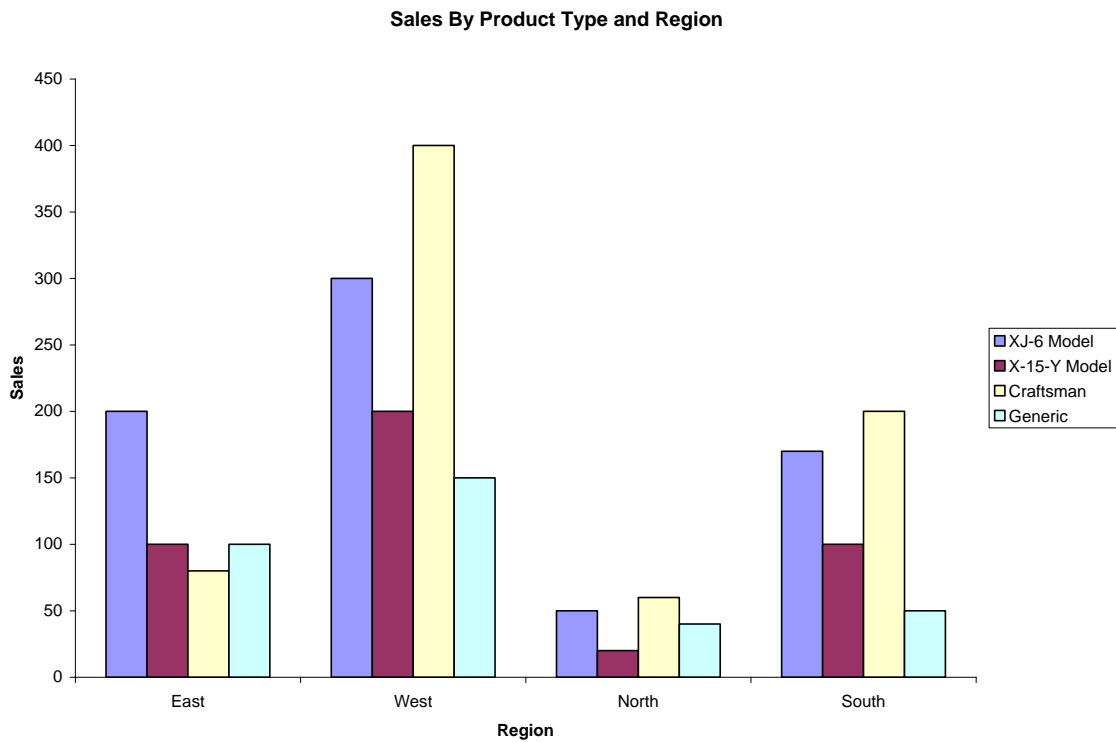
Another way to present the same data is:

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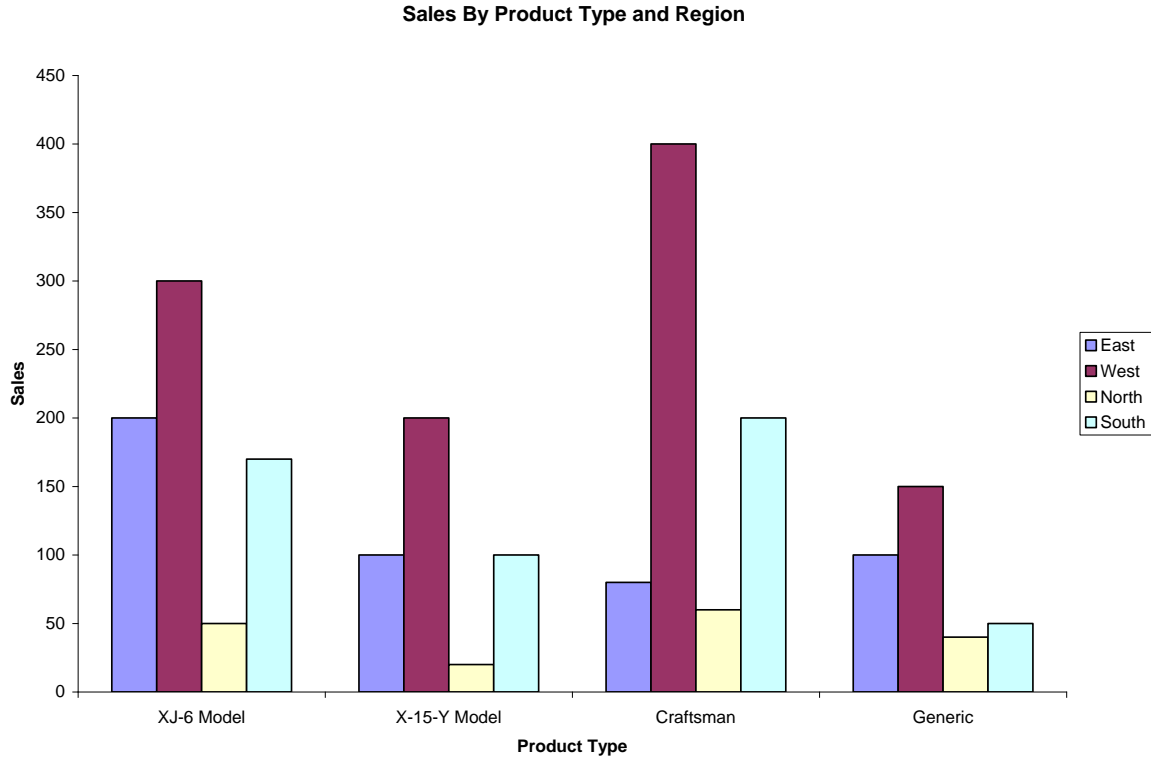
Still another possible way is called a "stacked" bar chart.

Sales By Product Type and Region
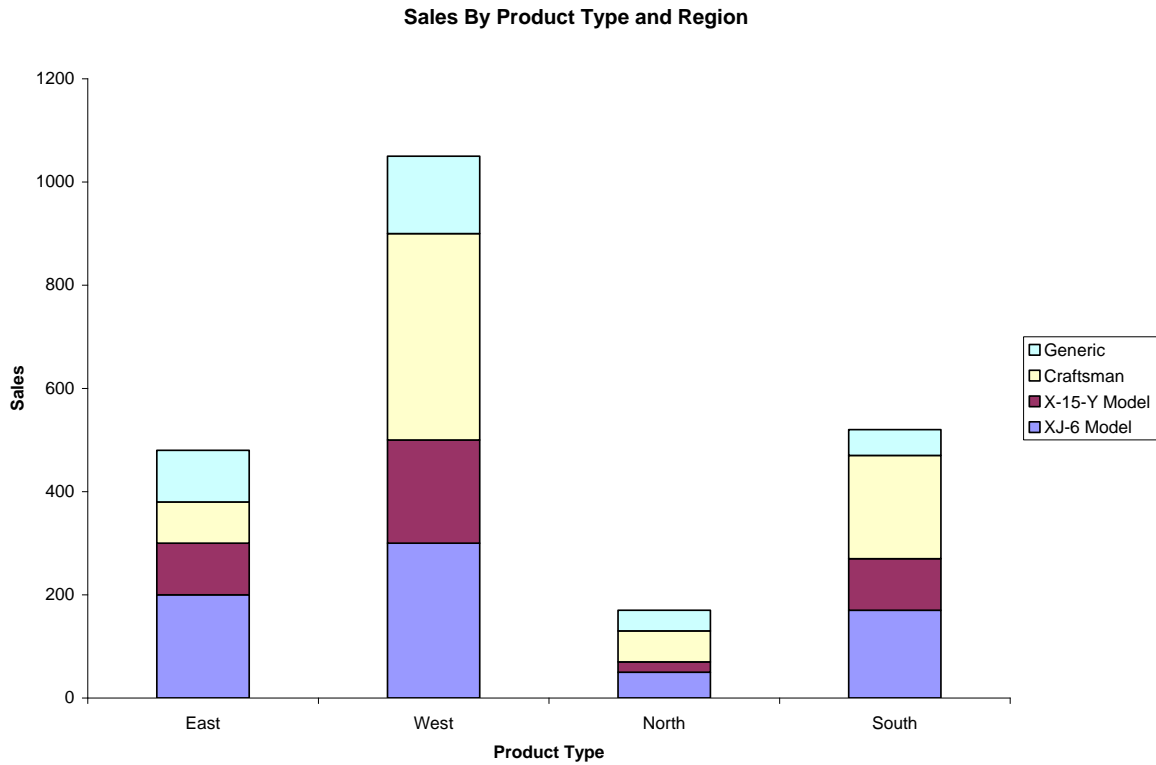
2.27. a.
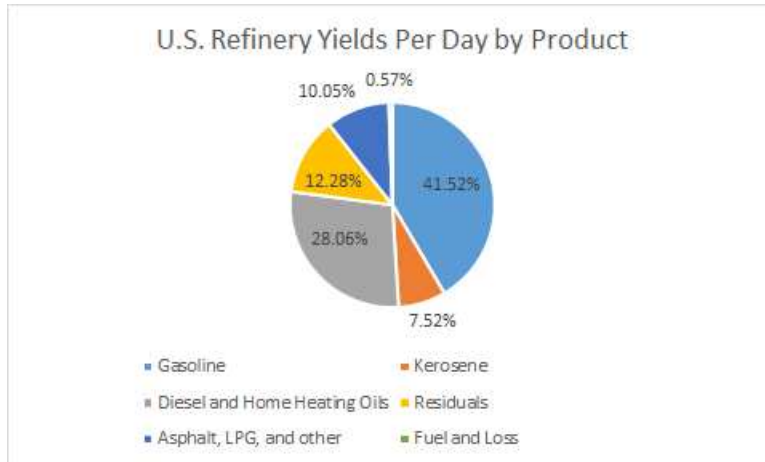
b.
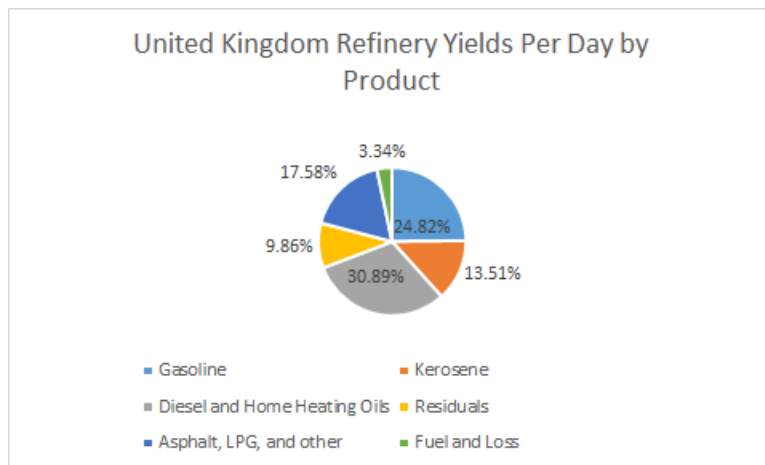
c.
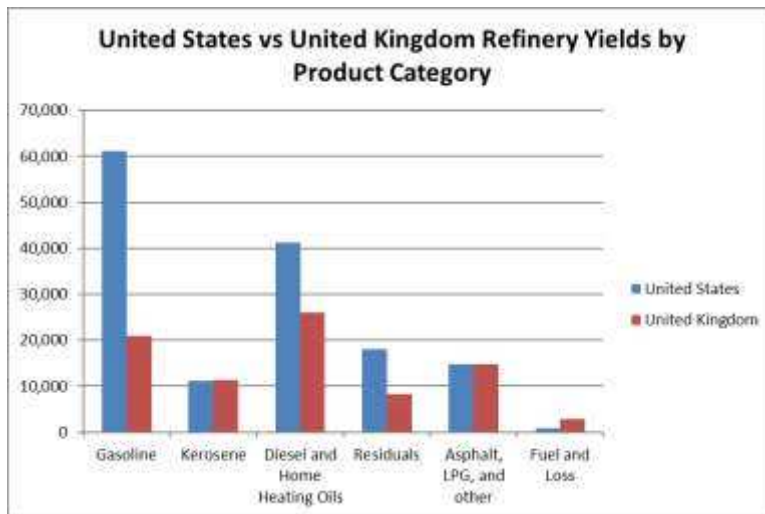
2.28. a. The pie chart with the data in percentage for the five major European countries' population in the year 2006:
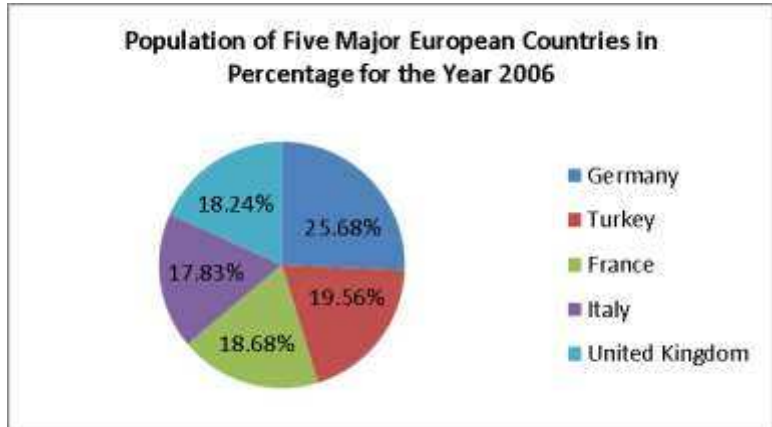
Population of Five Major European Countries in Percentage for the Year 2006

b. The pie chart with the data in percentage for the five major European countries' population in the year 2016:

Population of Five Major European Countries in Percentage for the Year 2017
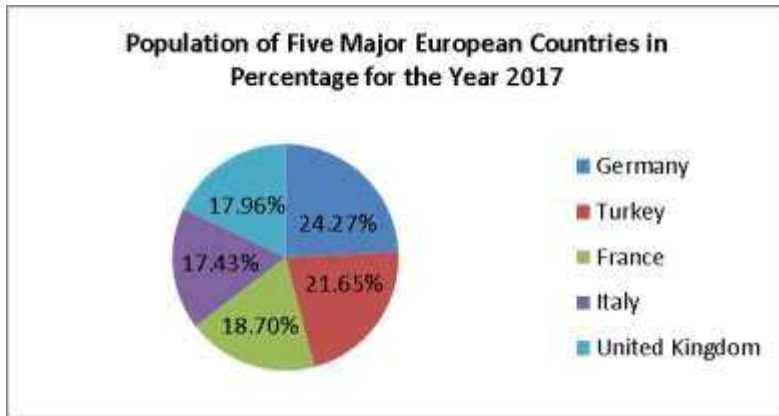
c. The bar chart that compares the five major European countries' population between the two years:

Five Major European Countries' Population in the Years 2006 and 2016
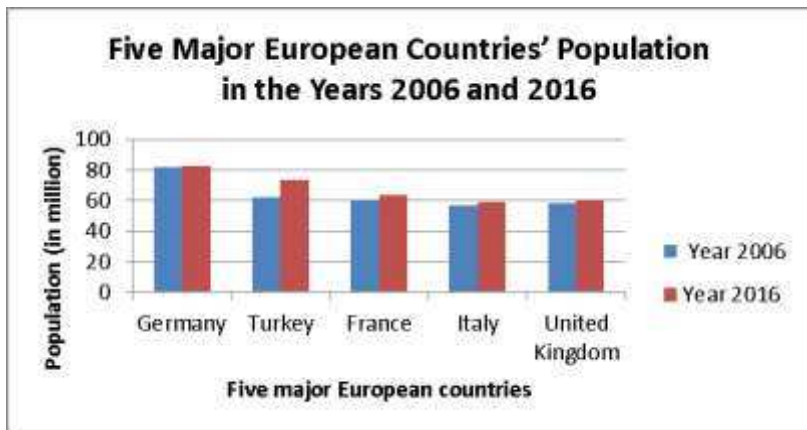
The population increased from 2006 to 2016 for all countries. The population of Turkey showed the maximum increase from the year 2006 to the year 2016 when compared with the other countries, whose population increase was nearly the same.

2.29 a.
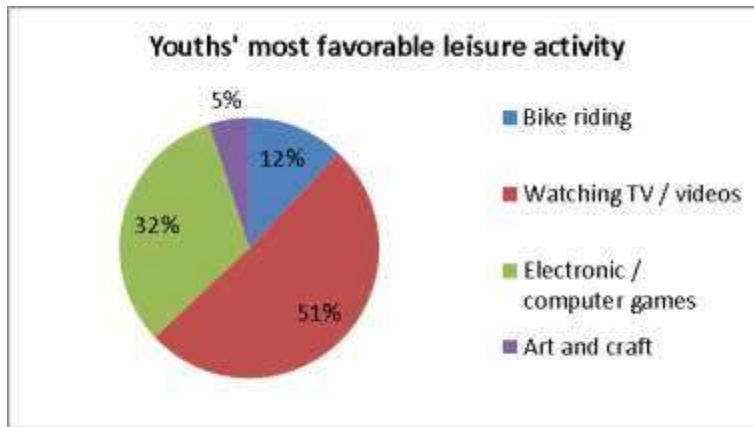
b.
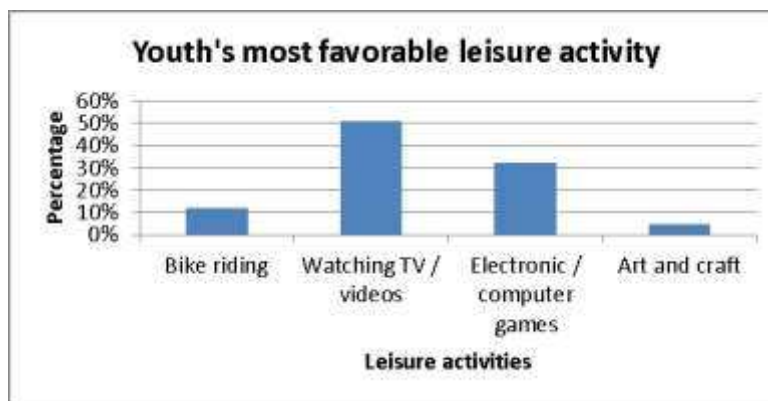
c. Pie charts are best to use when trying to compare parts of a whole and not when showing changes over time. Bar charts are used to compare values between different groups or to track changes over time.

2.30. a.

b.

Exported sugar amount (million tons)

March
April
May
June
July

22%
19%
18%
20%
21%
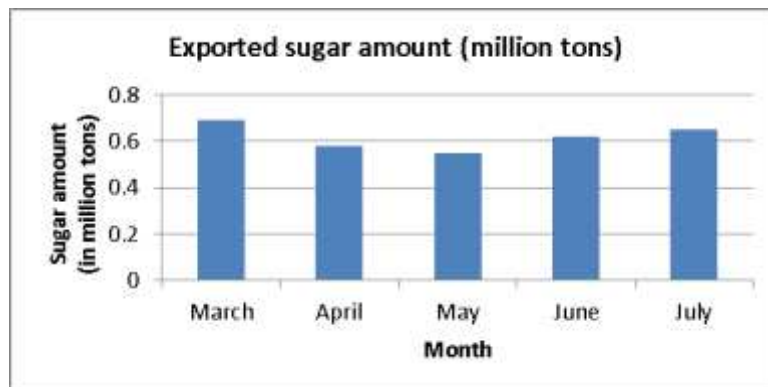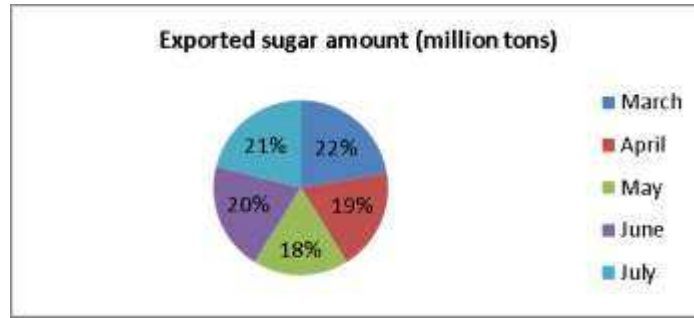
c.   The bar chart best represents the amount of sugar exported by White Santy for the past five months because it can track changes over the past five months.

2.31. a.

British Emigration to Five Select Destinations

Number of British emigrants (in 000)

Australia   Spain   USA   New Zealand   France

Destination

b.   The proportion for the destination with the most British emigrations = 52 / (52 + 34 + 26 + 21 + 28) = 52 / 161 = 0.323

2.32. 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 growingrapidly. It also appears that that growth is not linear.
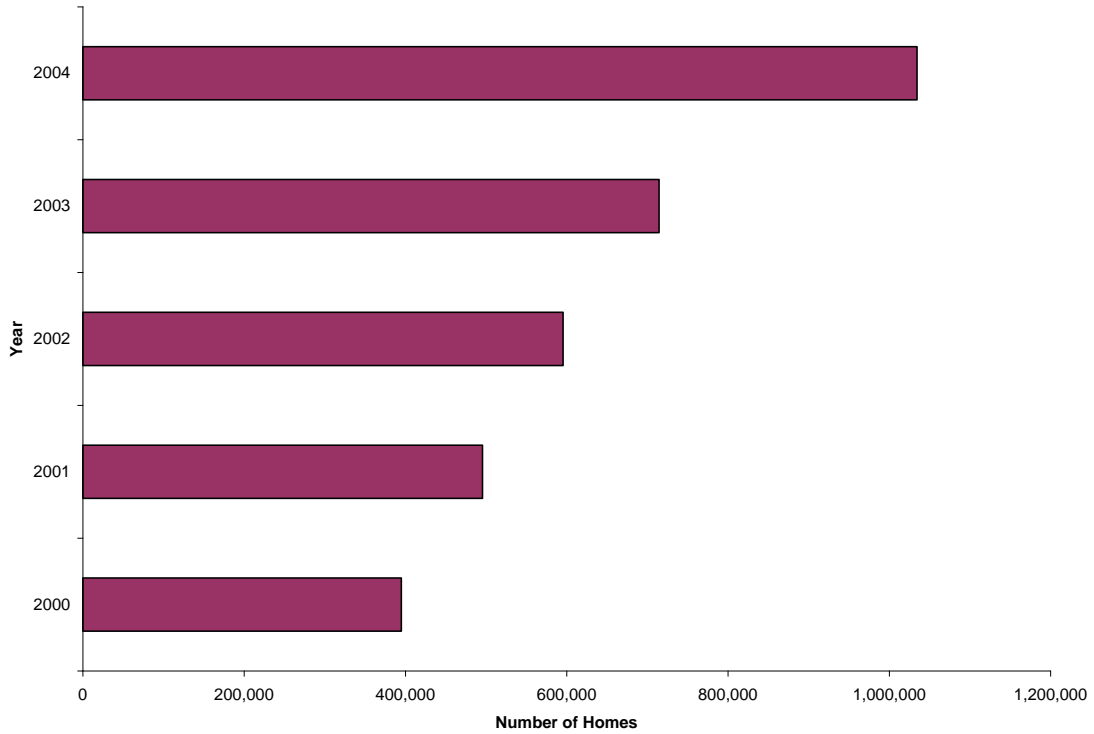
Step 4: Interpret the results.

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.

2.33.

Based on the graph, the profits for five days do not differ much, but there is a clear decrement on Wednesday and Monday. Aunty Soo's claim is probably true because the business did not show a good profit during the weekend.

2.34. A stacked bar chart is suitable for viewing the total and for performing a comparison.

Last year, the two packages earned approximately the same amount of profit, with package B having a slightly higher profit. Package B earned more at the start, and both packages earned approximately the same amount in the following two months.

2.35. a. The bar chart is shown below. The categories are the Global Segments and the measure for each category is the percent of total sales for the Global Segment.
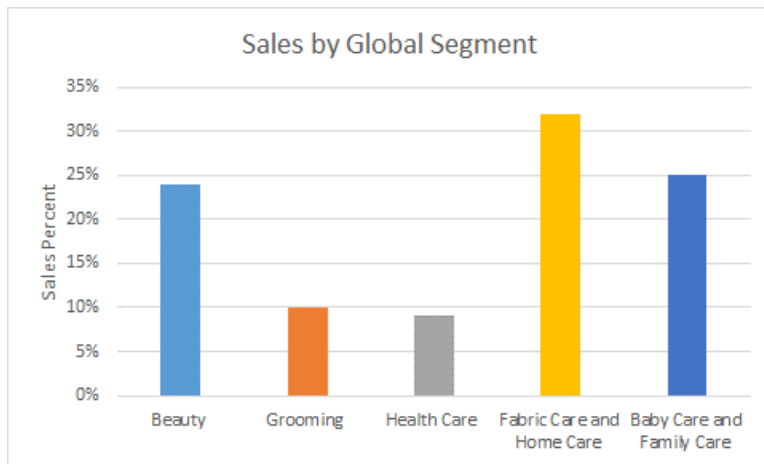
   b. The pie chart is shown below. The categories are the Global Segments and the measure is the proportion of each segment's total net sales.

2.36. 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.

2.37. 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.

2.38. a. The bar graph is

Seat Capacity of Bankrupt Airlines
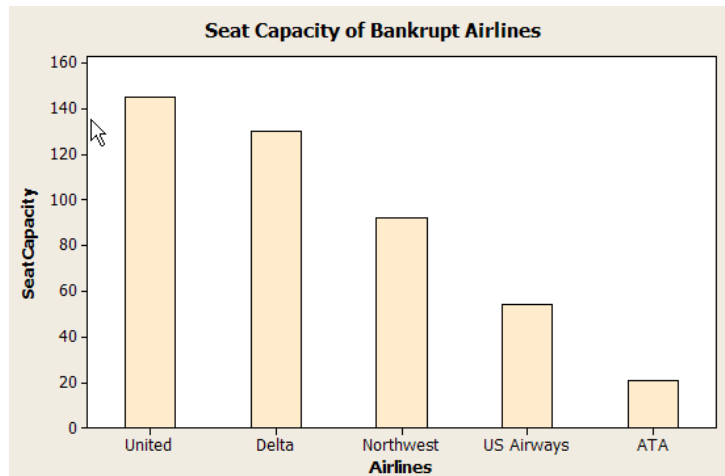
b. The percent equals the individual capacity divided by the total, e.g.
    United $\rightarrow$ percent $= (145/442)100\% = 32.81\%$, etc. This produces the following pie chart:
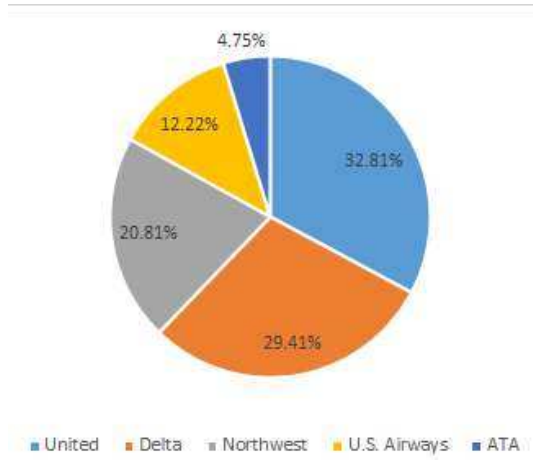
2.39. a.
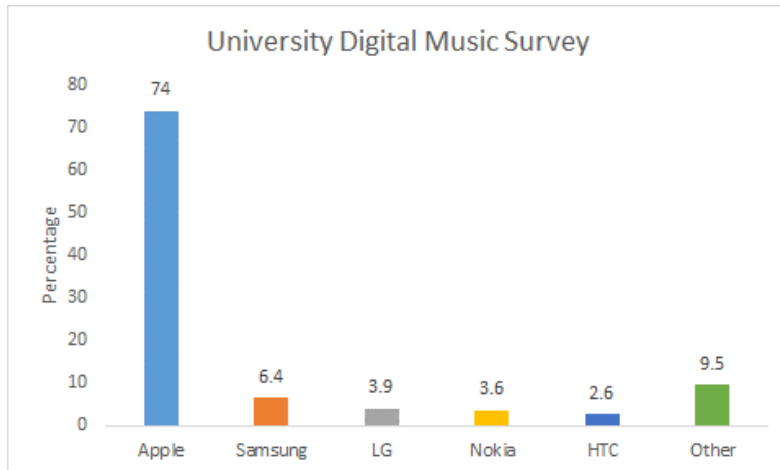
b.
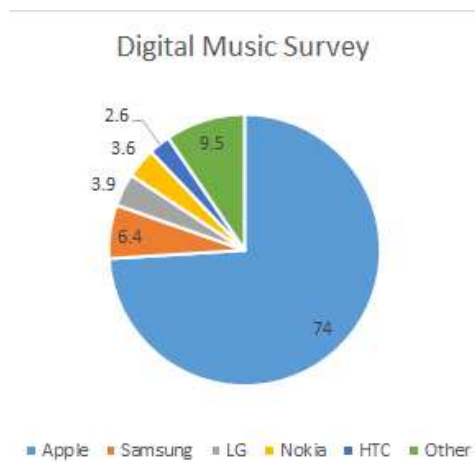
     c. A case can be made for either a bar chart or a pie chart. Pie charts are especially good at showing how the total is divided into parts. The bar chart is best to draw attention to specific results. In this case, a discussion might be centered on the relative large percentage attributable to Apple.

2.40.

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
```

2.41. 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
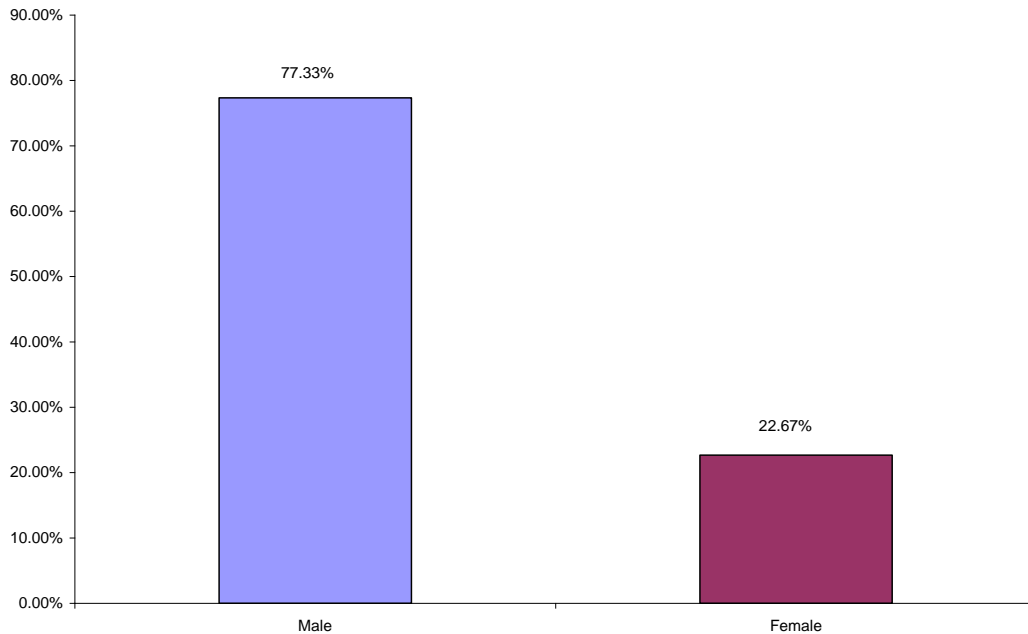
Step 4: Interpret the results.

This shows that a clear majority of credit card holders are males (77.33%)

2.42. 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
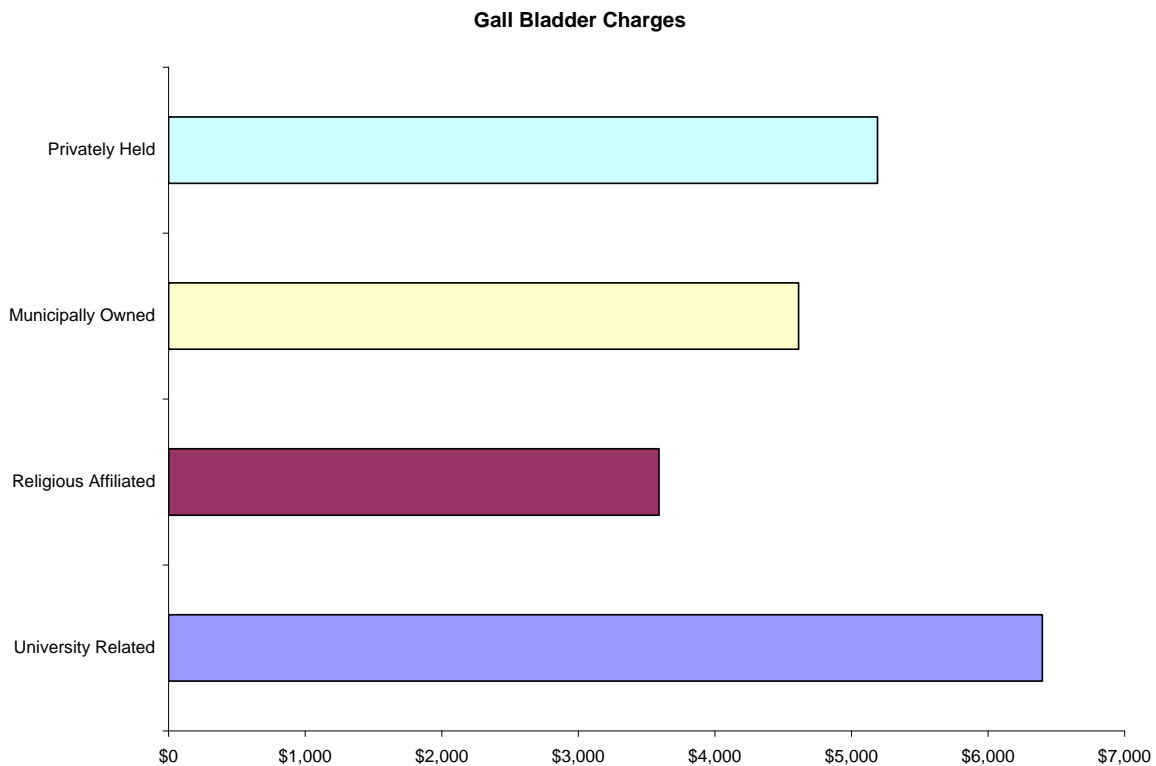
Step 4: Interpret the results.

The largest average charges occurred for gall bladder surgery appears to be in UniversityRelated hospitals and the lowest average appears to be in Religious Affiliated hospitals.
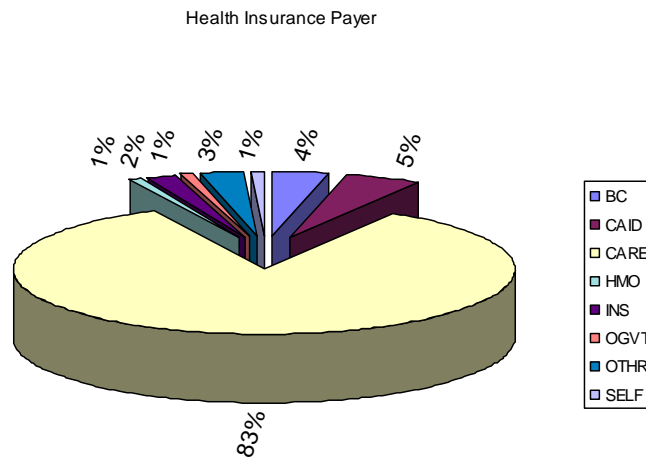
c. A pie chart is used to display the parts of a total. In this case the total charges of the four hospital types is not a meaningful number so a pie chart showing how that total is divided among the four hospital types would not be useful or appropriate.

2.43.   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 losses. Prior to
        this time, each year produced increased losses.
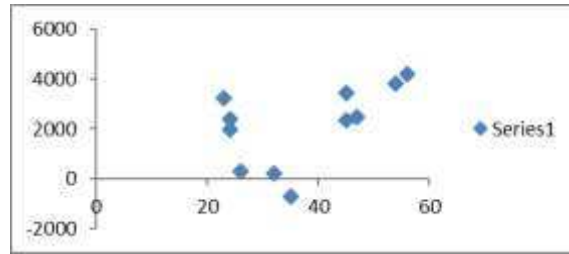
2.44.   a.

b.

Total Charges by Payer

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 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 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
     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

2.45. a.

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

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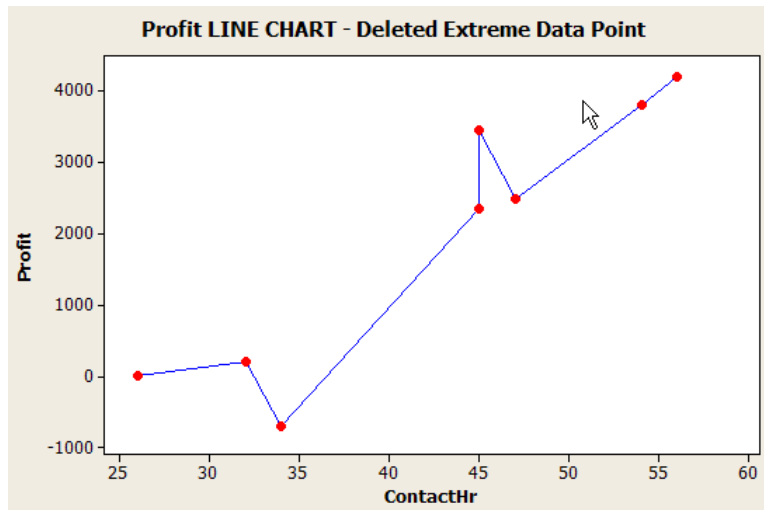
Having removed the extreme data points, the relationship between dependent and independent variables seems to be linear and positive.

2.46. 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
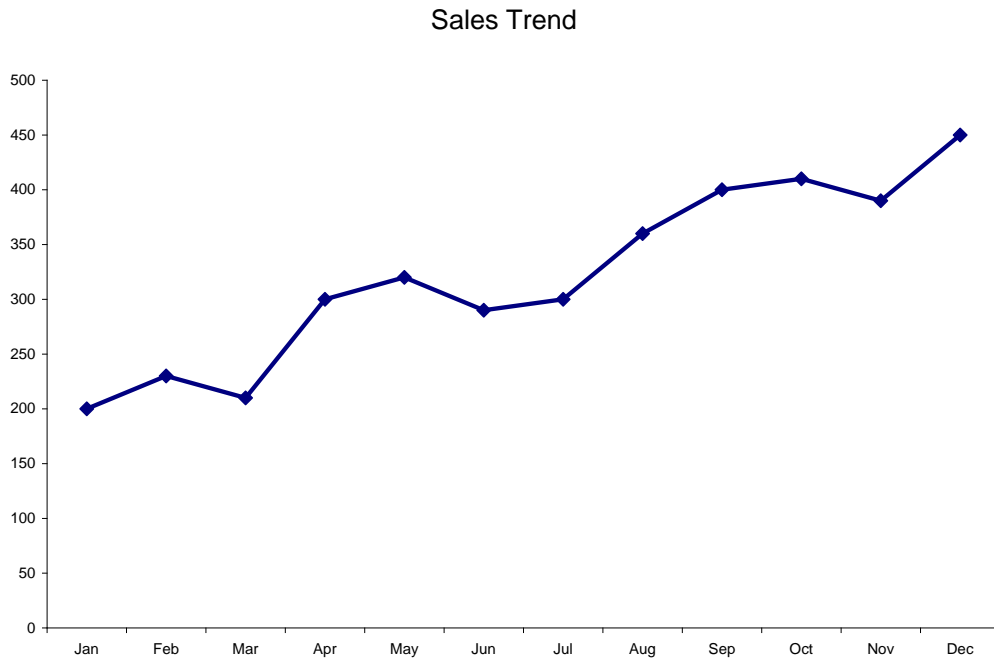
Sales Trend

The sales have trended upward over the past 12 months.

2.47.   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
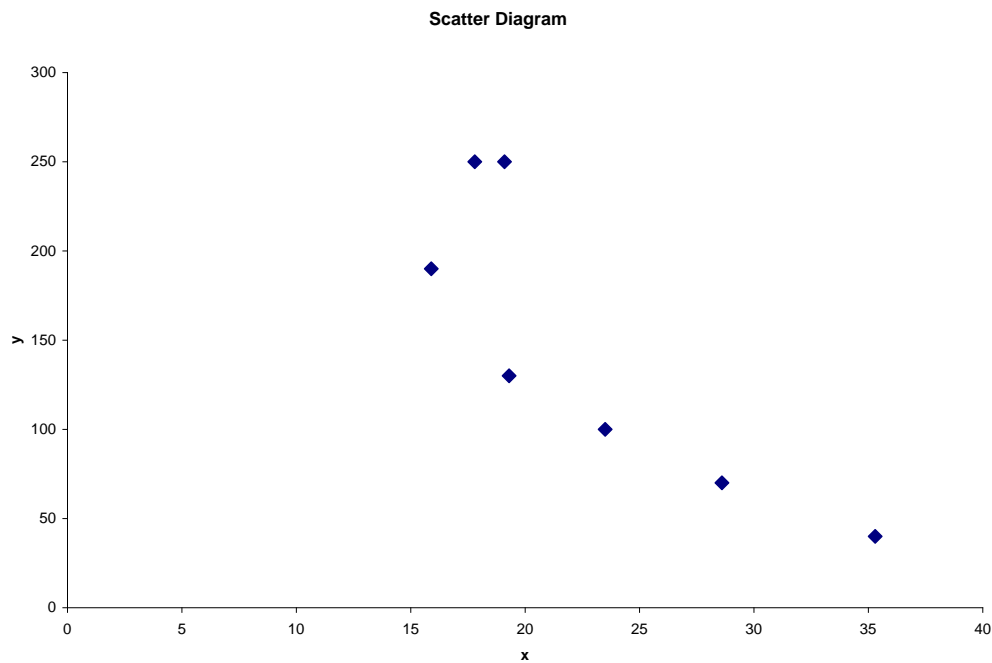
There is negative linear relationship between the two variables.

2.48. The time-series variable is Net Income ($ in millions) measured over 12 years with a maximum value of 172.5(million). The horizontal axis will have 12 time periods equally spaced. The vertical axis will start at 0 and go to a value exceeding 200. The vertical axis will also be divided into 10-unit increments. The line chart of the data is shown below.
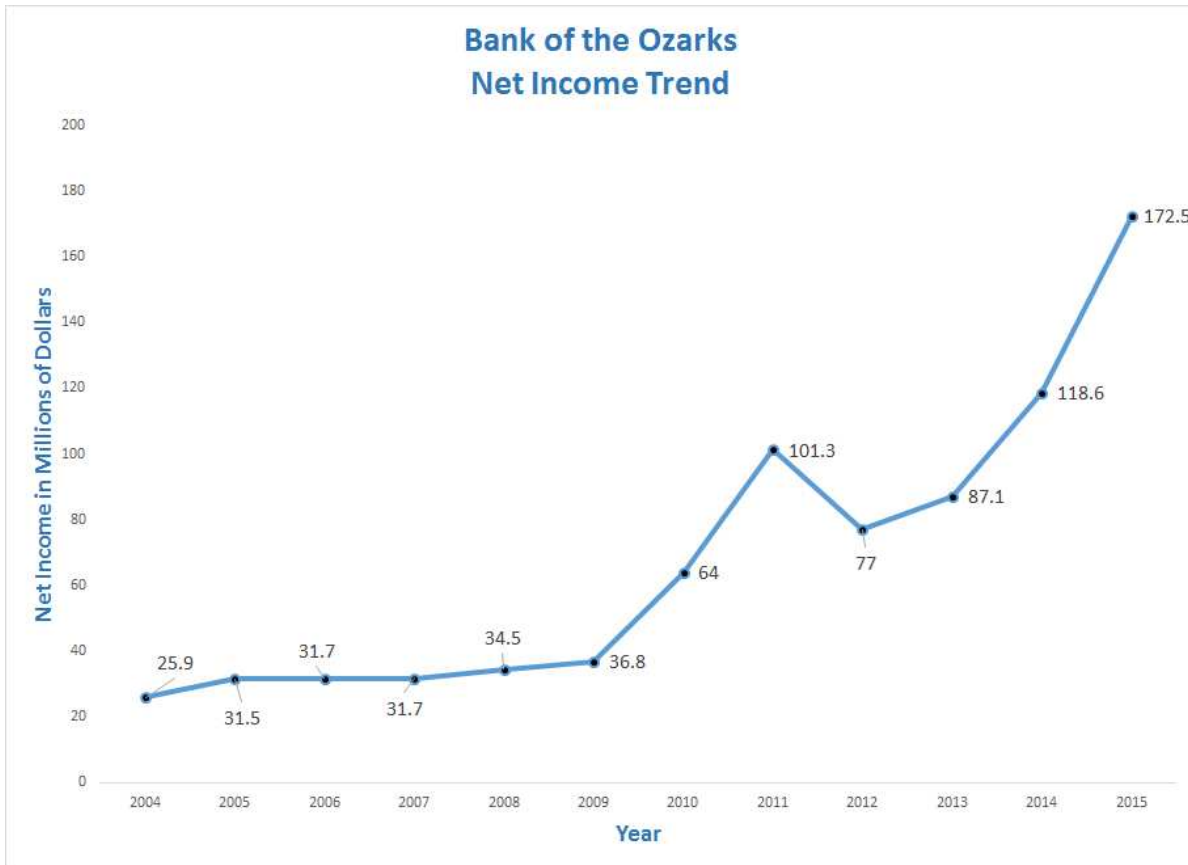
The line chart shows that Net Income has been increasing very steadily since 2005, but have increased more sharply since 2012.

2.49.
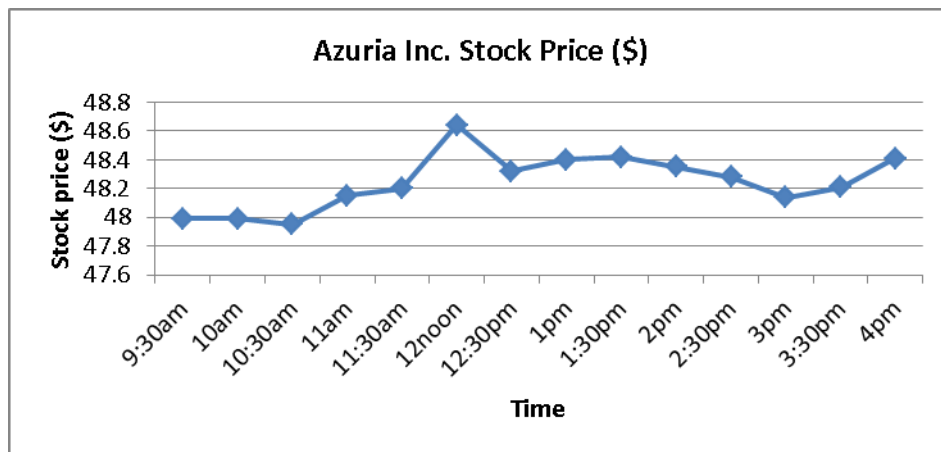
Azuria Inc.'s stock price shows a smooth increase from the beginning of the day until 11:30 a.m. Then it hits the highest peak of the day at noon and falls with a huge drop in the next half an hour. It continues to drop smoothly until 3 p.m. and shows an increment before the stock market closes.

2.50. a.
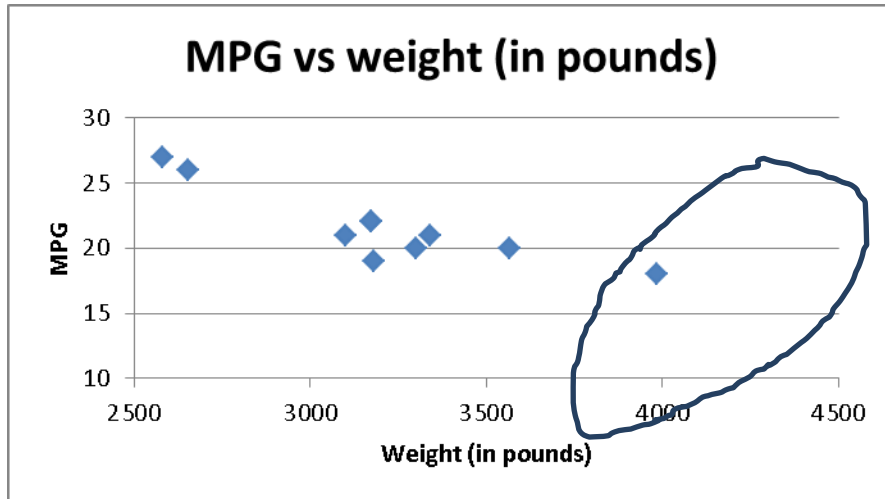
b.   There is a negative linear relationship between the car weight and its miles per gallon (MPG). This means that when the car weight increases, the miles per gallon for the car will decrease.
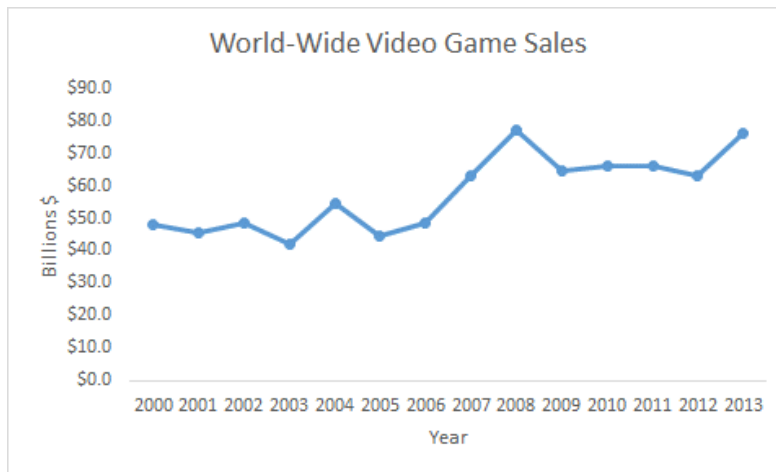
2.51. Step 1: Identify the time-series variable

The variable of interest is annual World-wide sales of video games

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

The line chart illustrates that over the 14 year period between 2000 and 2013, video game saleshave grown quite steadily from just below $50 billion to over $70 billion.

2.52.   a.

Number of Staff in a Clinic vs Number of Patients
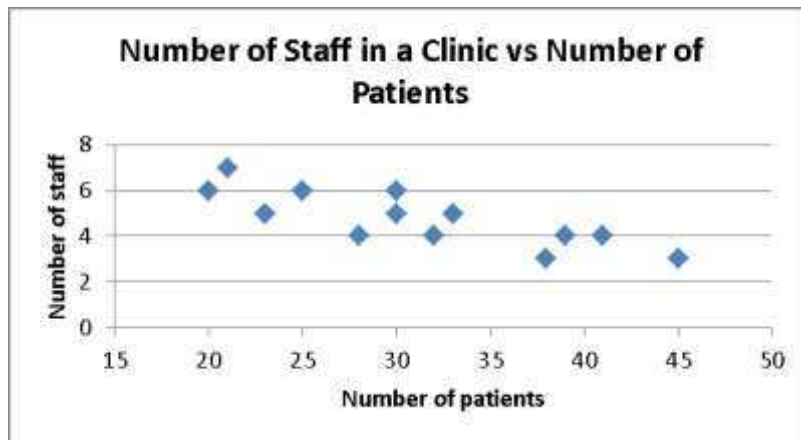
b.   The graph shows a negative linear relationship between the number of staff in a clinic and the number of patients visiting the clinic. If a decision is made based on the results presented by the graph, the doctor does not need to hire more staff to work in his clinic.

2.53.

S&P Oil Companies Combined Net Income
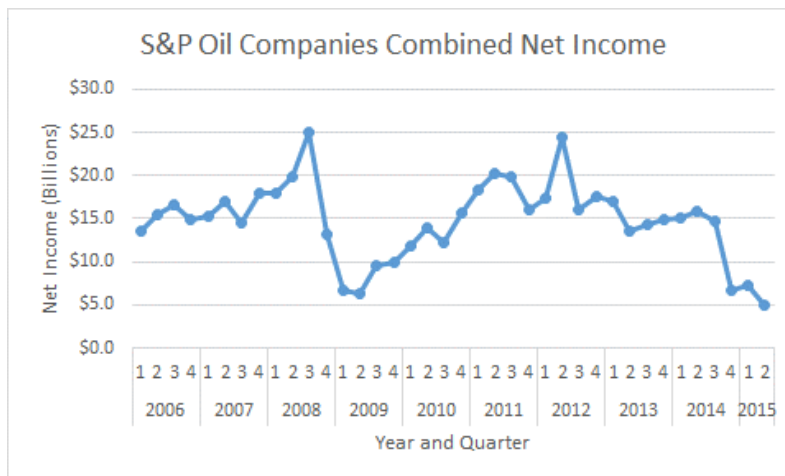
2.54.   Steps 1: Identify the two variables of interest
        In this example, there are two variables of interest, average home attendance and average road game attendance.
   Step 2: Identify the dependent and independent variables.
        Either one of these variables can be selected as the dependent variable. We willselect averageroad game attendance   Step 3: Establish the scales for the vertical and horizontal axes
   Step 4: Plot the joint values for the two variables by placing a point in the *x,y* space shown asfollows:
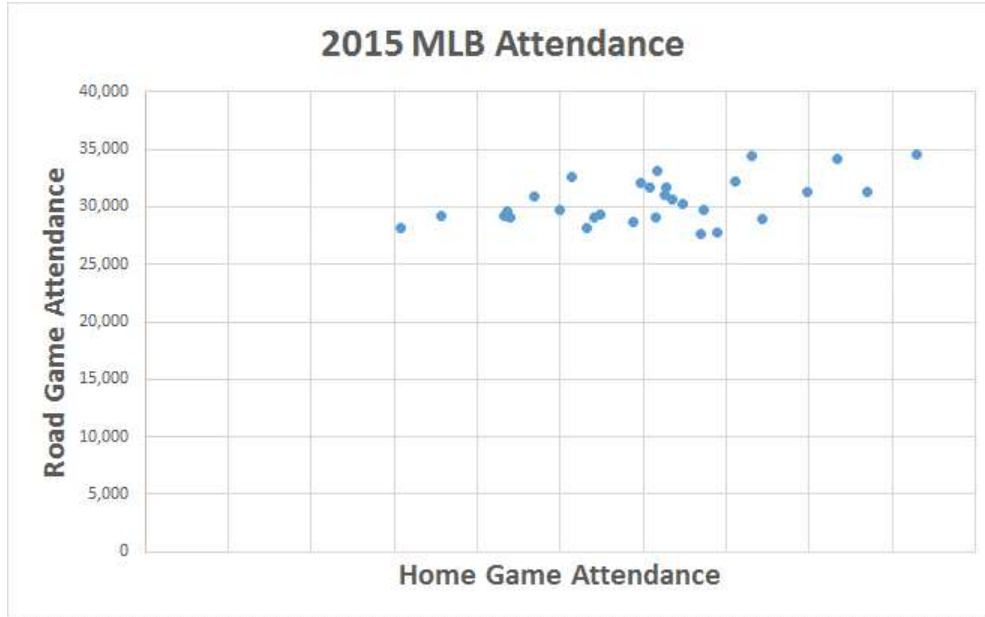
Based on the scatter diagram, it appears that there is a slight positive linear relationship between home and road attendance. However, the relationship is not perfect.

2.55.   Step 1: Identify the time-series variable

The variable of interest is number of customers

Step 2: Layout the Horizontal and Vertical Axis

The horizontal axis will be the year and the vertical axis is the number of customers

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

Since 1995, there has been a very steep growth in the number of customers over the time span.

2.56.   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:
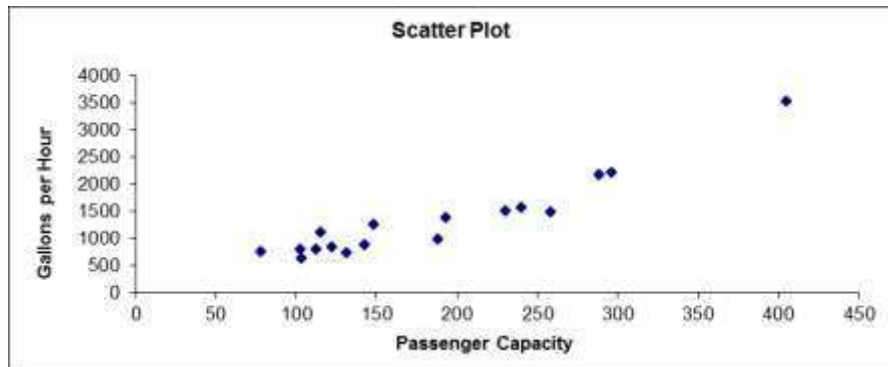
Based on the scatter diagram we see there is a strong positive linear relationship between passenger capacity and fuel consumption per hour.

2.57.   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.
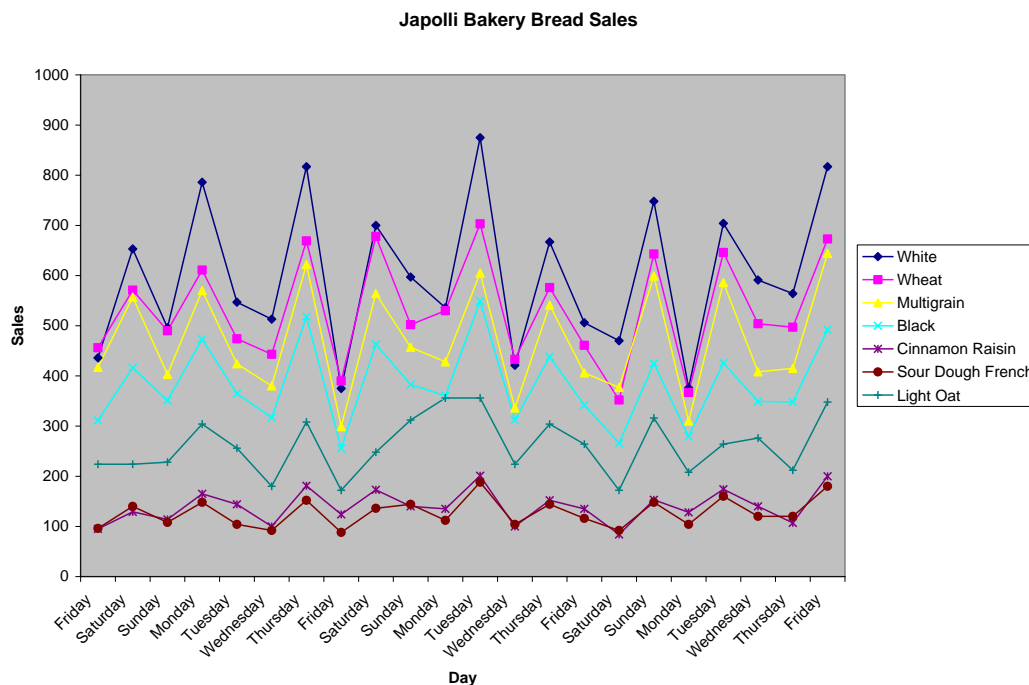
The graph illustrates a general pattern in the bread sales. Higher sales tend to occur for all types of bread on Saturdays, Mondays and Thursdays with Fridays typically the lowest.

2.58.   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

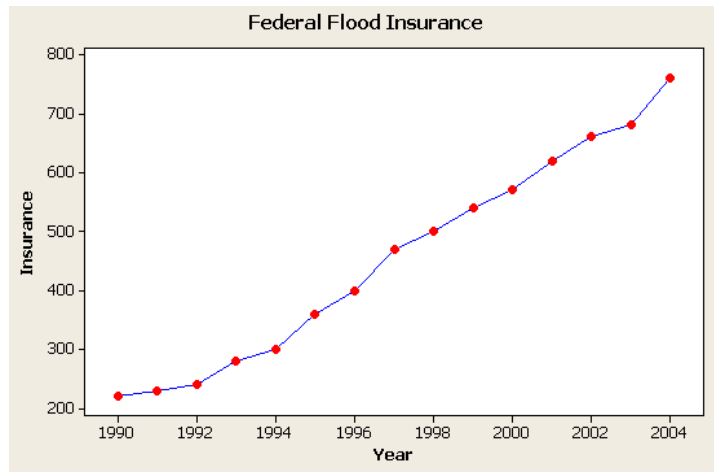The horizontal axis will be the year and the vertical axis is average price (See Step 3)

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.
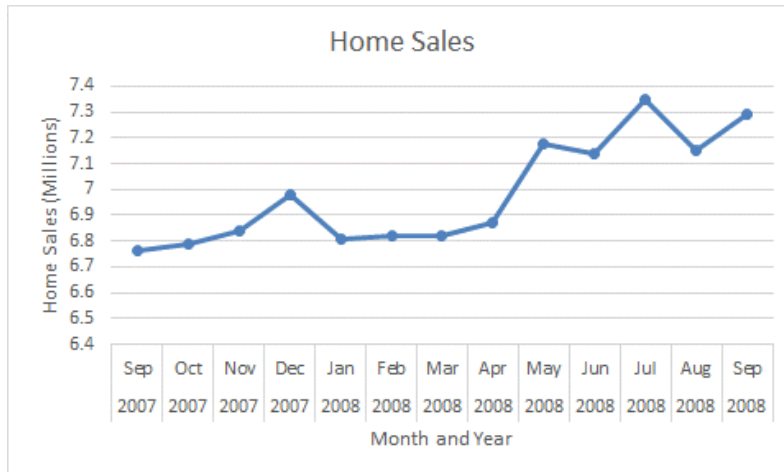
2.59.   a.

b. The relationship appears to be linear and positive.

c. The average equals the sum divided by the number of data points $= 6830/15 = 455.33$.
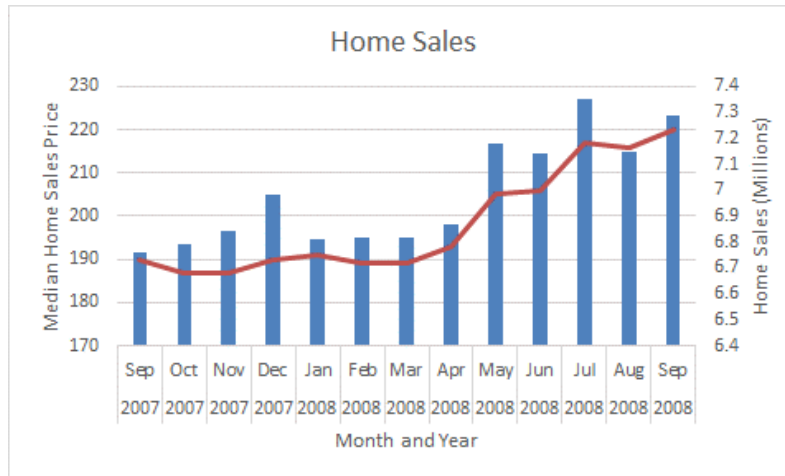
2.60. a.
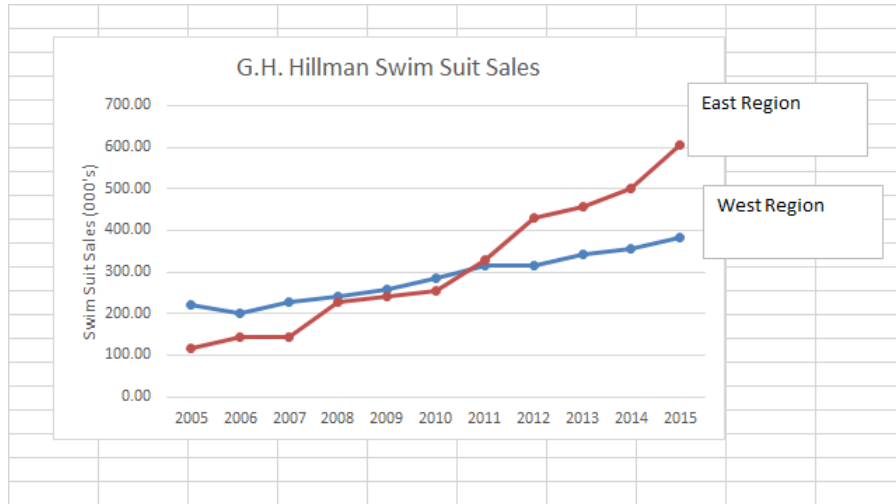
b. Both relationships seem to be linear in nature but the East region is growing faster in more recent years.

2.61. a.

b.

Note, a combo chart in Excel was used to show both data series since they have very different values.

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 Exercises

2.62.   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.

2.63.   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.

2.64.   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.

2.65.   The line chart records individual data values, and a line is created connecting the data points together to identify the movement of the data over a time interval. The scatter plot records a pair of data values and the best line is determined to identify the relationship between the two variables.
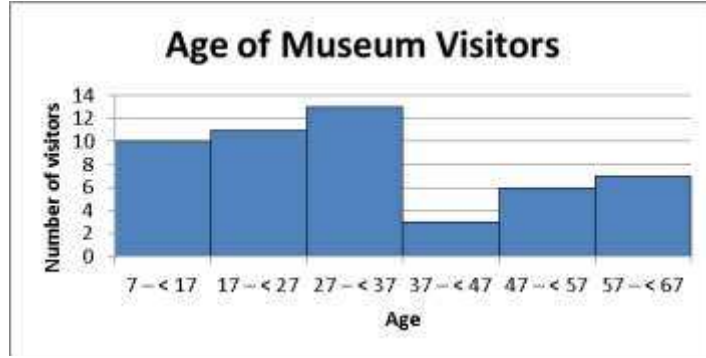
2.66.
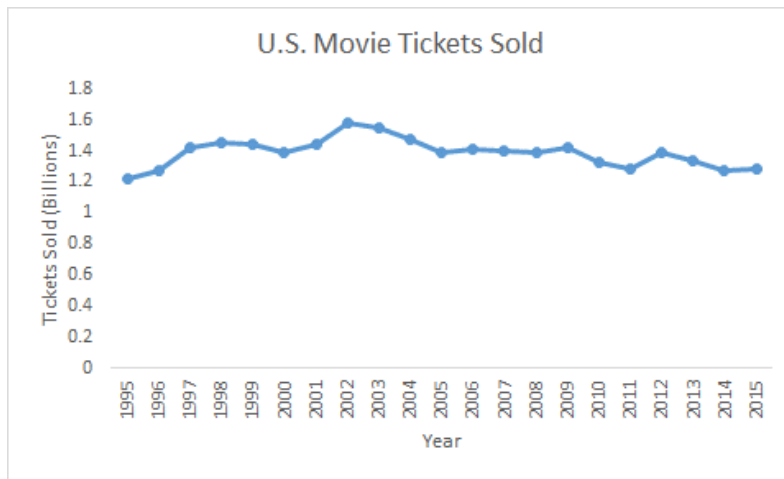
2.67.   a.
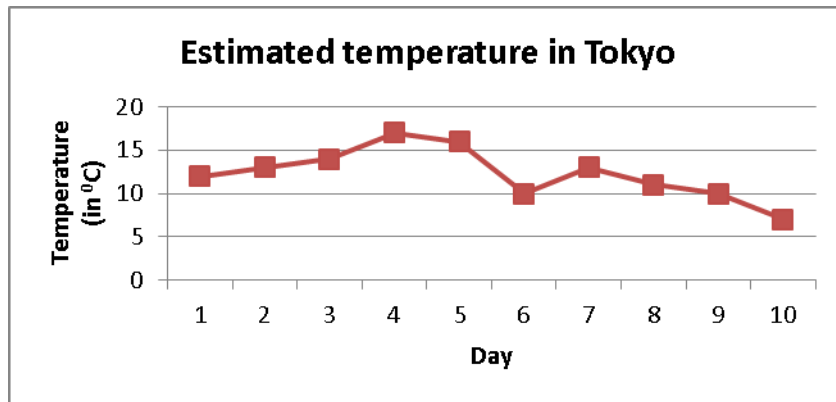
b.   The temperature shows a smooth increment for the first 4 days; then it sharply drops to 100C on day six. On day seven, the temperature goes up a bit but it drops smoothly for the next 3 days. Barbara should take more clothes since the temperatures are not stable during the 10 days of her business trip.

2.68.   a.   To determine the frequency distribution and histogram:

The number of classes, $2^k \geq n; 2^6 = 64$, which is greater than $n = 50$; so $k = 6$

The class width $= \dfrac{65 - 7}{6} = 9.6667$, round up to 10.

Age class	Frequency
$7 - < 17$	10
$17 - < 27$	11
$27 - < 37$	13
$37 - < 47$	3
$47 - < 57$	6
$57 - < 67$	7

Age of Museum Visitors

(histogram: Number of visitors vs. Age)

Age categories: 7 – < 17, 17 – < 27, 27 – < 37, 37 – < 47, 47 – < 57, 57 – < 67

b. The stem and leaf diagram

Stem	Leaves
0	7, 9
1	0, 1, 2, 3, 3, 4, 4, 4, 9
2	1, 3, 3, 3, 4, 4, 4, 4, 5, 5
3	2, 2, 3, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5
4	2, 6, 6, 7
5	3, 3, 4, 4, 6, 9
6	2, 3, 3, 3, 4, 5

Key: 5|3 = 53 years old

c. The frequency distribution for ticket price categories:

Ticket Price Category	Frequency
Free	6
$10	5
$15	5
$20	34
Total	50

Ticket Price Categories

*(pie chart: Free 12%, $10 10%, $15 10%, $20 68%)*
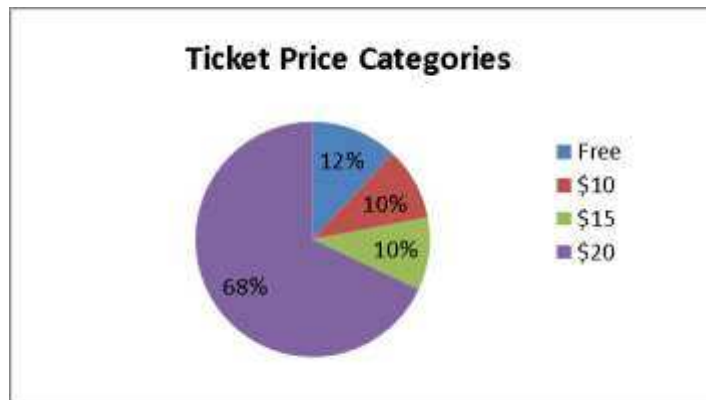
d.

2.69. a. The most appropriate graph for the data is the multiple bar chart:

b. It is clear that the household waste recycling center (HWRC) collected less waste than the civic amenity site (CA site) for the three items. Among the three items, the difference in the amount of waste was maximum for compost, followed by glass, and then paper and card (least).

2.70. a.
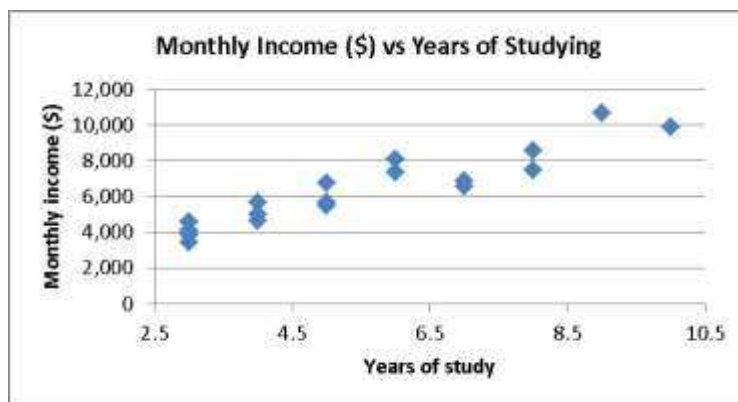
b. The graph shows a positive linear relationship between the number of years spent in studying after high school and the income. This means that when a person spends more years in studying after high school, they will earn more money.

2.71.   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	Relative Frequency
< 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

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

Ogive
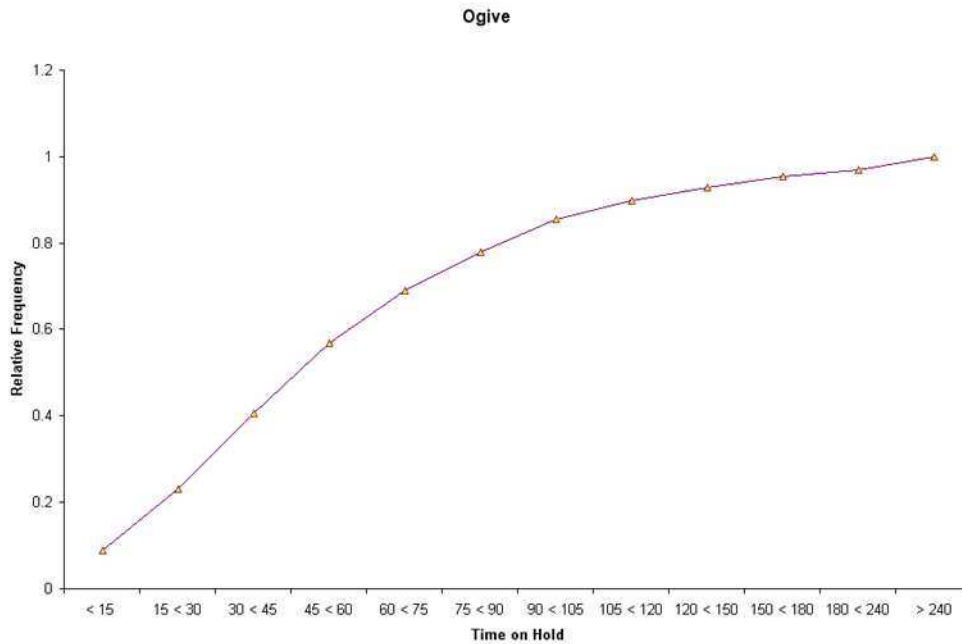
b.   The number of people who have to wait 120 seconds (2 minutes) or more is

$$158+124+87+153=522*\$30=\$15,660 \text{ month.}$$

2.72. a.
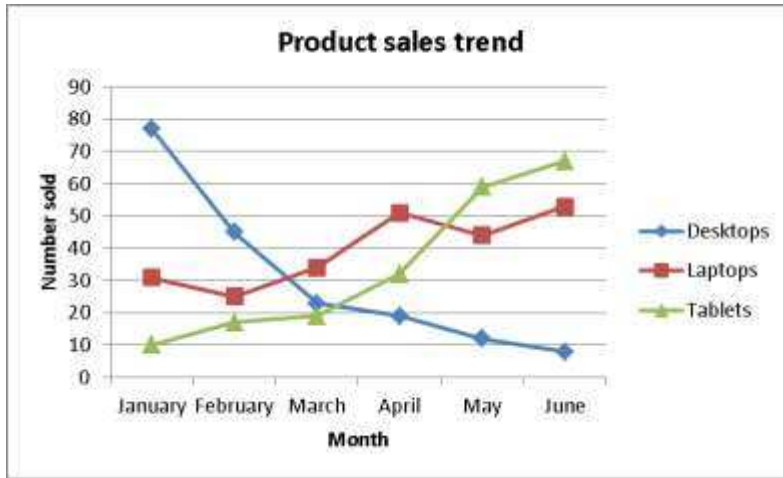
b. In the past six months, desktops had a serious decrement in sales; the company may need to pay extra attention to this trend. Conversely, tablets had a smooth increment in sales. The sale of laptops showed an up-and-down trend; however, the change was within a small range, and therefore, it is not a cause for worry.
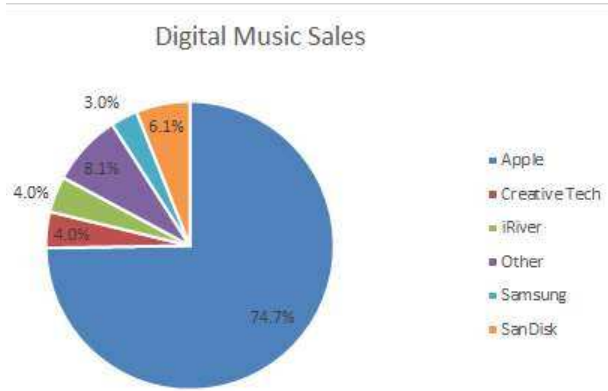
2.73. a.
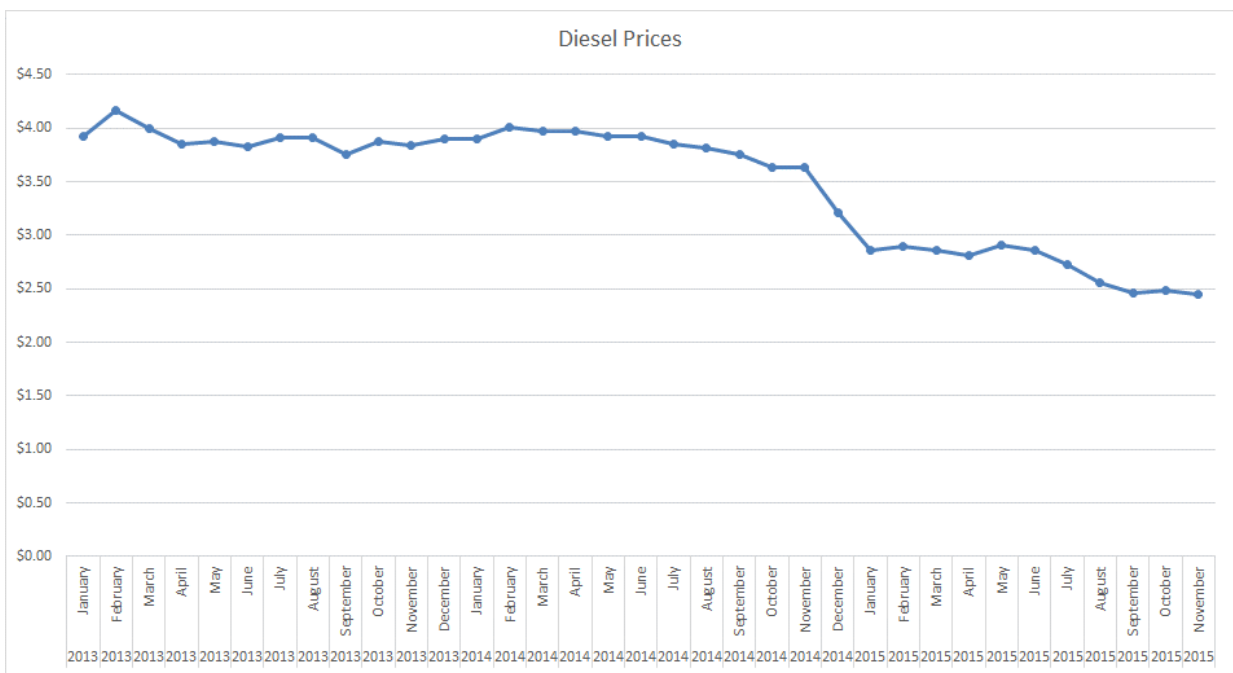
Student reports will vary but should discuss the negative trend in diesel prices over the time period given.

2.74.

Digital Music Sales

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

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

Determine the width: (310494 – 70464)/7 = 34,290. Round to 35,000

Classes	Frequency
70,000–104,999	43
105,000–139,999	34
140,000–174,999	13
175,000–209,999	5
210,000–244,999	2
245,000–274,999	1
280,000–314,999	2

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
175,000–209,999	5	0.05	0.95
210,000–244,999	2	0.02	0.97
245,000–274,999	1	0.01	0.98
280,000–314,999	2	0.02	1.00

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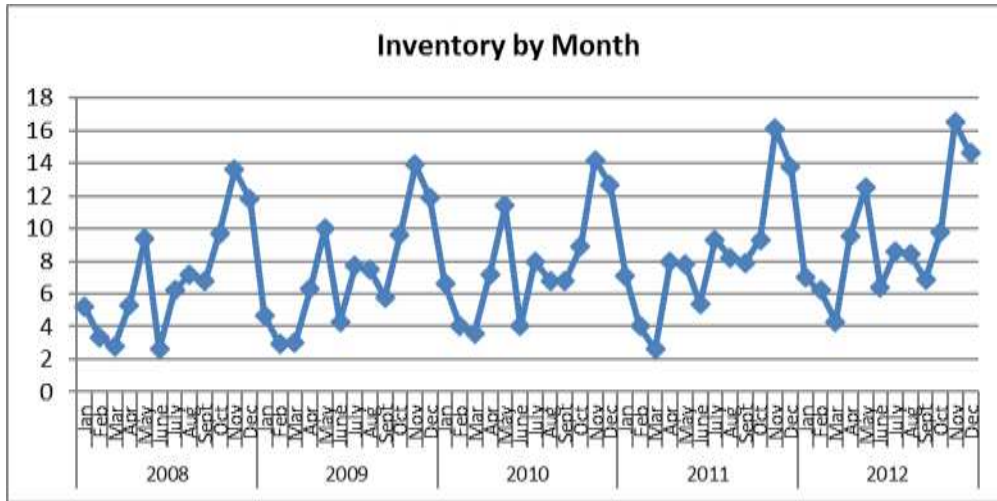
Inventory has been trending slightly up over the five years, but appears to be highly seasonal with predictable highs at certain points each year.

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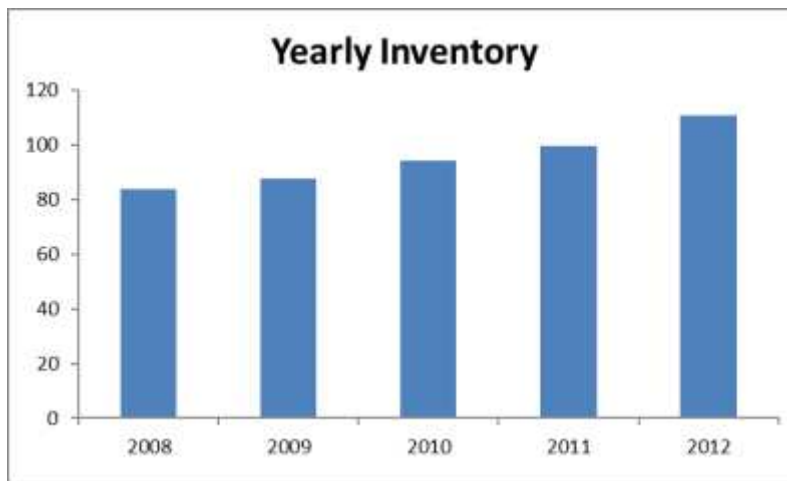
This bar chart is effective for showing the growth in total annual inventory over the five years. However, students should keep in mind that the sum of monthly inventory does not equate to how much inventory the store had on hand at the end of the year. Students might question why the store would graph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

2.77.   a.
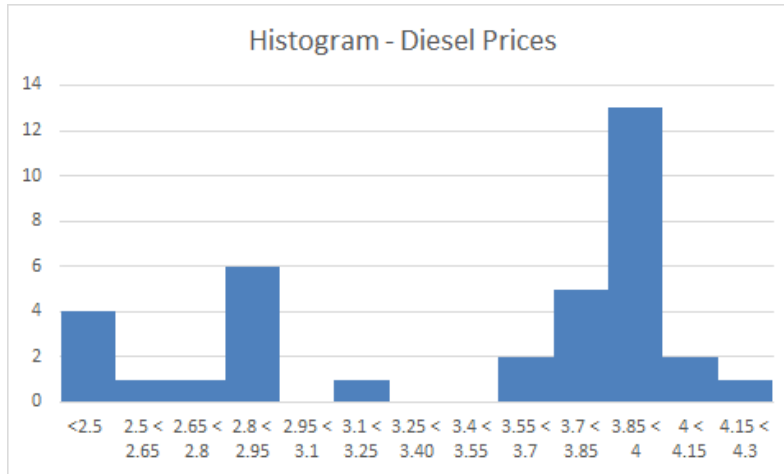
b.   Notice that three of the class interval have no observations. Since the numbers are averages taken across the United States, it is possible that the sampling technique, simply from randomness, didn't select prices in those ranges. It bears further investigation.