Business Statistics For Contemporary Decision Making 10th edition Black Solutions Manual

SKU: 9781119607458-SOLUTIONS

Chapter 2
Charts and Graphs

LEARNING OBJECTIVES

The overall objective of Chapter 2 is for you to master several techniques for summarizing and visualizing data, thereby enabling you to:

1. Construct a frequency distribution from a set of data
2. Construct different types of quantitative data graphs, including histograms, frequency polygons, ogives, dot plots, and stem-and-leaf plots, in order to interpret the data being graphed
3. Construct different types of qualitative data graphs, including pie charts, bar graphs, and Pareto charts, in order to interpret the data being graphed
4. Construct a cross-tabulation table and recognize basic trends in two-variable scatter plots of numerical data.
5. Construct a time-series graph and be able to visually identify any trends in the data.

CHAPTER OUTLINE

2.1 Frequency Distributions
- Class Midpoint
- Relative Frequency
- Cumulative Frequency

2.2 Quantitative Data Graphs
- Histograms
  - Using Histograms to Get an Initial Overview of the Data
- Frequency Polygons
- Ogives
- Dot Plots
- Stem-and-Leaf Plots

2.3 Qualitative Data Graphs
- Pie Charts
- Bar Graphs
- Pareto Charts

2.4 Charts and Graphs for Two Variables
- Cross Tabulation
- Scatter Plot

2.5 Visualizing Time-Series Data

KEY TERMS

Bar Charts
Bar Graph
Class Mark
Class Midpoint
Column Charts
Cross Tabulation
Cumulative Frequency
Dot Plot
Frequency Distribution
Frequency Polygon
Grouped Data
Histogram
Ogive
Pareto Chart
Pie Chart
Range
Relative Frequency
Scatter Plot
Stem-and-Leaf Plot
Time-Series Data
Ungrouped Data

STUDY QUESTIONS

1. The following data represents the number of printer ribbons used annually in a company by twenty-eight departments. This is an example of _____ data.

   8 4 5 10 6 5 4 6 3 4 4 6 1 12
   2 11 2 5 3 2 6 7 6 12 7 1 8 9

2. Below is a frequency distribution of ages of managers with a large retail firm. This is an example of _____ data.

Age	f
20-29	11
30-39	32
40-49	57
50-59	43
over 60	18

3. For best results, a frequency distribution should have between _____ and _____ classes.

4. The difference between the largest and smallest numbers is called the _____.

5. Consider the values below. In constructing a frequency distribution, the beginning Point of the lowest class should be at least as small as _____ and the endpoint of the highest class should be at least as large as _____.

   27 21 8 10 9 16 11 12 21 11 29 19 17 22 28 28 29 19 18 26 17 34 19
   16 20

6. The class midpoint can be determined by _____.

7-9 Examine the frequency distribution below:

class	frequency
5-under 10	56
10-under 15	43
15-under 20	21
20-under 25	11
25-under 30	12
30-under 35	8

7. The relative frequency for the class 15-under 20 is _____.

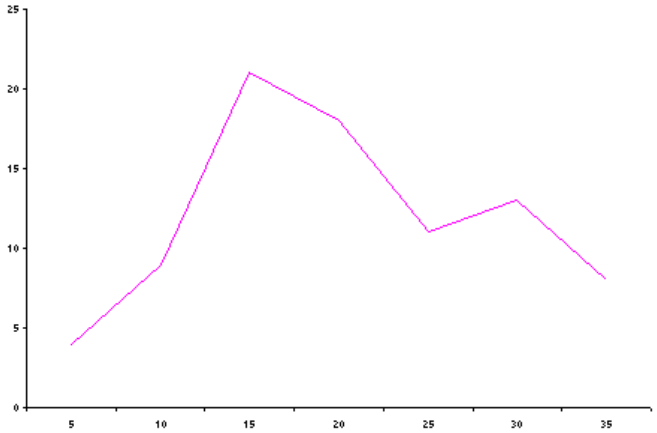
8. The cumulative frequency for the class 20-under 25 is _____.

9. The midpoint for the class 25-under 30 is _____.

10. The graphical depiction that is a type of vertical bar chart and is used to depict a frequency distribution is a _____.

11. The graphical depiction that utilizes cumulative frequencies is a _____.

12. The graph shown below is an example of a _____.



13. Consider the categories below and their relative amounts:

Category	Amount
A	112
B	319
C	57
D	148
E	202

If you were to construct a Pie Chart to depict these categories, then you would allot _____ degrees to category D.

14. A graph that is especially useful for observing the overall shape of the distribution of data points along with identifying data values or intervals for which there are groupings and gaps in the data is called a _____.

15. Given the values below, construct a stem and leaf plot using two digits for the stem.

346 340 322 339 342 332 338
357 328 329 346 341 321 332

16. A vertical bar chart that displays the most common types of defects that occur with a product, ranked in order from left to right, is called a _____.

17. A process that produces a two-dimensional table to display the frequency counts for two variables simultaneously is called a _____.

18. A two-dimensional plot of pairs of points often used to examine the relationship of two numerical variables is called a _____.

19. Time-series data are data gathered on a particular characteristic over a period of time at _____.

ANSWERS TO STUDY QUESTIONS

1. Raw or Ungrouped
2. Grouped
3. 5, 15
4. Range
5. 8, 34
6. Averaging the two class endpoints
7. 21/151 = .1391
8. 131
9. 27.5
10. Histogram
11. Ogive
12. Frequency Polygon
13. 148/838 of $360^o = 63.6^o$
14. Dot Plot
15. 32 1 2 8 9
    33 2 2 8 9
    34 0 1 2 6 6
    35 7
16. Pareto Chart
17. Cross Tabulation
18. Scatter Plot
19. Regular Intervals

SOLUTIONS TO THE ODD-NUMBERED PROBLEMS IN CHAPTER 2

2.1

a) One possible 5 class frequency distribution:

Class Interval	Frequency
0 - under 20	7
20 - under 40	15
40 - under 60	12
60 - under 80	12
80 - under 100	4
	50

b) One possible 10 class frequency distribution:

Class Interval	Frequency
10 - under 18	7
18 - under 26	3
26 - under 34	5
34 - under 42	9
42 - under 50	7
50 - under 58	3
58 - under 66	6
66 - under 74	4
74 - under 82	4
82 - under 90	2

c) The ten class frequency distribution gives a more detailed breakdown of temperatures, pointing out the smaller frequencies for the higher temperature intervals. The five class distribution collapses the intervals into broader classes making it appear that there are nearly equal frequencies in each class.

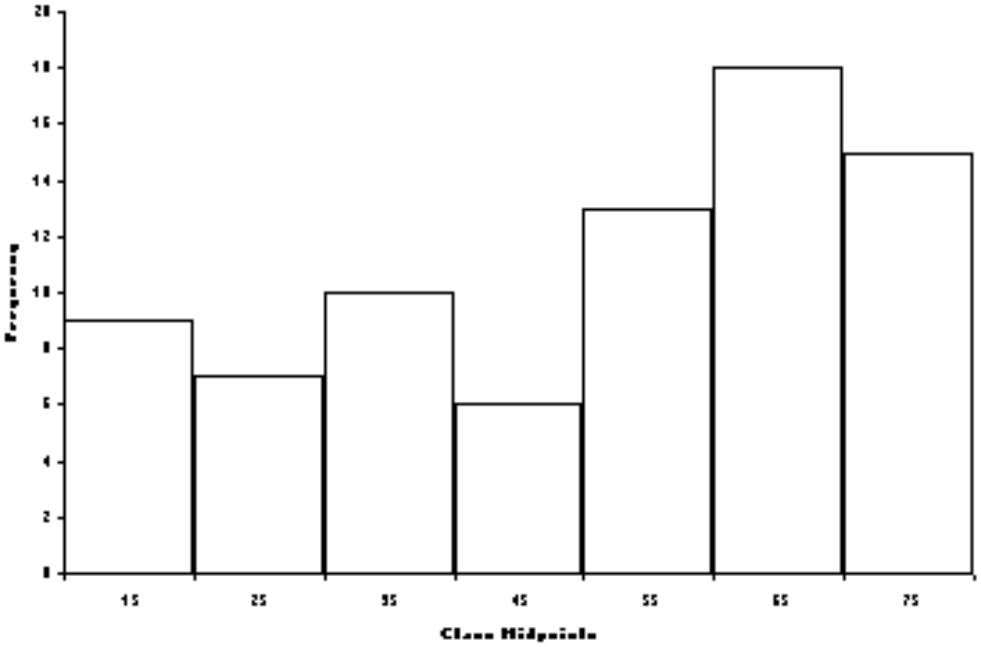
2.3

Class Interval	Frequency	Class Midpoint	Relative Frequency	Cumulative Frequency
0 - 5	6	2.5	6/86 = .0698	6
5 - 10	8	7.5	.0930	14
10 - 15	17	12.5	.1977	31
15 - 20	23	17.5	.2674	54
20 - 25	18	22.5	.2093	72
25 - 30	10	27.5	.1163	82
30 - 35	4	32.5	.0465	86
TOTAL	86		1.0000	

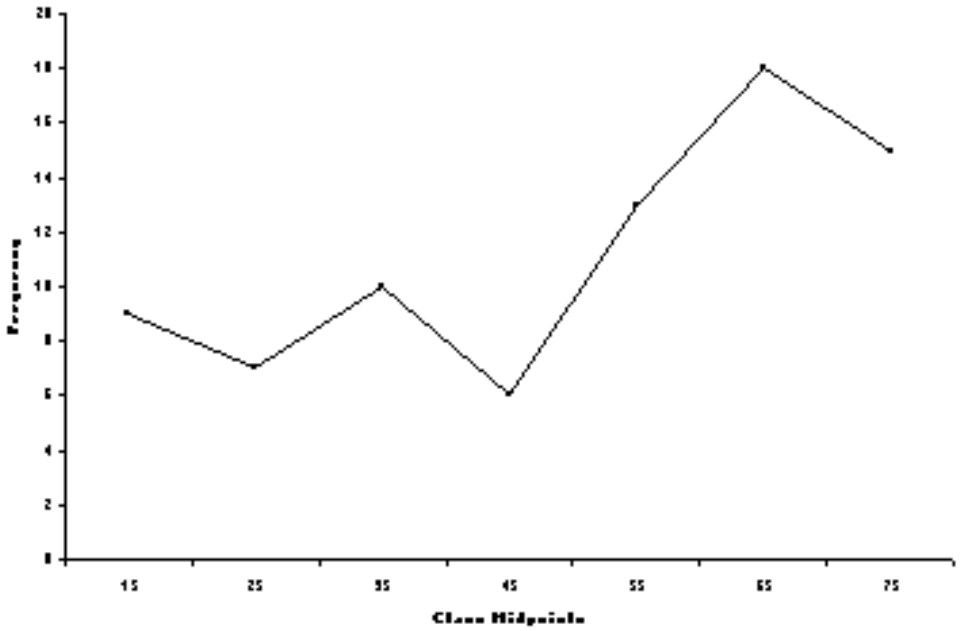
The relative frequency tells us that it is most probable that a customer is in the 15 - 20 category (.2674). Over two thirds (.6744) of the customers are between 10 and 25 years of age.

2.5 Some examples of cumulative frequencies in business:
sales for the fiscal year,
costs for the fiscal year,
spending for the fiscal year,
inventory build-up,
accumulation of workers during a hiring buildup,
production output over a time period.

2.7 Histogram:



Frequency Polygon:

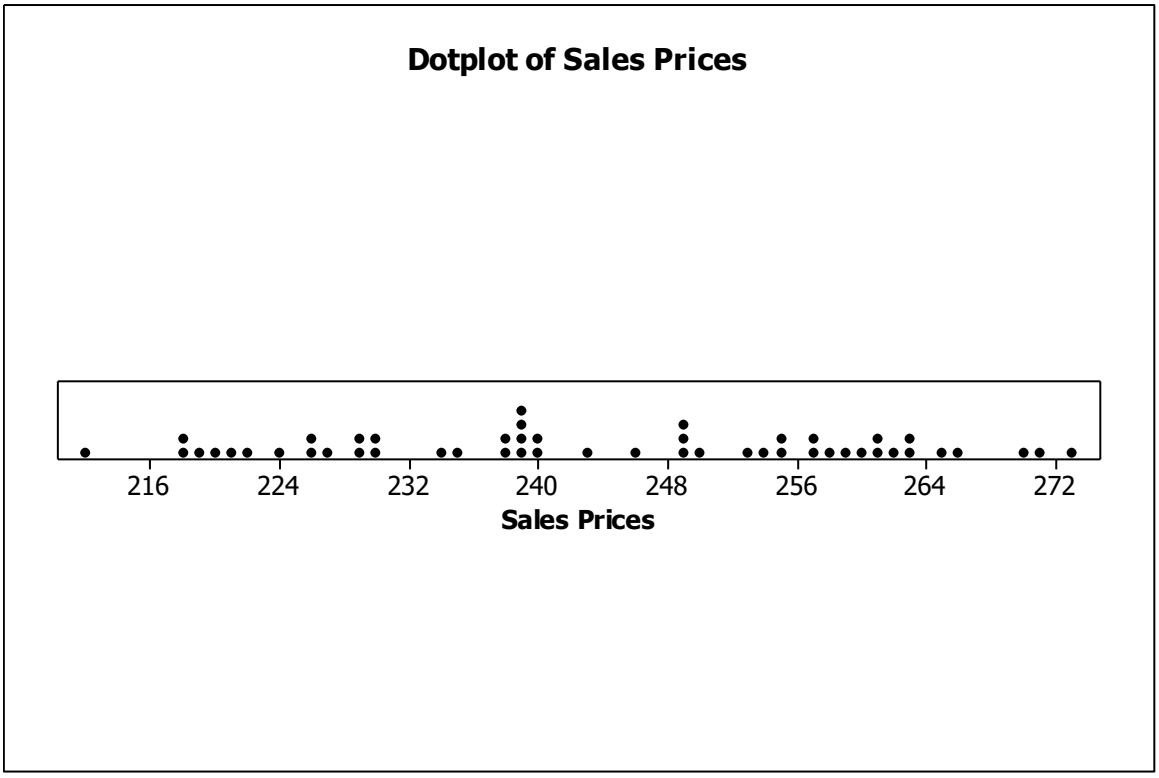


Comment: The histogram indicates that the number of calls per shift varies widely. However, the heavy numbers of calls per shift fall in the 50 to 80 range. Since these numbers occur quite frequently, staffing planning should be done with these number of calls in mind realizing from the rest of the graph that there may be shifts with as few as 10 to 20 calls.

2.9

STEM	LEAF
21	2 8 8 9
22	0 1 2 4 6 6 7 9 9
23	0 0 4 5 8 8 9 9 9 9
24	0 0 3 6 9 9 9
25	0 3 4 5 5 7 7 8 9
26	0 1 1 2 3 3 5 6
27	0 1 3

Dotplot



Both the stem and leaf plot and the dot plot indicate that sales prices vary quite a bit within the range of $212,000 and $273,000. It is more evident from the stem and leaf plot that there is a strong grouping of prices in the five price ranges from the $220's through the $260's.

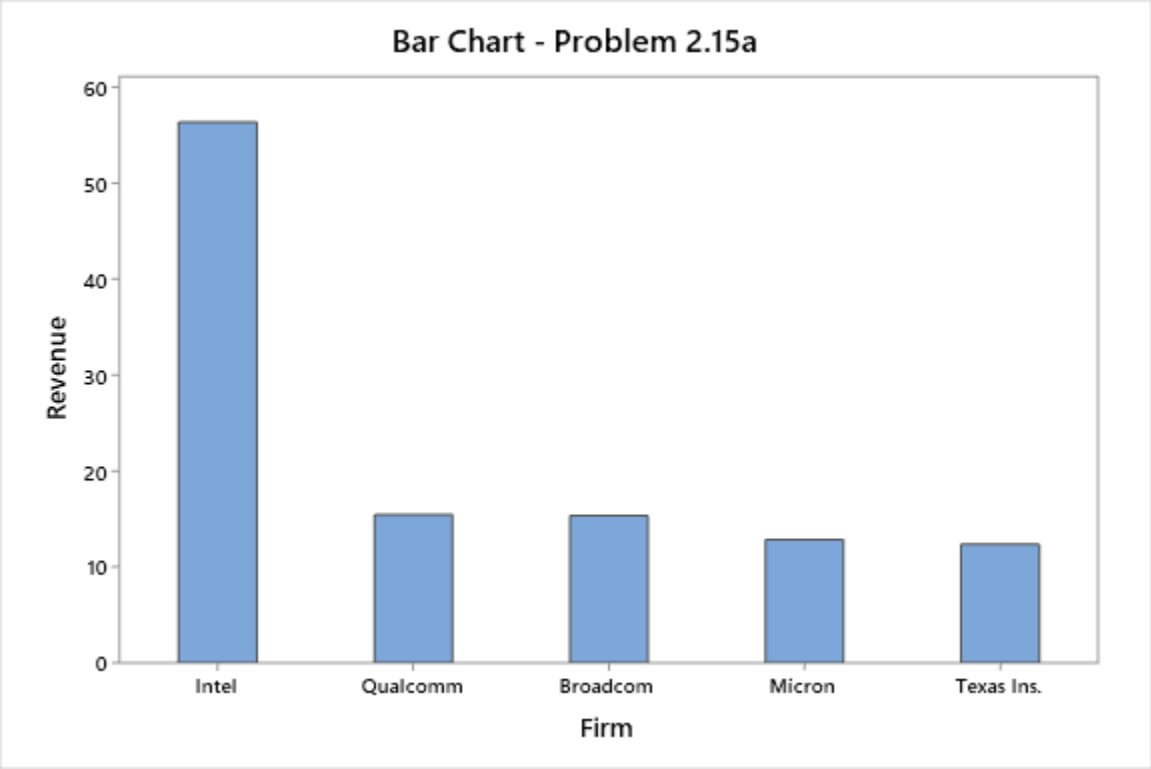
2.11 From the information given in the problem, we know that the busiest airport in the world is Atlanta's Hartsfield-Jackson International Airport which has almost 104 million passengers per year. The histogram shows that there are only two airports with between 95 and 105 million passengers and two other airports with between 85 and 95 million passengers. Eleven airports (more than one-third) served between 55 and 65 million passengers. Six airports served between 45 and 55 million and six other airports served 65 to 75 million.

2.13 From the stem and leaf display, the original raw data can be obtained. For example, the fewest number of cars washed on any given day are 25, 29, 29, 33, etc. The most cars washed on any given day are 141, 144, 145, and 147. The modal stems are 3, 4, and 10 in which there are 6 days with each of these numbers. Studying the left column of the Minitab output, it is evident that the median number of cars washed is 81. There are only two days in which 90 some cars are washed (90 and 95) and only two days in which 130 some cars are washed (133 and 137).

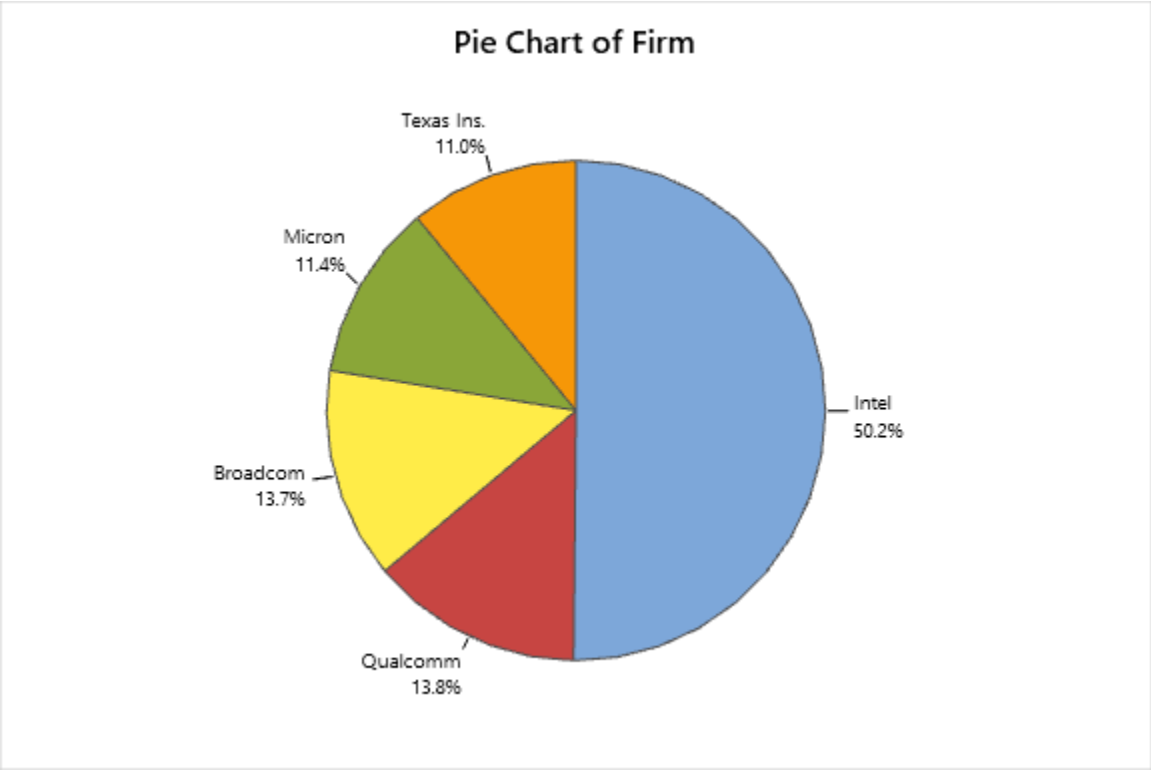
2.15

Firm	Proportion	Degrees
Intel Corp.	.502	180.7
Qualcomm	.138	49.7
Broadcom	.137	49.3
Micron Tech	.114	41.0
Texas Instru.	.110	39.6
TOTAL	1.001	360.3

a.) Bar Graph:



b.) Pie Chart:

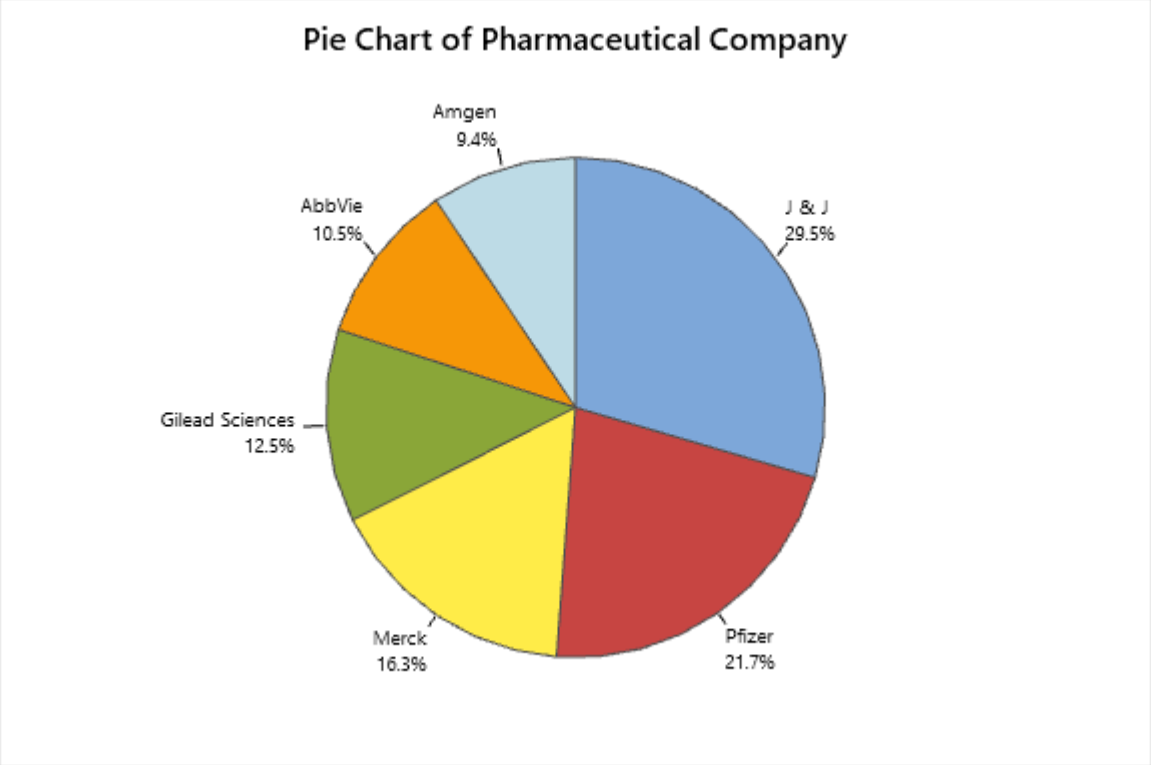


c.) While pie charts are sometimes interesting and familiar to observe, in this problem it is virtually impossible from the pie chart to determine the relative difference between Broadcom and Qualcom without the data labels. There is a similar problem between the slices of Micron and Texas Ins. In this particular problem, it is also difficult to discriminate differences between Qualcomm and Broadcom and between Micron and Texas Instrument. Both clearly show that Intel has about ½ the market share.

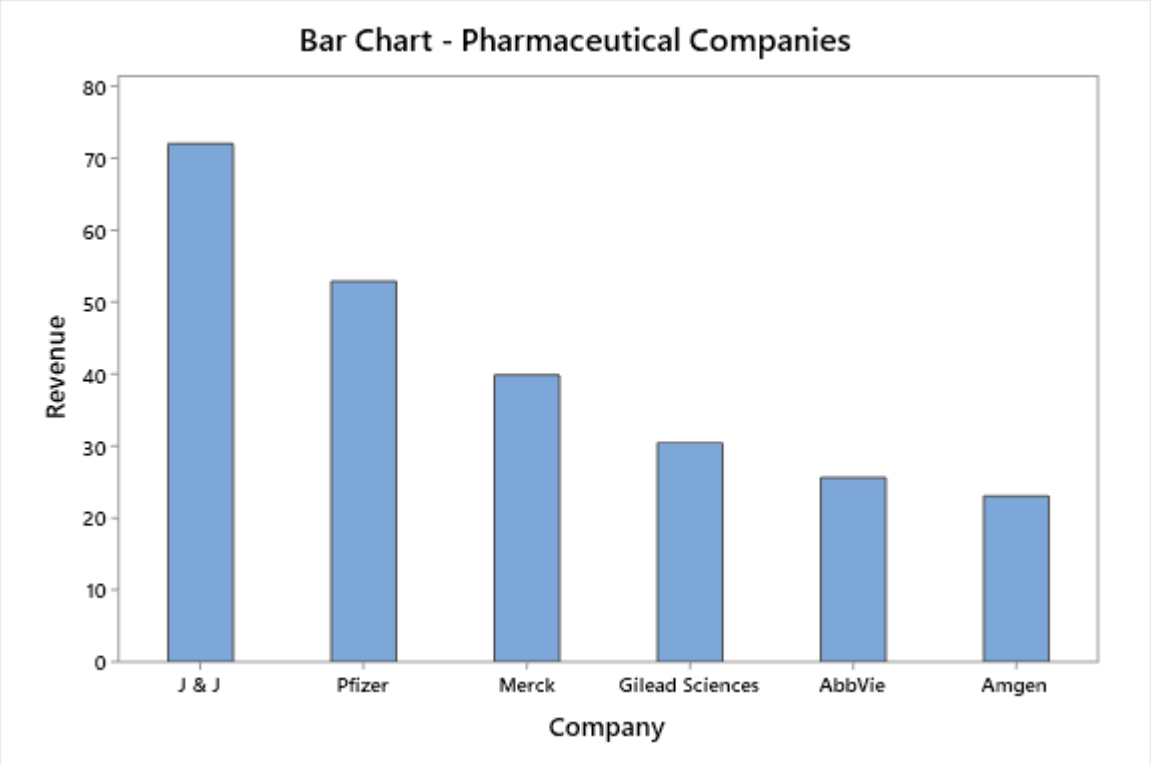
2.17

Brand	Proportion	Degrees
Johnson & Johnson	.295	106.3
Pfizer	.217	78.1
Merck	.163	58.8
Gilead Sciences	.125	44.9
AbbVie	.105	37.8
Amgen	.094	34.0
TOTAL	1.000	360.0

Pie Chart:

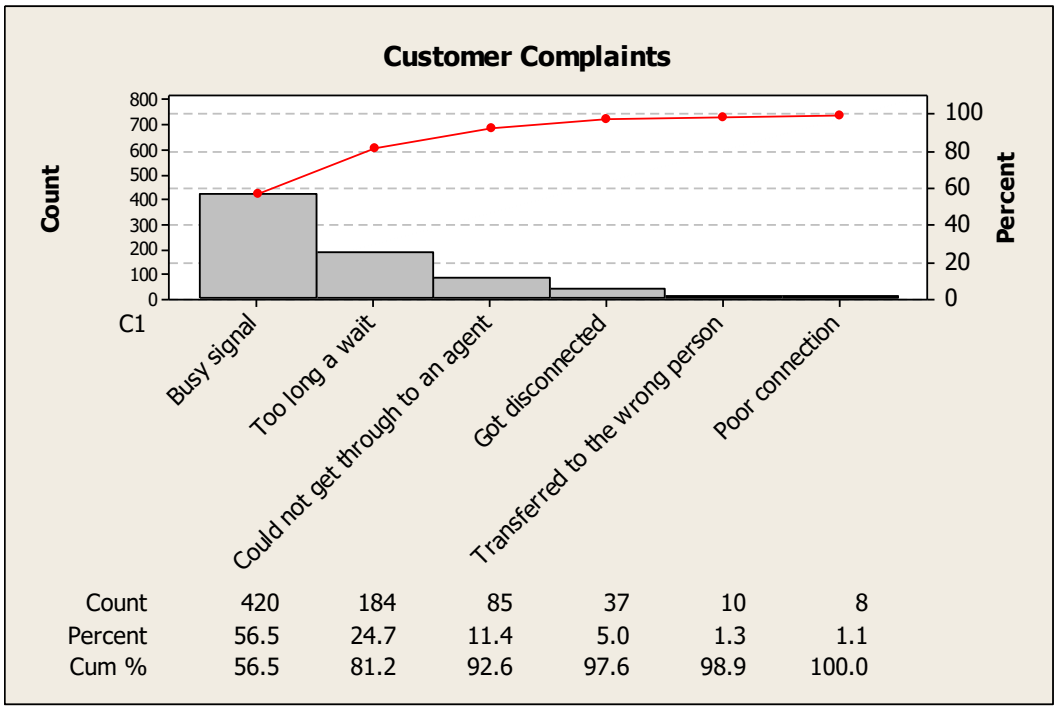


Bar Graph:

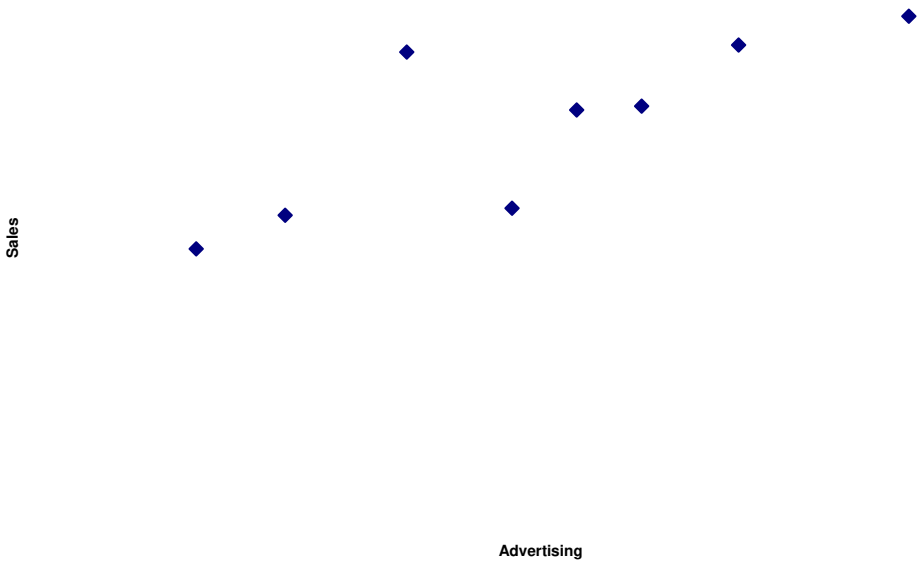


2.19

Complaint	Number	% of Total
Busy Signal	420	56.45
Too long a Wait	184	24.73
Could not get through	85	11.42
Got Disconnected	37	4.97
Transferred to the Wrong Person	10	1.34
Poor Connection	8	1.08
Total	744	99.99



2.21

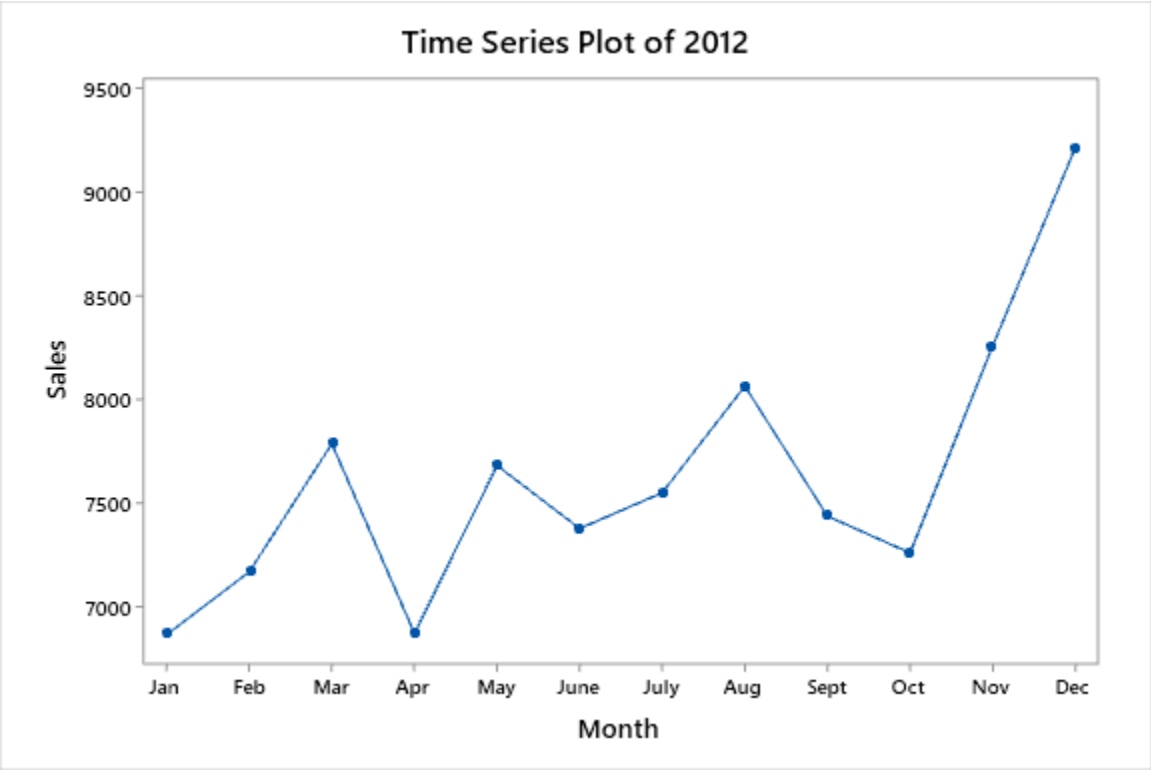


Generally, as advertising dollars increase, sales are increasing.

2.23 There is a slight tendency for there to be a few more absences as plant workers commute further distances. However, compared to the total number of workers in each category, these increases are relatively small (2.5% to 3.0% to 6.6%). Comparing workers who travel 4-10 miles to those who travel 0-3 miles, there is about a 2:1 ratio in all three cells indicating that for these two categories (0-3 and 4-10), number of absences is essentially independent of commute distance.

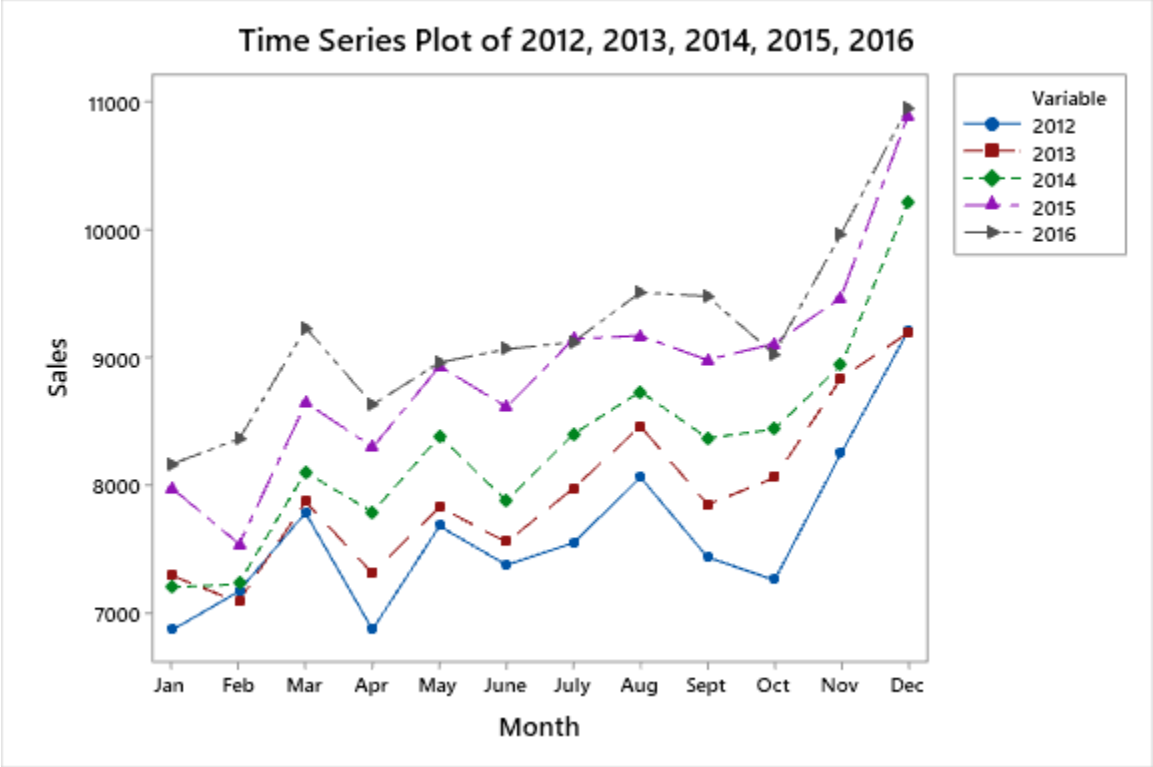
2.25

a.



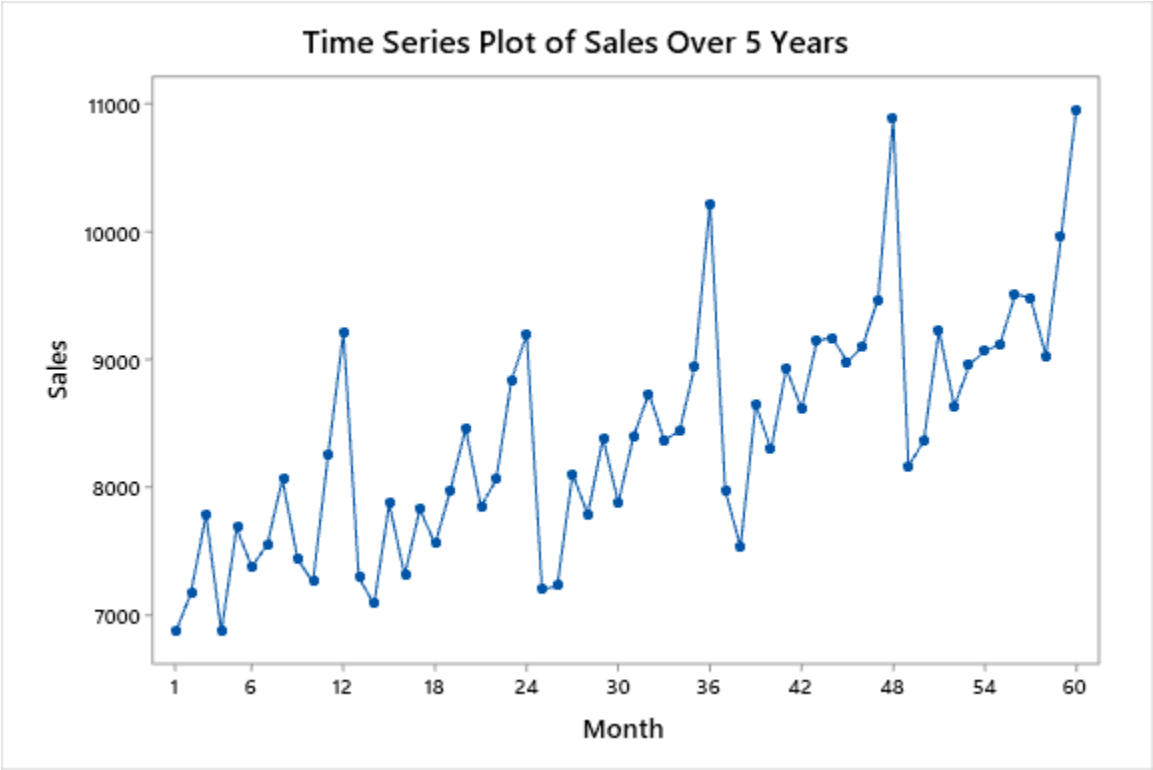
This time-series plot for the year 2012 shows up and down swings of sales until October with sales strongly increasing for November and December due to perhaps the holiday season. If true, we could be seeing possible seasonal effects of the holiday season followed by low sales in January.

b.



When placed together on one time-series plot, the graphs tend to show the same shape virtually every year. This underscores the notion that there is some seasonality in the data. Sales are down in January, April, June, and October. Sales are highest in December, increases every year from October through December, and then drops dramatically in January.

c.



With all five years of data shown chronicalogically, it is apparent that there is a repeating seasonal cycle within each year. In addition, however, there is a general trend of increasing sales over the five years.

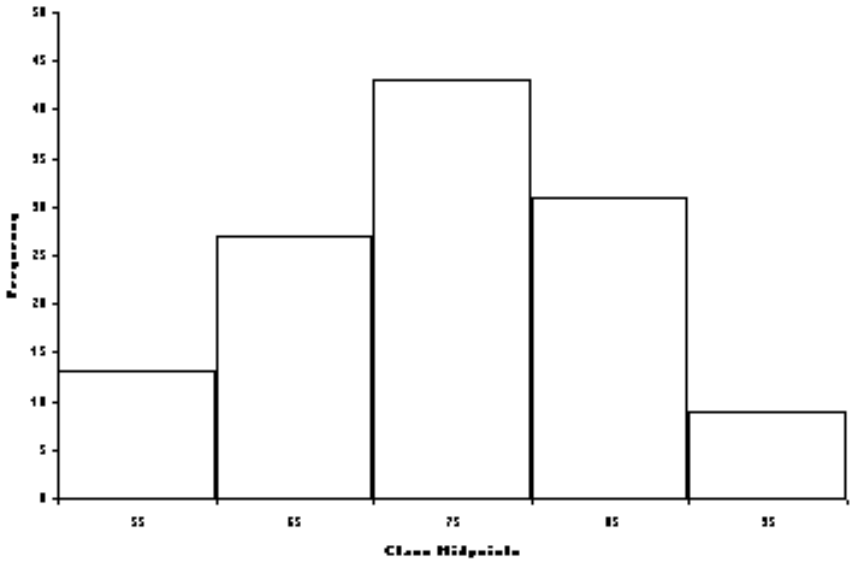
2.27

Class Interval	Frequencies
16 - under 23	6
23 - under 30	9
30 - under 37	4
37 - under 44	4
44 - under 51	4
51 - under 58	3
TOTAL	30

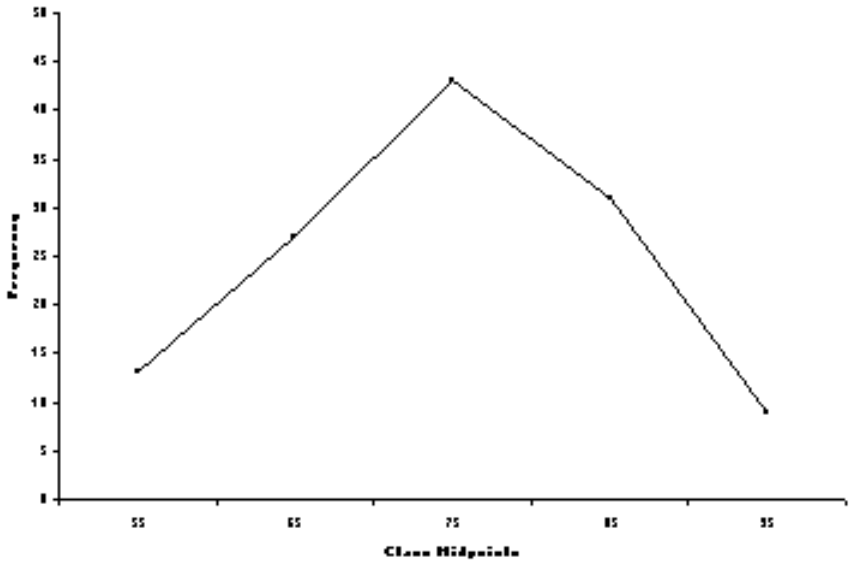
2.29

Class Interval	Frequencies
50 - under 60	13
60 - under 70	27
70 - under 80	43
80 - under 90	31
90 - under 100	9
TOTAL	123

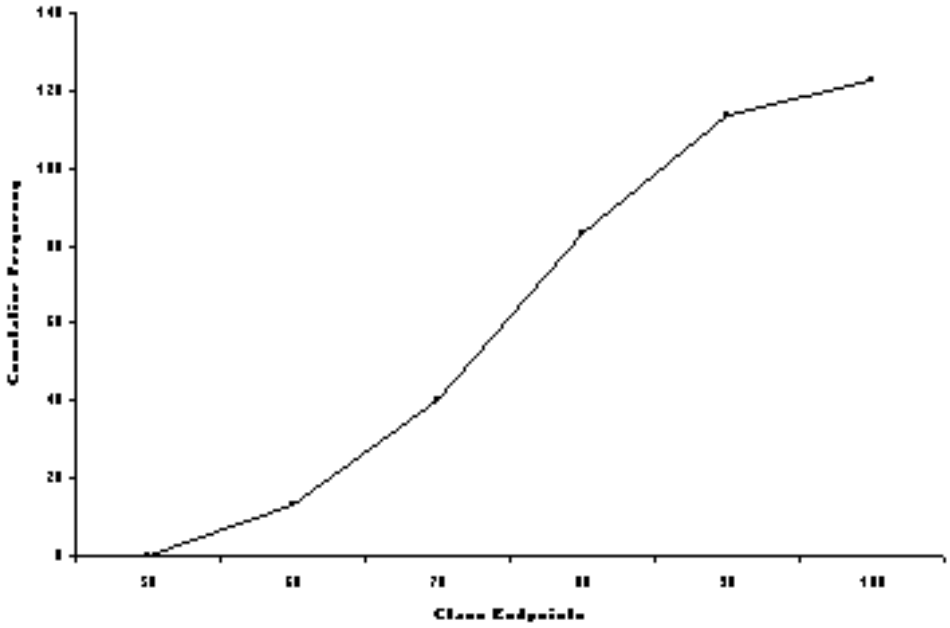
Histogram:



Frequency Polygon:



Ogive:

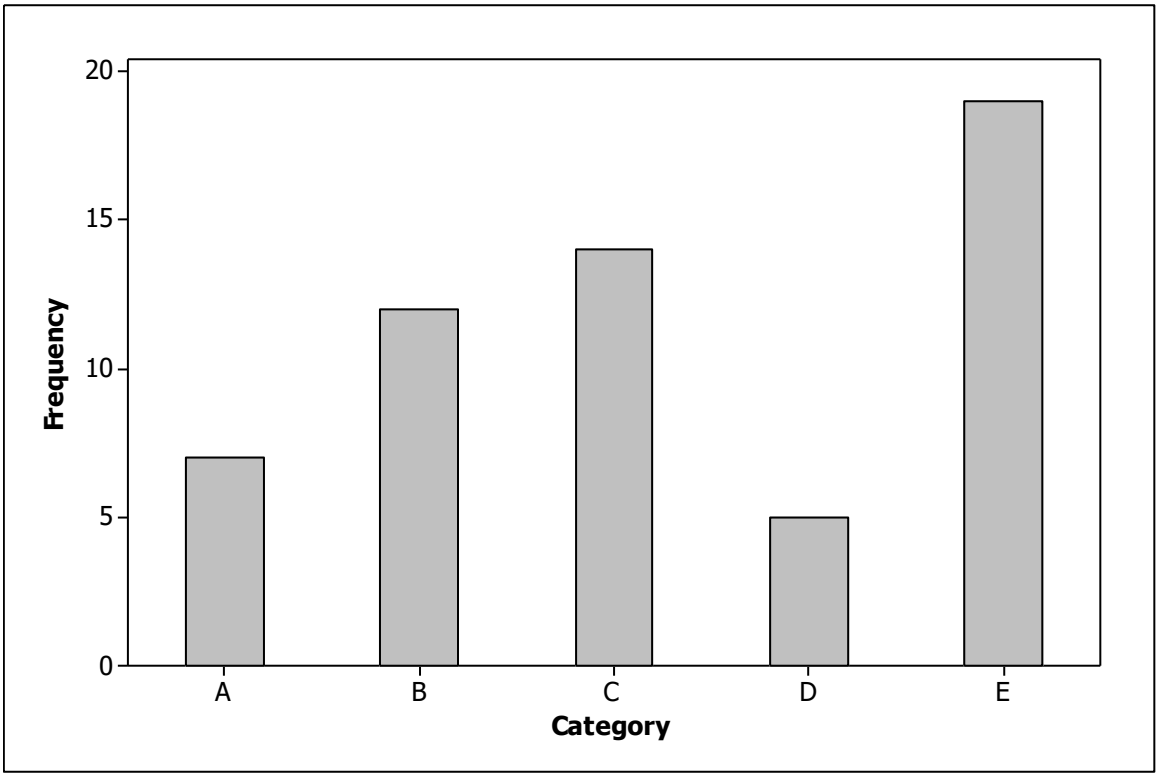


2.31

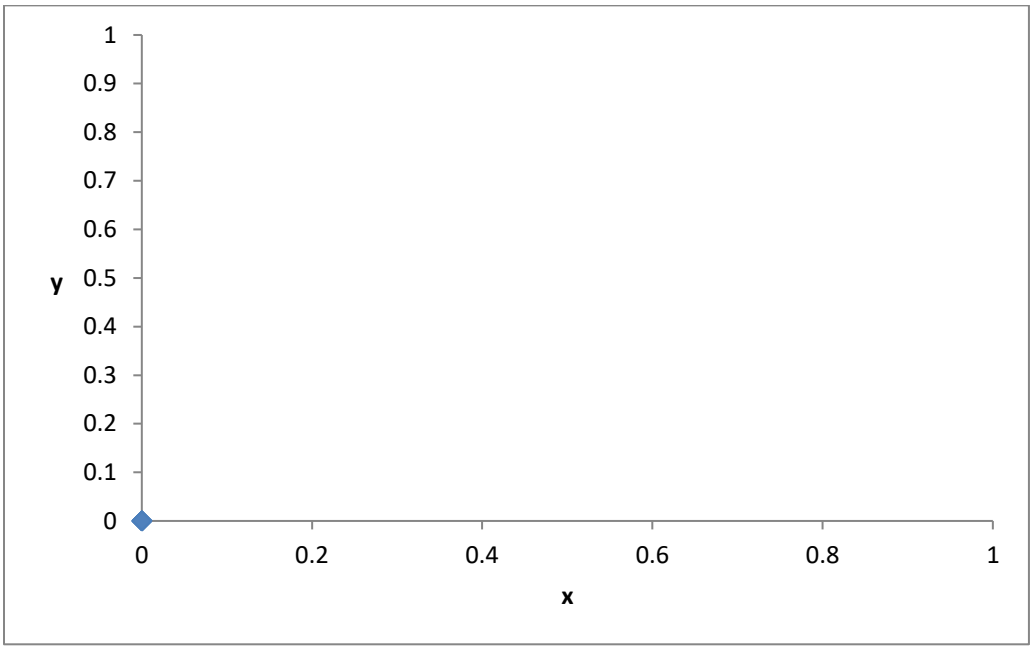
STEM	LEAF
28	4 6 9
29	0 4 8
30	1 6 8 9
31	1 2 4 6 7 7
32	4 4 6
33	5

2.33 Bar Graph:

Category	Frequency
A	7
B	12
C	14
D	5
E	19

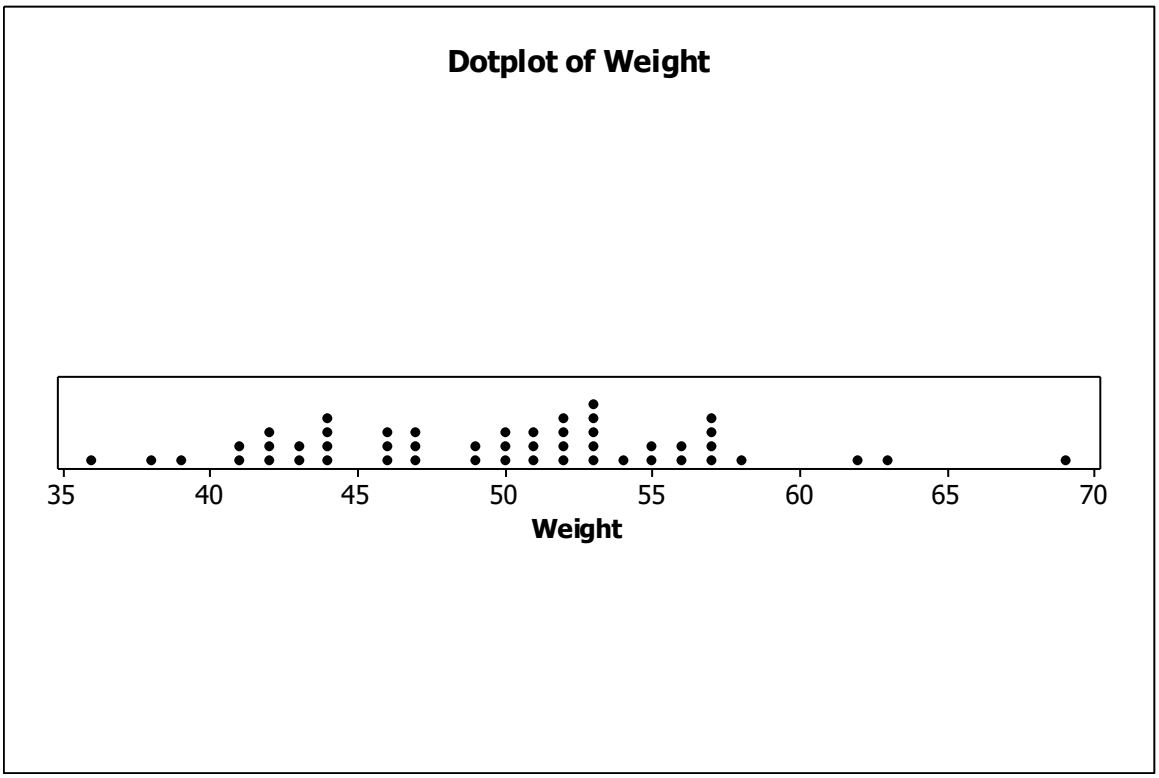


2.35 Scatter Plot



2.37 Whitcomb Company

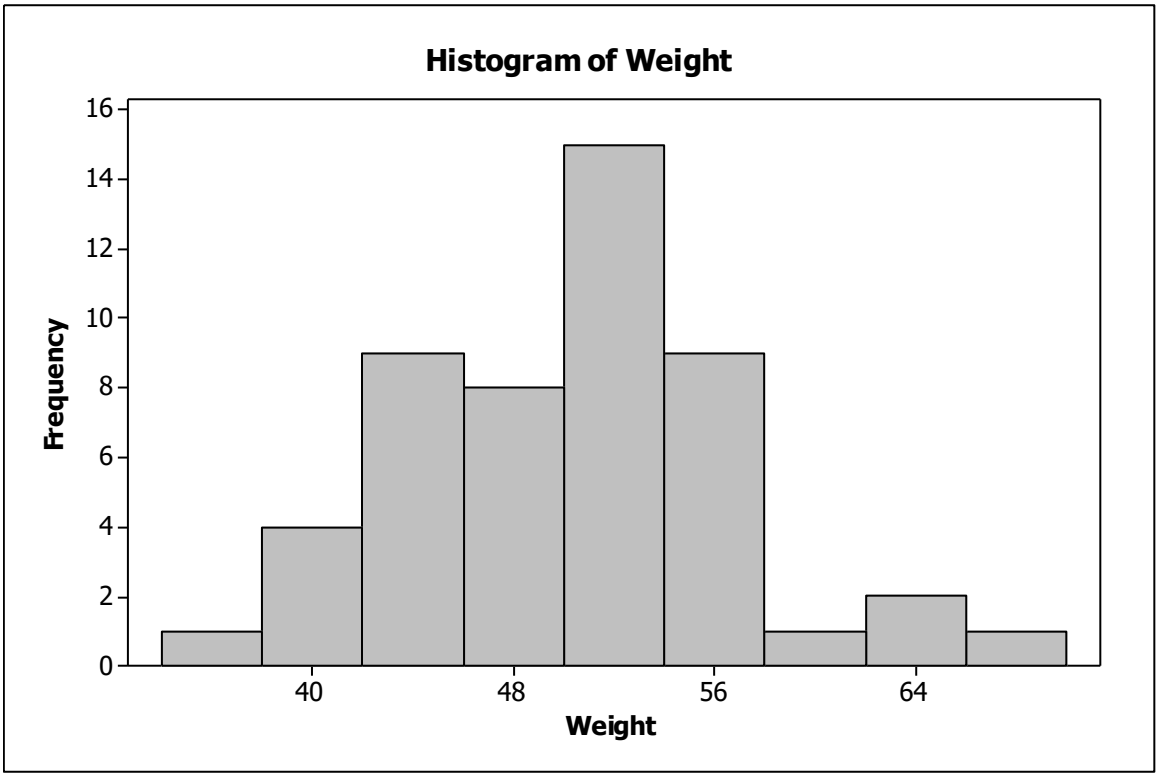
a.) Dot Plot. The metal ring is supposed to weigh around 50 ounces. The dot plot shows that most of the rings weigh between 41 and 58 ounces with the highest number being 53 (in Chapter 3 we will name this as the mode) and other weights piling up at 44, 52, and 57 ounces. There are some extreme values (later in the text termed "outliers") at 36, 38, 39, 62, 63, and even 69 ounces.



b.) Frequency Distribution and Histogram. Shown first here is a frequency distribution with class intervals containing widths of 5 ounces.

Class Interval	Frequency	Cumulative Frequency
32 - under 37	1	1
37 - under 42	4	5
42 - under 47	12	17
47 - under 52	11	28
52 - under 57	14	42
57 - under 62	5	47
62 - under 67	2	49
67 - under 72	1	50
TOTAL	50	

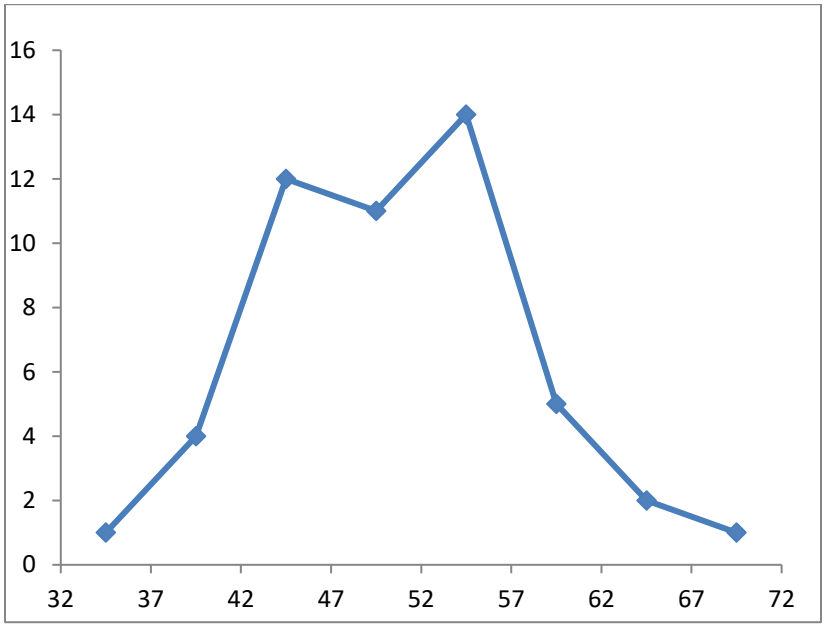
Shown below is a histogram of the weights.



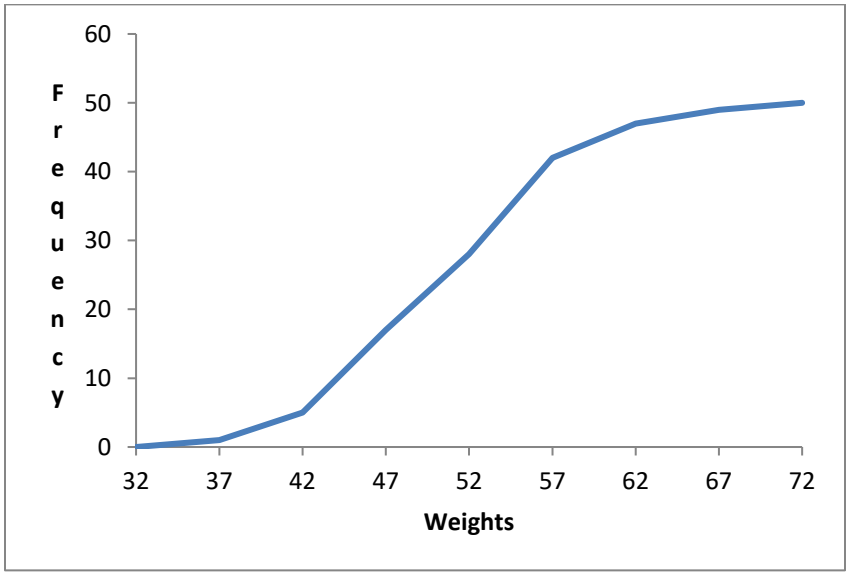
From the histogram, we can see that the bulk of the weights pile up between about 42 ounces and 58 ounces. The distribution appears to have one high point (in Chapter 3 we will call it the modal class – 50 to 54).

c.) Frequency Polygon and Ogive

Frequency Polygon:



Ogive of weights:



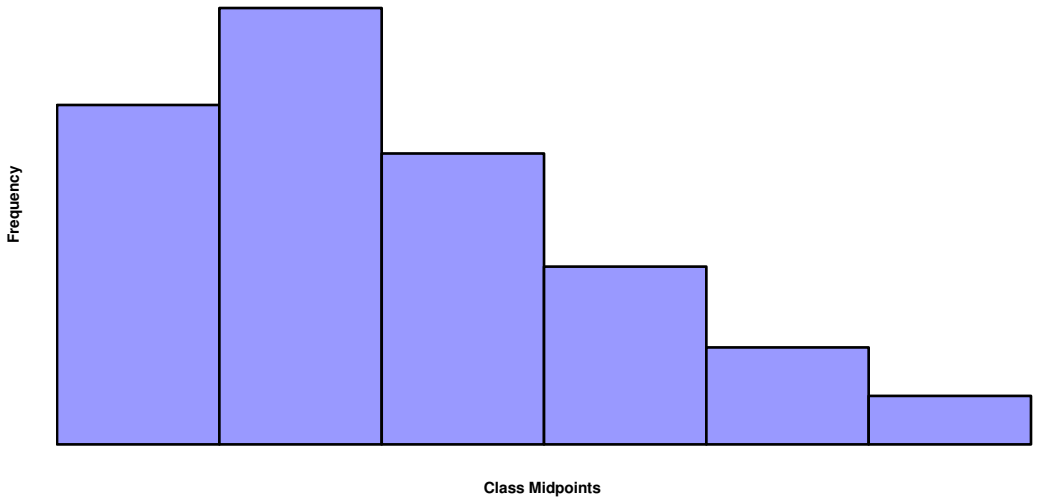
From the frequency polygon, we can see that the highest frequency was in the 52 – 57 class followed by the 42 - 47 class with the 47-52 class not far behind. Frequency values fell dramatically on either side of these three classes. This is underscored by the ogive which shows a relatively steep slope from 42 through 57 (large increases) preceded by and followed by relatively flat slopes indicating only small increases.

2.39 Examining the shape of the distribution, the commute times generally appear in the shape of what we will refer to as the normal curve or bell-shaped curve in Chapter 3. That is, the graph is relatively symmetrical and "piles up" near the middle. Most of the commute times are between about 27 minutes and 53 minutes. A few report commute times of as little as about 10 minutes and a few report times of over 70 minutes. The highest number report commute times of about 40 minutes which seems to be in the middle of the data.

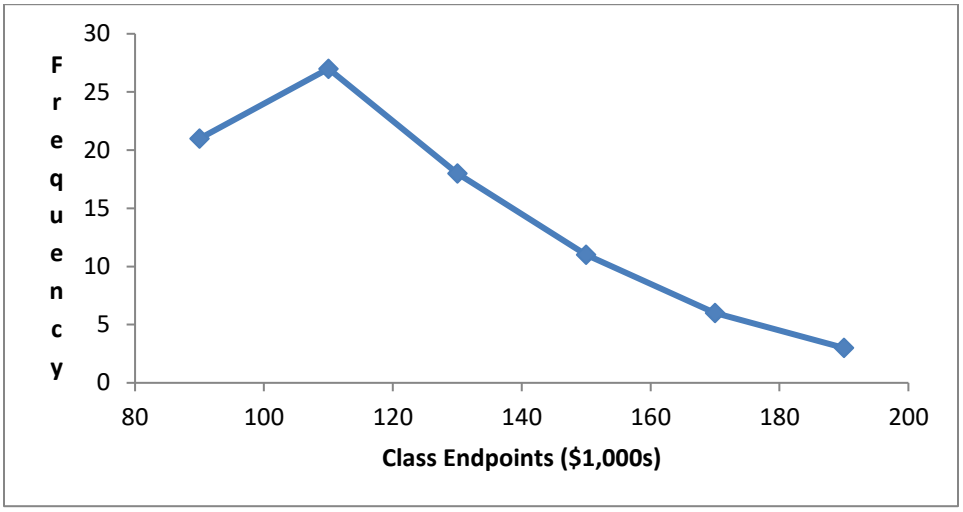
2.41

Asking Price	Frequency	Cumulative Frequency
$ 80,000 - under $ 100,000	21	21
$ 100,000 - under $ 120,000	27	48
$ 120,000 - under $ 140,000	18	66
$ 140,000 - under $ 160,000	11	77
$ 160,000 - under $ 180,000	6	83
$ 180,000 - under $ 200,000	3	86
	86	

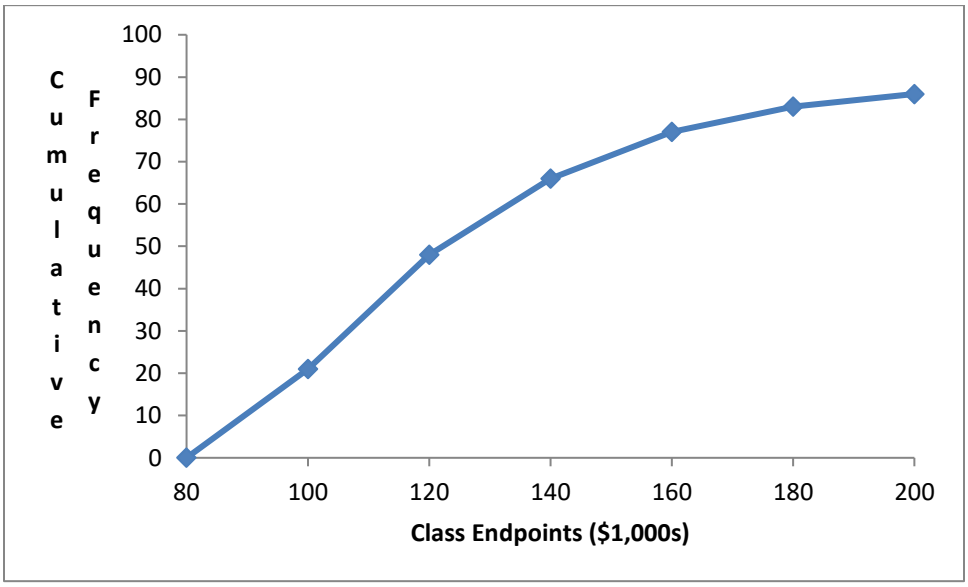
Histogram:



Frequency Polygon:



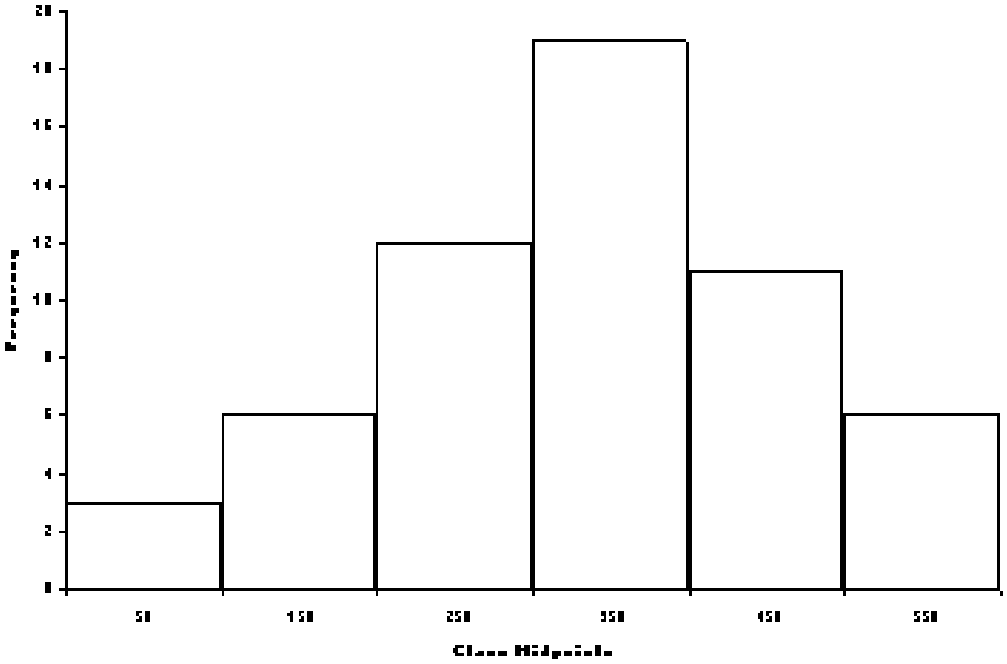
Ogive:



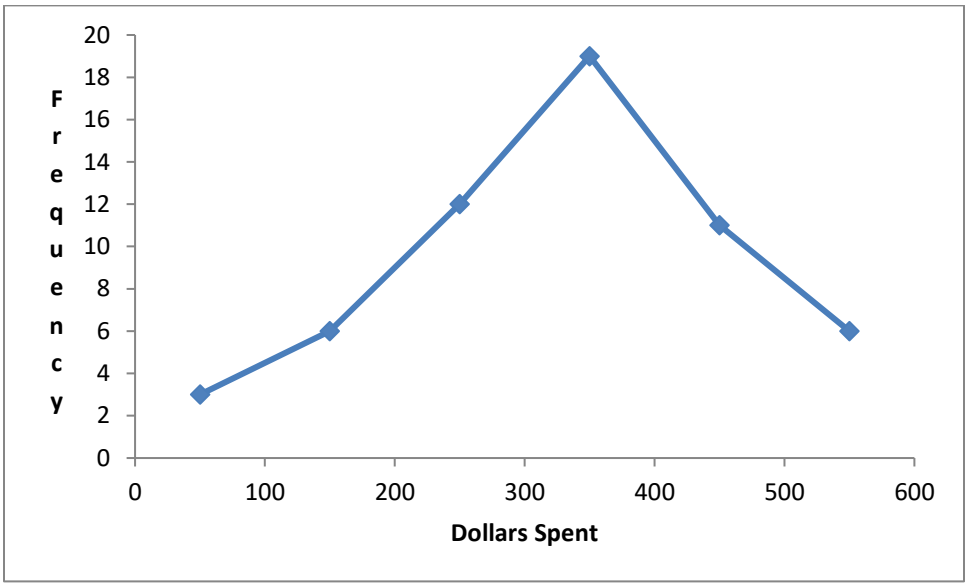
2.43

Amount Spent on Prenatal Care	Frequency	Cumulative Frequency
$ 0 - under $100	3	3
$100 - under $200	6	9
$200 - under $300	12	21
$300 - under $400	19	40
$400 - under $500	11	51
$500 - under $600	6	57
	57	

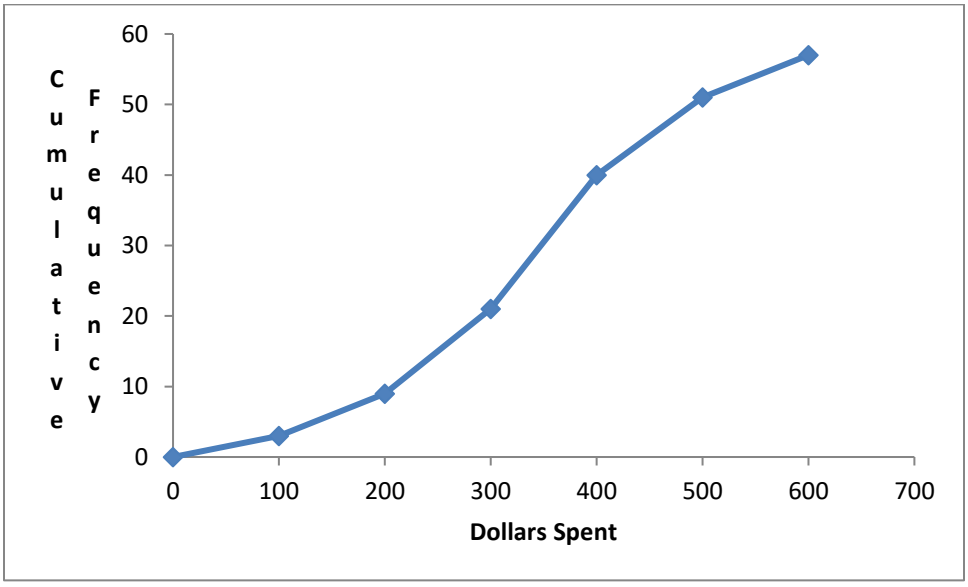
Histogram:



Frequency Polygon:



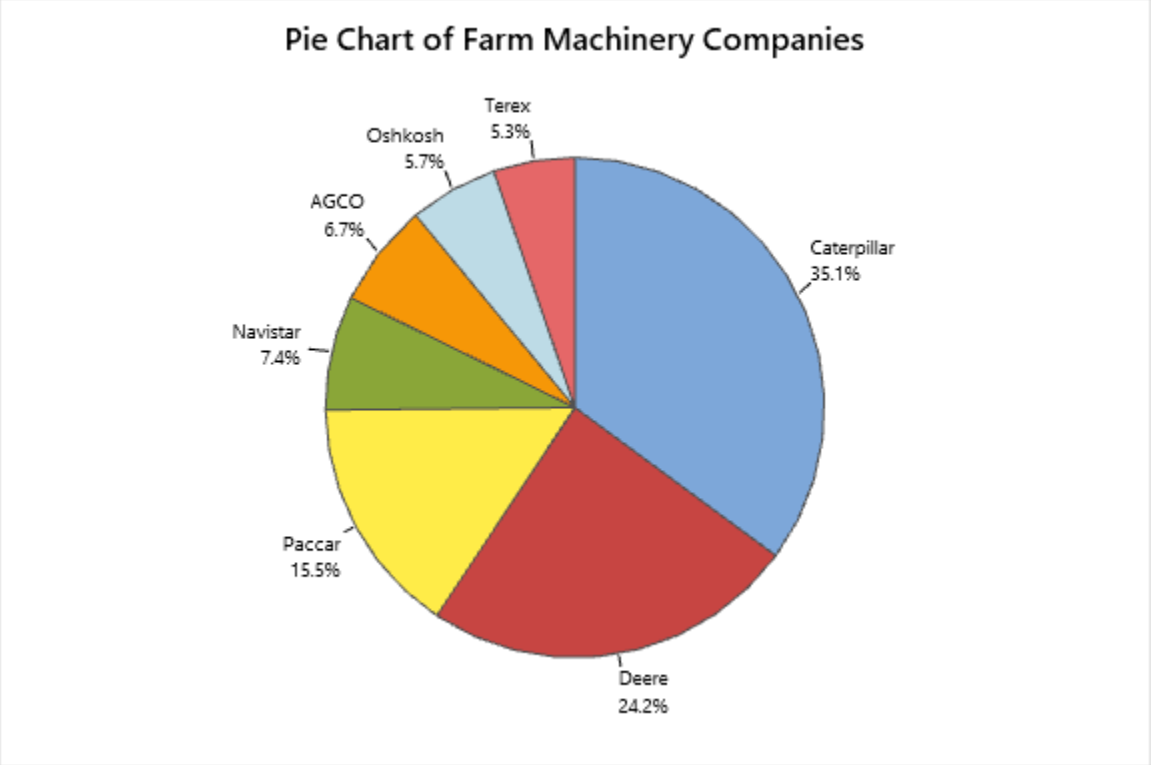
Ogive:



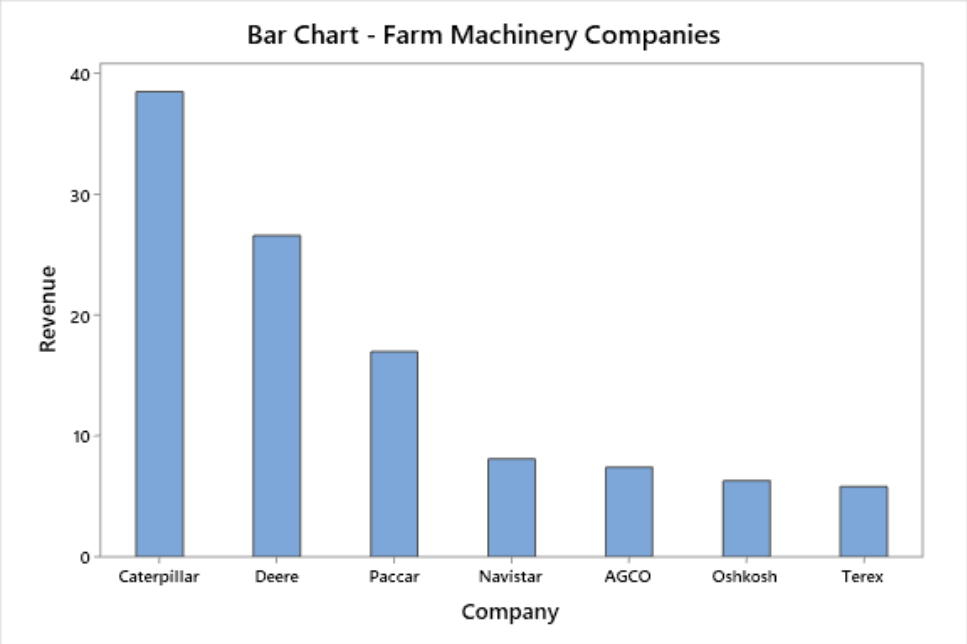
2.45

Company	Revenue	Proportion	Degrees
Caterpillar	38.5	.35	126.3
Deere	26.6	.24	87.3
Paccar	17.0	.15	55.8
Navistar International	8.1	.07	26.6
AGCO	7.4	.07	24.3
Oshkosh	6.3	.06	20.7
Terex	5.8	.05	19.0
TOTAL	109.7	1.00	360.0

a.) Pie Chart:

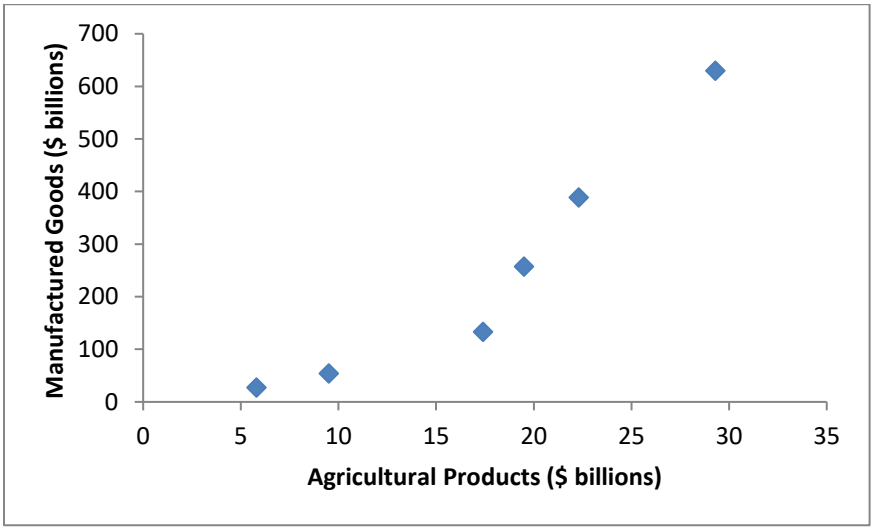


b.) Bar Chart:



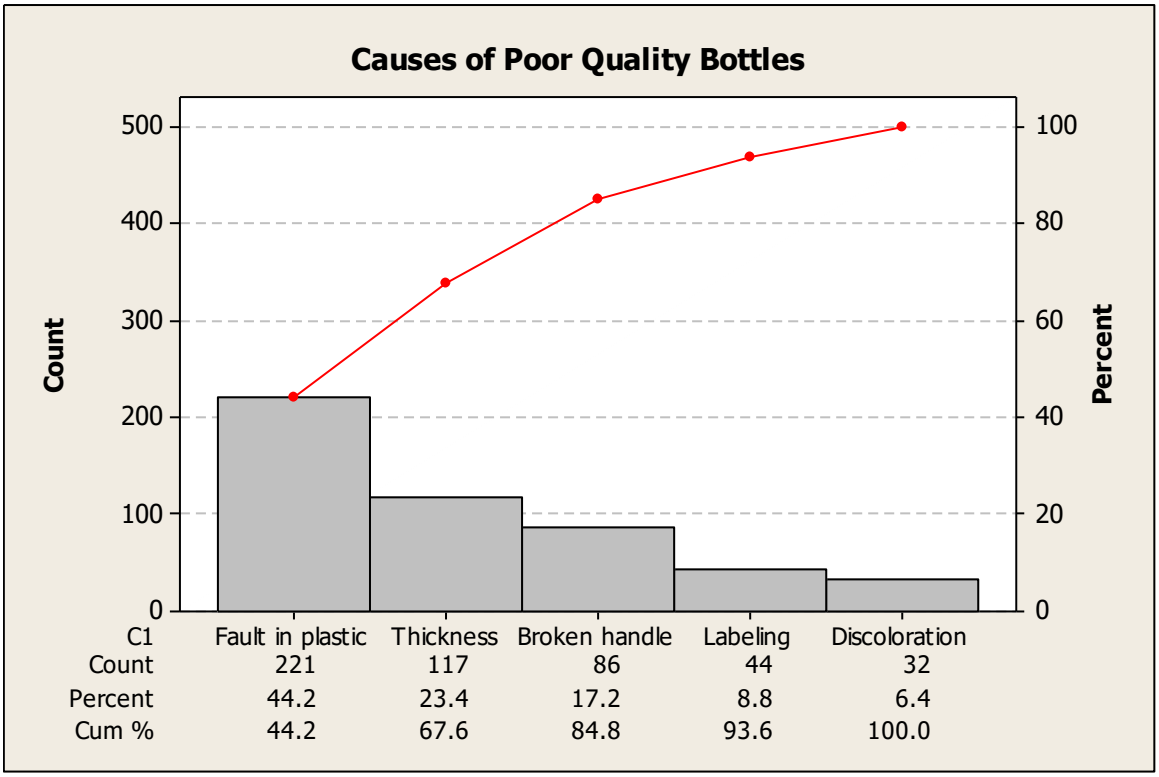
c.) This problem points out the advantage of the bar chart over the pie chart. In the bar chart it is more evident that the Navistar sales are greater than AGCO which are greater than Oshkosh which are greater than Terex. In the pie chart, without the labels, differentiating between these four company revenues would be difficult.

2.47



It can be observed that as the U.S. import of agricultural products increased, the U.S. import of manufactured goods also increased. As a matter of fact, a non-linear association may exist between the two variables.

2.49



One of the main purposes of a Pareto chart is that it has the potential to help prioritize quality initiatives by ranking the top problems in order starting with the most frequently occurring problem. Thus, all things being equal, in attempting to improve the quality of plastic bottles, a quality team would begin with studying why there is a fault in plastic and determining how to correct for it. Next, the quality team would study thickness issues followed by causes of broken handles. Assuming that each problem takes a comparable time and effort to solve, the quality team could make greater strides sooner by following the items shown in the Pareto chart from left to right.

2.51 The distribution of household income is bell-shaped with an average of about $ 90,000 and a range of from $ 30,000 to $ 140,000.

2.53 We can conclude from the pie chart that of the six specialties graphed, emergency medicine is the smallest and internal medicine is the largest. General practice is so close in slice size to internal medicine that without the labels,it would be difficult to distinguish between the two. Pediatrics is number three but only about ½ the size of each of the first two largest. Anesthesiology and OB-GYN are in a virtual tie.

2.55 There appears to be a relatively strong positive relationship between the NASDAQ-100 and the DJIA. Note that generally as the DJIA became higher, the NASDAQ-100 tended to also get higher. Except for a few outliers, the slope of the graph was generally the same over time.

Chapter 2
Visualizing Data with Charts and Graphs

LEARNING OBJECTIVES

The overall objective of Chapter 2 is for you to master several techniques for summarizing and visualizing data, thereby enabling you to:

1. Construct a frequency distribution from a set of data
2. Construct different types of quantitative data graphs, including histograms, frequency polygons, ogives, dot plots, and stem-and-leaf plots, in order to interpret the data being graphed
3. Construct different types of qualitative data graphs, including pie charts, bar graphs, and Pareto charts, in order to interpret the data being graphed
4. Construct a cross-tabulation table and recognize basic trends in two-variable scatter plots of numerical data.
5. Construct a time-series graph and be able to visually identify any trends in the data.

CHAPTER TEACHING STRATEGY

Chapter 1 brought to the attention of students the exponential growth in data available to decision makers to assist them in producing better business outcomes. Such data come rapidly in a wide variety and often in great volume. In chapter 2, we begin to tackle the problem of trying to summarize and present data in a meaningful manner. To this end, in Chapter 2 the student is exposed to a plethora of graphical and visualizing techiques. Getting an initial overview of the data is an important first step. This often takes place by the analyst studying a graphical representation of the data. In addition, finding hidden information and discovering new relationships in the data through visualization can be very helpful to business decision makers. Remember, that it has been said that data visualization is perhaps the most useful component of business analytics and that utilizing visual techniques to convey information gleaned from the data allows for a broader audience with a wider range of backgrounds to view and understand its meaning.

One mechanism for data summarization is the frequency distribution which is essentially a way of organizing ungrouped or raw data into grouped data. It is important to realize that there is considerable art involved in constructing a frequency distribution. There are nearly as many possible frequency distributions for a problem as there are students in a class. Students should begin to think about the receiver or user of their statistical product. For example, what class widths and class endpoints would be most

familiar and meaningful to the end user of the distribution? How can the data best be communicated and summarized using the frequency distribution?

Chapter 2 presents various ways to depict data using graphs. The student should view these graphical techniques as tools for use in communicating characteristics of the data in an effective manner. Most business students will have some type of management opportunity in their field before their career ends. The ability to make effective presentations and communicate their ideas in succinct, clear ways is an asset. Through the use of graphics packages and such techniques as frequency polygons, ogives, histograms, and pie charts, the manager can enhance his/her personal image as a communicator and decision-maker – which also are characteristics of a good business analyst. In addition, emphasize that the final product (the frequency polygon, etc.) is just the beginning. Students should be encouraged to study the graphical output to recognize business trends, highs, lows, etc. and realize that the ultimate goal for these tools is their usage in decision making.

CHAPTER OUTLINE

2.1 Frequency Distributions
- Class Midpoint
- Relative Frequency
- Cumulative Frequency

2.2 Quantitative Data Graphs
- Histograms
  - Using Histograms to Get an Initial Overview of the Data
- Frequency Polygons
- Ogives
- Dot Plots
- Stem-and-Leaf Plots

2.3 Qualitative Data Graphs
- Pie Charts
- Bar Graphs
- Pareto Charts

2.4 Charts and Graphs for Two Variables
- Cross Tabulation
- Scatter Plot

2.5 Visualizing Time-Series Data

KEY TERMS

Bar Charts
Bar Graph
Class Mark
Class Midpoint
Column Charts
Cross Tabulation
Cumulative Frequency
Dot Plot
Frequency Distribution
Frequency Polygon
Grouped Data
Histogram
Ogive
Pareto Chart
Pie Chart
Range
Relative Frequency
Scatter Plot
Stem-and-Leaf Plot
Time-Series Data
Ungrouped Data

SOLUTIONS TO PROBLEMS IN CHAPTER 2

2.1

a) One possible 5 class frequency distribution:

Class Interval	Frequency
0 - under 20	7
20 - under 40	15
40 - under 60	12
60 - under 80	12
80 - under 100	4
	50

b) One possible 10 class frequency distribution:

Class Interval	Frequency
10 - under 18	7
18 - under 26	3
26 - under 34	5
34 - under 42	9
42 - under 50	7
50 - under 58	3
58 - under 66	6
66 - under 74	4
74 - under 82	4
82 - under 90	2

c) The ten class frequency distribution gives a more detailed breakdown of temperatures, pointing out the smaller frequencies for the higher temperature intervals. The five class distribution collapses the intervals into broader classes making it appear that there are nearly equal frequencies in each class.

2.2 a) One possible frequency distribution is the one below with 12 classes and class intervals of 2.

Class Interval	Frequency
39 - under 41	2
41 - under 43	1
43 - under 45	5
45 - under 47	10
47 - under 49	18
49 - under 51	13
51 - under 53	15
53 - under 55	15
55 - under 57	7
57 - under 59	9
59 - under 61	4
61 – under 63	1

b) The distribution reveals that only 13 of the 100 boxes of raisins contain 50 ± 1 raisin (49 -under 51). However, 71 of the 100 boxes of raisins contain between 45 and 55 raisins. It shows that there are five boxes that have 9 or more extra raisins (59-61 and 61-63) and two boxes that have 9-11 less raisins (39-41) than the boxes are supposed to contain.

2.3

Class Interval	Frequency	Class Midpoint	Relative Frequency	Cumulative Frequency
0 - 5	6	2.5	6/86 = .0698	6
5 - 10	8	7.5	.0930	14
10 - 15	17	12.5	.1977	31
15 - 20	23	17.5	.2674	54
20 - 25	18	22.5	.2093	72
25 - 30	10	27.5	.1163	82
30 - 35	4	32.5	.0465	86
TOTAL	86		1.0000	

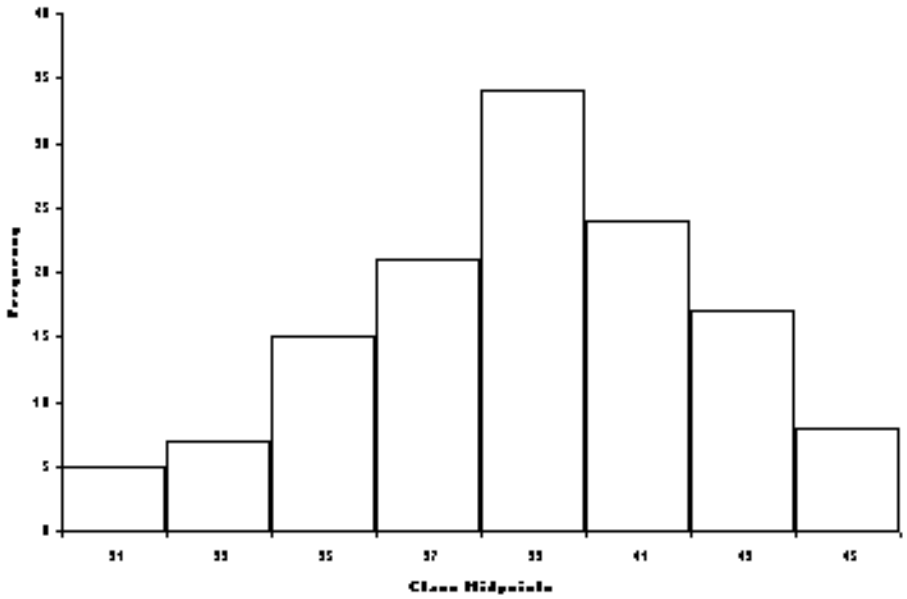
The relative frequency tells us that it is most probable that a customer is in the 15 - 20 category (.2674). Over two thirds (.6744) of the customers are between 10 and 25 years of age.

2.4

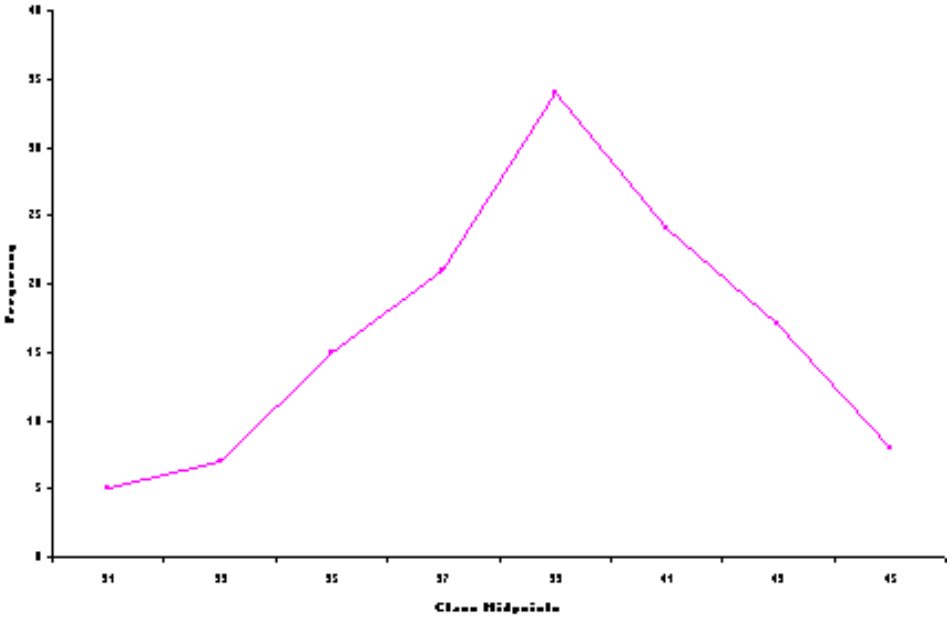
Class Interval	Frequency	Class Midpoint	Relative Frequency	Cumulative Frequency
0-2	218	1	.436	218
2-4	207	3	.414	425
4-6	56	5	.112	481
6-8	11	7	.022	492
8-10	8	9	.016	500
TOTAL	500		1.000	

2.5 Some examples of cumulative frequencies in business:
sales for the fiscal year,
costs for the fiscal year,
spending for the fiscal year,
inventory build-up,
accumulation of workers during a hiring buildup,
production output over a time period.

2.6 Histogram:

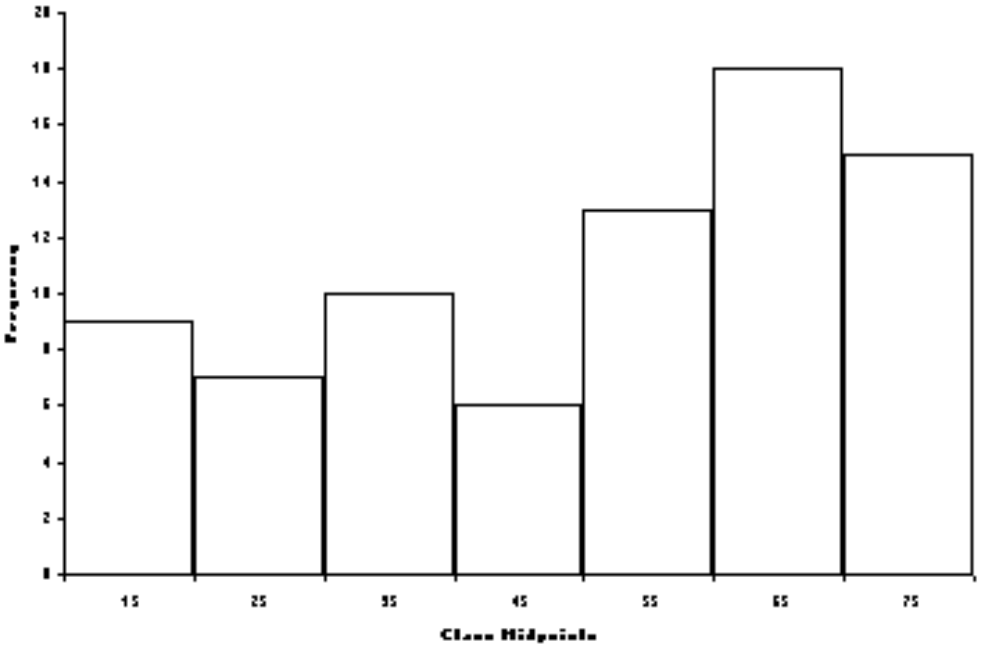


Frequency Polygon:

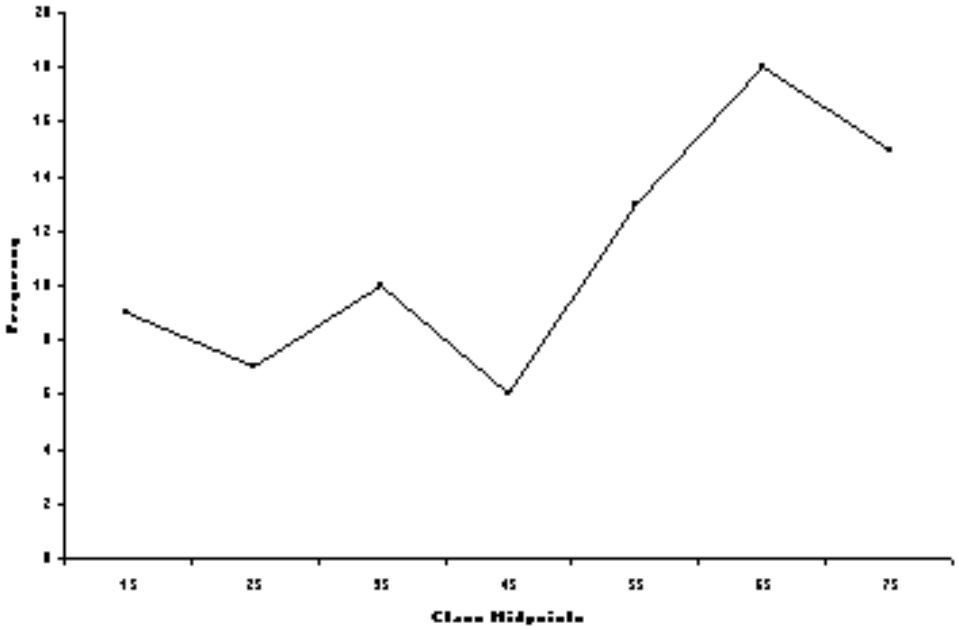


Comment: The assembly times "pile up" near the middle of the graphs indicating that many of the assembly times are between 36 and 42 minutes.

2.7 Histogram:

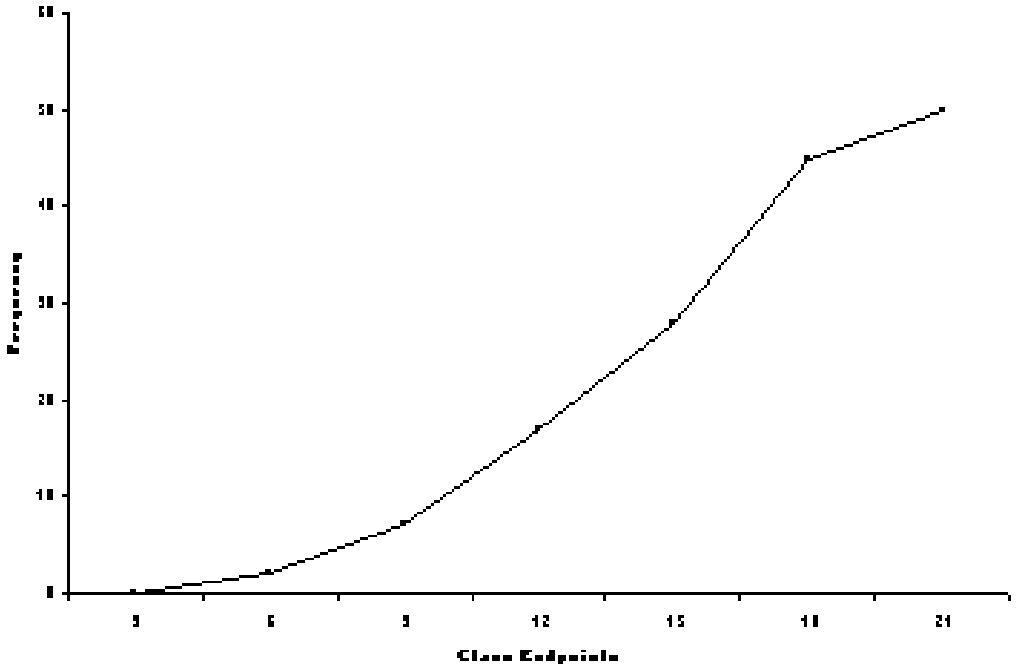


Frequency Polygon:



Comment: The histogram indicates that the number of calls per shift varies widely. However, the heavy numbers of calls per shift fall in the 50 to 80 range. Since these numbers occur quite frequently, staffing planning should be done with these number of calls in mind realizing from the rest of the graph that there may be shifts with as few as 10 to 20 calls.

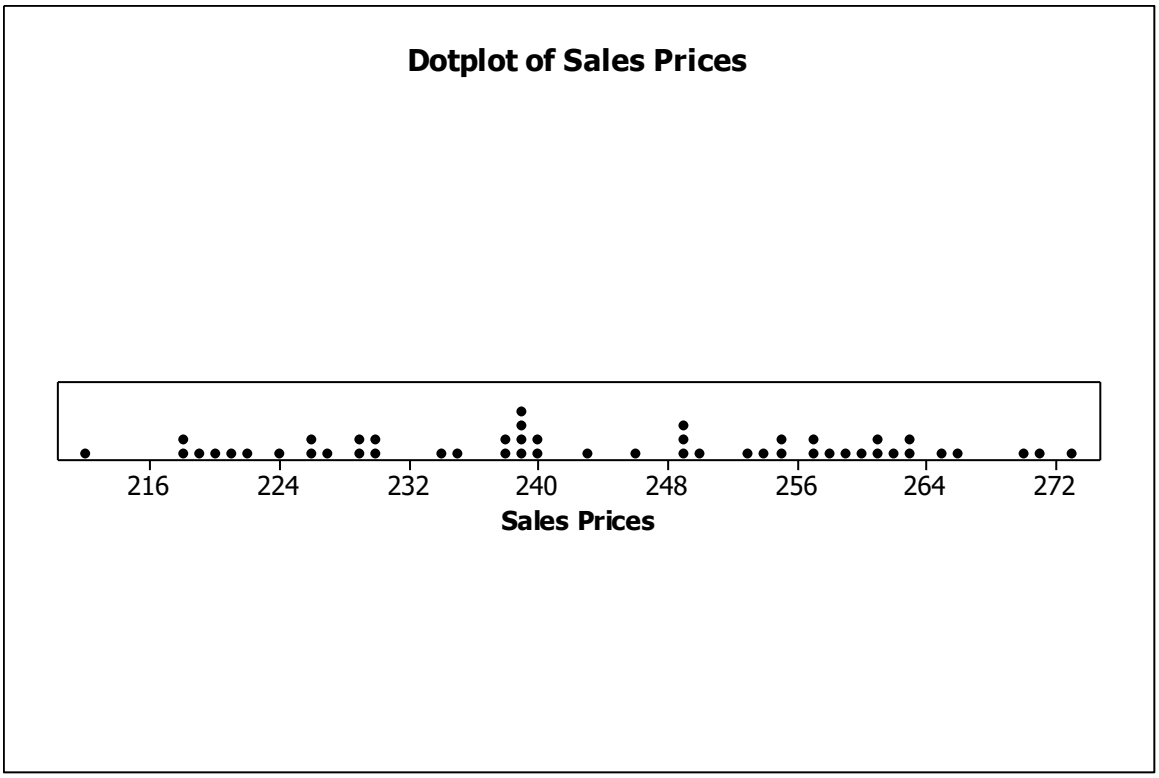
2.8 Ogive:



2.9

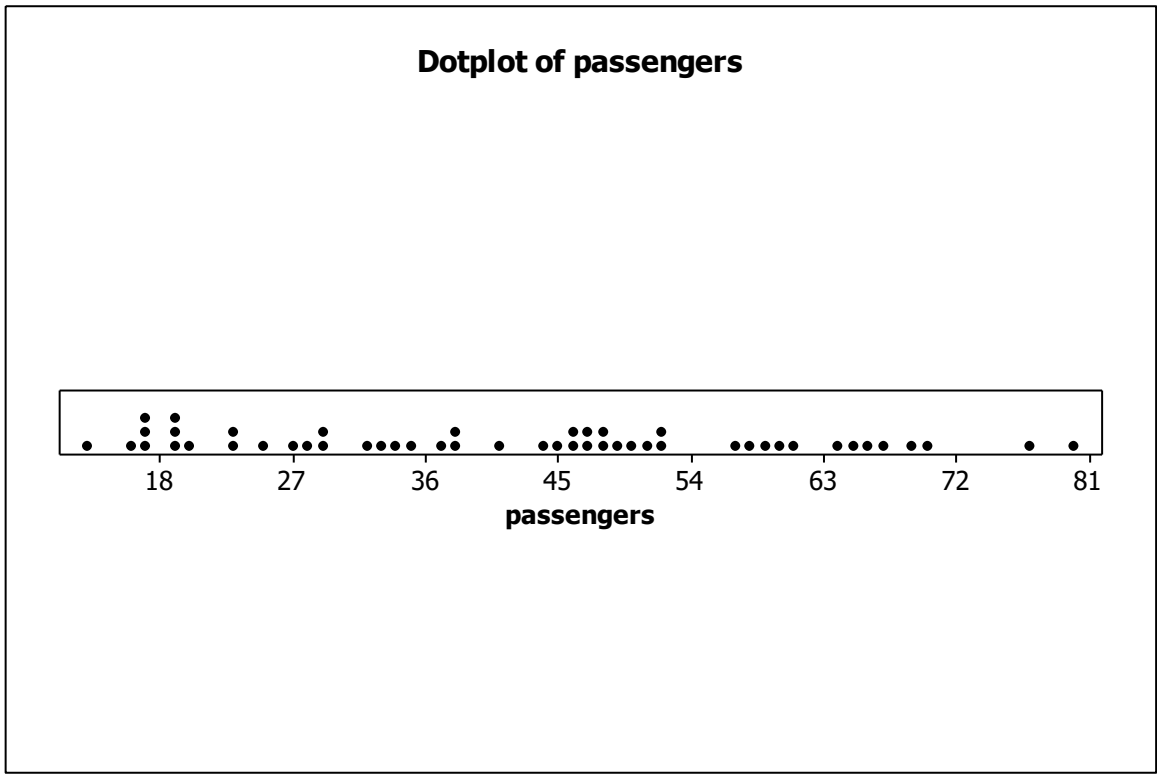
STEM	LEAF
21	2 8 8 9
22	0 1 2 4 6 6 7 9 9
23	0 0 4 5 8 8 9 9 9 9
24	0 0 3 6 9 9 9
25	0 3 4 5 5 7 7 8 9
26	0 1 1 2 3 3 5 6
27	0 1 3

Dotplot



Both the stem and leaf plot and the dot plot indicate that sales prices vary quite a bit within the range of $212,000 and $273,000. It is more evident from the stem and leaf plot that there is a strong grouping of prices in the five price ranges from the $220's through the $260's.

2.10 a)



b)

STEM	LEAF
1	3 6 7 7 7 9 9 9
2	0 3 3 5 7 8 9 9
3	2 3 4 5 7 8 8
4	1 4 5 6 6 7 7 8 8 9
5	0 1 2 2 7 8 9
6	0 1 4 5 6 7 9
7	0 7
8	0

The stem and leaf plot shows that the number of passengers per flight were relatively evenly distributed between the high teens through the sixties. Rarely was there a flight with at least 70 passengers. The category of 40's contained the most flights (10).

2.11 From the information given in the problem, we know that the busiest airport in the world is Atlanta's Hartsfield-Jackson International Airport which has almost 104 million passengers per year. The histogram shows that there are only two airports with between 95 and 105 million passengers and two other airports with between 85 and 95 million passengers. Eleven airports (more than one-third) served between 55 and 65 million passengers. Six airports served between 45 and 55 million and six other airports served 65 to 75 million.

2.12 The dotplot shows that all but one state have less than 100,000 farms. There is only one state with more than 105,000 farms and that state has around 240,000 farms. Most of the states have between a few hundred and about 90,000 farms. There are six states that appear to have around 5,000 farms (modal number). Five states have about 10,000 farms, five states have about 25,000 farms, and five states have about 35,000 farms. The actual mean for the data (including the one state with the very large number of farms) is 41,200 (not shown in graph). While this mean is influenced by the one state with an exceedingly large number of farms, it falls relatively near the middle of the other 49 states and it fairly representative of the other states.

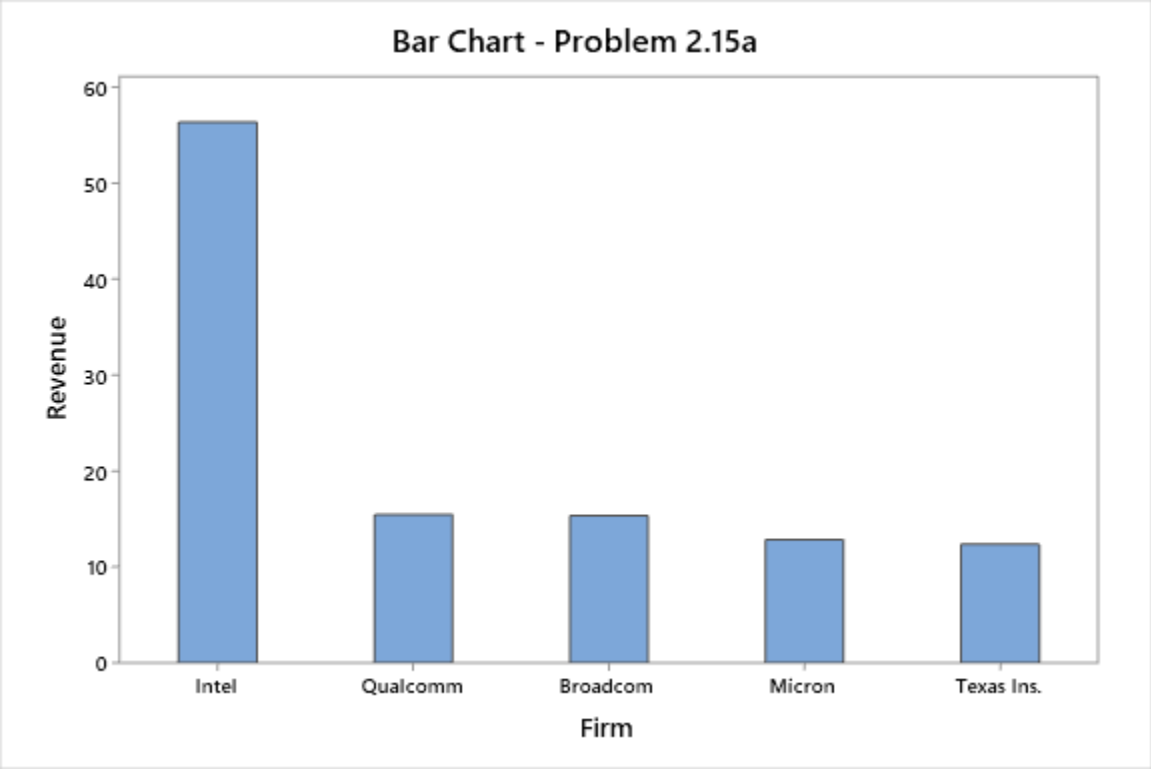
2.13 From the stem and leaf display, the original raw data can be obtained. For example, the fewest number of cars washed on any given day are 25, 29, 29, 33, etc. The most cars washed on any given day are 141, 144, 145, and 147. The modal stems are 3, 4, and 10 in which there are 6 days with each of these numbers. Studying the left column of the Minitab output, it is evident that the median number of cars washed is 81. There are only two days in which 90 some cars are washed (90 and 95) and only two days in which 130 some cars are washed (133 and 137).

2.14 The dot plot tells us several things. Out of 103 pots, 25 of them (amost ¼) contained only 10 King crabs. From the dot plot, it is possible to see that 59 (or more than ½) have 30 or fewer crabs. Almost ¾ of the pots have fewer than 60 crabs. A quick observation of the graph shows that only a very small number (about 5%) have as many as 100 crabs.

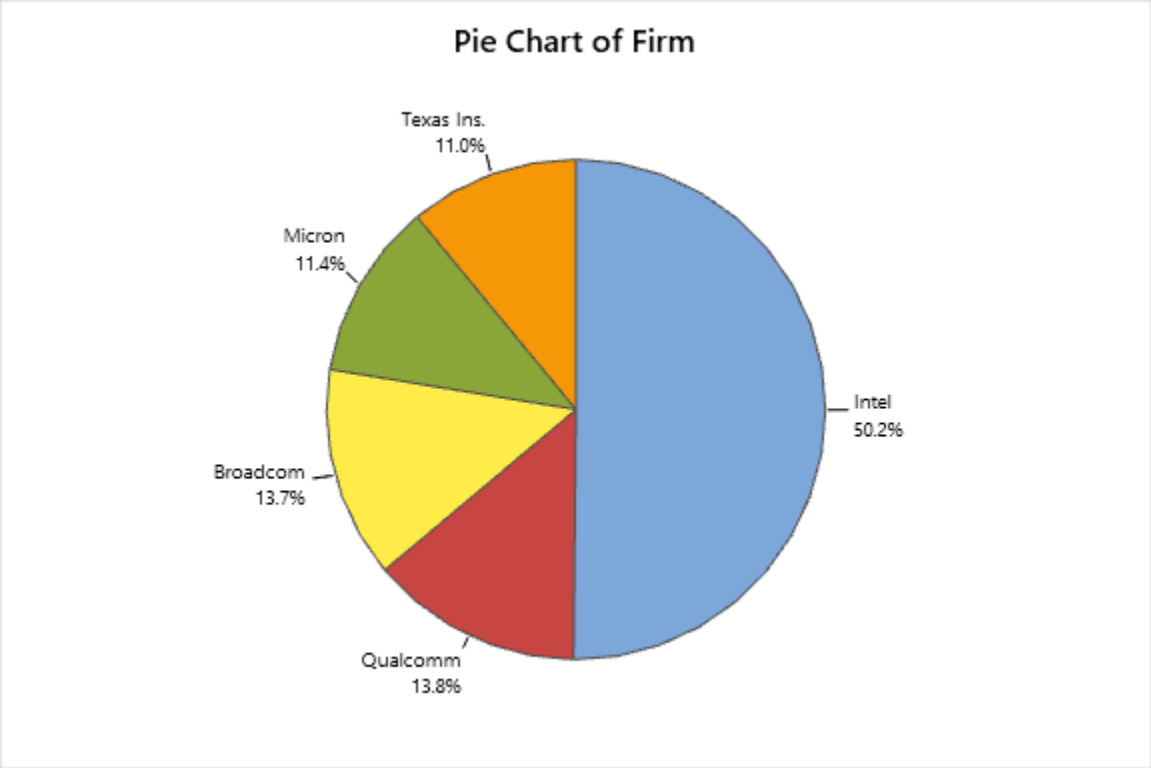
2.15

Firm	Proportion	Degrees
Intel Corp.	.502	180.7
Qualcomm	.138	49.7
Broadcom	.137	49.3
Micron Tech	.114	41.0
Texas Instru.	.110	39.6
TOTAL	1.001	360.3

a.) Bar Graph:



b.) Pie Chart:

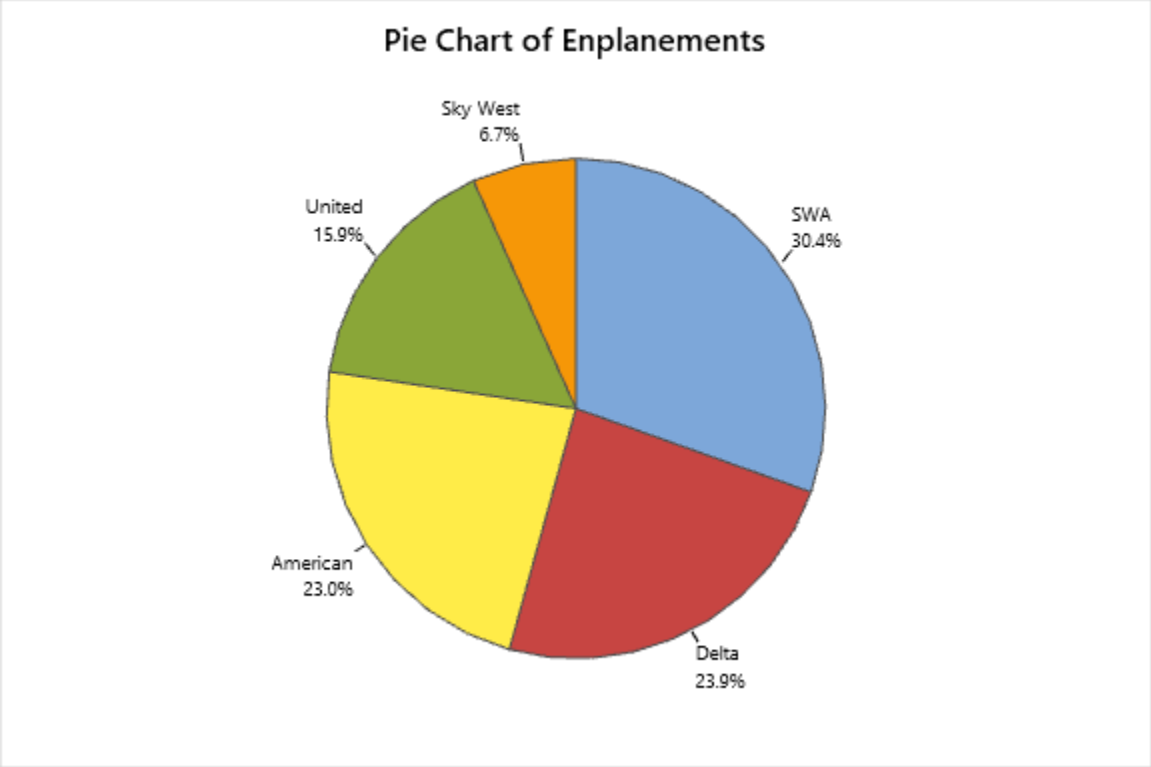


c.) While pie charts are sometimes interesting and familiar to observe, in this problem it is virtually impossible from the pie chart to determine the relative difference between Broadcom and Qualcom without the data labels. There is a similar problem between the slices of Micron and Texas Ins. In this particular problem, it is also difficult to discriminate differences between Qualcomm and Broadcom and between Micron and Texas Instrument. Both clearly show that Intel has about ½ the market share.

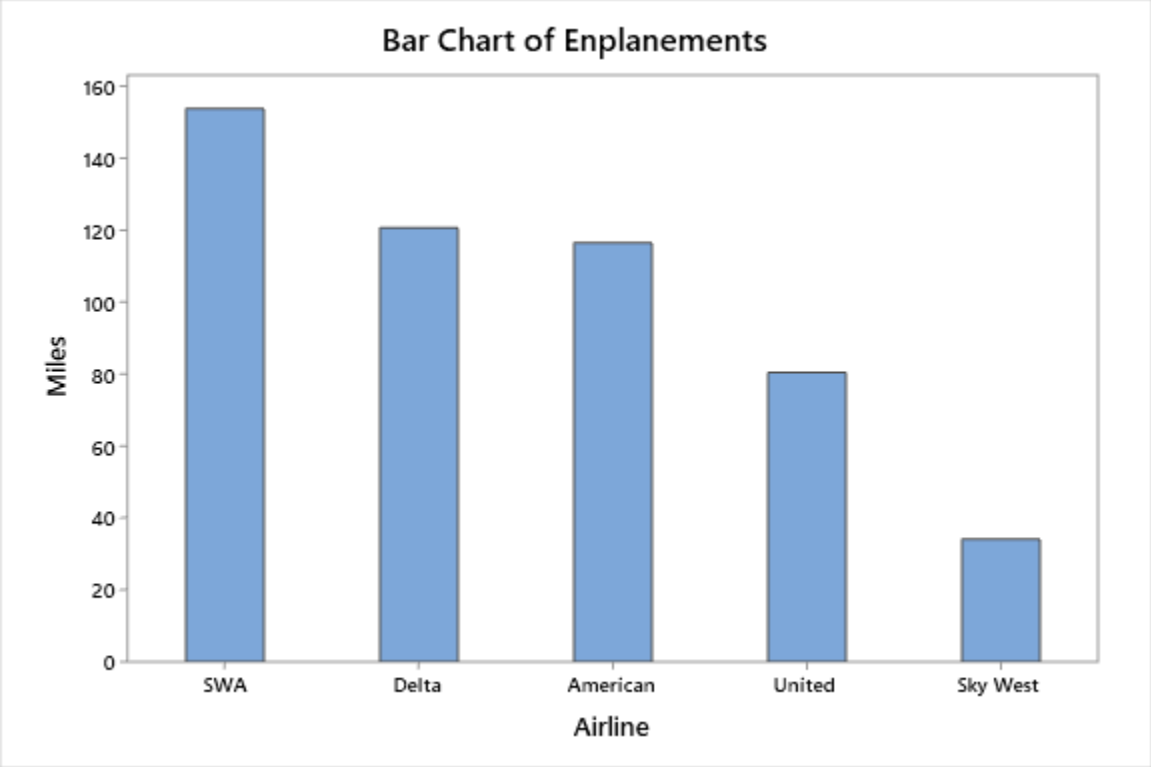
2. 16

Company	Proportion	Degrees
Southwest Airlines	.304	109.5
Delta	.239	86.0
American Airlines	.230	83.0
United Airlines	.159	57.3
SkyWest	.067	24.3
TOTAL	1.000	360.0

Pie Chart:



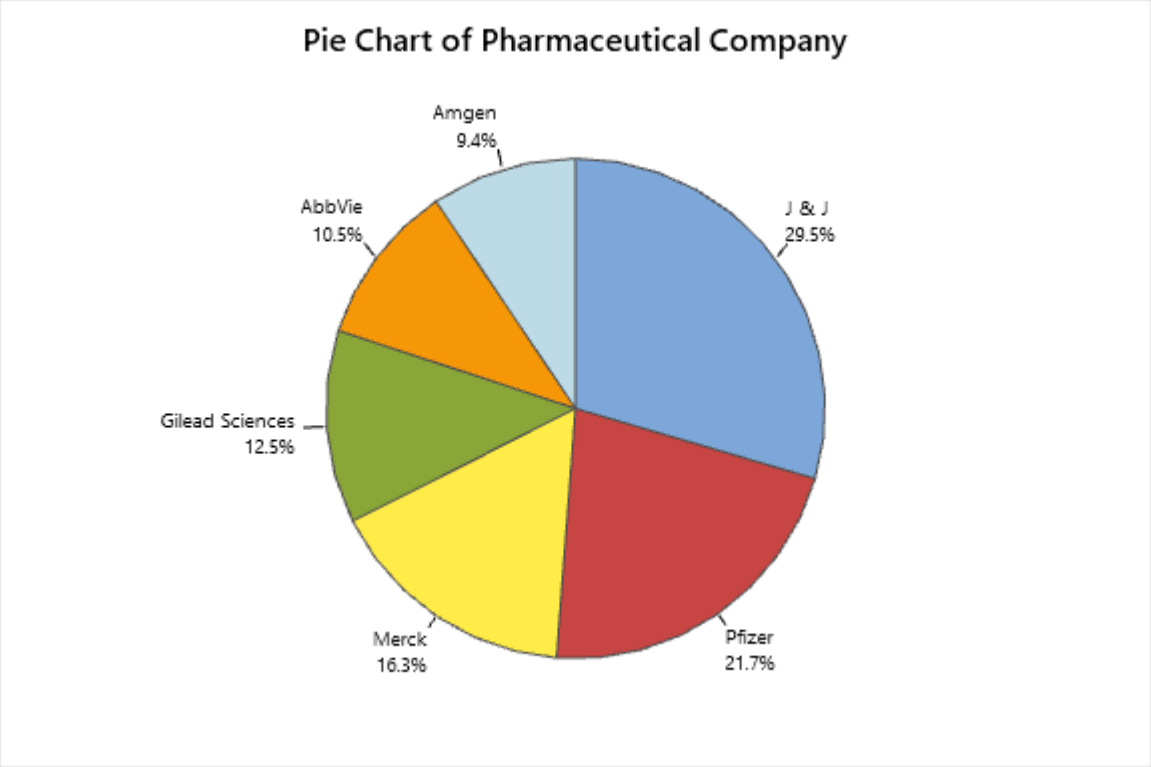
Bar Graph:



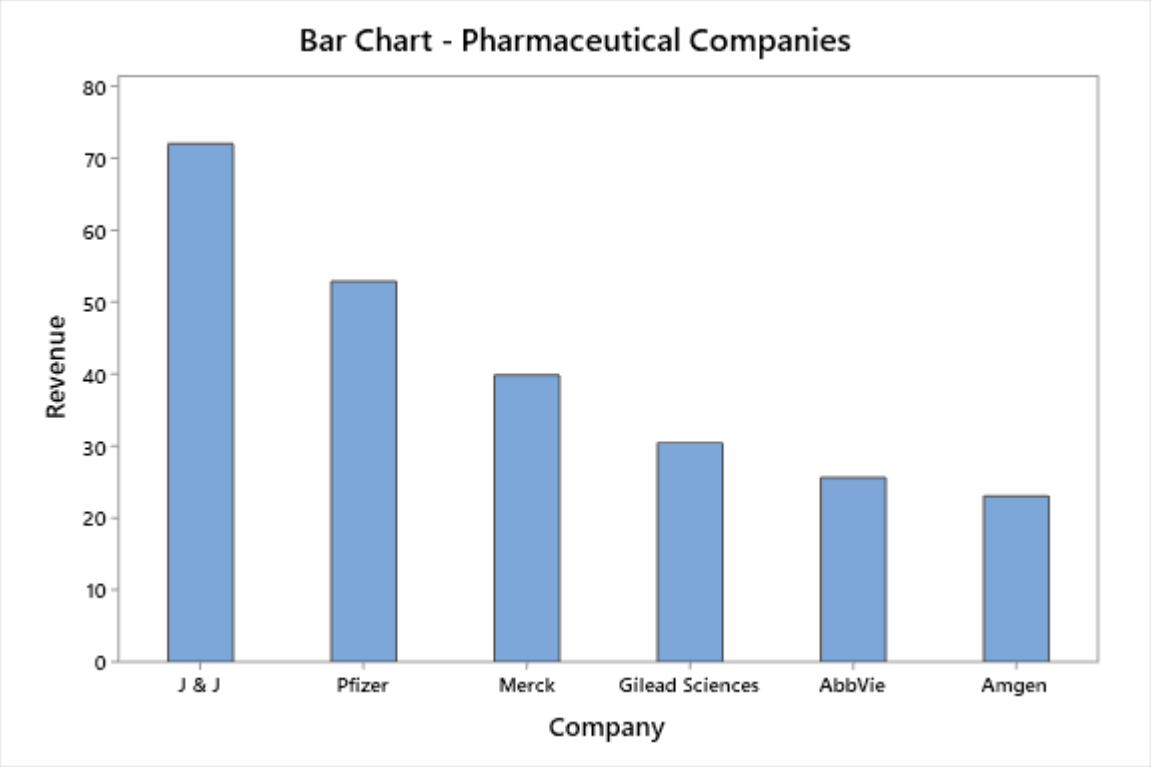
2.17

Brand	Proportion	Degrees
Johnson & Johnson	.295	106.3
Pfizer	.217	78.1
Merck	.163	58.8
Gilead Sciences	.125	44.9
AbbVie	.105	37.8
Amgen	.094	34.0
TOTAL	1.000	360.0

Pie Chart:



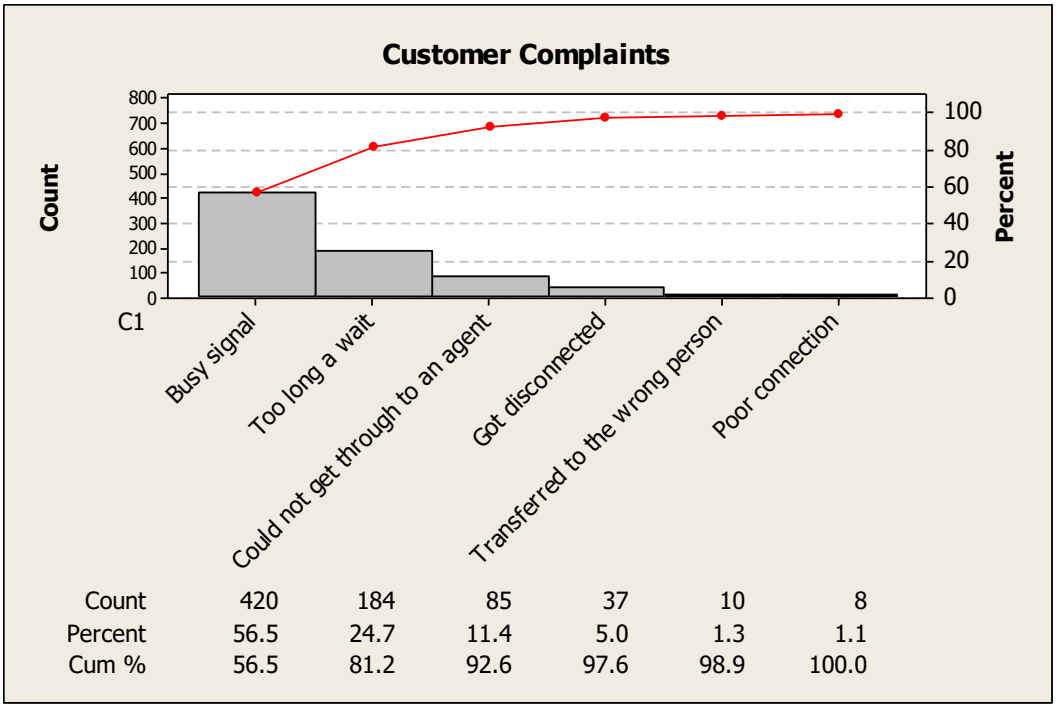
Bar Graph:



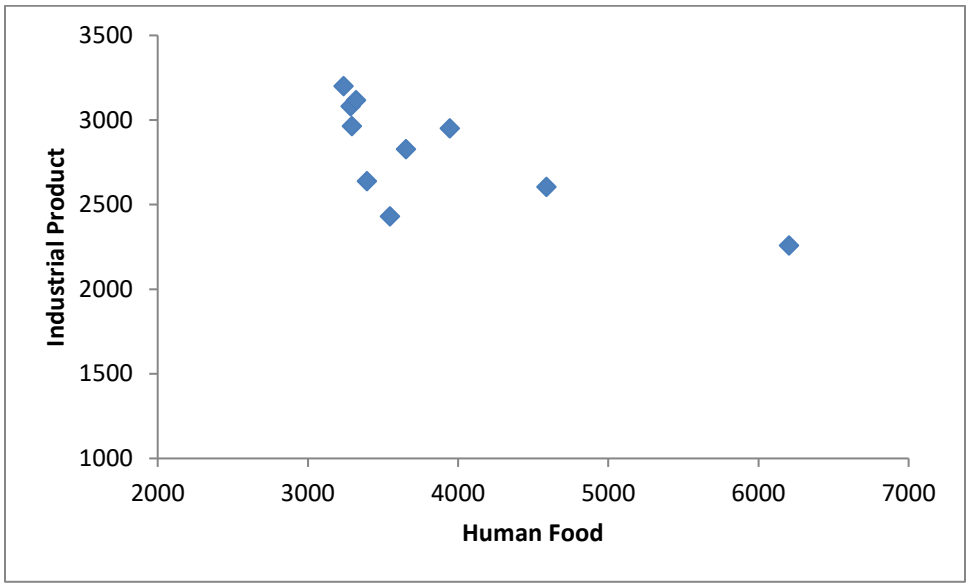
2.18 The bar chart shows that of all the currencies considered here, the Euro is strongest against the U.S. dollar (each Euro is worth $1.14). The India rupee is weakest against the U.S. dollar with a worth of only about one cent. The Canadian dollar is worth seventy-six cents compared to the U.S. dollar and the New Zealand dollar is worth about sixty-six U.S. cents. The Malaysia ringgit is worth twenty-four U.S. cents, the UAE dirham is worth about twenty-seven U.S. cents, the Chinese yuan is worth fifteen U.S. cents, and the Mexican peso is worth about five U.S. cents.

2.19

Complaint	Number	% of Total
Busy Signal	420	56.45
Too long a Wait	184	24.73
Could not get through	85	11.42
Got Disconnected	37	4.97
Transferred to the Wrong Person	10	1.34
Poor Connection	8	1.08
Total	744	99.99

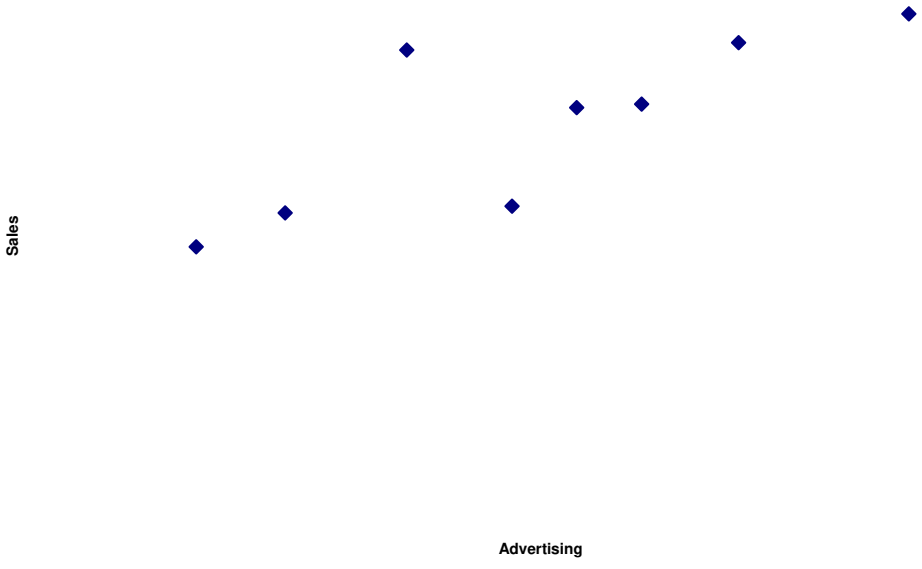


2.20



Generally, as the amount of fish caught for human food increases, the amount used for industrial products tends to decrease.

2.21



Generally, as advertising dollars increase, sales are increasing.

2.22 It appears from the graph that as job satisfaction decreases, there is an increase in tardiness. Thus, there appears to be an inverse relationship between job satisfaction and tardiness. The scatter plot also shows that when employees are highly satisfied, the level of tardiness is low.

2.23 There is a slight tendency for there to be a few more absences as plant workers commute further distances. However, compared to the total number of workers in each category, these increases are relatively small (2.5% to 3.0% to 6.6%). Comparing workers who travel 4-10 miles to those who travel 0-3 miles, there is about a 2:1 ratio in all three cells indicating that for these two categories (0-3 and 4-10), number of absences is essentially independent of commute distance.

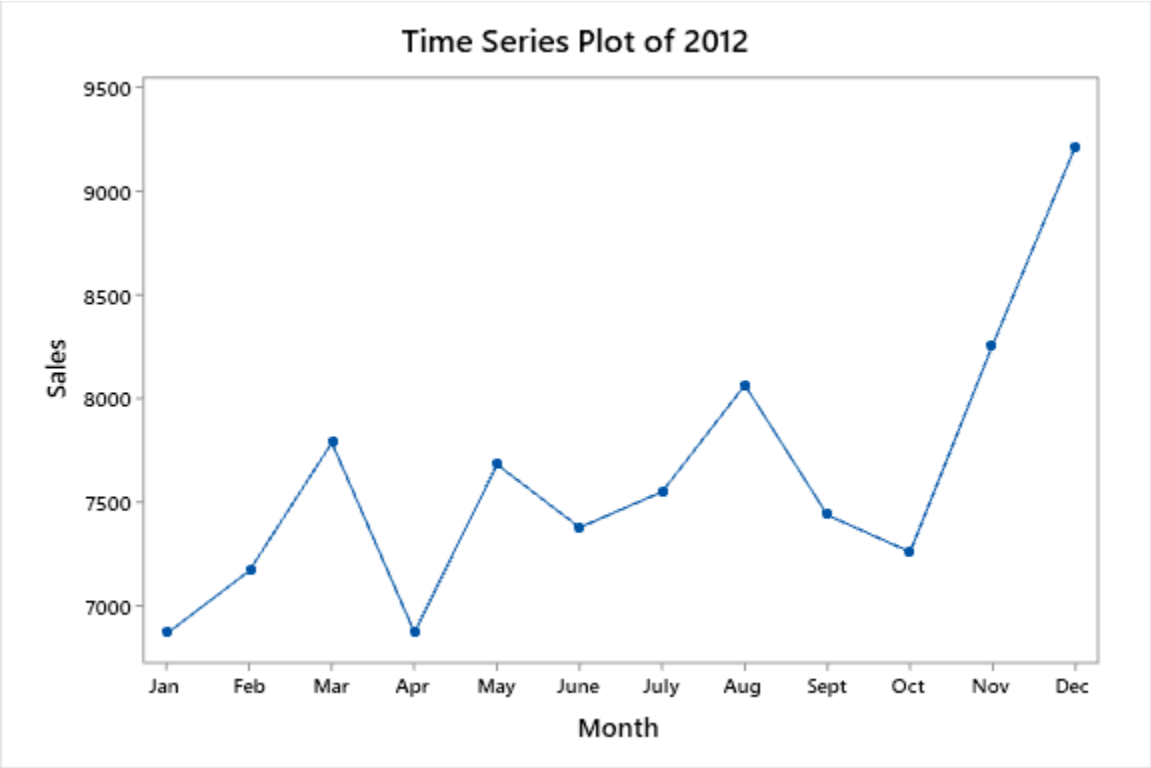
2.24

		Level of Education	
		High School	College
Rating of Service	Acceptable	9	6
	Unacceptable	2	8

It appears that a much higher proportion of high school level customers rate the service as acceptable than as unacceptable (9 to 2 ratio or about 4.5 times as many). On the other hand, more of the college educated customers rated the service as unacceptable than as acceptable. From this, we can conclude that the lower the level of education, the more acceptable is the service.

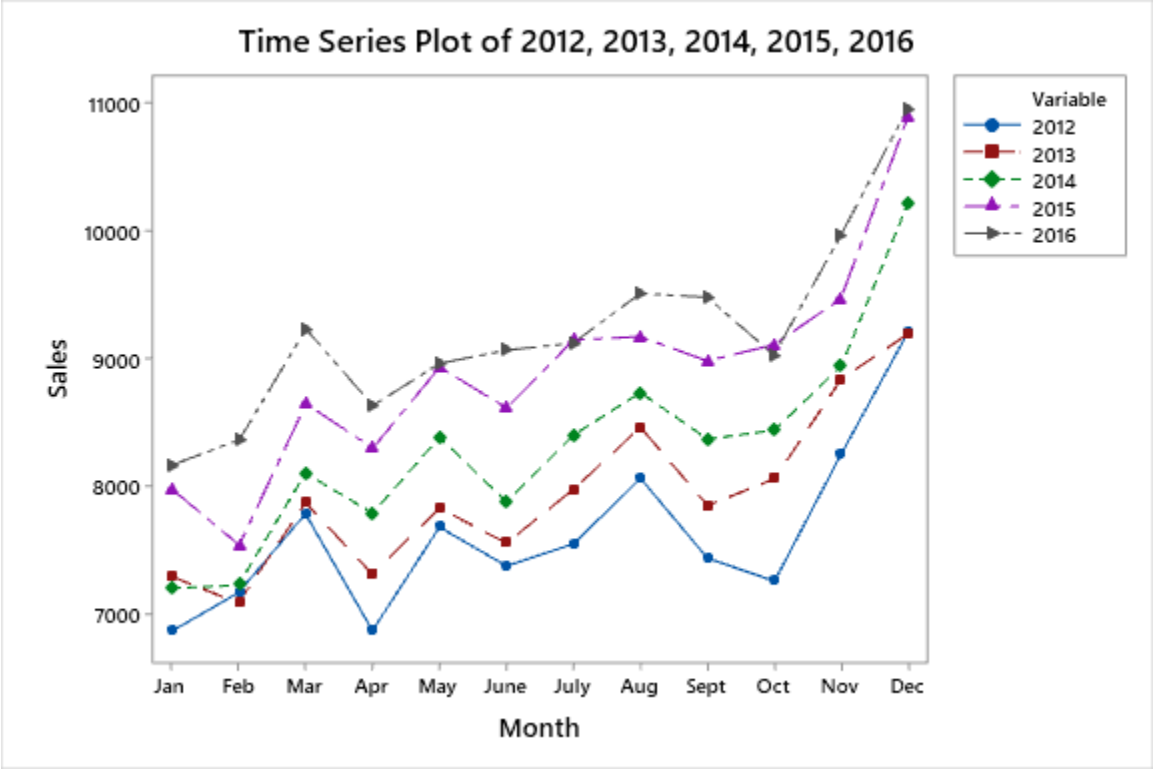
2.25

a.



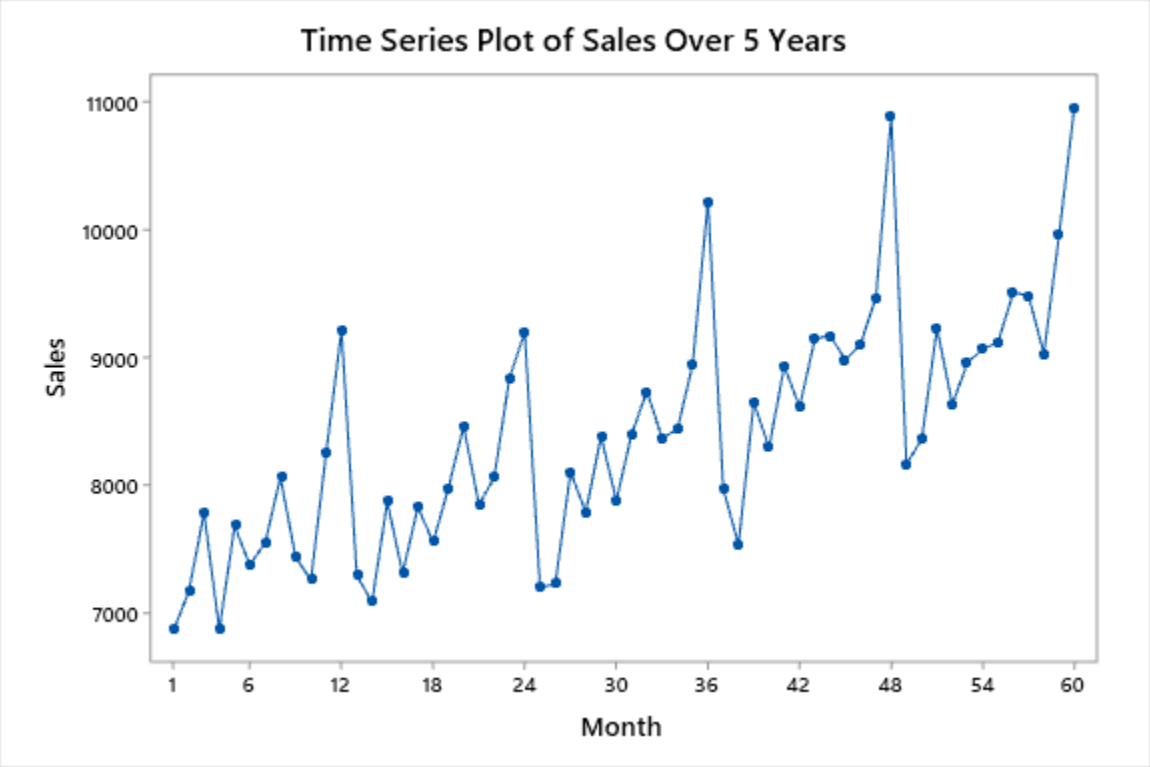
This time-series plot for the year 2012 shows up and down swings of sales until October with sales strongly increasing for November and December due to perhaps the holiday season. If true, we could be seeing possible seasonal effects of the holiday season followed by low sales in January.

b.



When placed together on one time series plot, the graphs tend to show the same shape virtually every year. This underscores the notion that there is some seasonality in the data. Sales are down in January, April, June, and October. Sales are highest in December, increases every year from October through December, and then drops dramatically in January.

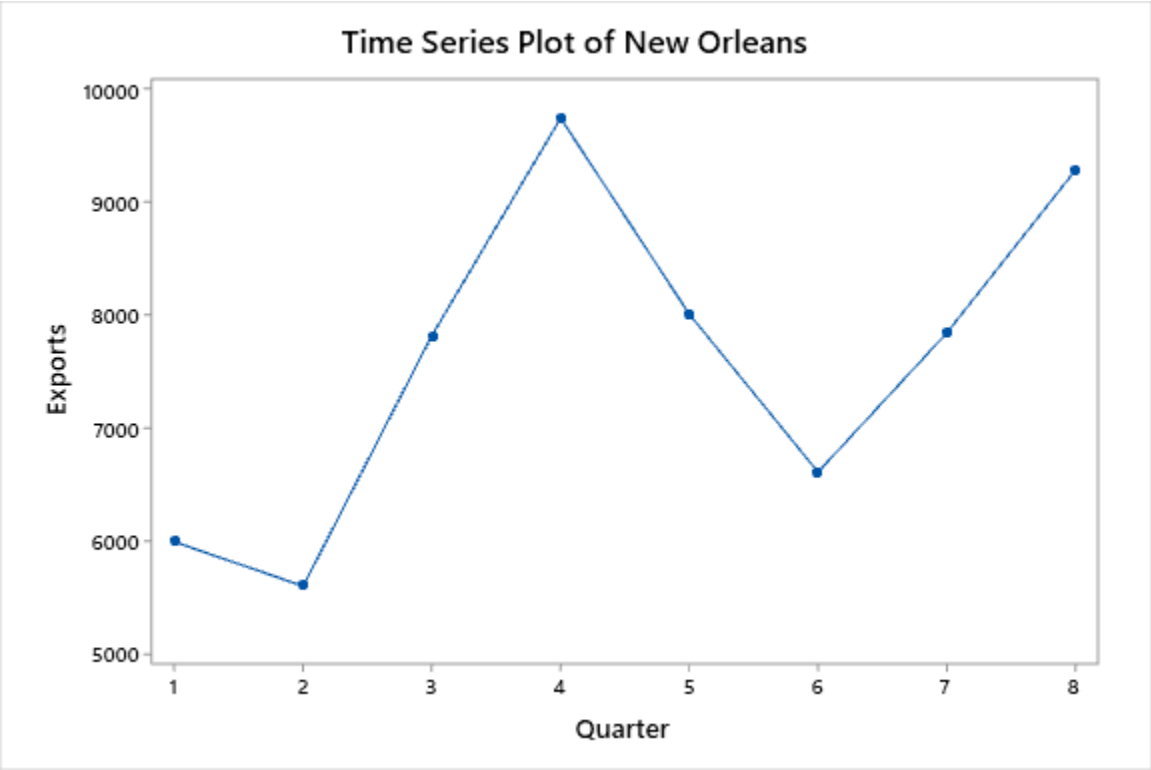
c.



With all five years of data shown chronicalogically, it is apparent that there is a repeating seasonal cycle within each year. In addition, however, there is a general trend of increasing sales over the five years.

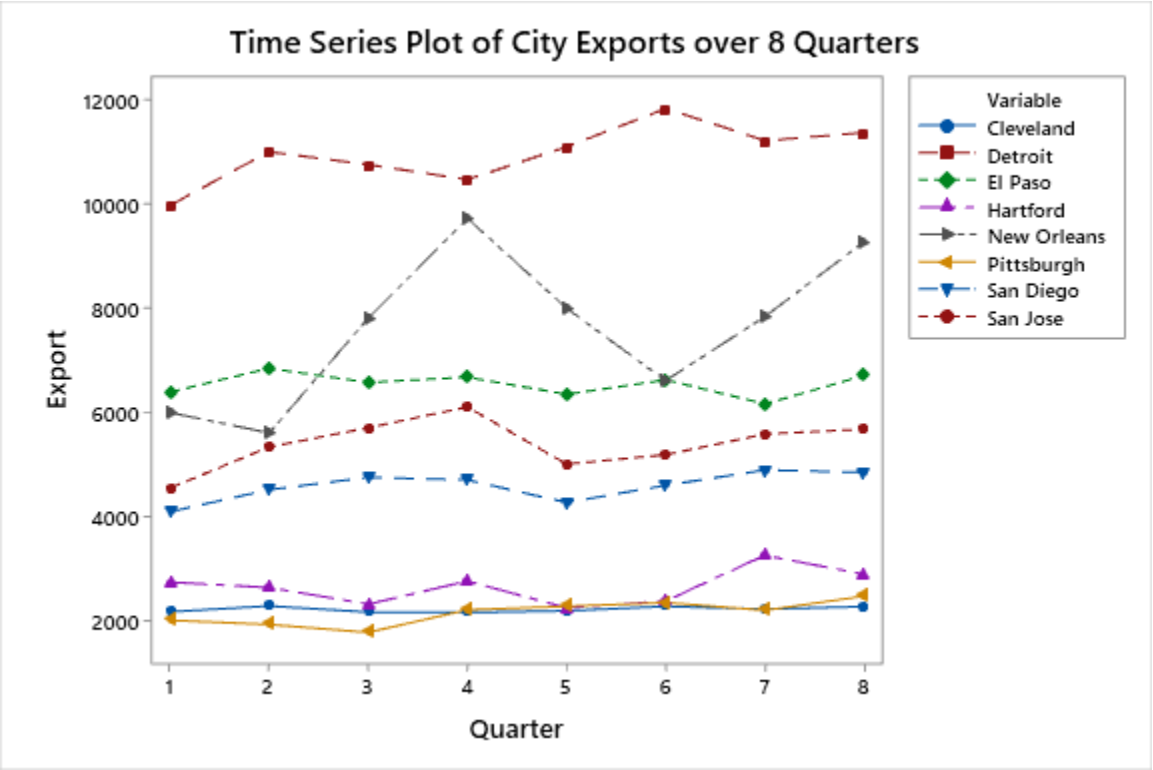
2.26

a.



Examining the exports for New Orleans over these two years, it appears that there are some seasonal effects going on within a year. For example, quarter 4 of each year yields the highest exports and quarter 2 yields the lowest exports. Overall, there seems to be little or no trend in exports over time just a pattern of ups and downs.

b.



Many of the cities actually have relatively constant exports over these 8 quarters with few significant ups and downs. New Orleans (second highest at the end) seems to be the most volatile of the cities in terms of exports. Detroit consistently has the highest exports.

2.27

Class Interval	Frequencies
16 - under 23	6
23 - under 30	9
30 - under 37	4
37 - under 44	4
44 - under 51	4
51 - under 58	3
TOTAL	30

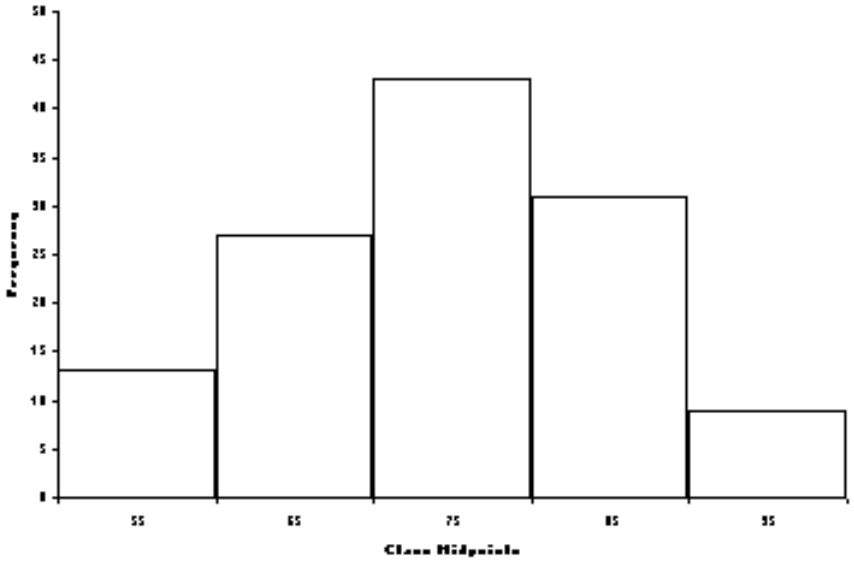
2.28

Class Interval	Frequency	Midpoint	Rel. Freq.	Cum. Freq.
20 - under 25	17	22.5	.207	17
25 - under 30	20	27.5	.244	37
30 - under 35	16	32.5	.195	53
35 - under 40	15	37.5	.183	68
40 - under 45	8	42.5	.098	76
45 - under 50	6	47.5	.073	82

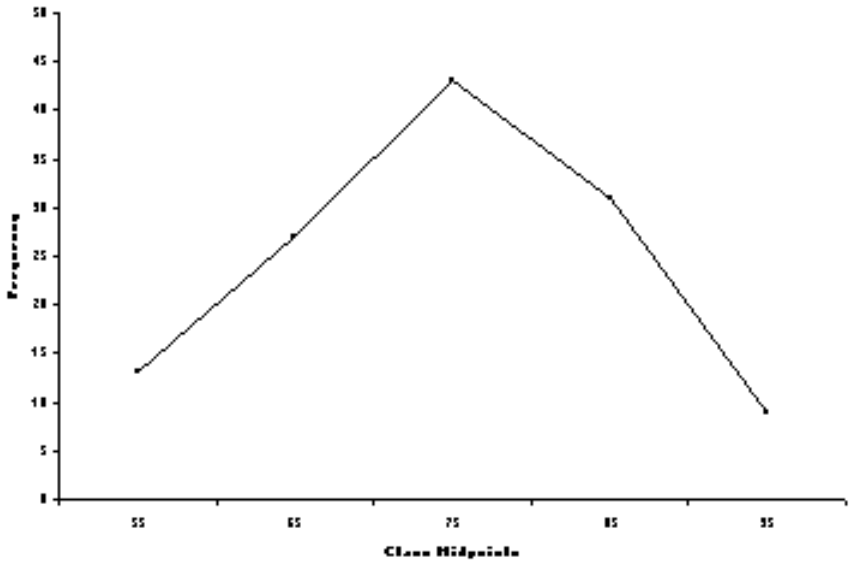
2.29

Class Interval	Frequencies
50 - under 60	13
60 - under 70	27
70 - under 80	43
80 - under 90	31
90 - under 100	9
TOTAL	123

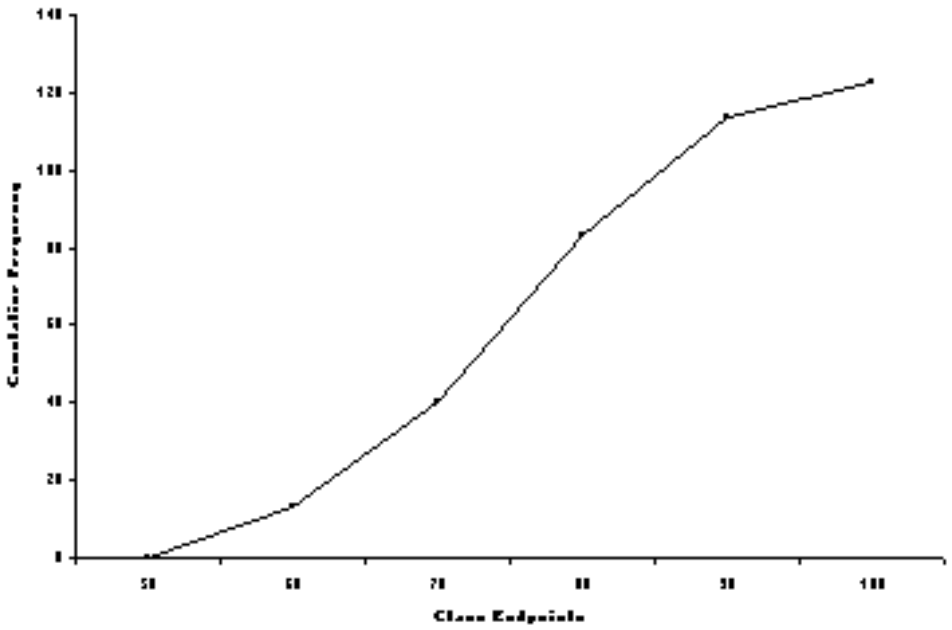
Histogram:



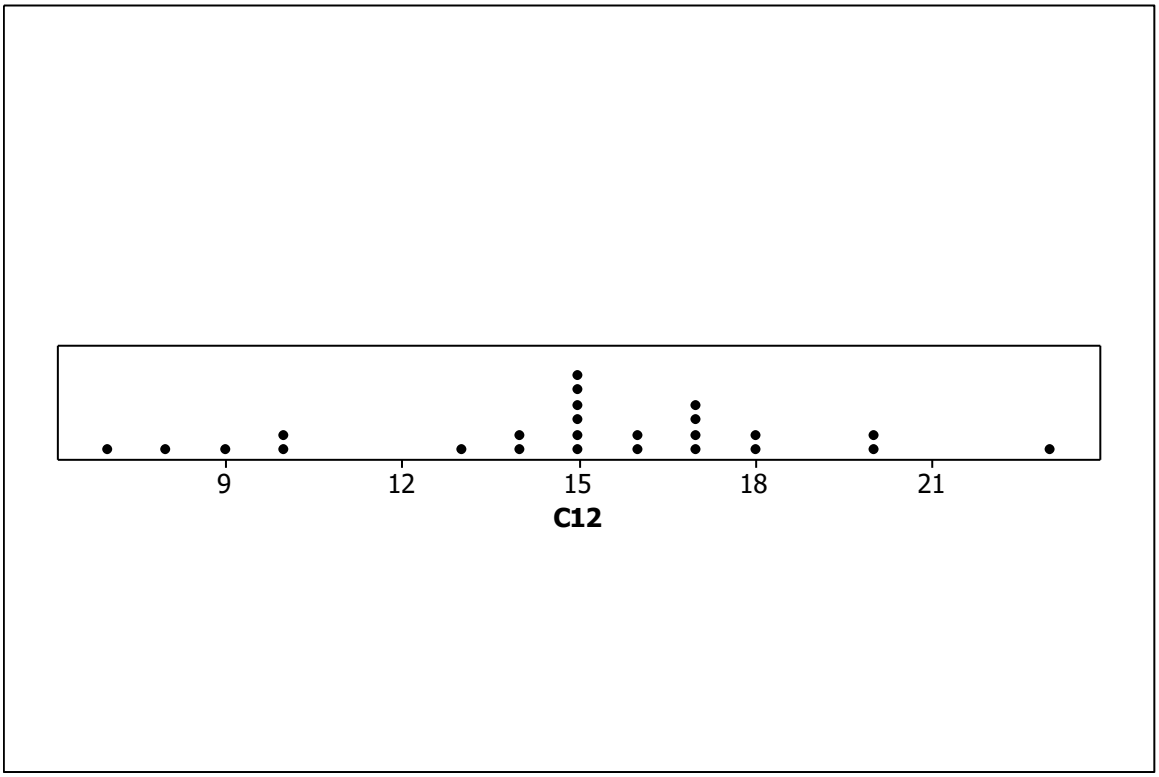
Frequency Polygon:



Ogive:



2.30 Dot Plot



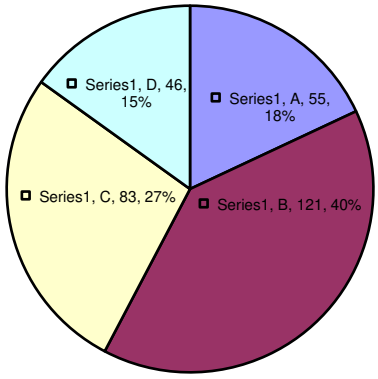
2.31

STEM	LEAF
28	4 6 9
29	0 4 8
30	1 6 8 9
31	1 2 4 6 7 7
32	4 4 6
33	5

2.32

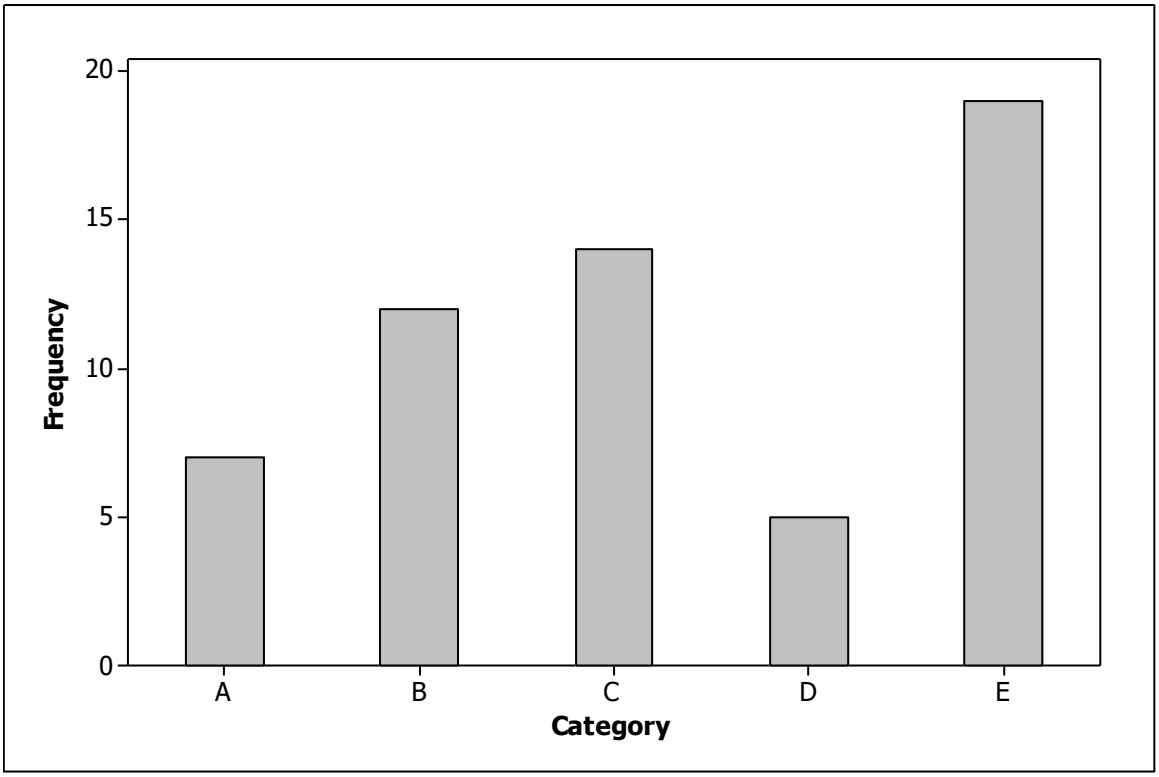
Label	Value	Proportion	Degrees
A	55	.180	65
B	121	.397	143
C	83	.272	98
D	46	.151	54
TOTAL	305	1.000	360

Pie Chart:



2.33 Bar Graph:

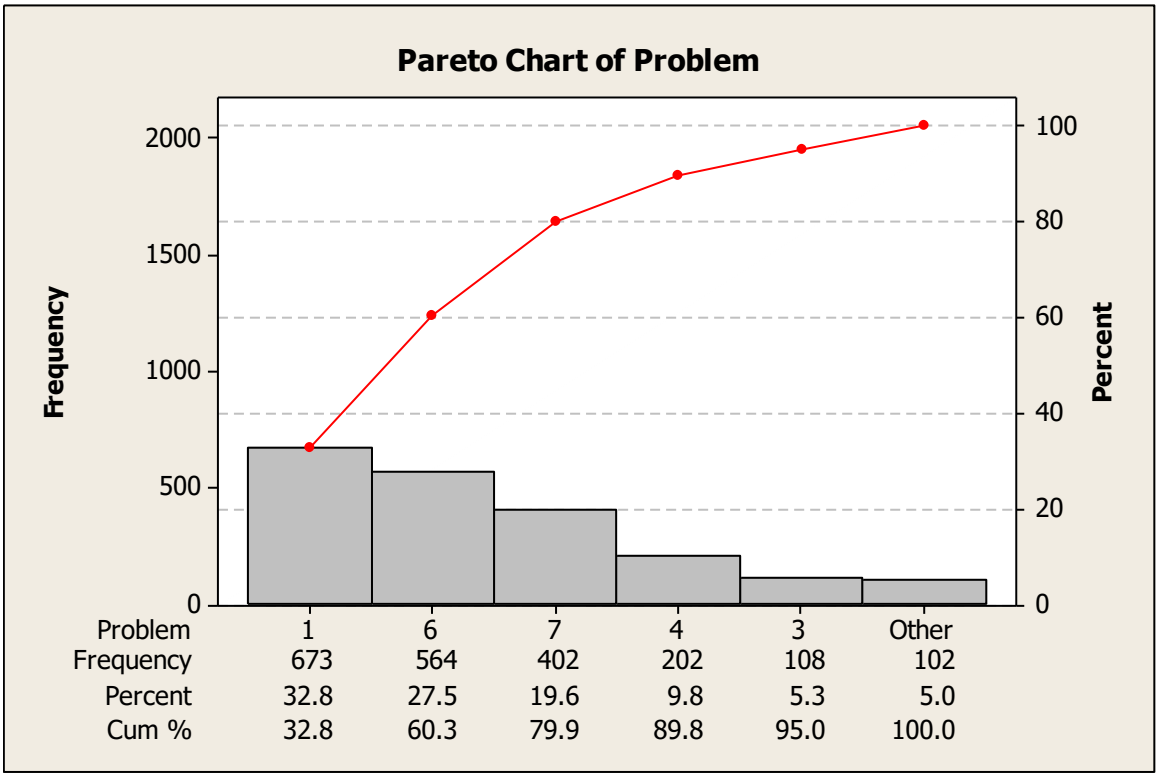
Category	Frequency
A	7
B	12
C	14
D	5
E	19



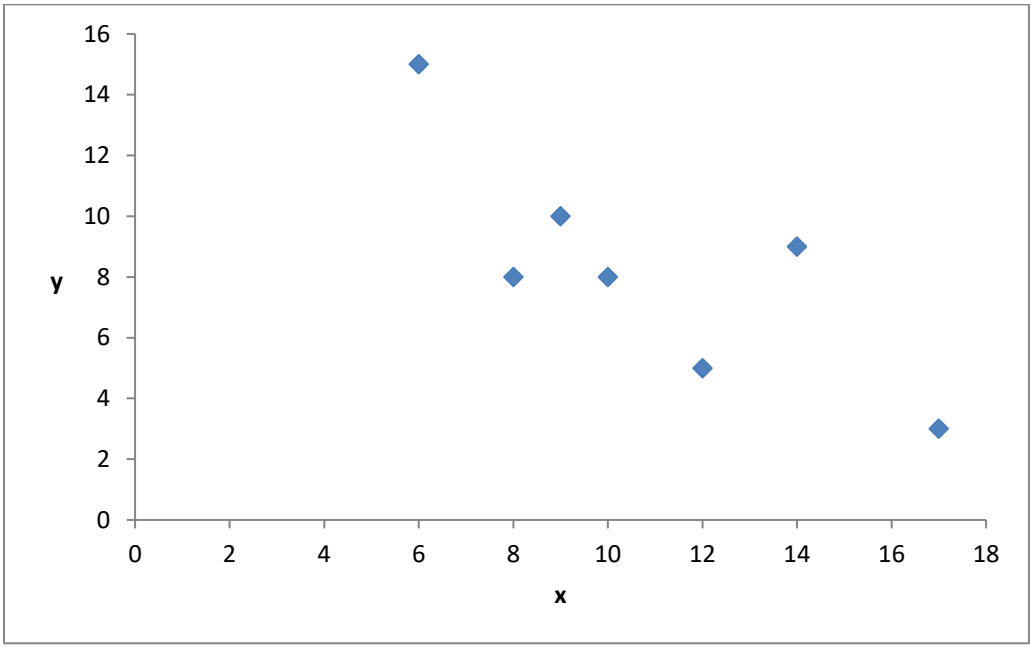
2.34

Problem	Frequency	Percent of Total
1	673	32.81
2	29	1.41
3	108	5.27
4	202	9.85
5	73	3.56
6	564	27.50
7	402	19.60
	2051	

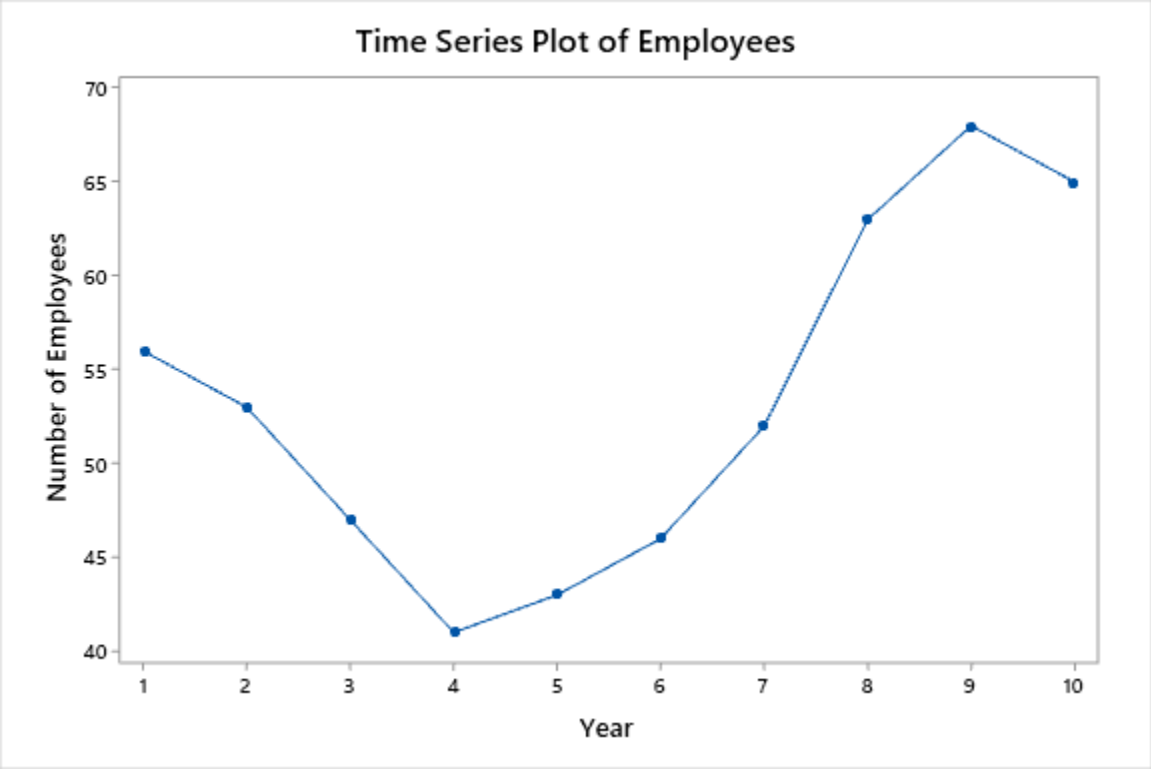
Pareto Chart:



2.35 Scatter Plot

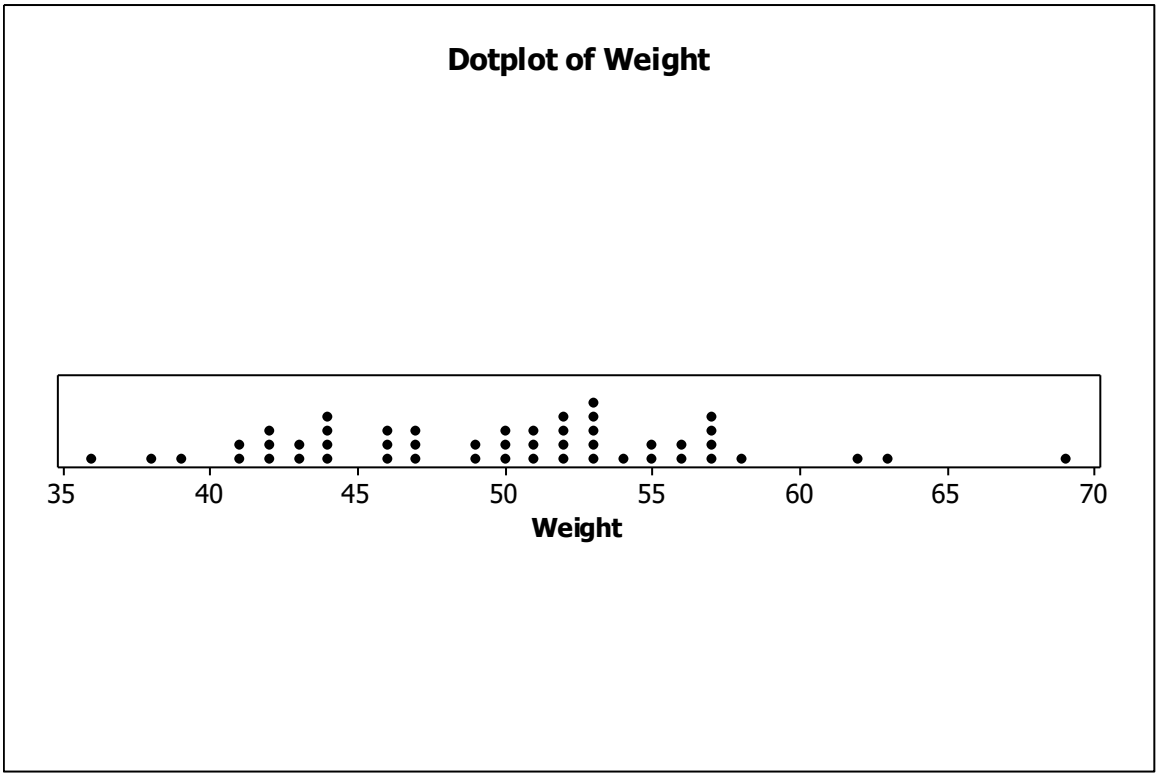


2.36



2.37 Whitcomb Company

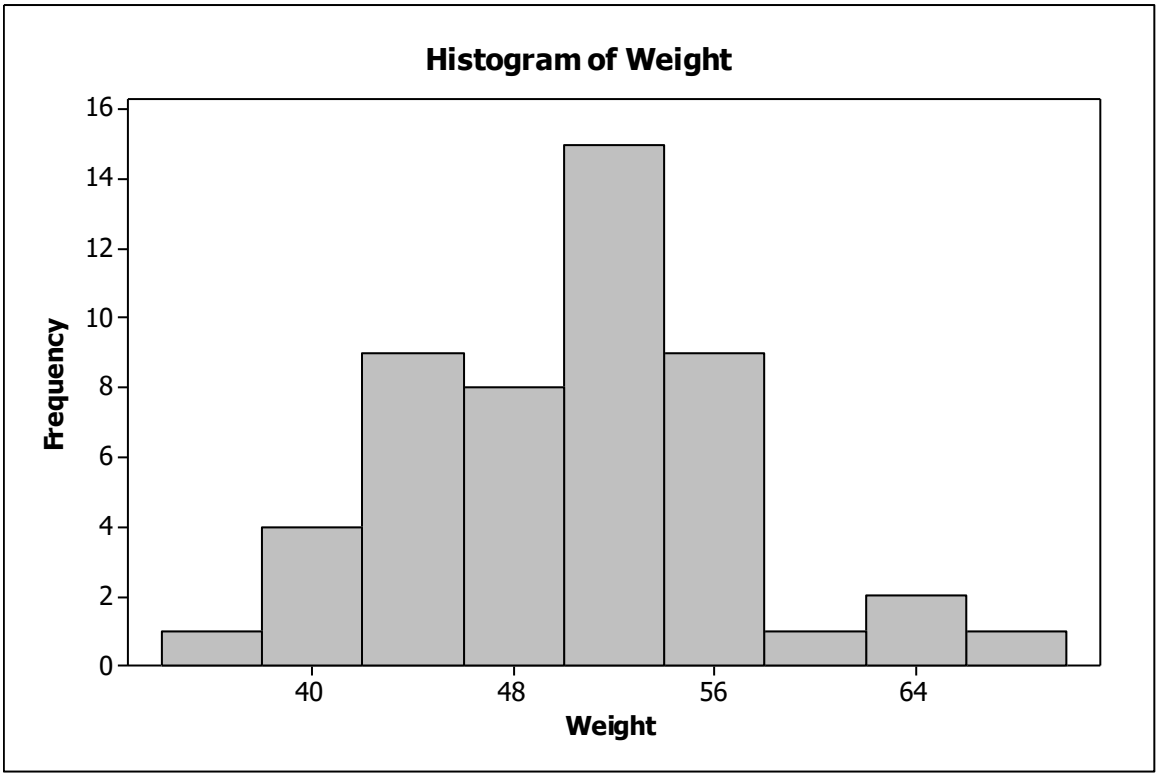
a.) Dot Plot. The metal ring is supposed to weigh around 50 ounces. The dot plot shows that most of the rings weigh between 41 and 58 ounces with the highest number being 53 (in Chapter 3 we will name this as the mode) and other weights piling up at 44, 52, and 57 ounces. There are some extreme values (later in the text termed "outliers") at 36, 38, 39, 62, 63, and even 69 ounces.



b.) Frequency Distribution and Histogram. Shown first here is a frequency distribution with class intervals containing widths of 5 ounces.

Class Interval	Frequency	Cumulative Frequency
32 - under 37	1	1
37 - under 42	4	5
42 - under 47	12	17
47 - under 52	11	28
52 - under 57	14	42
57 - under 62	5	47
62 - under 67	2	49
67 - under 72	1	50
TOTAL	50	

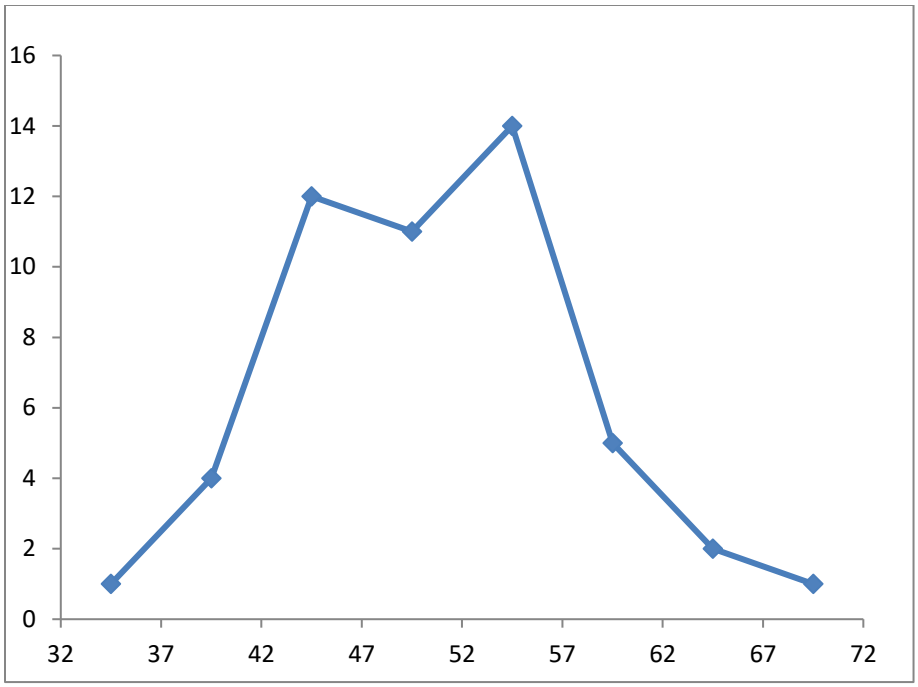
Shown below is a histogram of the weights.



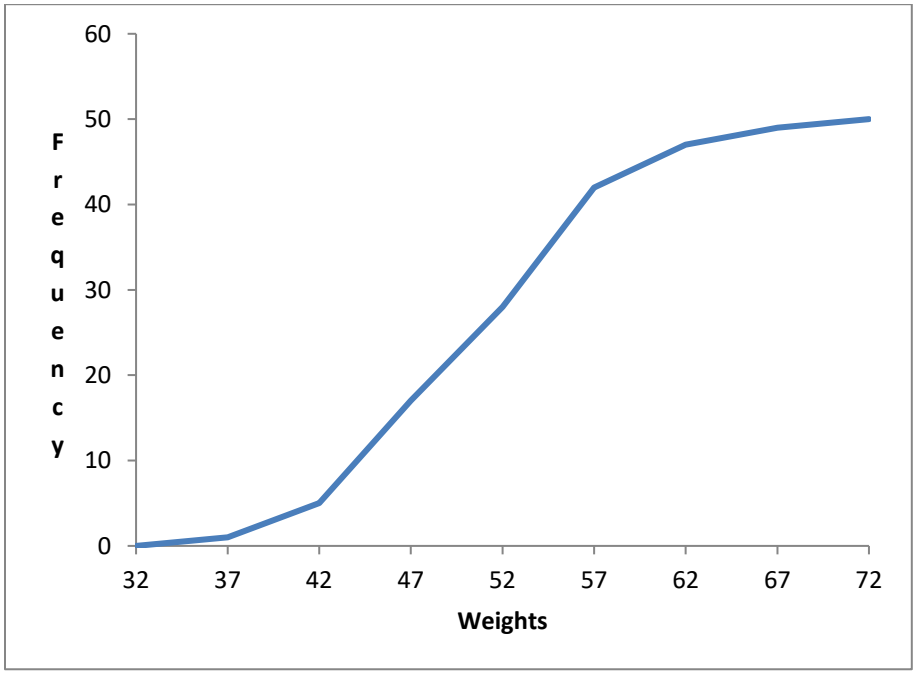
From the histogram, we can see that the bulk of the weights pile up between about 42 ounces and 58 ounces. The distribution appears to have one high point (in Chapter 3 we will call it the modal class – 50 to 54).

c.) Frequency Polygon and Ogive

Frequency Polygon:



Ogive of weights:



From the frequency polygon, we can see that the highest frequency was in the 52 – 57 class followed by the 42 - 47 class with the 47-52 class not far behind. Frequency values fell dramatically on either side of these three classes. This is underscored by the ogive which shows a relatively steep slope from 42 through 57 (large increases) preceded by and followed by relatively flat slopes indicating only small increases.

2.38

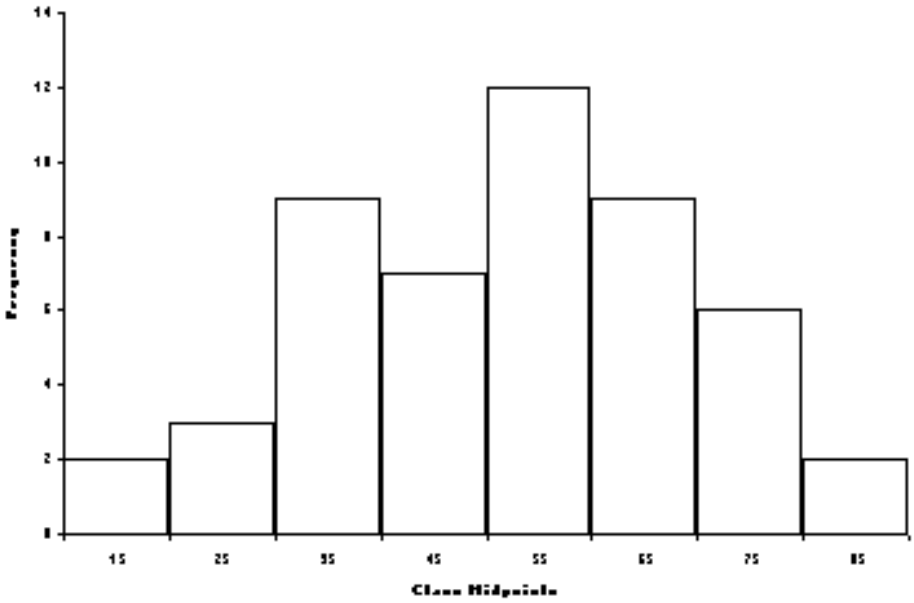
Class Interval	Frequency	Class Midpoint	Relative Frequency	Cumulative Frequency
20 – 25	8	22.5	8/53 = .1509	8
25 – 30	6	27.5	.1132	14
30 – 35	5	32.5	.0943	19
35 – 40	12	37.5	.2264	31
40 – 45	15	42.5	.2830	46
45 – 50	7	47.5	.1321	53
TOTAL	53		.9999	

2.39 Examining the shape of the distribution, the commute times generally appear in the shape of what we will refer to as the normal curve or bell-shaped curve in Chapter 3. That is, the graph is relatively symmetrical and "piles up" near the middle. Most of the commute times are between about 27 minutes and 53 minutes. A few report commute times of as little as about 10 minutes and a few report times of over 70 minutes. The highest number report commute times of about 40 minutes which seems to be in the middle of the data.

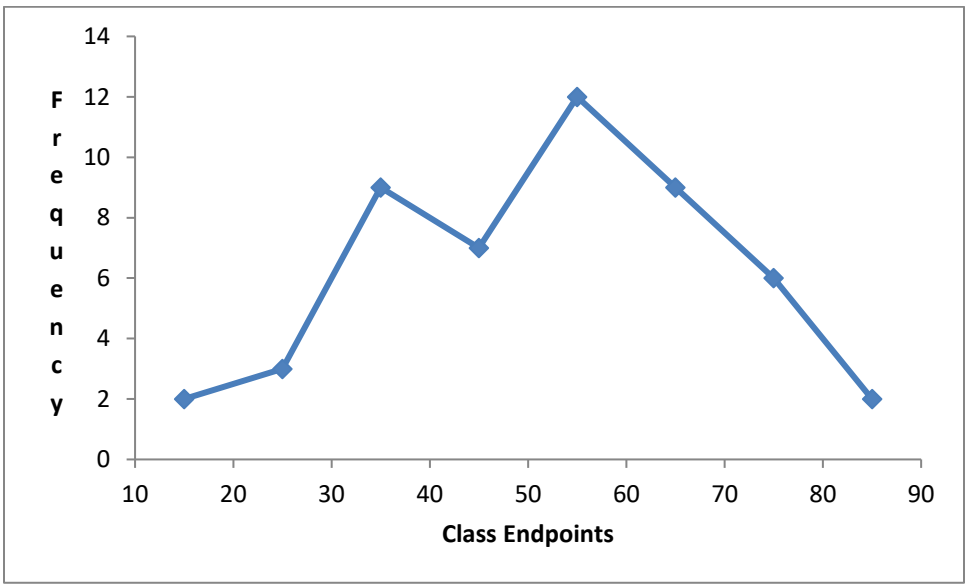
2.40 Frequency Distribution:

Class Interval	Frequency
10 - under 20	2
20 - under 30	3
30 - under 40	9
40 - under 50	7
50 - under 60	12
60 - under 70	9
70 - under 80	6
80 - under 90	2
	50

Histogram:



Frequency Polygon:



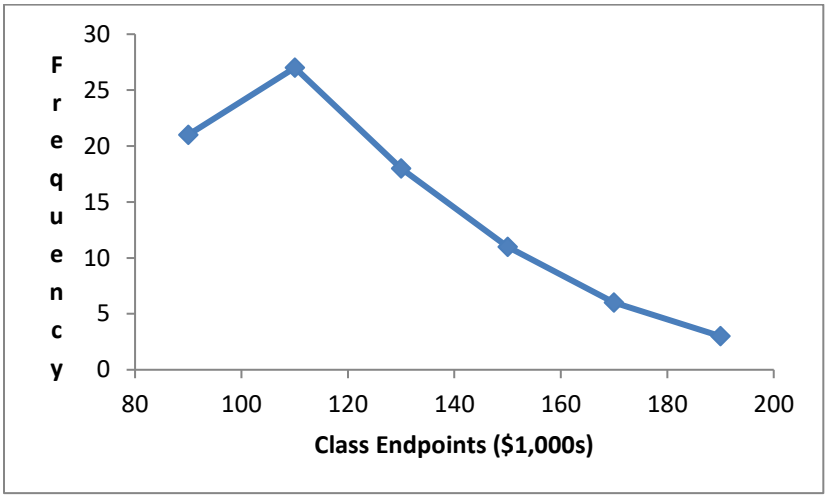
The normal distribution appears to peak near the center and diminish towards the end intervals.

2.41

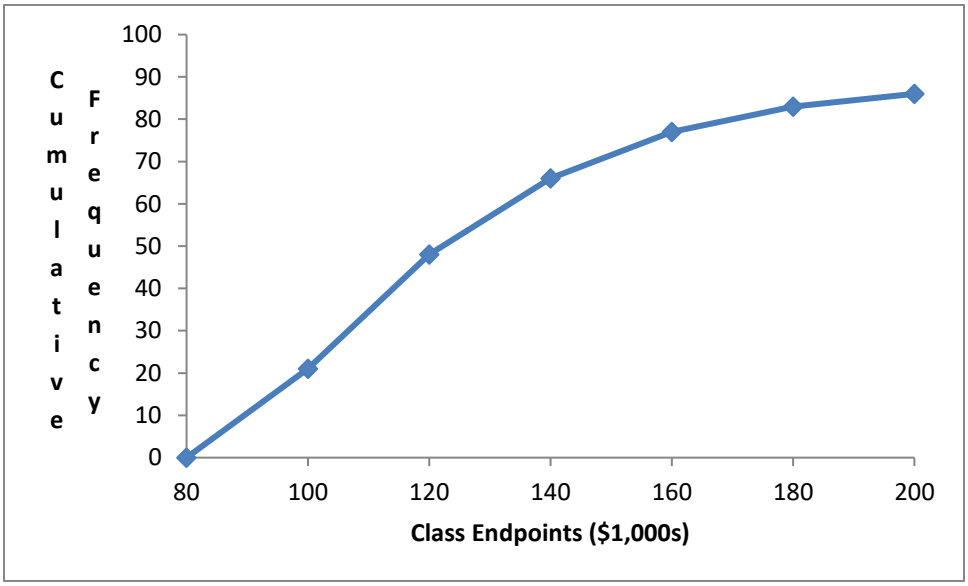
Asking Price	Frequency	Cumulative Frequency
$ 80,000 - under $ 100,000	21	21
$ 100,000 - under $ 120,000	27	48
$ 120,000 - under $ 140,000	18	66
$ 140,000 - under $ 160,000	11	77
$ 160,000 - under $ 180,000	6	83
$ 180,000 - under $ 200,000	3	86
	86	

Histogram:

Frequency Polygon:



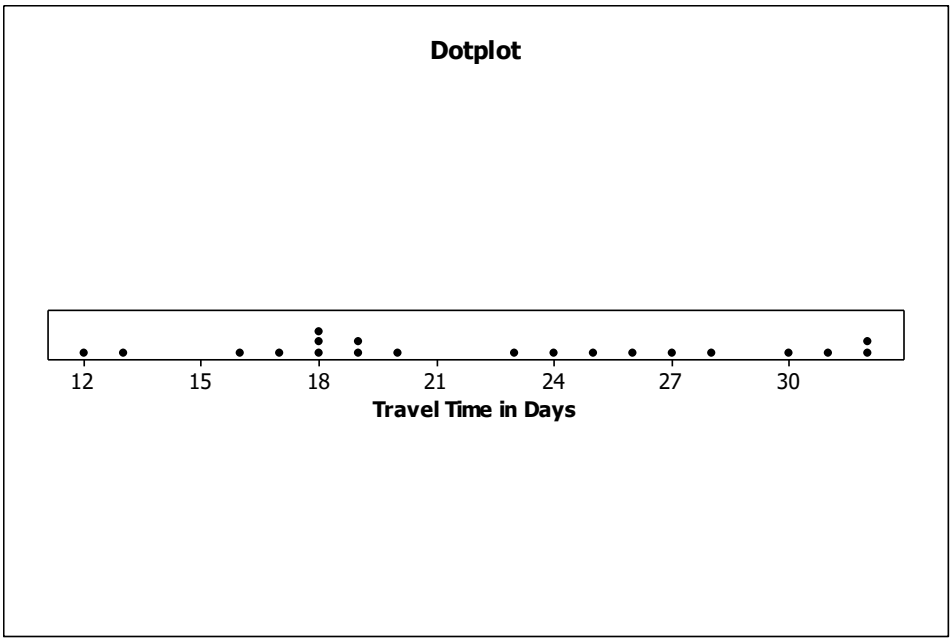
Ogive:



2.42 a.) Stem and Leaf Plot

STEM	LEAF
1	2, 3, 6, 7, 8, 8, 8, 9, 9
2	0, 3, 4, 5, 6, 7, 8
3	0, 1, 2, 2

b.) Dot Plot



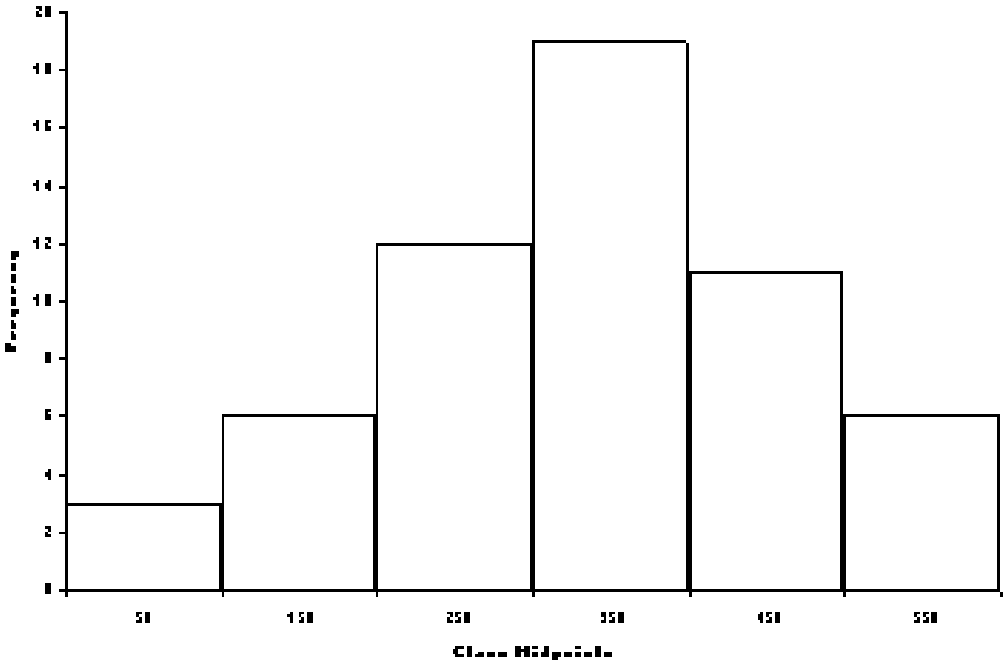
c.) Comments:

Both the dot plot and the stem and leaf plot show that the travel times are relatively evenly spread out between 12 days and 32 days. The stem and leaf plot shows that the most travel times fall in the 12 to 19 day interval followed by the 20 to 28 day interval. Only four of the travel times were thirty or more days. The dot plot show that 18 days is the most frequently occurring travel time (occurred three times).

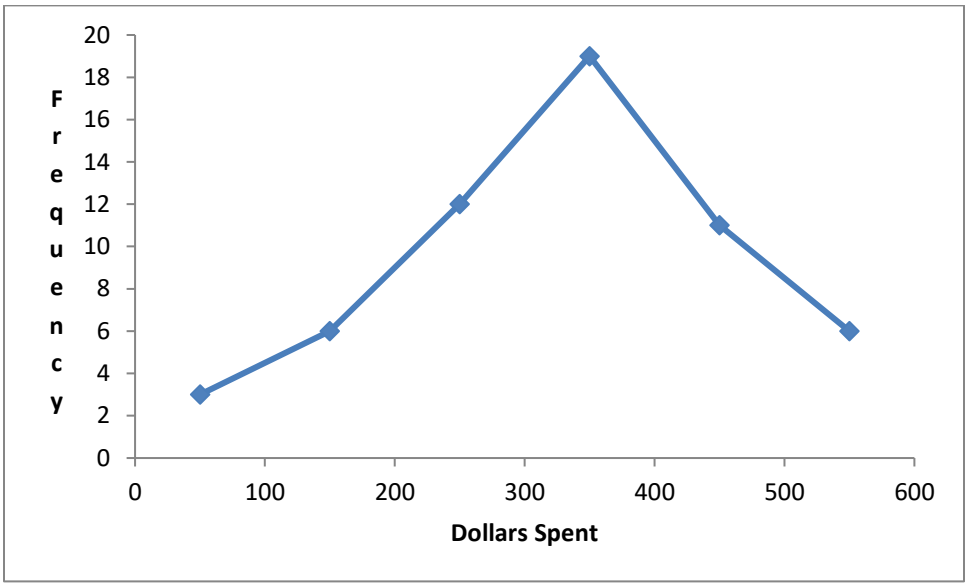
2.43

Amount Spent on Prenatal Care	Frequency	Cumulative Frequency
$ 0 - under $100	3	3
$100 - under $200	6	9
$200 - under $300	12	21
$300 - under $400	19	40
$400 - under $500	11	51
$500 - under $600	6	57
	57	

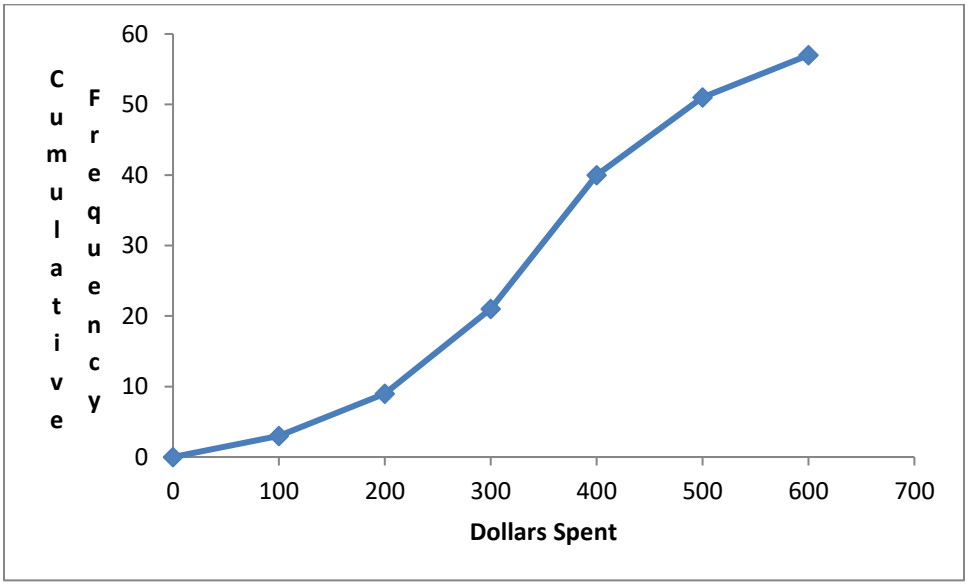
Histogram:



Frequency Polygon:



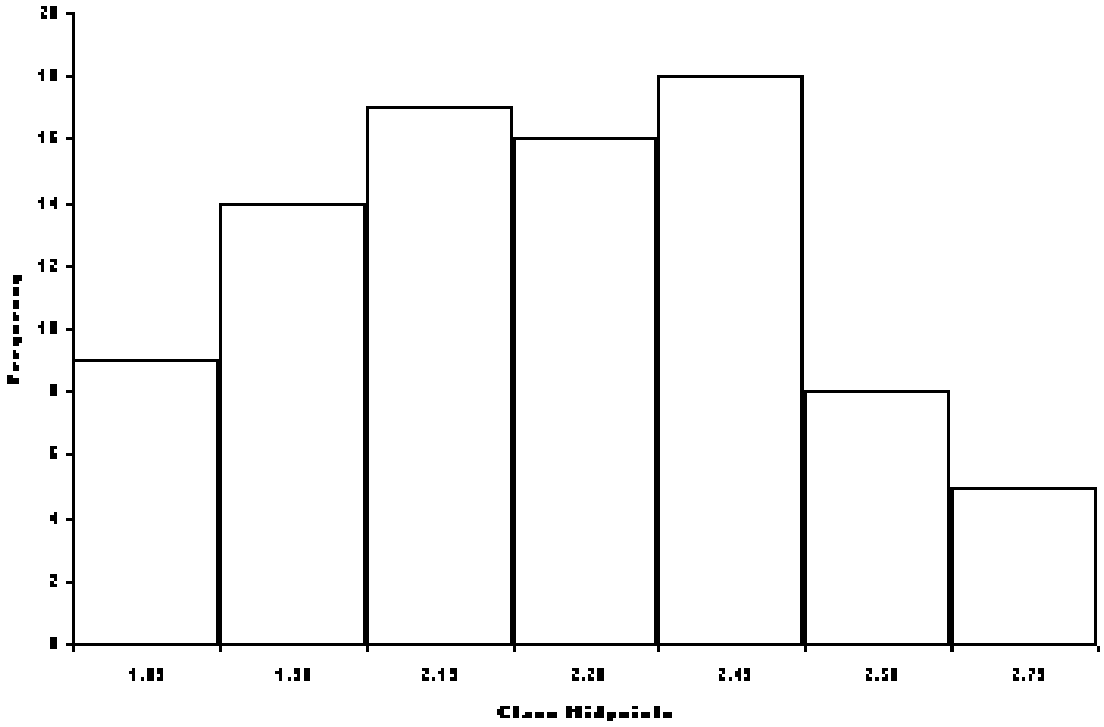
Ogive:



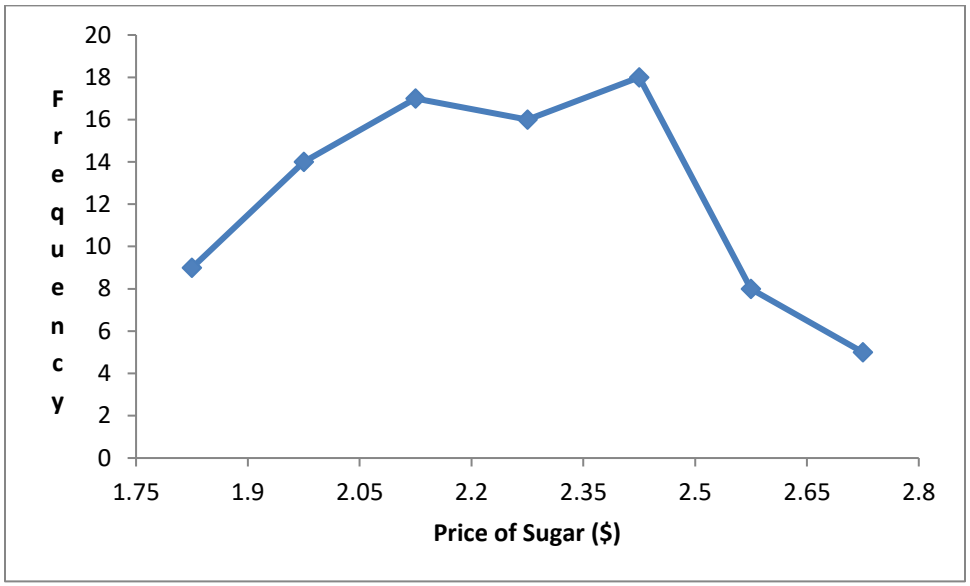
2.44

Price	Frequency	Cumulative Frequency
$1.75 - under $1.90	9	9
$1.90 - under $2.05	14	23
$2.05 - under $2.20	17	40
$2.20 - under $2.35	16	56
$2.35 - under $2.50	18	74
$2.50 - under $2.65	8	82
$2.65 - under $2.80	5	87
	87	

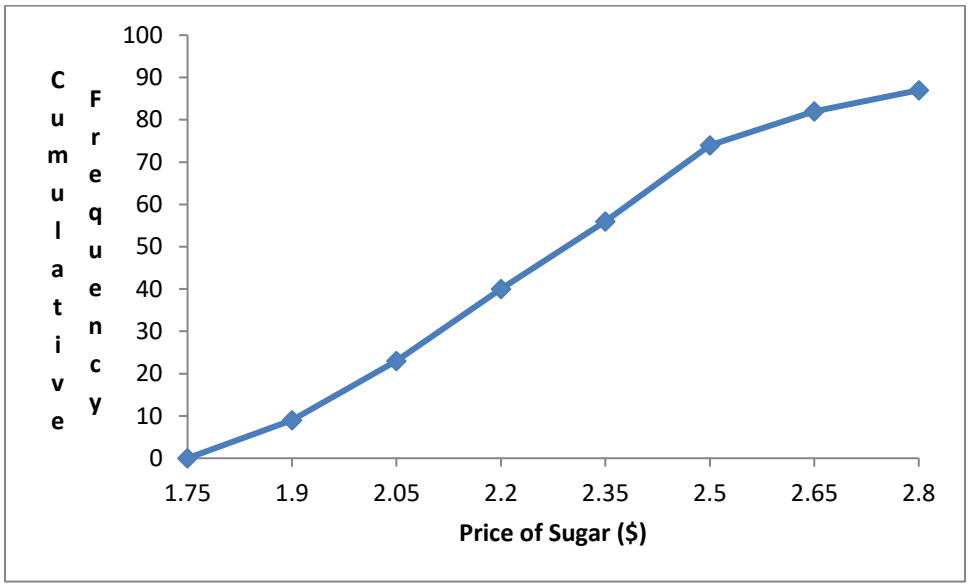
Histogram:



Frequency Polygon:



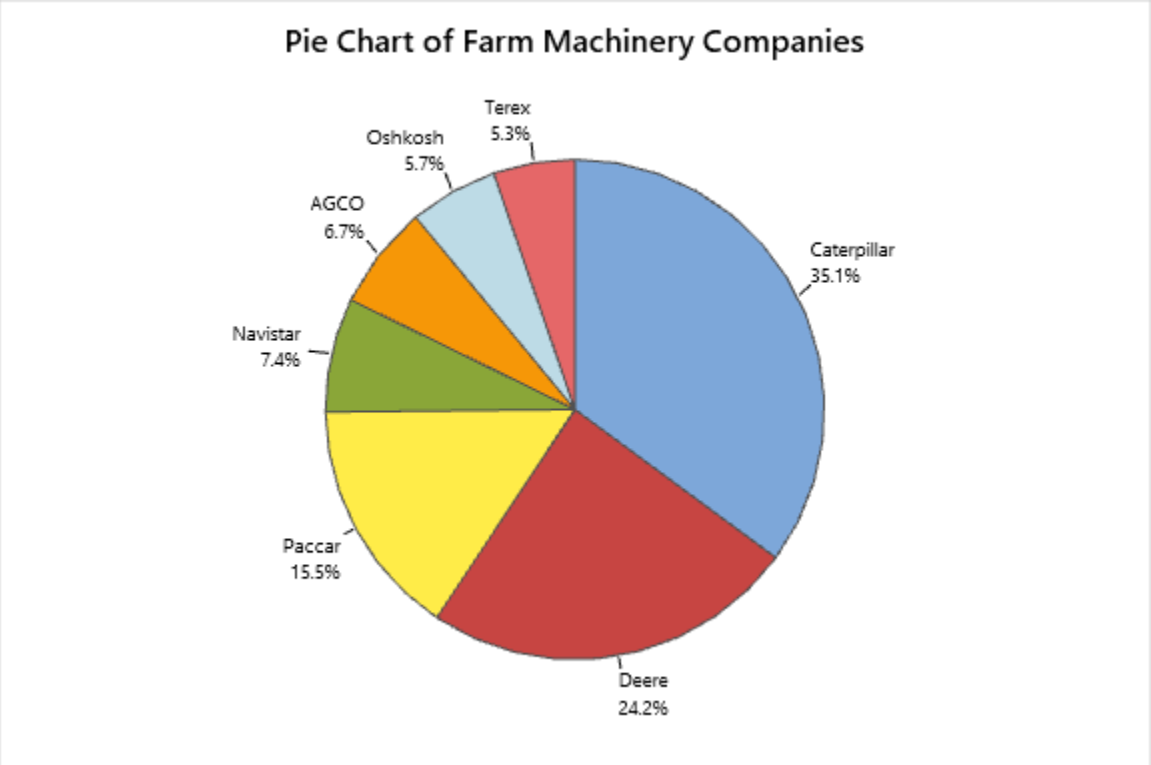
Ogive:



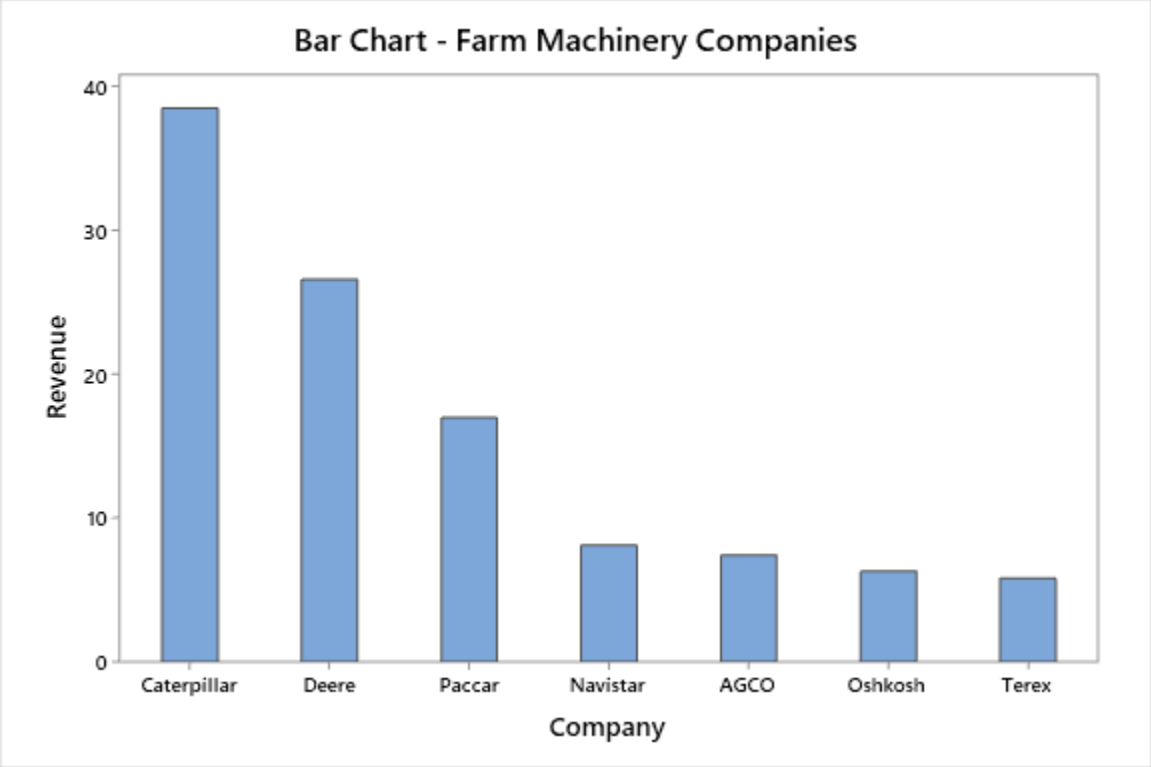
2.45

Company	Revenue	Proportion	Degrees
Caterpillar	38.5	.35	126.3
Deere	26.6	.24	87.3
Paccar	17.0	.15	55.8
Navistar International	8.1	.07	26.6
AGCO	7.4	.07	24.3
Oshkosh	6.3	.06	20.7
Terex	5.8	.05	19.0
TOTAL	109.7	1.00	360.0

a.) Pie Chart:

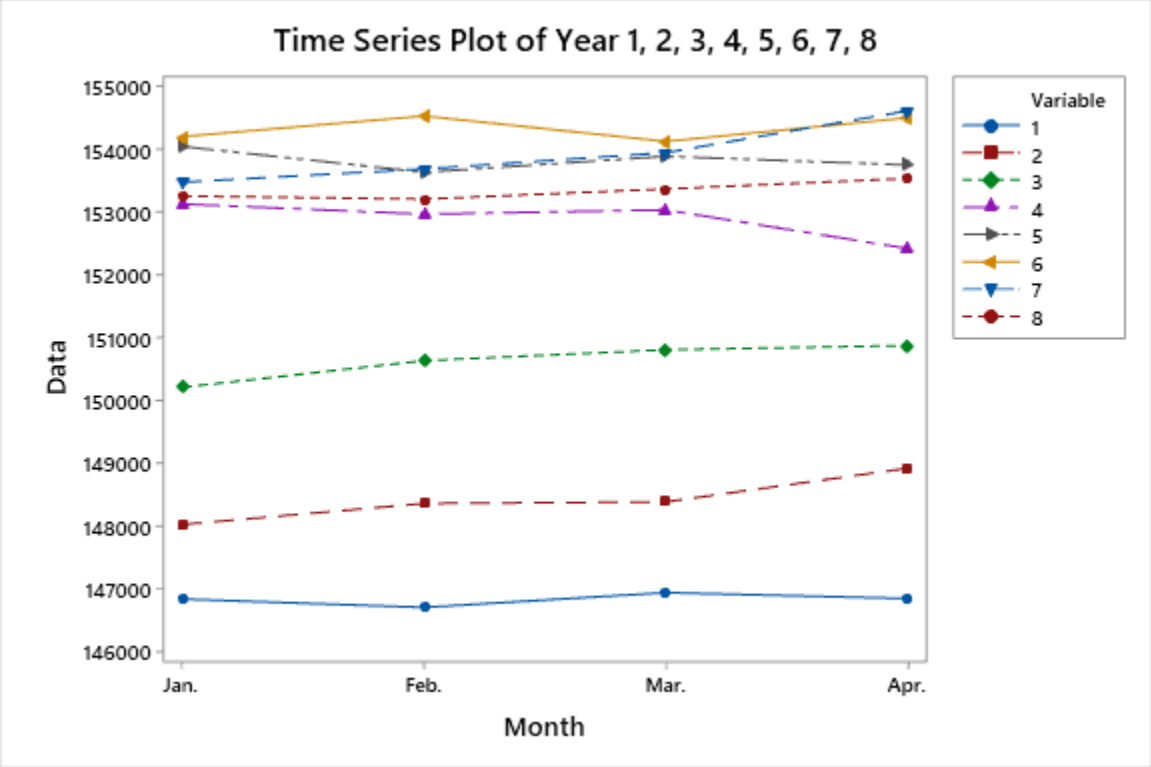


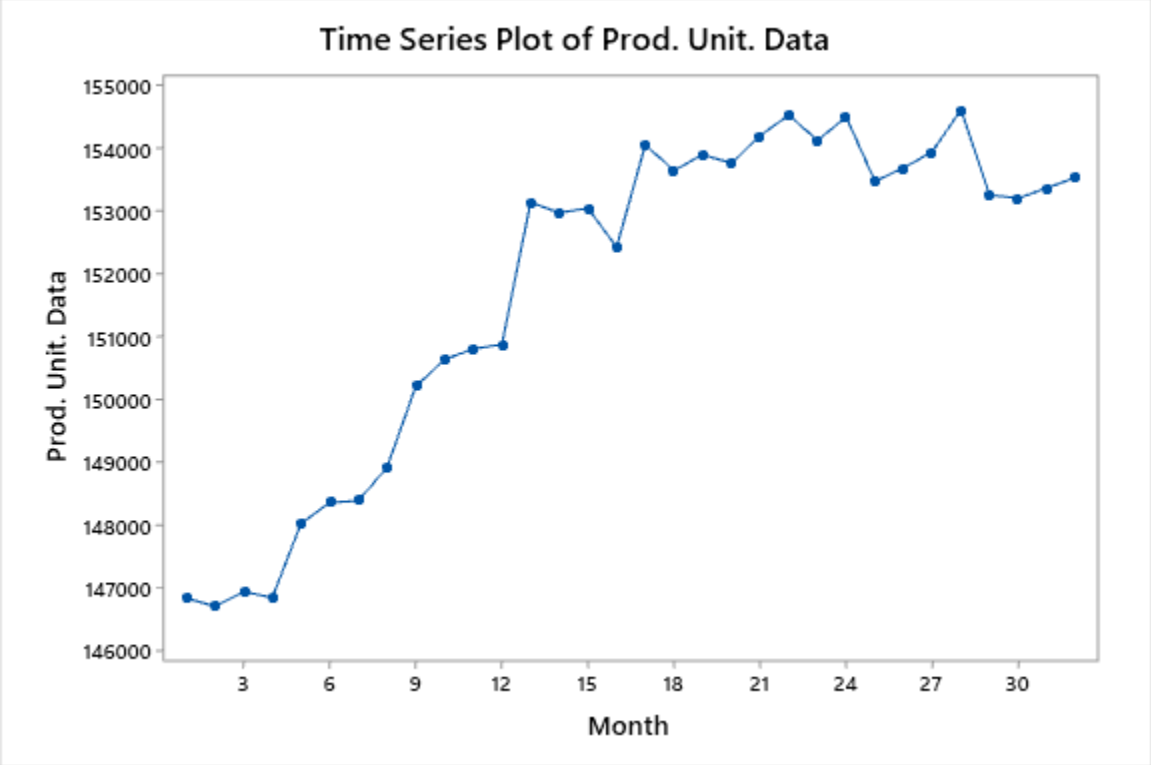
b.) Bar Chart:



c.) This problem points out the advantage of the bar chart over the pie chart. In the bar chart it is more evident that the Navistar sales are greater than AGCO which are greater than Oshkosh which are greater than Terex. In the pie chart, without the labels, differentiating between these four company revenues would be difficult.

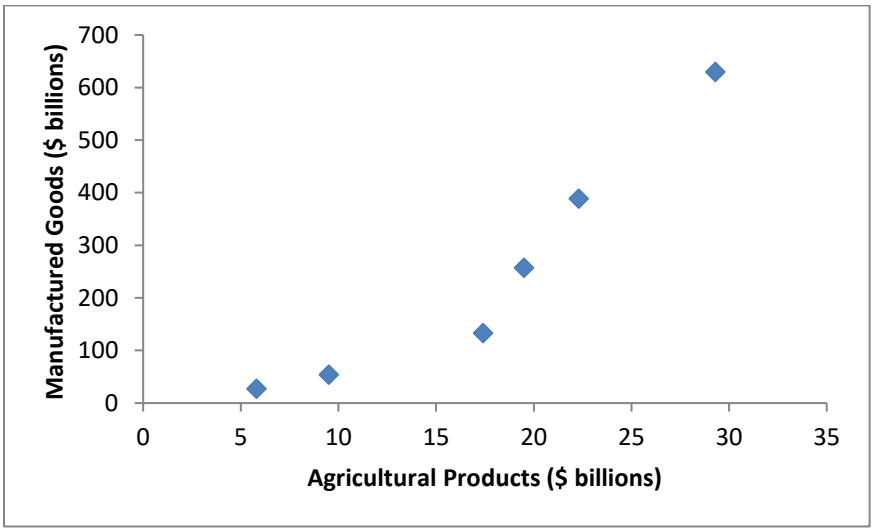
2.46





In the first graph with each year separate, you can see similar trends within each year and production unit data are relatively constant. The fact that the early years are on the bottom and ensuing years seem to move up tells us that there may be some sort of trend over the years. The second graph with all 8 years together shows that, indeed, there is some upward bound trend at least for the first 4 years or so.

2.47

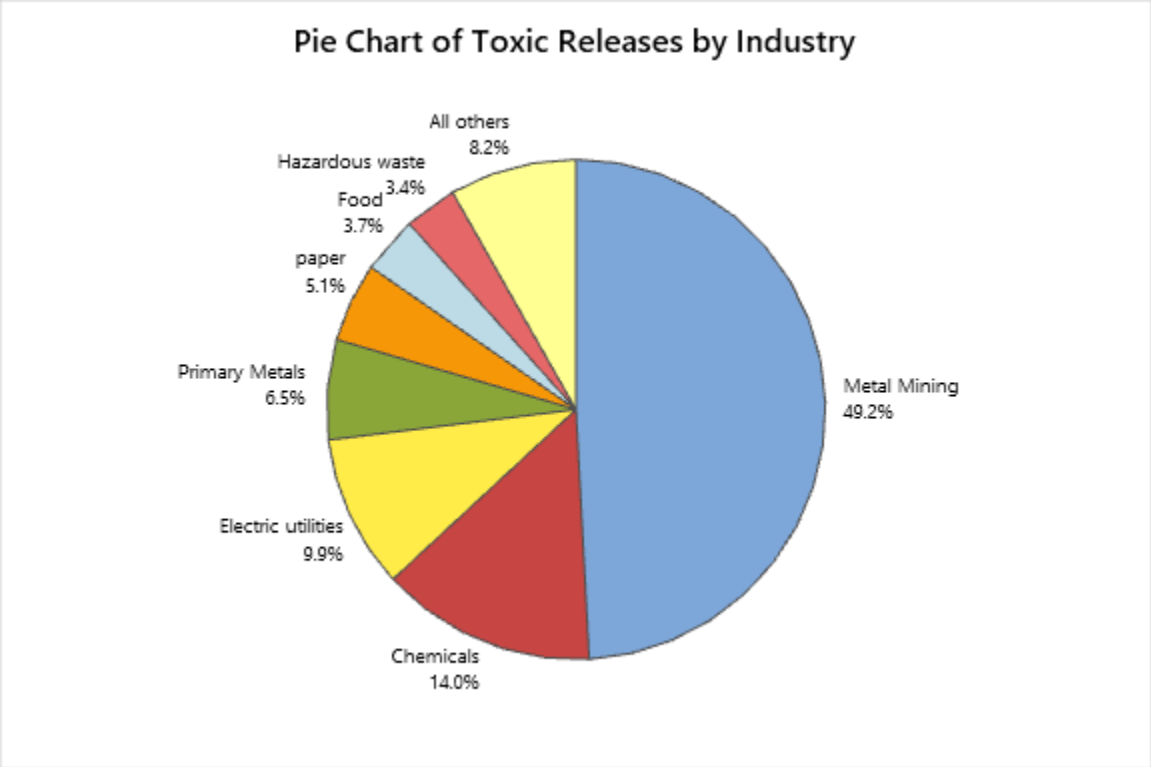


It can be observed that as the U.S. import of agricultural products increased, the U.S. import of manufactured goods also increased. As a matter of fact, a non-linear association may exist between the two variables.

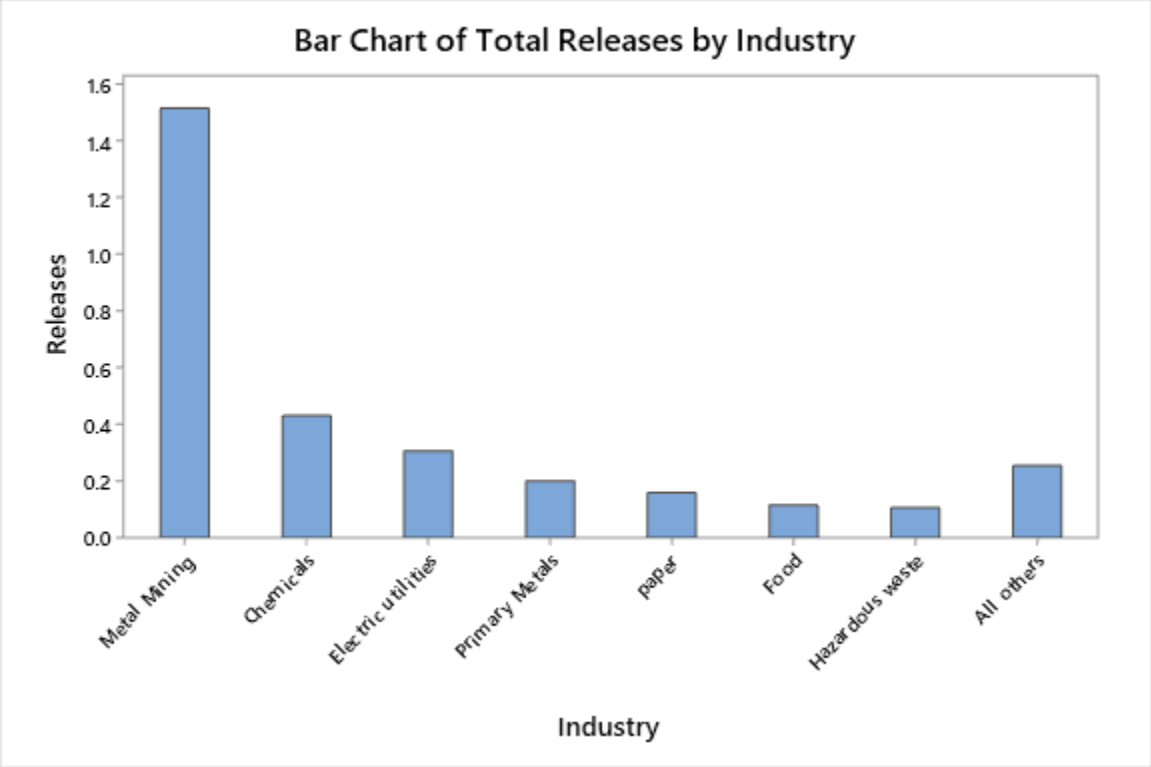
2.48

Industry	Total Release (billions lbs.)	Proportion	Degrees
Metal Mining	1.514	.49	177.0
Chemicals	0.430	.14	50.3
Electric Utilities	0.306	.10	35.8
Primary Metals	0.199	.06	23.3
Paper	0.158	.05	18.5
Food	0.114	.04	13.3
Hazardous Waste	0.105	.03	12.3
All Others	0.254	.08	29.7
TOTAL	3.080	0.99	360.2

Pie Chart:

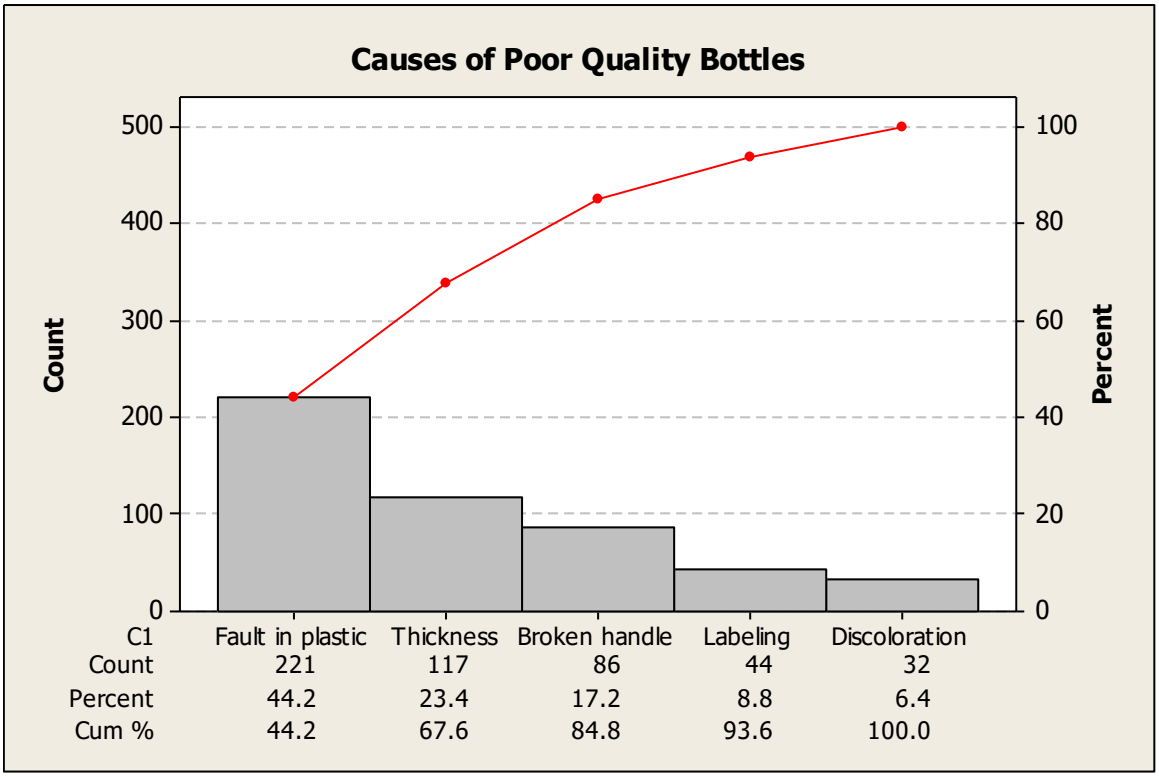


Bar Graph:



Of the several comparisons of pie charts and bar charts made in problems associated with this chapter, the differences between the two in this problem are the least clear. For example, in this bar chart the two industries, food and hazardous waste are so close in total releases that it is hard to differentiate them in either the bar chart or the pie chart (if the labels were not there). An advantage that the pie chart has is that pie charts can be pleasing to view. All in all, especially in this problem, the choice between the two is mostly a matter of personal (or professional) preference.

2.49



One of the main purposes of a Pareto chart is that it has the potential to help prioritize quality initiatives by ranking the top problems in order starting with the most frequently occurring problem. Thus, all things being equal, in attempting to improve the quality of plastic bottles, a quality team would begin with studying why there is a fault in plastic and determining how to correct for it. Next, the quality team would study thickness issues followed by causes of broken handles. Assuming that each problem takes a comparable time and effort to solve, the quality team could make greater strides sooner by following the items shown in the Pareto chart from left to right.

2.50

STEM	LEAF
42	12 16 24 32 99 99
43	04 28 39 46 61 88
44	20 40 59
45	12
46	53 54
47	30 34 58
48	22 34 66 78
49	63
50	48 49 90
51	66
52	21 54 57 63 91
53	38 66 66
54	31 78
55	56
56	69
57	37 50
58	31 32 58 73
59	19 23

2.51 The distribution of household income is bell-shaped with an average of about $ 90,000 and a range of from $ 30,000 to $ 140,000.

2.52 There is an especially heavy concentration of values between about 24 and 33. There is somewhat of a gap between 18 and 24 but an especially large gap between 52 and 66. Sixty-six appears to be an outlier.

2.53 We can conclude from the pie chart that of the six specialties graphed, emergency medicine is the smallest and internal medicine is the largest. General practice is so close in slice size to internal medicine that without the labels,it would be difficult to distinguish between the two. Pediatrics is number three but only about ½ the size of each of the first two largest. Anesthesiology and OB-GYN are in a virtual tie.

2.54 The fewest number of audits is 21, 36, and 39. The most audits is 27. More companies (8) performed 27 audits than any other number. Thirty-five companies performed between 12 and 19 audits. Only 7 companies performed 40 or more audits.

2.55 There appears to be a relatively strong positive relationship between the NASDAQ-100 and the DJIA. Note that generally as the DJIA became higher, the NASDAQ-100 tended to also get higher. Except for a few outliers, the slope of the graph was generally the same over time.

File: ch02, Chapter 2: Charts and Graphs

True/False

1. Frequency distribution is a summary of data presented in the form of class intervals and frequencies.
Ans: True
Title: Practice Question 1
Level of Difficulty: Easy
Learning Objective: LO 1
Feedback True: Correct! Frequency distribution is a summary of data presented in the form of class intervals and frequencies.
Feedback False: Incorrect. Frequency distribution is a summary of data presented in the form of class intervals and frequencies.

2. Data that have been organized into a frequency distribution are called ungrouped data.
Ans: False
Title: Practice Question 2
Level of Difficulty: Easy
Learning Objective: LO 1
Feedback True: Incorrect. Data that have been organized into a frequency distribution are not called ungrouped data.
Feedback False: Correct! Data that have been organized into a frequency distribution are not called ungrouped data.

3. The sum of the frequencies of a grouped data set is always equal to one.
Ans: False
Title: Practice Question 3
Level of Difficulty: Medium
Learning Objective: LO 1
Feedback True: Incorrect. The sum of the frequencies of a grouped data set is not always equal to one.
Feedback False: Correct! The sum of the frequencies of a grouped data set is not always equal to one.

4. A cumulative frequency distribution is also called an ogive.
Ans: True
Title: Practice Question 4
Level of Difficulty: Medium
Learning Objective: LO 2

Feedback True: Correct! A cumulative frequency distribution is also called an ogive.
Feedback False: Incorrect. A cumulative frequency distribution is also called an ogive.

5. A stem and leaf plot is ~~a~~ constructed by separating the digits for each number of the data into two groups.
Ans: True
Title: Practice Question 5
Level of Difficulty: Easy
Learning Objective: LO 2
Feedback True: Correct! Stem and leaf plot is constructed by separating the digits into two groups.
Feedback False: Incorrect. Stem and leaf plot is constructed by separating the digits into two groups.

6. A vertical bar chart that displays the frequency of observations ranked in order of occurrence from left to right is called a relative frequency histogram.
Ans: False
Title: Practice Question 6
Level of Difficulty: Medium
Learning Objective: LO 3
Feedback True: Incorrect. A vertical bar chart that displays the frequency of observations ranked in order of occurrence from left to right is not called a relative frequency histogram.
Feedback False: Correct! A vertical bar chart that displays the frequency of observations ranked in order of occurrence from left to right is not called a relative frequency histogram.

7. A pie chart is a circular depiction of data where the area of the whole pie represents 100% of the data and slices of the pie represents a percentage breakdown of the sublevels.
Ans: True
Title: Practice Question 7
Level of Difficulty: Easy
Learning Objective: LO 3
Feedback True: Correct! A pie chart is a circular depiction of data.
Feedback False: Incorrect. A pie chart is a circular depiction of data.

8. Cross tabulation is a process for producing a two-dimensional table that displays the frequency counts for two variables simultaneously.
Ans: True
Title: Practice Question 8
Level of Difficulty: Easy
Learning Objective: LO 4

Feedback True: Correct! Cross tabulation produces a two-dimensional table that displays the frequency counts for two variables simultaneously.
Feedback False: Incorrect. Cross tabulation produces a two-dimensional table that displays the frequency counts for two variables simultaneously.

Multiple Choice

9. Given a small dataset: 2.3, 7.1, 6.3, 5.8, 4.2, what is the range for this data?
a) 1.9
b) 2.5
c) 4.8
d) 2.3
Ans: c, 4.8
Title: Practice Question 9
Level of Difficulty: Easy
Learning Objective: LO 1
Feedback A: Incorrect. The range for this data is 4.8.
Feedback B: Incorrect. The range for this data is 4.8.
Feedback C: Correct! The range for this data is 4.8.
Feedback D: Incorrect. The range for this data is 4.8.

10. A useful tool for grouping data is _____.
a) a frequency distribution
b) group dynamics
c) segmentation
d) constructive statistics
Ans: a, a frequency distribution
Title: Practice Question 10
Level of Difficulty: Easy
Learning Objective: LO 1
Feedback A: Correct! A frequency distribution is a useful tool for grouping data.
Feedback B: Incorrect. A frequency distribution is a useful tool for grouping data.
Feedback C: Incorrect. A frequency distribution is a useful tool for grouping data.
Feedback D: Incorrect. A frequency distribution is a useful tool for grouping data.

11. A frequency distribution of a particular data can be presented as a table with two columns: one is the "Class Interval" and the other is the _____.
a) class width
b) class range
c) class frequency
d) class mid-point
Ans: c, class frequency
Title: Practice Question 11

Level of Difficulty: Medium
Learning Objective: LO 1
Feedback A: Incorrect. A frequency distribution of a particular data can be presented as a table with two columns: one is the "Class Interval" and the other is the class frequency.
Feedback B: Incorrect. A frequency distribution of a particular data can be presented as a table with two columns: one is the "Class Interval" and the other is the class frequency.
Feedback C: Correct! A frequency distribution of a particular data can be presented as a table with two columns: one is the "Class Interval" and the other is the class frequency.
Feedback D: Incorrect. A frequency distribution of a particular data can be presented as a table with two columns: one is the "Class Interval" and the other is the class frequency.

12. In a frequency distribution, which of the following is sometimes referred to as the class mark?
a) relative frequency
b) cumulative frequency
c) class midpoint
d) median class frequency
e) modal class frequency
Ans: c, class midpoint
Title: Practice Question 12
Level of Difficulty: Medium
Learning Objective: LO 1
Feedback A: Incorrect. In a frequency distribution, the class midpoint is sometimes referred to as the class mark.
Feedback B: Incorrect. In a frequency distribution, the class midpoint is sometimes referred to as the class mark.
Feedback C: Correct! In a frequency distribution, the class midpoint is sometimes referred to as the class mark.
Feedback D: Incorrect. In a frequency distribution, the class midpoint is sometimes referred to as the class mark.
Feedback E: Incorrect. In a frequency distribution, the class midpoint is sometimes referred to as the class mark.

13. For the following frequency distribution, what is the relative frequency for the class "30-under 40"?

Class interval	Frequency
10-under 20	20
20-under 30	16
30-under 40	30
40-under 50	18

a) 0.24
b) 0.36
c) 0.46

d) 0.21
e) 0.22
Ans: b, 0.36
Title: Practice Question 13
Level of Difficulty: Easy
Learning Objective: LO 1
Feedback A: Incorrect. The relative frequency for the class "30-under 40" is 0.36.
Feedback B: Correct! The relative frequency for the class "30-under 40" is 0.36.
Feedback C: Incorrect. The relative frequency for the class "30-under 40" is 0.36.
Feedback D: Incorrect. The relative frequency for the class "30-under 40" is 0.36.
Feedback E: Incorrect. The relative frequency for the class "30-under 40" is 0.36.

14. The range of a given dataset is 12.5, if 7 classes are used to construct a frequency distribution, what is width of each class is approximately?
a) 1.58
b) 1.79
c) 1.56
d) 1.75
Ans: b, 1.79
Title: Practice Question 14
Level of Difficulty: Easy
Learning Objective: LO 1
Feedback A: Incorrect. The width of each class is approximately 1.79.
Feedback B: Correct! The width of each class is approximately 1.79.
Feedback C: Incorrect. The width of each class is approximately 1.79.
Feedback D: Incorrect. The width of each class is approximately 1.79.

15. A frequency table about degrees of 40 CEOs was given:

Degree	Frequency
Bachelors	8
Laws	4
Masters	4
MBA	20
None	2
PhD	2
Total	40

What is the relative frequency of CEOs with PhD degrees?
a) 5%
b) 10%
c) 50%
d) 100%
Ans: a, 5%
Title: Practice Question 15

Level of Difficulty: Easy
Learning Objective: LO 1
Feedback A: Correct! The relative frequency of CEOs with PhD degrees is 5%.
Feedback B: Incorrect. The relative frequency of CEOs with PhD degrees is 5%.
Feedback C: Incorrect. The relative frequency of CEOs with PhD degrees is 5%.
Feedback D: Incorrect. The relative frequency of CEOs with PhD degrees is 5%.

16. Two techniques for grouping data are _____ and _____.
a) scatter plot and Pareto chart
b) frequency distribution and stem-and-leaf plot
c) class intervals and class frequencies
d) histogram and pie chart
Ans: b, frequency distribution and stem-and-leaf plot
Title: Practice Question 16
Level of Difficulty: Easy
Learning Objective: LO 1, LO 2
Feedback A: Incorrect. Two techniques for grouping data are frequency distribution and stem-and-leaf plot.
Feedback B: Correct! Two techniques for grouping data are frequency distribution and stem-and-leaf plot.
Feedback C: Incorrect. Two techniques for grouping data are frequency distribution and stem-and-leaf plot.
Feedback D: Incorrect. Two techniques for grouping data are frequency distribution and stem-and-leaf plot.

17. The stem in a stem-and-leaf plot of numbers with five digits, the stem has three digits. For the number, 45637, the leaf is _____.
a) 456
b) 37
c) 637
d) 63
e) 74
Ans: b, 37
Title: Practice Question 17
Level of Difficulty: Medium
Learning Objective: LO 2
Feedback A: Incorrect. The leaf is 37.
Feedback B: Correct! The leaf is 37.
Feedback C: Incorrect. The leaf is 37.
Feedback D: Incorrect. The leaf is 37.

18. A histogram is vertical bar chart on an X-Y plane, in which the labels along horizontal (or X) axis represent the class end points and the bars along the vertical (or Y) axis show the _____.
a) class width
b) class range
c) class frequency
d) class mid-point
Ans: c, class frequency
Title: Practice Question 18
Level of Difficulty: Hard
Learning Objective: LO 2
Feedback A: Incorrect. The bars along the vertical (or Y) axis show the class frequency.
Feedback B: Incorrect. The bars along the vertical (or Y) axis show the class frequency.
Feedback C: Correct! The bars along the vertical (or Y) axis show the class frequency.
Feedback D: Incorrect. The bars along the vertical (or Y) axis show the class frequency.

19. Relative frequencies for a set of data must add up to _____.
a) the total frequency
b) the cumulative frequency
c) one
d) hundred
Ans: c, one
Title: Practice Question 19
Level of Difficulty: Medium
Learning Objective: LO 2
Feedback A: Incorrect. Relative frequencies for a set of data must add up to one.
Feedback B: Incorrect. Relative frequencies for a set of data must add up to one.
Feedback C: Correct! Relative frequencies for a set of data must add up to one.
Feedback D: Incorrect. Relative frequencies for a set of data must add up to one.

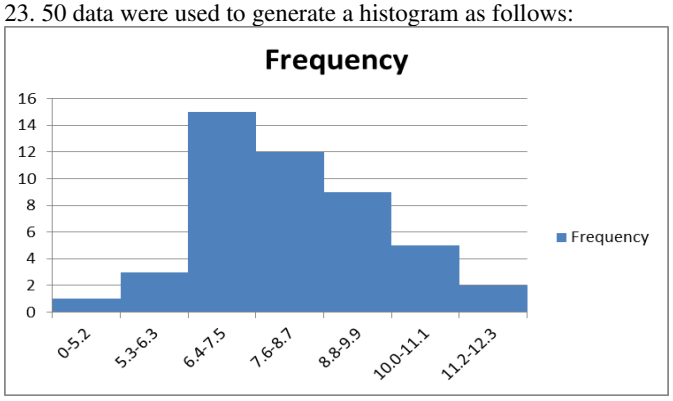
20. A type of vertical bar chart that is used to depict a frequency distribution is called a _____.
a) frequency polygon
b) histogram
c) dot plot
d) ogive
Ans: b, histogram
Title: Practice Question 20
Level of Difficulty: Easy
Learning Objective: LO 2
Feedback A: Incorrect. A type of vertical bar chart that is used to depict a frequency distribution is called a histogram.
Feedback B: Correct! A type of vertical bar chart that is used to depict a frequency distribution is called a histogram.

Feedback C: Incorrect. A type of vertical bar chart that is used to depict a frequency distribution is called a histogram.
Feedback D: Incorrect. A type of vertical bar chart that is used to depict a frequency distribution is called a histogram.

21. One of the ways by which graphical displays such as histogram are used to "lie with statistics" is by _____.
a) using multiple colors
b) using 8 classes instead of 10
c) using grouped data instead of raw data
d) changing the scales of the axes
Ans: d, changing the scales of the axes
Title: Practice Question 21
Level of Difficulty: Medium
Learning Objective: LO 2
Feedback A: Incorrect. One of the ways by which graphical displays such as histogram are used to "lie with statistics" is by changing the scale of the axes.
Feedback B: Incorrect. One of the ways by which graphical displays such as histogram are used to "lie with statistics" is by changing the scale of the axes.
Feedback C: Incorrect. One of the ways by which graphical displays such as histogram are used to "lie with statistics" is by changing the scale of the axes.
Feedback D: Correct! One of the ways by which graphical displays such as histogram are used to "lie with statistics" is by changing the scale of the axes.

22. An advantage of stem-and-leaf plots compared to most frequency distributions is _____.
a) grouped data are sufficient and the values of original data are not required
b) they can be used for botanical data unlike other charts
c) the values of original raw data are retained
d) different colors used for stem and leaf makes it attractive
Ans: c, the values of original raw data are retained
Title: Practice Question 22
Level of Difficulty: Medium
Learning Objective: LO 2
Feedback A: Incorrect. An advantage of stem-and-leaf plots compared to most frequency distributions is the values of original raw data are retained.
Feedback B: Incorrect. An advantage of stem-and-leaf plots compared to most frequency distributions is the values of original raw data are retained.
Feedback C: Correct! An advantage of stem-and-leaf plots compared to most frequency distributions is the values of original raw data are retained.
Feedback D: Incorrect. An advantage of stem-and-leaf plots compared to most frequency distributions is the values of original raw data are retained.

23. 50 data were used to generate a histogram as follows:



What is the approximate percentage of the data between (including) 6.4 to 8.7?
a) 30%
b) 54%
c) 24%
d) 36%
Ans: b, 54%
Title: Practice Question 23
Level of Difficulty: Medium
Learning Objective: LO 2
Feedback A: Incorrect. The approximate percentage of the data between (including) 6.4 to 8.7 is 54%.
Feedback B: Correct! The approximate percentage of the data between (including) 6.4 to 8.7 is 54%.
Feedback C: Incorrect. The approximate percentage of the data between (including) 6.4 to 8.7 is 54%.
Feedback D: Incorrect. The approximate percentage of the data between (including) 6.4 to 8.7 is 54%.

24. Consider the stem-and-leaf display shown here.

Stem	Leaf
0	12
1	456
2	01135
3	12334689
4	2358
5	113

How many observations were in the original data set?
a) 21
b) 22
c) 253
d) 24

Formatted: Not Highlight

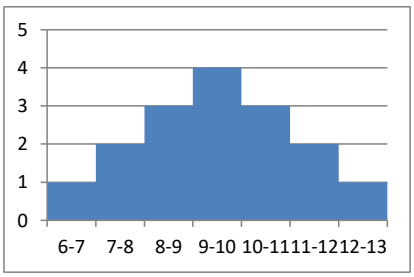
Ans: c, 25

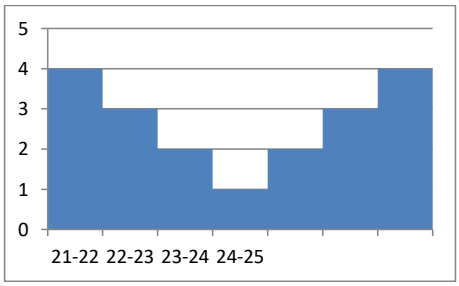
Commented [MJ1]: Maybe I am doing something wrong, but I definitely see 25 observations. I'm not sure how one could get 23.

Title: Practice Question 24
Level of Difficulty: Easy
Learning Objective: LO 2
Feedback A: Incorrect. There are 25 observations in the original data set.
Feedback B: Incorrect. There are 25 observations in the original data set.
Feedback C: Correct! There are 25 observations in the original data set.
Feedback D: Incorrect. There are 25 observations in the original data set.

25. In the bottom row of the stem-and-leaf display, identify the stem, and the numbers in the original data set represented by this stem and its leaves.
a) stem: 5; original numbers: 15, 15, 35
b) stem: 5; original numbers: 51, 51, 53
c) stem: 1, 1, 3; original numbers: 15, 15, 35
d) stem: 1, 1, 3; original numbers: 51, 51, 53
Ans: b, stem: 5; original numbers: 51, 51, 53
Title: Practice Question 25
Level of Difficulty: Medium
Learning Objective: LO 2
Feedback A: Incorrect. The correct original numbers are 51, 51, 53.
Feedback B: Correct! The correct original numbers are 51, 51, 53.
Feedback C: Incorrect. The correct original numbers are 51, 51, 53.
Feedback D: Incorrect. The correct original numbers are 51, 51, 53.

26. Consider the histograms below





Which of the following statements are true?
I. Both data sets are symmetric.
II. Both data sets have the same range.
a) I only
b) II only
c) I and II
d) Neither is true
Ans: c, I and II
Title: Practice Question 26
Level of Difficulty: Medium

Learning Objective: LO 2
Feedback A: Incorrect. The correct statements are I and II.
Feedback B: Incorrect. The correct statements are I and II.
Feedback C: Correct! The correct statements are I and II.
Feedback D: Incorrect. The correct statements are I and II.

27. Given the following table on the data of 200 patients in a hospital about their blood groups, what is the angle (number of degrees) of AB type blood in a pie chart?

Blood Group	Frequency
A	60
B	70
AB	
O	45
Total	

a) 108°
b) 100°
c) 45°
d) 81°
e) 125°
Ans: c, 45°
Title: Practice Question 27
Level of Difficulty: ~~Medium~~Hard
Learning Objective: LO 3 ~~2.3: Construct different types of qualitative data graphs, including pie charts, bar graphs, and Pareto charts, in order to interpret the data being graphed.~~
Feedback A: Incorrect. The angle of AB type blood in a pie chart is 45°
Feedback B: Incorrect. The angle of AB type blood in a pie chart is 45°
Feedback C: Correct! The angle of AB type blood in a pie chart is 45°
Feedback D: Incorrect. The angle of AB type blood in a pie chart is 45°

28. What is the relative frequency of type B in the previous question?
a) 0.25
b) 0.35
c) 0.5
d) 0.64
Ans: b, 0.35
Title: Practice Question 28
Level of Difficulty: Medium
Learning Objective: LO 2
Feedback A: Incorrect. The relative frequency of type B is .35.
Feedback B: Correct! The relative frequency of type B is .35.
Feedback C: Incorrect. The relative frequency of type B is .35.
Feedback D: Incorrect. The relative frequency of type B is .35.

29. In sciences and technology, use of pie charts is limited compared to other types of graphs such as a histogram, because _____.
a) it is difficult to draw pie charts
b) pie charts require data to be measured on interval or ratio scale
c) it is more difficult to interpret the relative size of angles in a pie chart than judge the height of rectangles in a histogram
d) pie chart cannot be used for grouped data
Ans: c, it is more difficult to interpret the relative size of angles in a pie chart than judge the height of rectangles in a histogram
Title: Practice Question 29
Level of Difficulty: Hard
Learning Objective: LO 3
Feedback A: Incorrect. In sciences and technology, use of pie charts is limited compared to other types of graphs such as a histogram, because it is more difficult to interpret the relative size of angles in a pie chart than judge the height of rectangles in a histogram.
Feedback B: Incorrect. In sciences and technology, use of pie charts is limited compared to other types of graphs such as a histogram, because it is more difficult to interpret the relative size of angles in a pie chart than judge the height of rectangles in a histogram.
Feedback C: Correct! In sciences and technology, use of pie charts is limited compared to other types of graphs such as a histogram, because it is more difficult to interpret the relative size of angles in a pie chart than judge the height of rectangles in a histogram.
Feedback D: Incorrect. In sciences and technology, use of pie charts is limited compared to other types of graphs such as a histogram, because it is more difficult to interpret the relative size of angles in a pie chart than judge the height of rectangles in a histogram.

30. A chart commonly used in quality management to separate the most important types of defects from trivial ones is called a _____.
a) Deming Chart
b) Quality Control Chart
c) Taguchi Chart
d) Pareto Chart
Ans: d, Pareto Chart
Title: Practice Question 30
Level of Difficulty: Easy
Learning Objective: LO 3
Feedback A: Incorrect. A chart commonly used in quality management to separate the most important types of defects from trivial ones is called a Pareto Chart.
Feedback B: Incorrect. A chart commonly used in quality management to separate the most important types of defects from trivial ones is called a Pareto Chart.
Feedback C: Incorrect. A chart commonly used in quality management to separate the most important types of defects from trivial ones is called a Pareto Chart.
Feedback D: Correct! A chart commonly used in quality management to separate the most important types of defects from trivial ones is called a Pareto Chart.

31. Which of the following always examine the relationship between two variables?
a) Pie Chart
b) Cross Tabulation Table
c) Scatter Plot
d) Pareto Chart
e) none of the above
f) both b and c
Ans: f, both b and c
Title: Practice Question 31
Level of Difficulty: Easy
Learning Objective: LO 4
Feedback A: Incorrect. Cross Tabulation Tables and Scatter Plots always examine the relationship between two variables.
Feedback B: Incorrect. Cross Tabulation Tables and Scatter Plots always examine the relationship between two variables.
Feedback C: Incorrect. Cross Tabulation Tables and Scatter Plots always examine the relationship between two variables.
Feedback D: Incorrect. Cross Tabulation Tables and Scatter Plots always examine the relationship between two variables.
Feedback E: Incorrect. Cross Tabulation Tables and Scatter Plots always examine the relationship between two variables.
Feedback F: Correct! Cross Tabulation Tables and Scatter Plots always examine the relationship between two variables.

32. Which of the following is/are true of Cross Tabulation Tables?
a) Cross tabulation is a process for producing a two-dimensional table that displays the frequency counts for two variables simultaneously.
b) Ordinal data may be displayed in a Cross Tabulation Table.
c) Cross Tabulation Tables indicate the relative frequency of observations within a class.
d) both a and b
e) all of the above
Ans: d, both a and b
Title: Practice Question 32
Level of Difficulty: Medium
Learning Objective: LO 4
Feedback A: Incorrect. Both a and b are correct.
Feedback B: Incorrect. Both a and b are correct.
Feedback C: Incorrect. Both a and b are correct.
Feedback D: Correct! Both a and b are correct.
Feedback E: Incorrect. Both a and b are correct.

33. Which of the following would be a valid purpose for a Cross Tabulation Tables?
a) display number of people in each income bracket by age range
b) calculate the average math score by grade level
c) correlate average height with average weight of a group of people
d) all of the above
e) none of the above
Ans: a, display income bracket by age range
Title: Practice Question 33
Level of Difficulty: Medium
Learning Objective: LO 4
Feedback A: Correct! Displaying income brackets by age range would be a valid purpose of a Cross Tabulation Table.
Feedback B: Incorrect. Only a is correct. Displaying income brackets by age range would be a valid purpose of a Cross Tabulation Table.
Feedback C: Incorrect. Only a is correct. Displaying income brackets by age range would be a valid purpose of a Cross Tabulation Table.
Feedback D: Incorrect. Only a is correct. Displaying income brackets by age range would be a valid purpose of a Cross Tabulation Table.
Feedback E: Incorrect. Only a is correct. Displaying income brackets by age range would be a valid purpose of a Cross Tabulation Table.

Formatted: Font: Bold, Complex Script Font: Bold

34. Which of the following is/are true of Time Series Data?
a) it uses historical data to estimate what might happen in the future
b) it shows trends over time
c) it can be used to produce a forecast
d) it can be in the form of a pie chart
e) all of the above
f) a, b, and c only
Ans: f, a, b, and c only
Title: Practice Question 34
Level of Difficulty: Medium
Learning Objective: LO 5
Feedback A: Incorrect. a, b, and c are correct.
Feedback B: Incorrect. a, b, and c are correct.
Feedback C: Incorrect. a, b, and c are correct.
Feedback D: Incorrect. a, b, and c are correct.
Feedback E: Incorrect. a, b, and c are correct.
Feedback F: Correct! a, b, and c are correct.

Formatted: Font: Bold

35. Which of the following statements regarding Time Series Data is/are false?
a) data needs to be cleaned such that measurements are taken on regular time intervals
b) data must be arranged according to time of occurrence

c) data is often plotted as a line chart
d) ethical considerations are not applicable since the data is unbiased
e) all of the above are true statements
Ans: d, ethical considerations are not applicable since the data is unbiased
Title: Practice Question 35
Level of Difficulty: Easy
Learning Objective: LO 5
Feedback A: Incorrect. ~~d~~Ethical considerations are not applicable since the data is unbiased is a false statement.
Feedback B: Incorrect. ~~d~~Ethical considerations are not applicable since the data is unbiased is a false statement.
Feedback C: Incorrect. ~~d~~ Ethical considerations are not applicable since the data is unbiased is a false statement.
Feedback D: Correct! Ethical considerations are not applicable since the data is unbiased is a false statement.
Feedback E: Incorrect. Ethical considerations are not applicable since the data is unbiased~~d~~ is a false statement.

Formatted: Font: Bold, Complex Script Font: Bold

36. Time Series Data falls under which category of Business Analytics?
a) descriptive analytics
b) predictive analytics
c) prescriptive analytics
d) non-parametric statistics
e) none of the above
Ans: a, descriptive analytics
Title: Practice Question 36
Level of Difficulty: Easy
Learning Objective: LO 5
Feedback A: Correct! Time Series Data falls under the descriptive analytics category of Business Analytics.
Feedback B: Incorrect. Time Series Data falls under the descriptive analytics category of Business Analytics.
Feedback C: Incorrect. Time Series Data falls under the descriptive analytics category of Business Analytics.
Feedback D: Incorrect. Time Series Data falls under the descriptive analytics category of Business Analytics.
Feedback E: Incorrect. Time Series Data falls under the descriptive analytics category of Business Analytics.

Formatted: Font: Bold, Complex Script Font: Bold

Big Data Case Answers for Chapter 2

1. Cross Tabulation Tables:

a.) Region and Control – the following crosstabulation table was constructed using the Pivot Table in Excel. In the database, both Region and Control have been converted to categorical values that were then used in the creation of this table. The analyst can see the number of hospitals in the various regions, the number of hospitals under the various types of control, and the intersection of each. Pie charts and other visuals can be used to display the various types of control, for example, in the Mid Atlantic region.

Count of Control	**Column Labels**					
Row Labels	**Gov Fed**	**Gov NonFed**	**InvestorOwned ForProfit**	**NonGov NonProfit**	**Other Nonprofit**	**Grand Total**
East North Central	7	36	46	38	135	262
East South Central	2	42	34	10	48	136
Mid Atlantic	8	20	18	11	111	168
Mountain	6	28	29	15	82	160
New England	3	6	10	3	78	100
Pacific	2	48	31	38	140	259
South Atlantic	9	44	46	10	143	252
West North Central	6	54	32	27	95	214
West South Central	3	64	233	41	108	449
Grand Total	**46**	**342**	**479**	**193**	**940**	**2000**

b.) Region and Service – This Excel-produced crosstabulation table between region and service was created using the Pivot Table. It shows which types of service dominate the hospital database (General Medical & Surgery, followed by Psychiatric, and Long-term Acute), and how the types of service breakdown by Region.

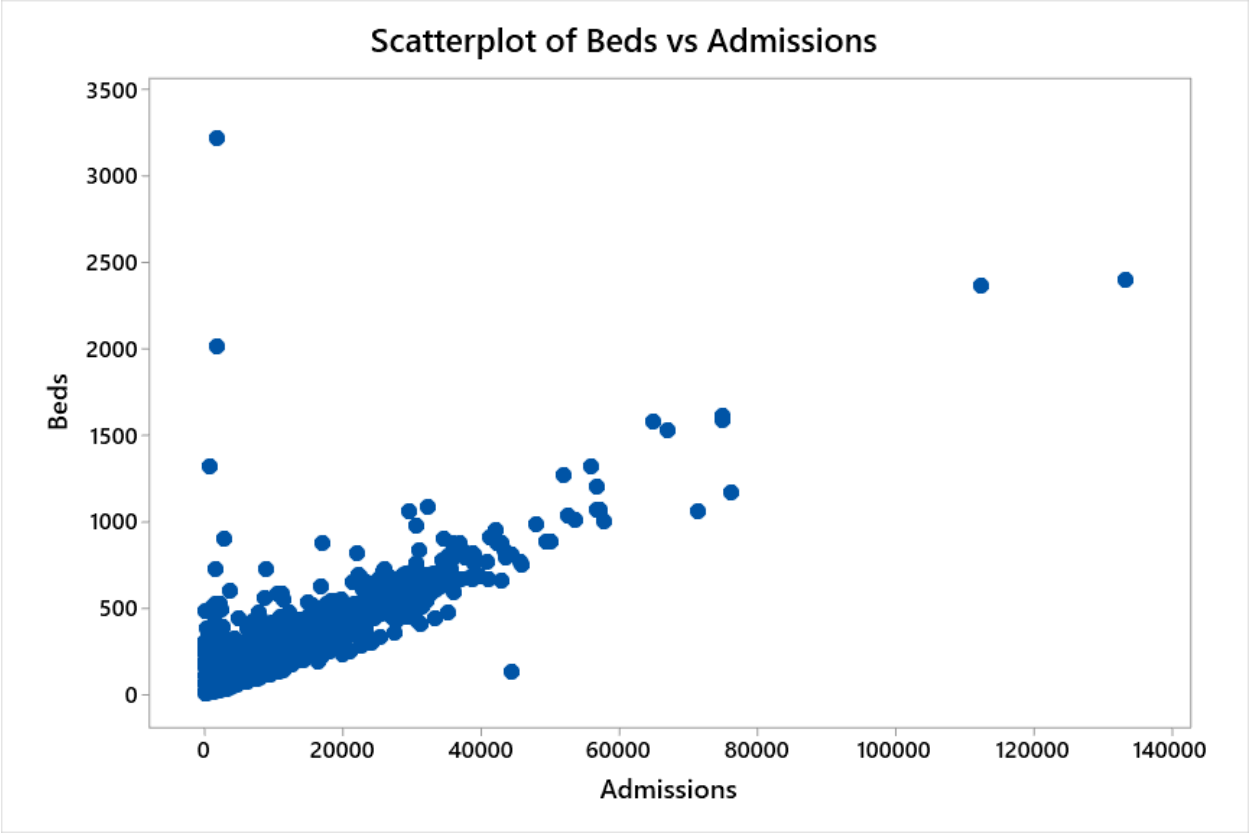
Row Labels	East North Central	East South Central	Mid Atlantic	Mountain	New England	Pacific	South Atlantic	West North Central	West South Central	Grand Total
Alcoh & Chem Dependency	3						3	1		7
Cancer	2		2			3	1		1	9
Children's	10	3	6	5	5	14	4	12	22	81
Chronic Disease					1		1			2
Gen Med & Surg	170	101	112	127	65	218	179	147	197	1316
Heart	3								4	7
Intellectual Disab								2		2
Long-term Acute	17	6	9	3	5	3	18	17	86	164
Obstetrics & Gyn	1	1	1		1	1			3	8
Orthopedic	2		2	1	1		1	1	4	12
Other Specialty			2	1	1	1	5	2	1	13
Psychiatric	38	14	19	12	13	13	20	18	58	205
Rehab	15	11	14	9	8	4	19	11	58	149
Surgical	1		1	2		2	1	3	15	25
Grand Total	**262**	**136**	**168**	**160**	**100**	**259**	**252**	**214**	**449**	**2000**

c.) Service and Control – This Excel-produced crosstabulation table between service and control was created using the Pivot Table. It shows which types of service dominate the hospital database (General Medical & Surgery, followed by Psychiatric, Long-term Acute, and Rehab), and how the types of service breakdown by type of control. For example, Government Federal Control (mostly veteran's hospitals) are mainly General Medical & Surgery hospitals. Whereas, the modal control of rehab hospitals is Owned by Investors and are For Profit.

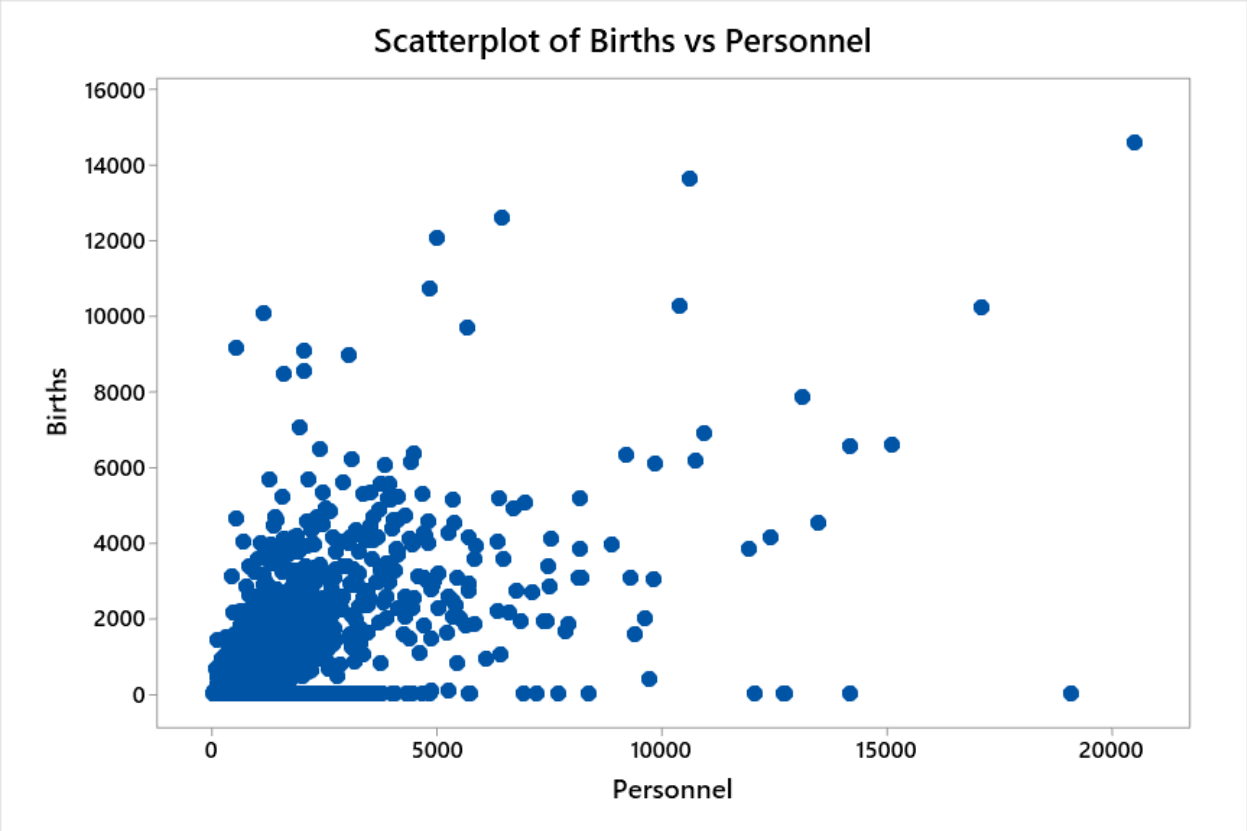
Count of Control	**Column Labels**					
Row Labels	**Gov Fed**	**Gov NonFed**	**InvestorOwned ForProfit**	**NonGov NonProfit**	**Other Nonprofit**	**Grand Total**
Alcoh & Chem Dependency		1	3		3	7
Cancer		3			6	9
Children's		2	11	2	66	81
Chronic Disease		1			1	2
Gen Med & Surg	43	249	127	167	730	1316
Heart			6		1	7
Intellectual Disab		2				2
Long-term Acute		6	113	13	32	164
Obstetrics & Gynecology			3		5	8
Orthopedic			8		4	12
Other Specialty	1	2			10	13
Psychiatric	2	70	81	4	48	205
Rehab		4	107	6	32	149
Surgical		2	20	1	2	25
Grand Total	**46**	**342**	**479**	**193**	**940**	**2000**

2. Four Different Scatter Plots – As an example, presented here are four different scatter plots from this AHA large database.

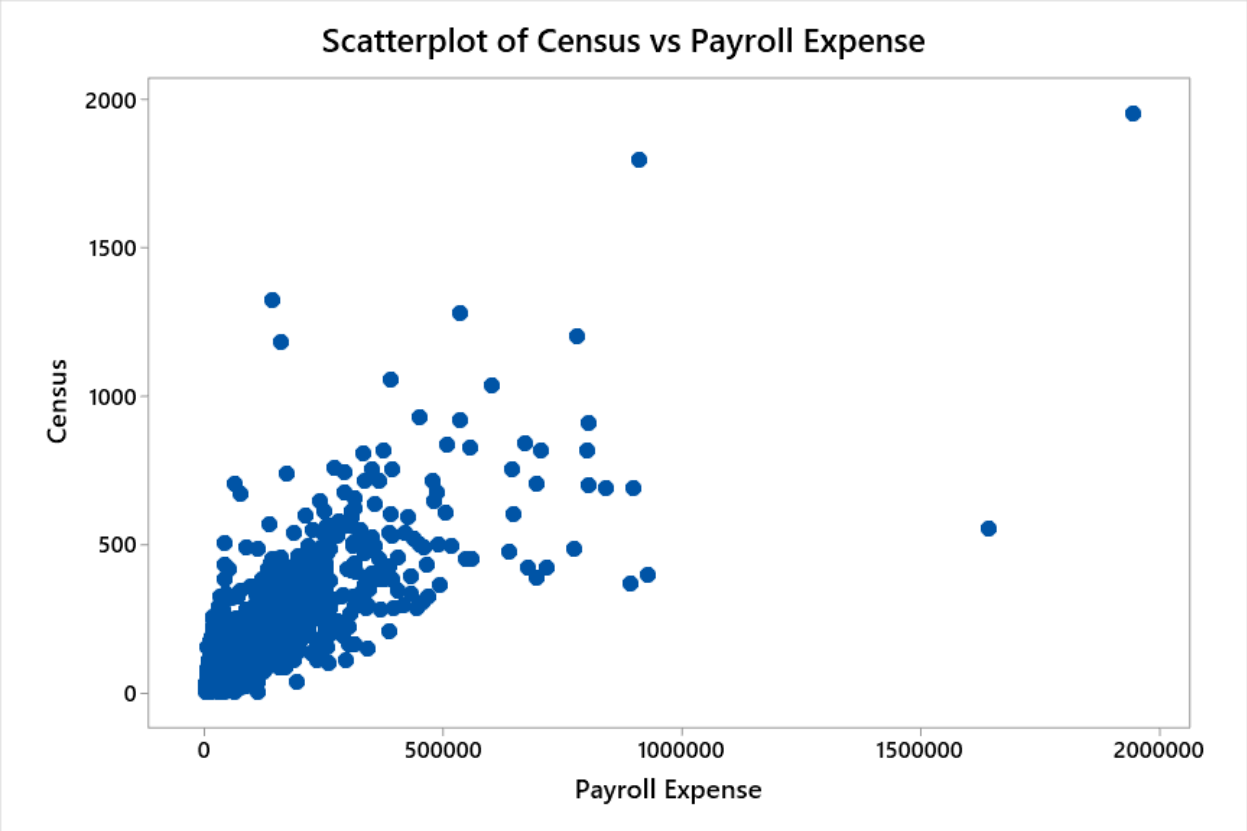
a.) Shown here is a Minitab-produced scatterplot of Number of Beds vs. Number of Annual Admissions. Of the four scatter plots, this one between beds and admissions, most indicates a potential strong correlation.



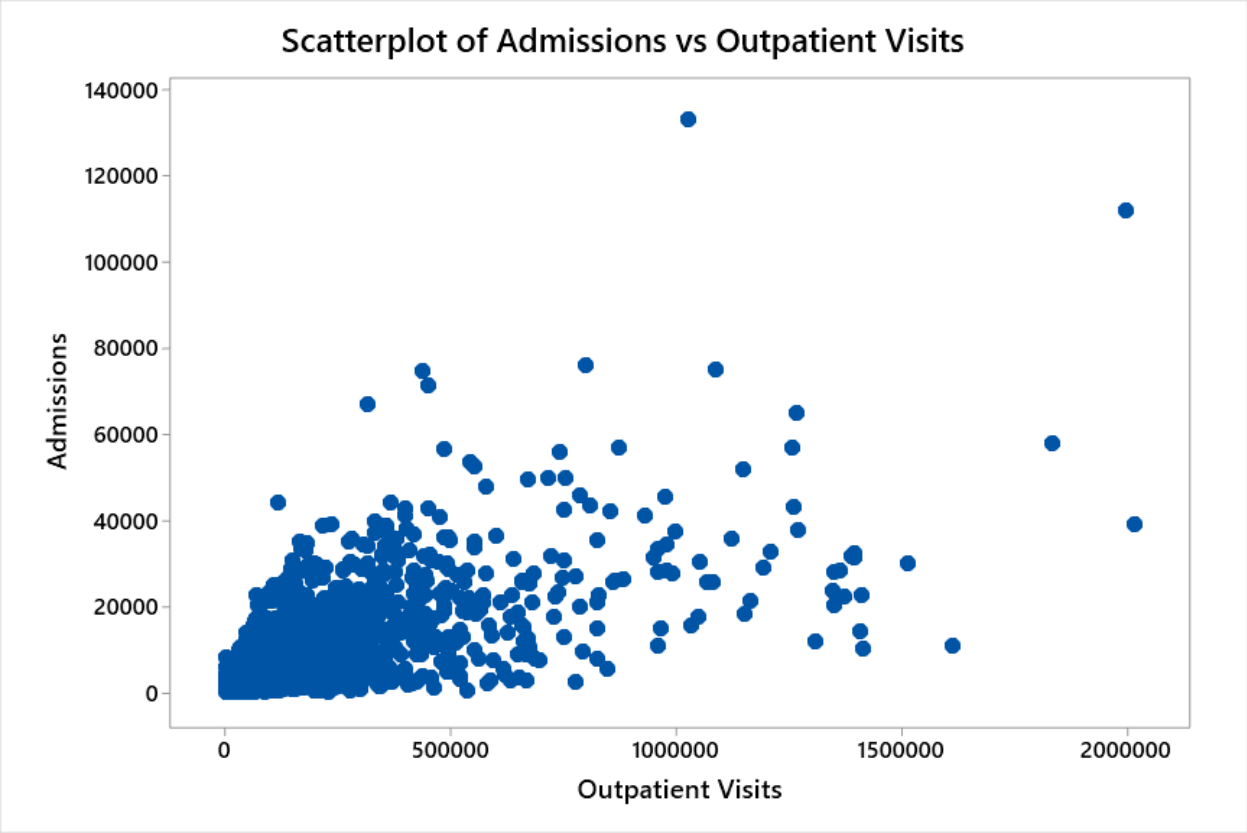
b.) Shown here is a Minitab-produced scatterplot of Number of Births vs. Number of Personnel. This scatter plot indicates that the relationship between births and personnel is probably the weakest of the four scatter plots.



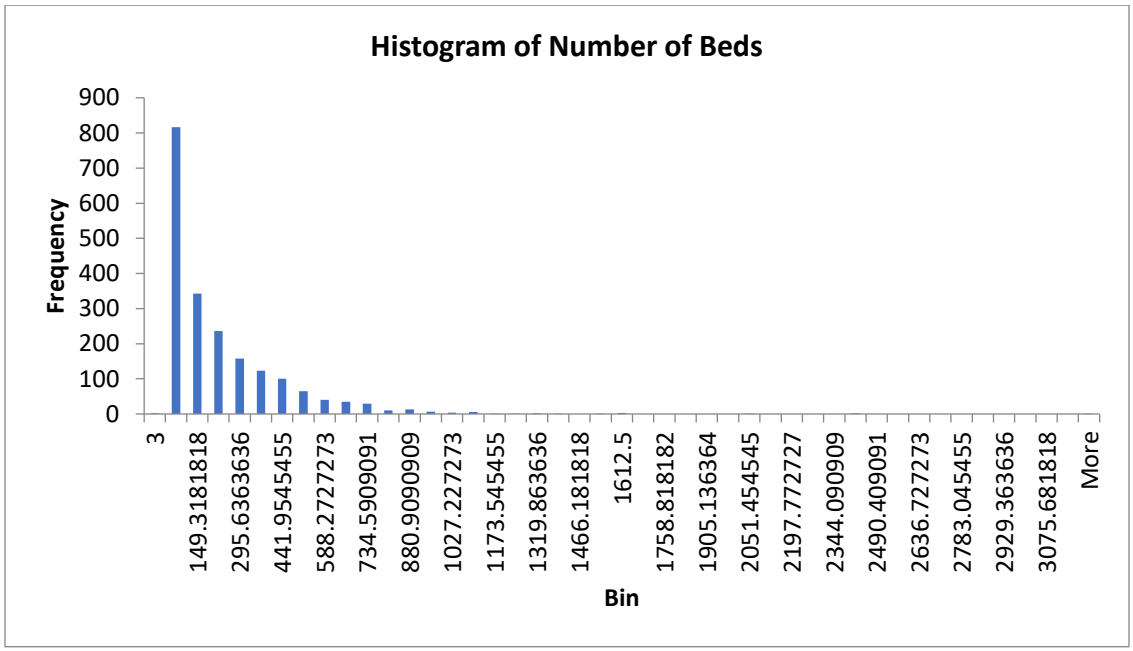
c.) Shown here is a Minitab-produced scatterplot of Census vs. Payroll Expense. This scatter plot indicates that the relationship between census and payroll expense is stronger than births vs. personnel and weaker than beds vs. admissions.



d.) Shown here is a Minitab-produced scatterplot of Admissions vs. Outpatient Visits. This scatter plot indicates that the relationship between admissions and outpatient visits is about as weak as for births vs. personnel.

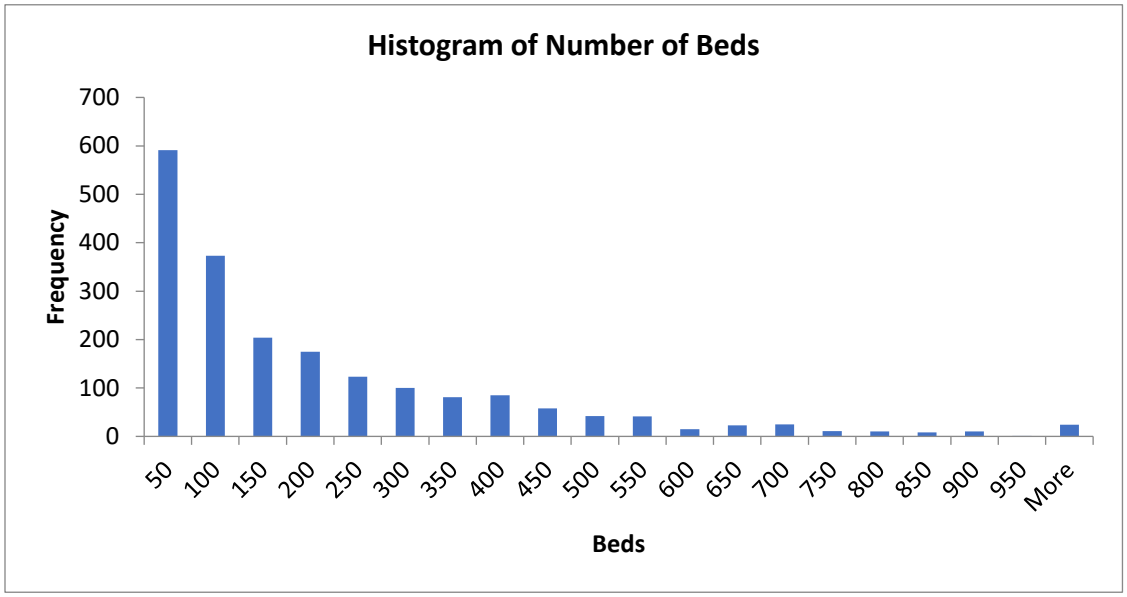


3. Histograms of Number of Beds. Shown first is a histogram of Number of Beds constructed from Excel. For this histogram, Excel created the bins.

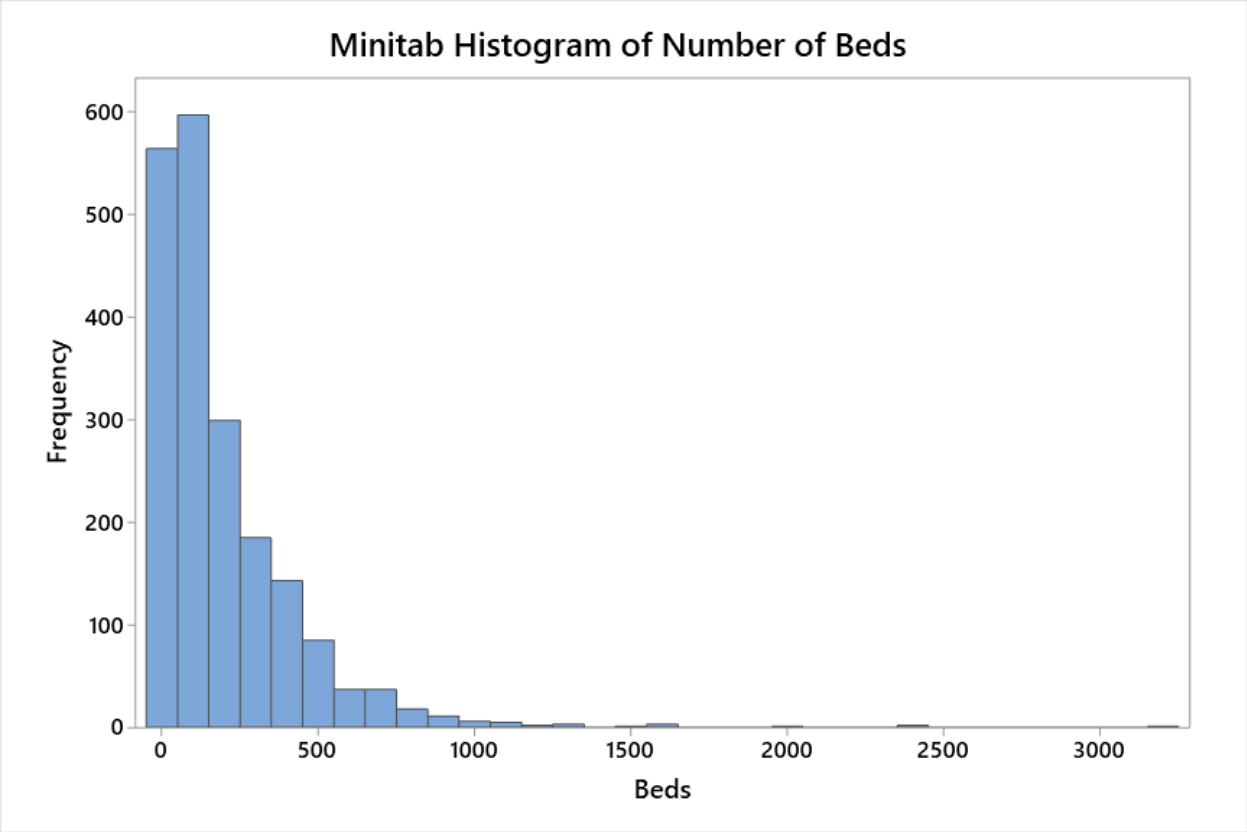


This shows that a few large hospitals are pulling the class intervals way to the right. There is a large "spike" of hospitals probably around 50 beds.

Next, we provided Excel with bins of 50, 100, 150, 200, 250, …, 950. The resulting histogram makes the distribution a little easier to see the breaks in number of beds classes with many of the hospital having less than 150 beds.



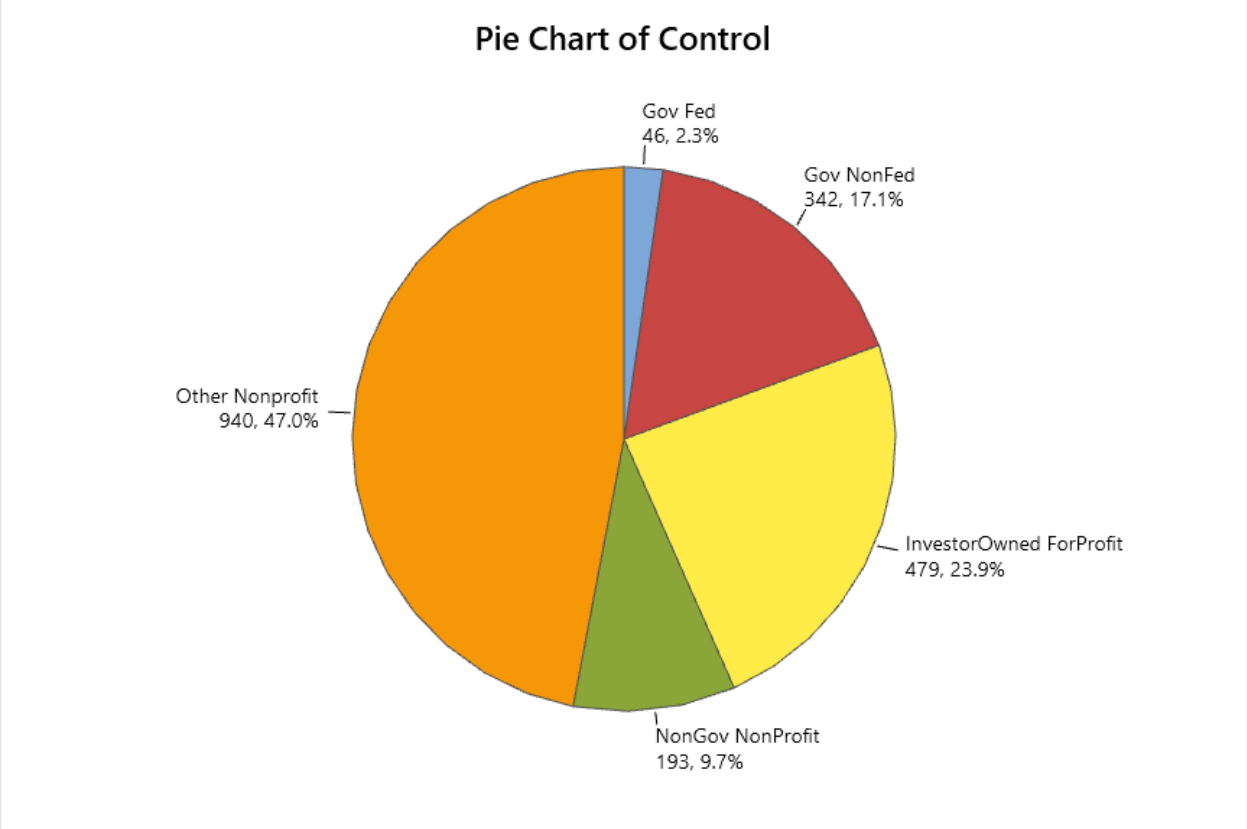
Shown next is a histogram of Number of Beds created in Minitab.



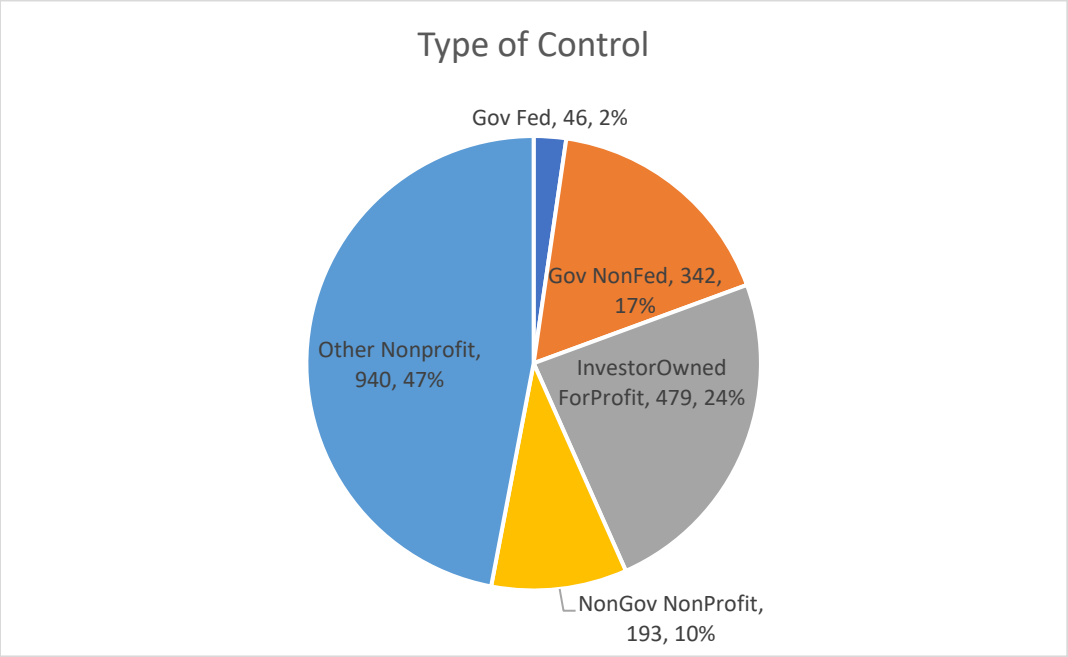
The student should be able to study their histogram and reach some conclusions about the size of the hospitals in this database. Obviously, we can see that they "pile-up" on the left of the graph and have a long thin "tail" to the right which is indicative of a graph that is skewed to the right (an opportunity to introduce a chapter 3 concept visually). Most of the hospitals in the database have less than 500 beds. In addition, one can see from these different histograms that the bin range (class endpoints) makes a difference in how the data are visualized.

4. Construct a pie chart of the variable, Control.

From Minitab,



From Excel,

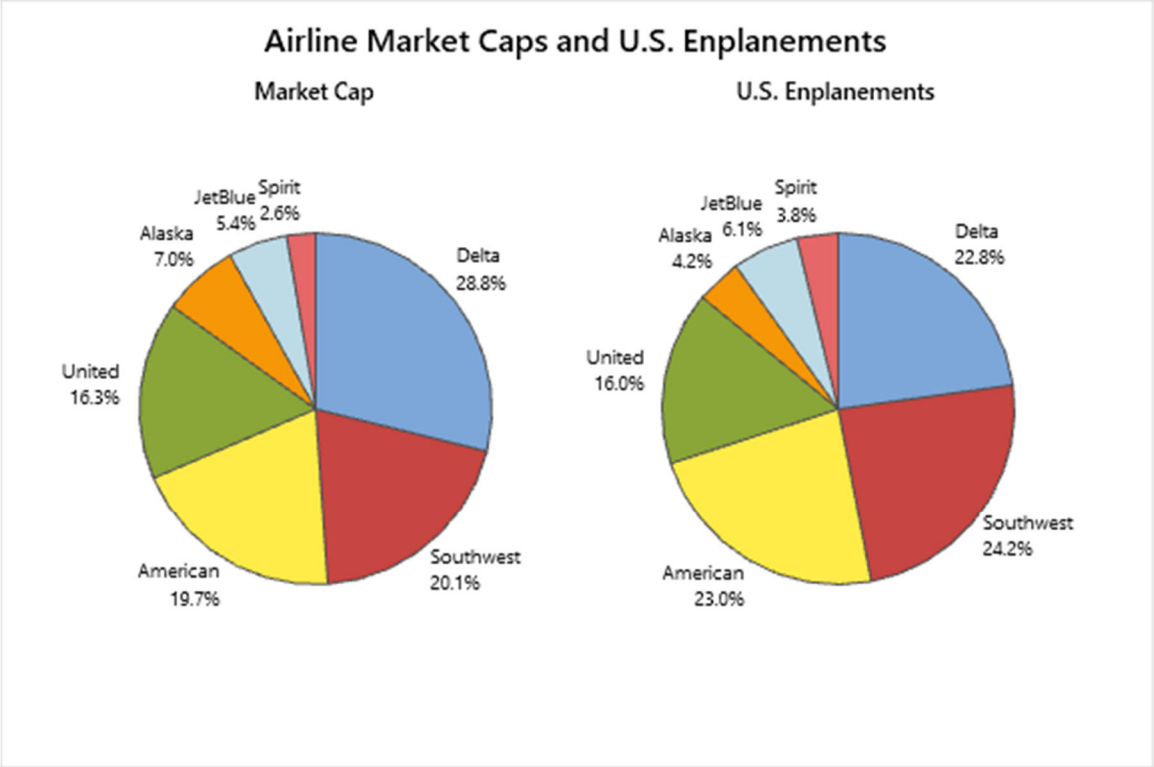


The pie charts show that almost half of all the hospitals in this database are “Other Nonprofit” (47%). Next are investor owned for-profit hospitals at 24% followed by nonfederal government hospitals (state, county, city, etc.) at 17%,. Note that there were 46 federal government hospitals (2%). About 90% of these federal government hospitals were veterans’ hospitals (this information is not contained in the database but is known by the author) and there were a couple of hospitals that were dedicated to native Americans.

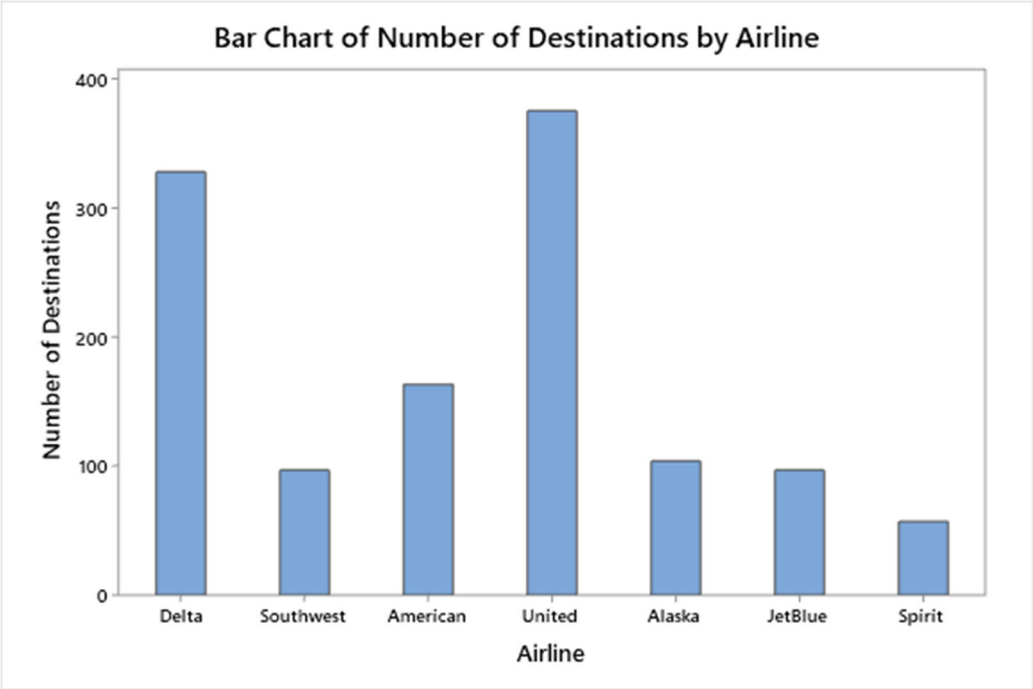
Chapter 2
Southwest Airlines

1. Shown below are a sampling of visuals that could be generated to present the data on Market Cap, U.S. Enplanements, and Number of Destinations for Seven U.S. airlines in Minitab, Tableau, or Excel. The student is encouraged to explore their own visualization options such as these and then write a brief report comparing the airlines on each of the three variables and overall.

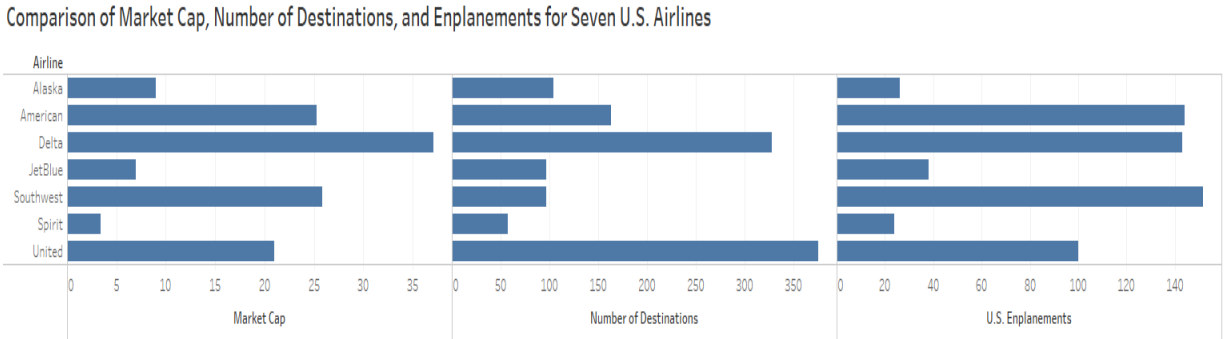
   Shown below is a Minitab graph with pie charts on two of the three variables. It is also possible to display all three variables in one graphic so as to compare the airlines on all three.



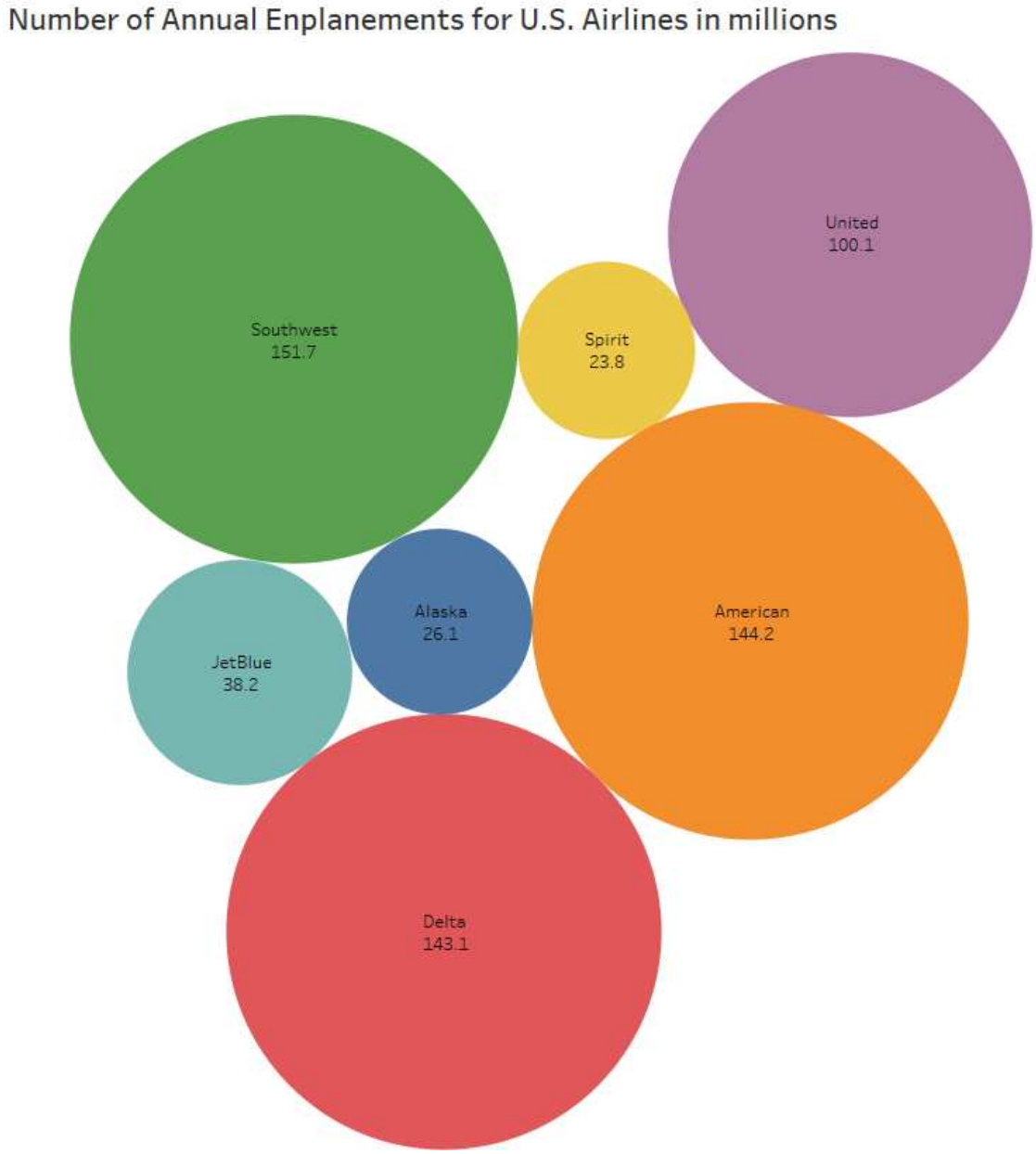
Shown below is a bar chart (in Minitab) of the number of U.S. destinations of the top seven U.S. airlines.



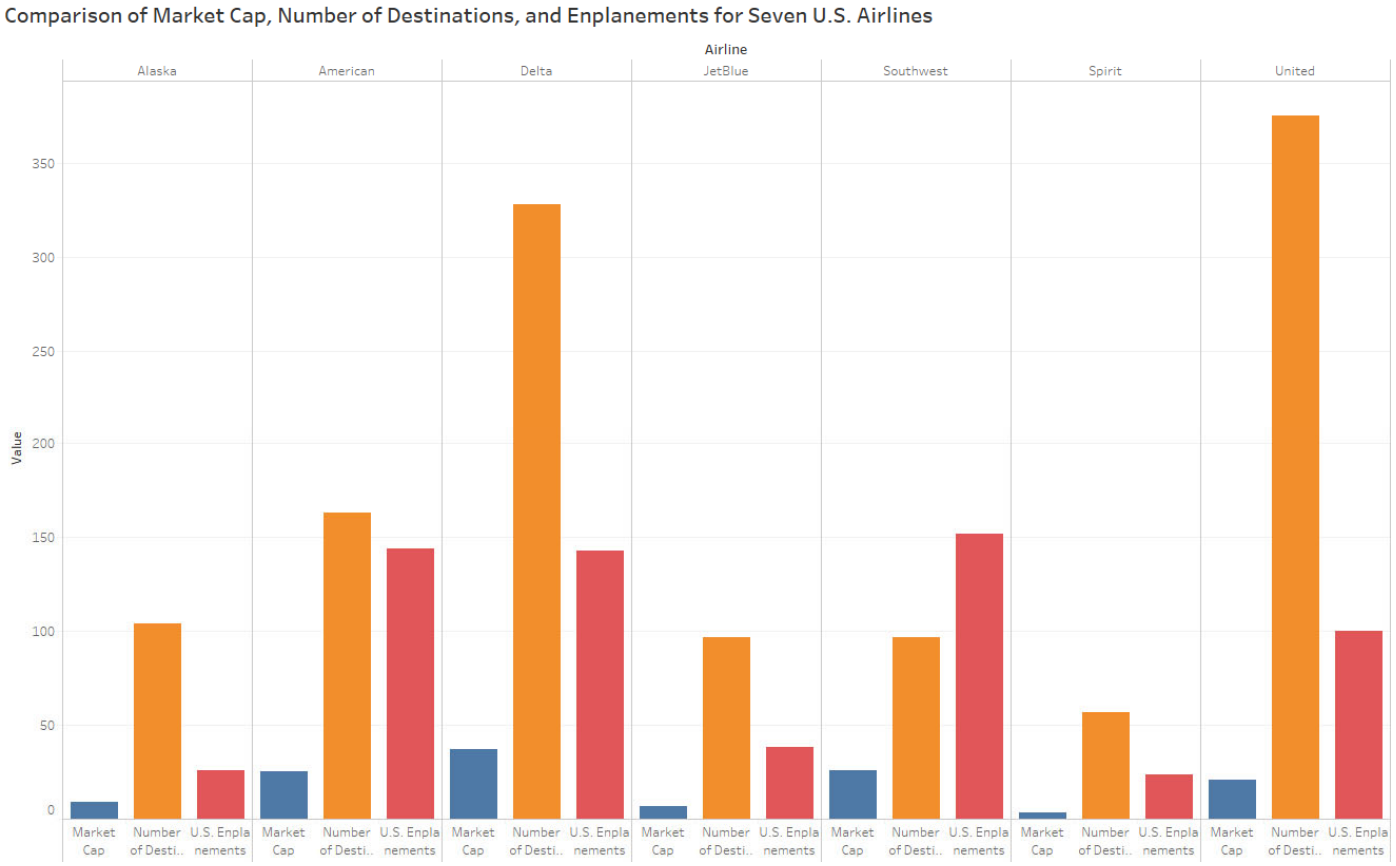
Below is a Tableau produced horizontal bar chart depicting all seven airlines on all three variables. Students can be asked to explain why this depiction might be better than say three separate bar charts like the one displayed above.



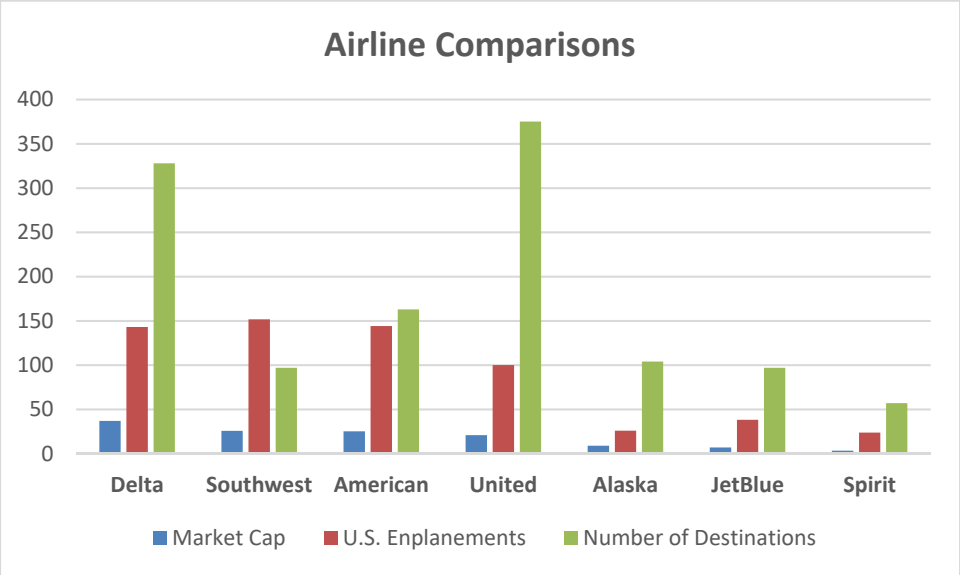
Shown next is a Tableau bubble chart of the number of annual U.S. enplanement for the seven airlines. Have the student share what might be an advantage of a bubble chart. Is it more esthetically pleasing? What are the advantages/disadvantages of such a bubble chart?



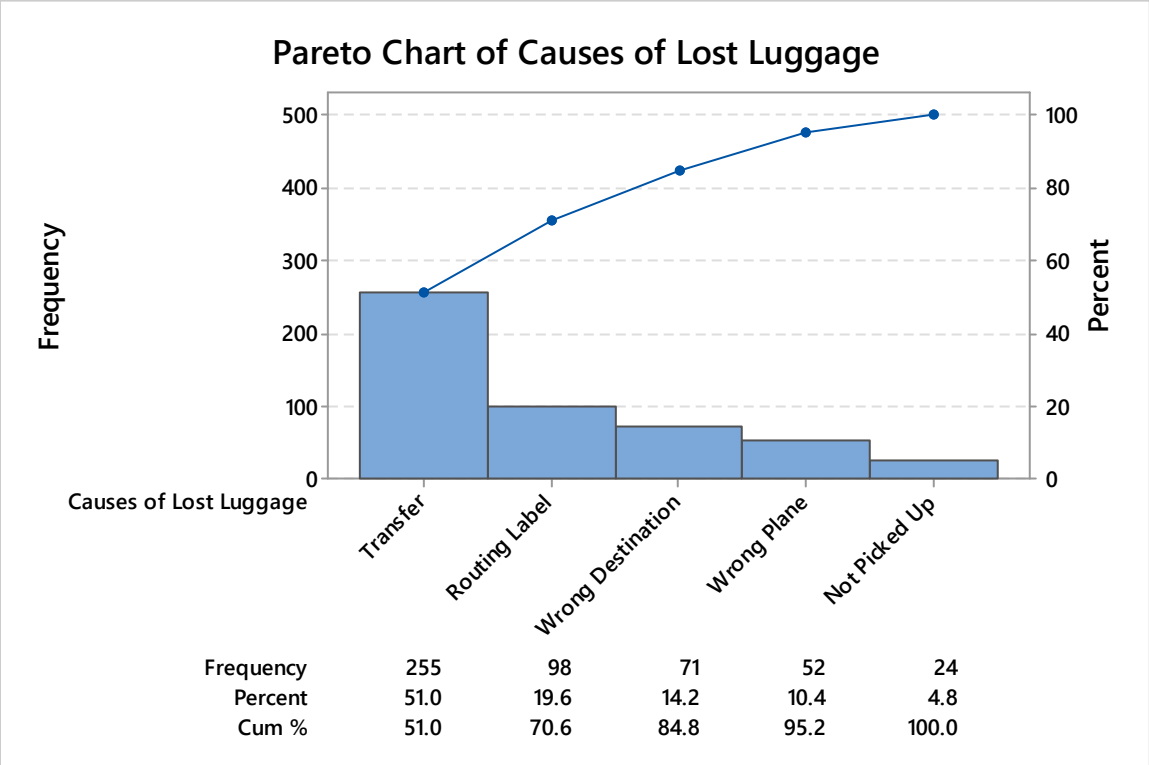
Below is a Tableau-produced vertical bar chart showing all seven airlines on all three variables. Needless-to-say, the graph is a bit "cramped" and difficult to see. However, if shown on a larger format or in a landscape orientation, it might be an effective way to display all the data in one place.



Shown next is a very clean and clear Excel bar graph displaying all three variables on all of the airlines. This graph was created when the author clicked and dragged over the data in Excel and let Excel choose a recommended chart.

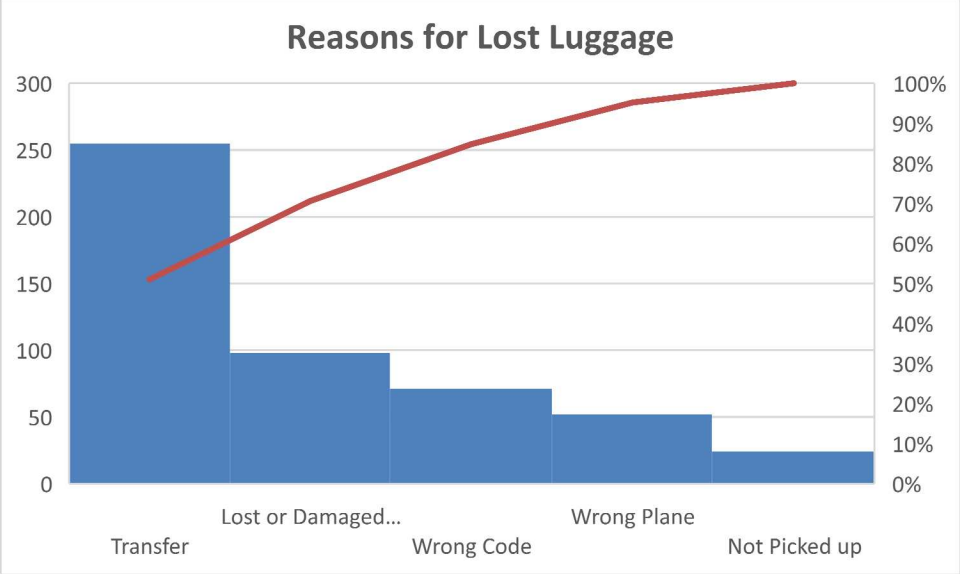


2. The Pareto Chart is a graphical technique for displaying problem causes. Question 2 of this case presents various causes for the problem of lost airline passenger luggage. Shown here is a Minitab-produced Pareto Chart for the frequency of various causes associated with lost luggage.

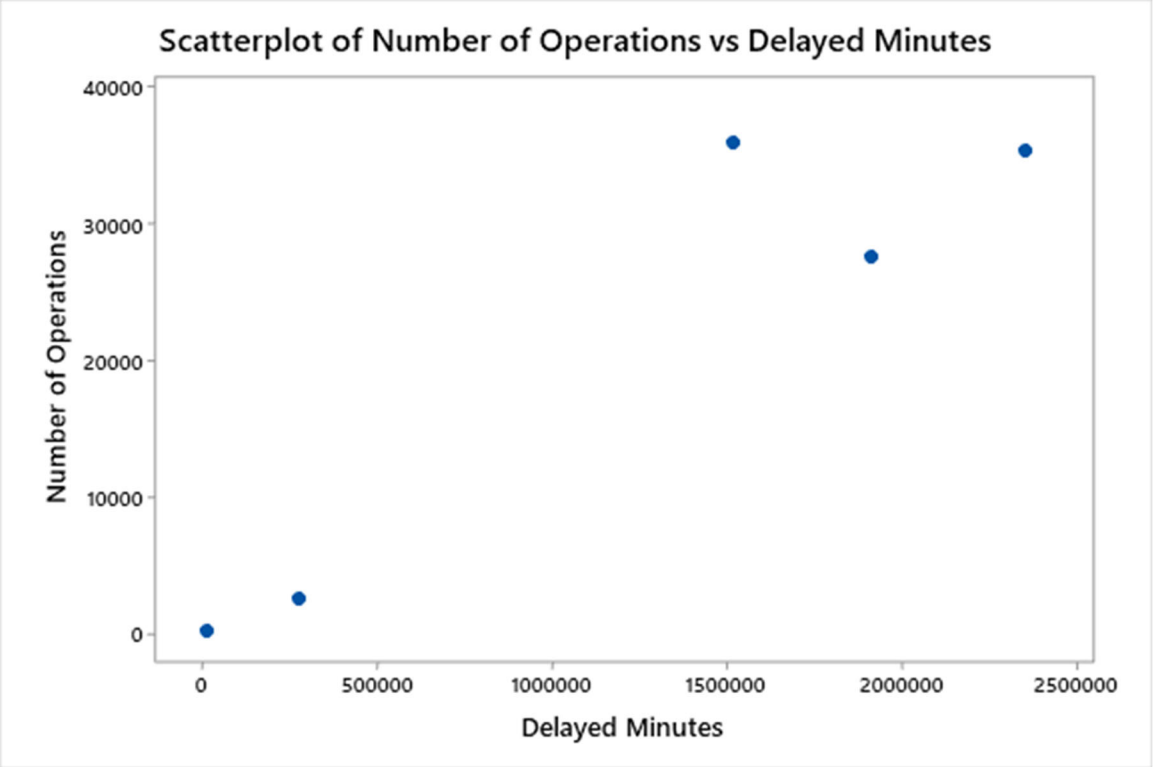


The vertical bars of a Pareto Chart display the most common types of causes ranked in order of occurrence from left to right. This Pareto Chart shows that the number one cause of lost luggage is transfer from one plane to another (51% of lost luggage) followed by a damaged or lost routing label (an addition 20%). If an airline is undertaking an effort to significantly reduce the numbers of lost luggage, then they should start by tackling the "transfer" issue. In the quality improvement circles, this would be referred to as starting with the "low lying fruit". That is, if the transfer problem could be completely solved, an airline could reduce the numbers of lost luggage by 50%.

Shown next is an Excel-produced Pareto Chart that is similar to the Minitab version.

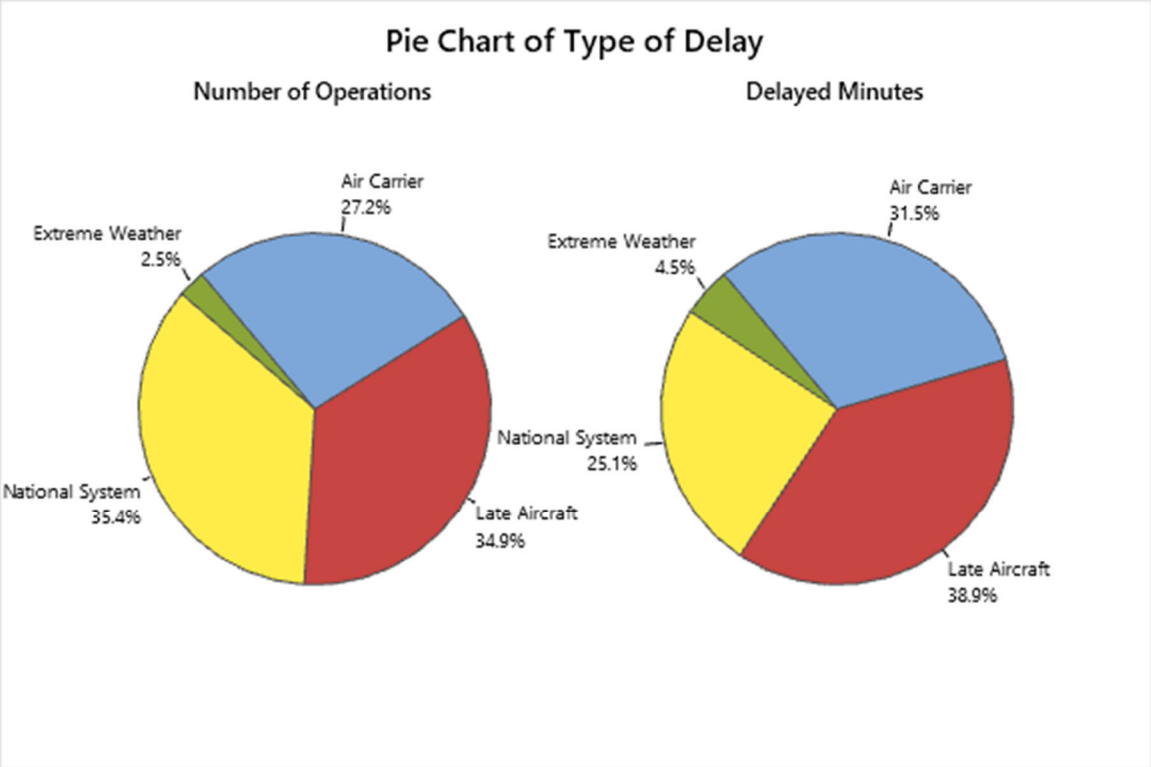


3. This question asks the student to analyze, through visualization, the Bureau of Transportation data on flight delays by all U.S. carriers in 2018. Note that there are only two distinct variables in these data, Number of Operations and Delayed Minutes. The other two variables, % of Total Operations and % of Total Delayed Minutes, yield essentially the same relative data as Number of Operations and Delayed Minutes. In constructing a scatter plot, the student must select two variables and eliminate the row of Total Operations from the data. The resulting scatter plot in Minitab is:

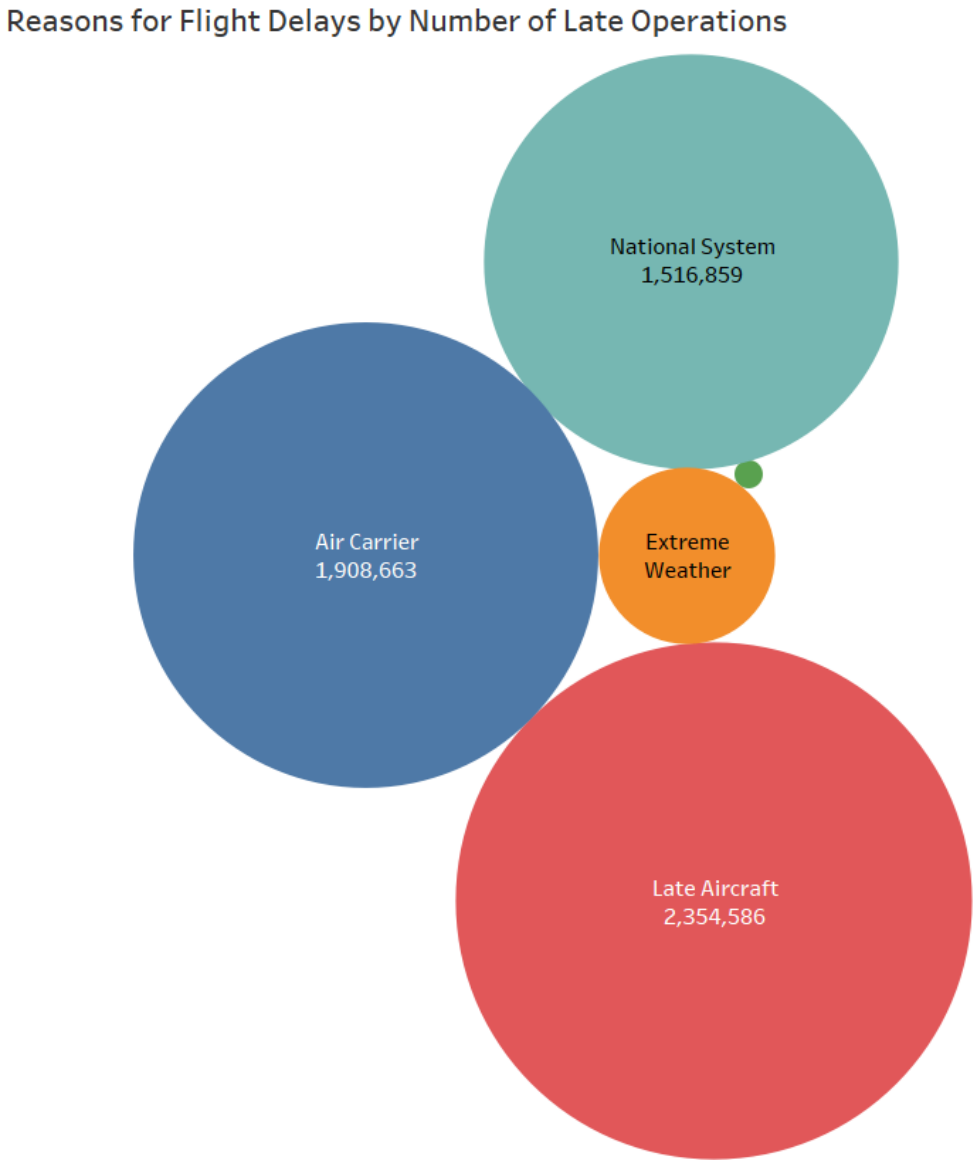


Note that there is a high correlation between Number of Operations (with delays) and Delayed Minutes. The exception seems to be the National Aviation System delay which has the most Operations (with delays) but only the third most total Delayed Minutes. It would be interesting to have students explain why they think this is.

Shown next is a Minitab-produced graph that displays pie charts for both Number of Operations and Delayed Minutes. Because Security delay was so small, it was omitted from the pie charts. Side by side pie charts allow the business analyst to see that there are some relative differences (slice sizes) between Number of Late Operations and Delayed Minutes particularly for national system delays and late aircraft delays. For example, late aircraft counts for a smaller percent of delayed operations than it does for delayed minutes. Thus, while late aircraft may be a smaller number than say national system delays, it results in a greater percentage of delayed minutes.



One could also use a bubble chart like the one presented below to display such delay information.



Shown next is an Excel-produced horizontal bar chart for minutes of flights delays by reasons for flights delays. It is easy to see that aircraft arriving late is the greatest cause.