Chapter 2
Charts and Graphs

LEARNING OBJECTIVES

The overall objective of Chapter 2 is for you to master several techniques for summarizing and depicting 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.

CHAPTER TEACHING STRATEGY

Chapter 1 brought to the attention of students the wide variety and amount of data available in the world of business. In chapter 2, we confront the problem of trying to begin to summarize and present the data in a meaningful manner. 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?

The second part of 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. 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

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
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	

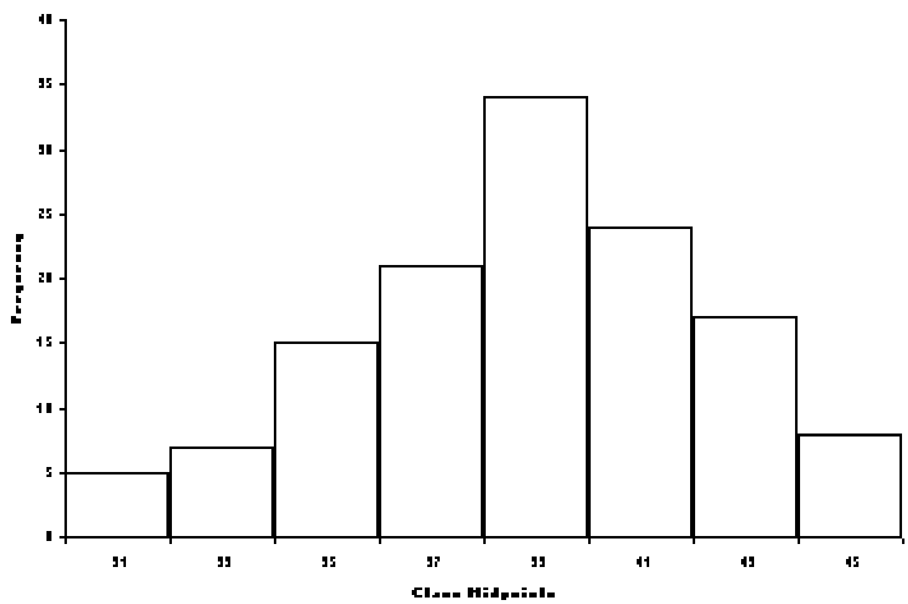
The relative frequency tells us that it is most probable that a customer is in the 15 - 20 category (.2674). Over two thirds (.6744) of the customers are between 10 and 25 years of age.

2.4

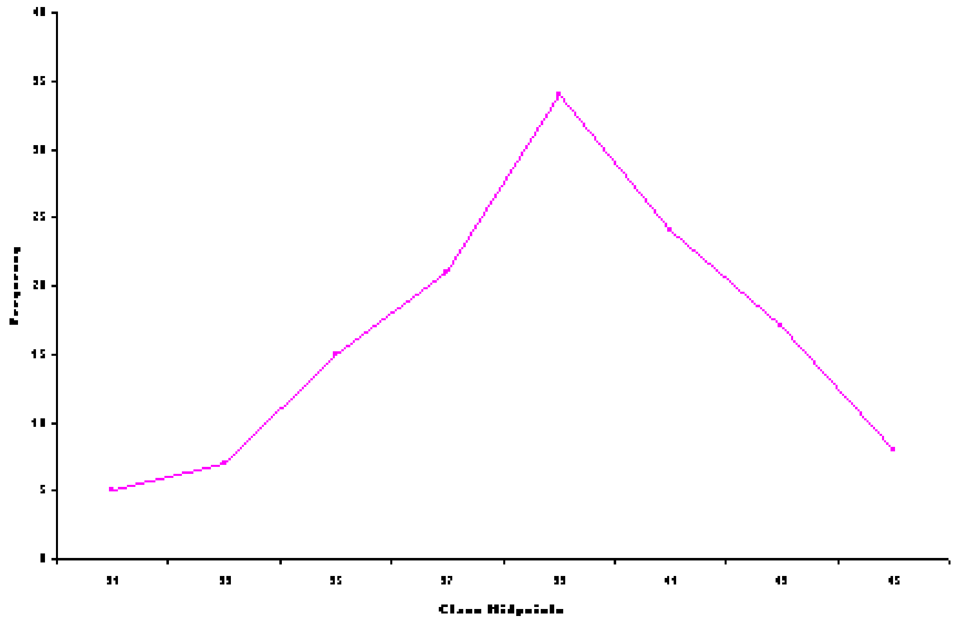
Class Interval	Frequency	Class Midpoint	Relative Frequency	Cumulative Frequency
0-2	218	1	.436	218
2-4	207	3	.414	425
4-6	56	5	.112	481
6-8	11	7	.022	492
8-10	8	9	.016	500
TOTAL	500		1.000	

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:

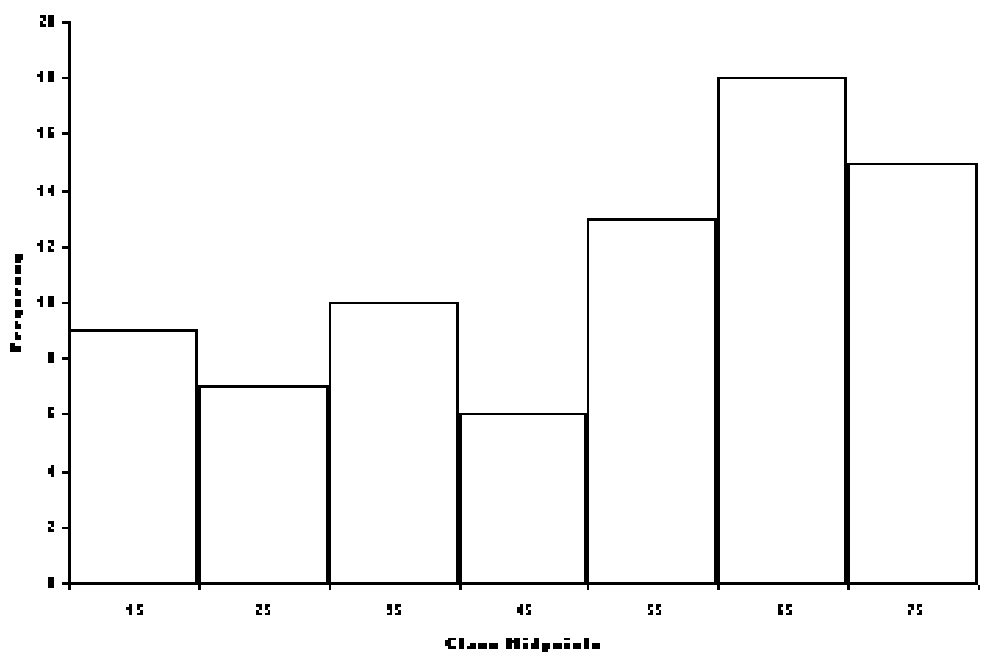


Frequency Polygon:

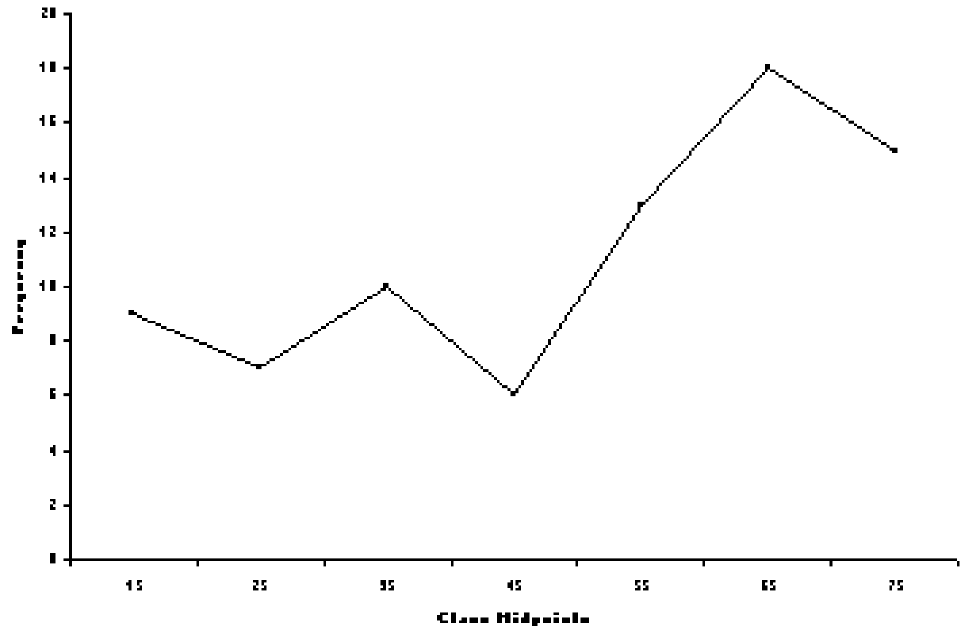


Comment: The assembly times "pile up" near the middle of the graphs indicating that many of the assembly times are between 36 and 42 minutes.

2.7 Histogram:

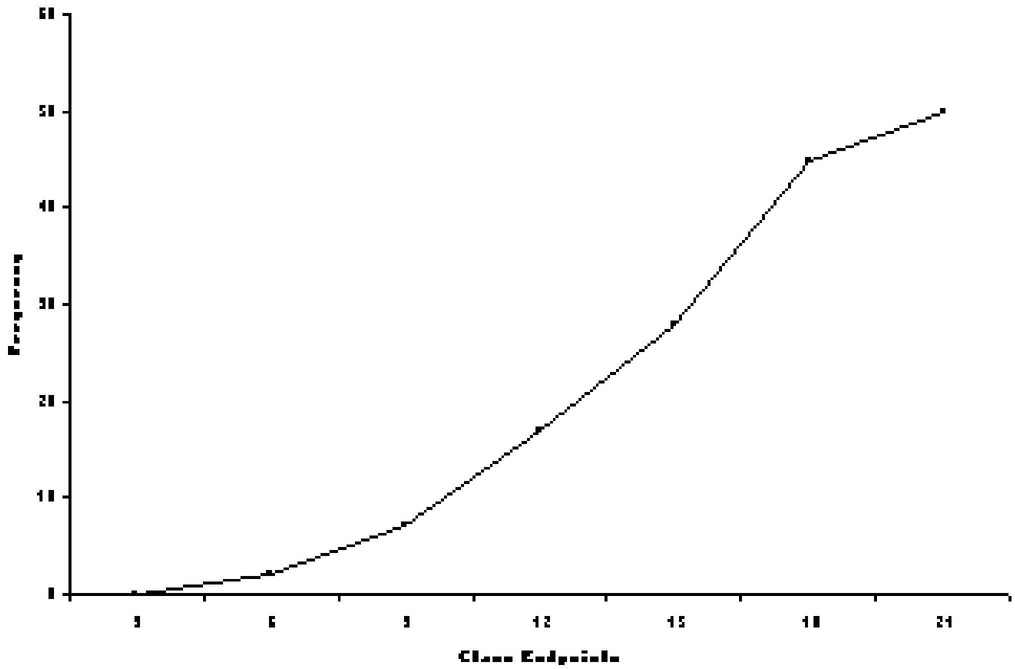


Frequency Polygon:



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.

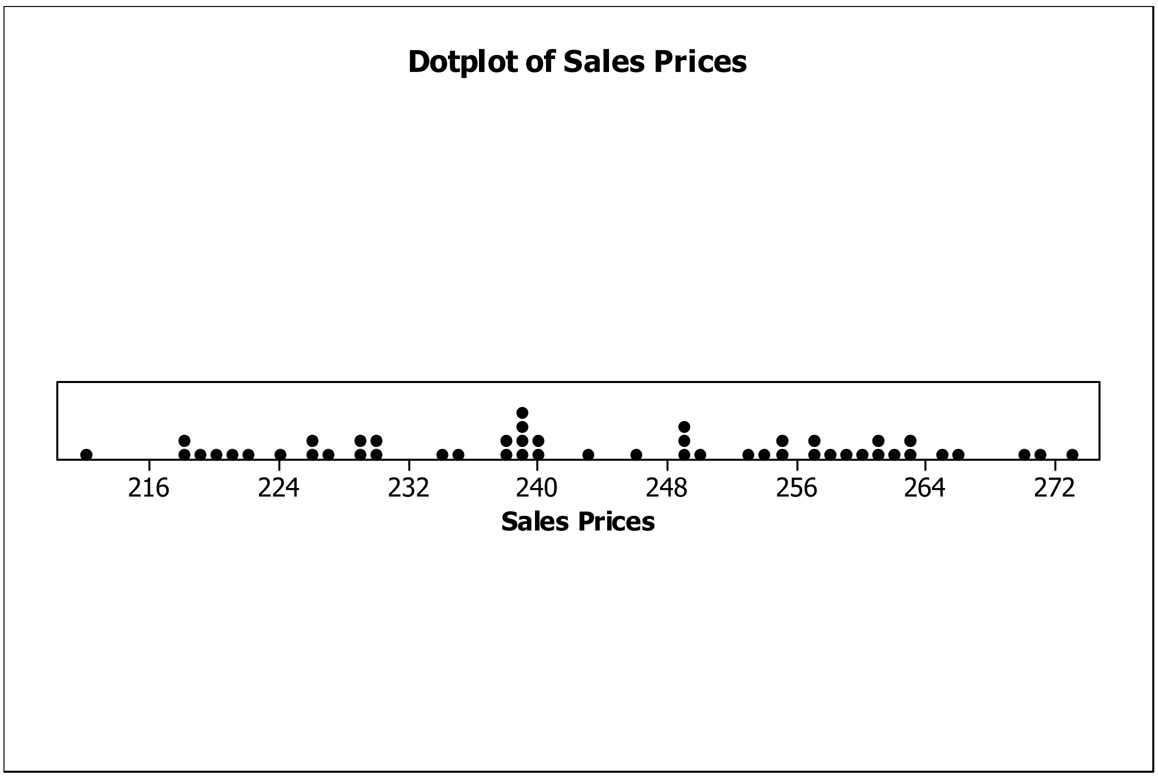
2.8 Ogive:



2.9

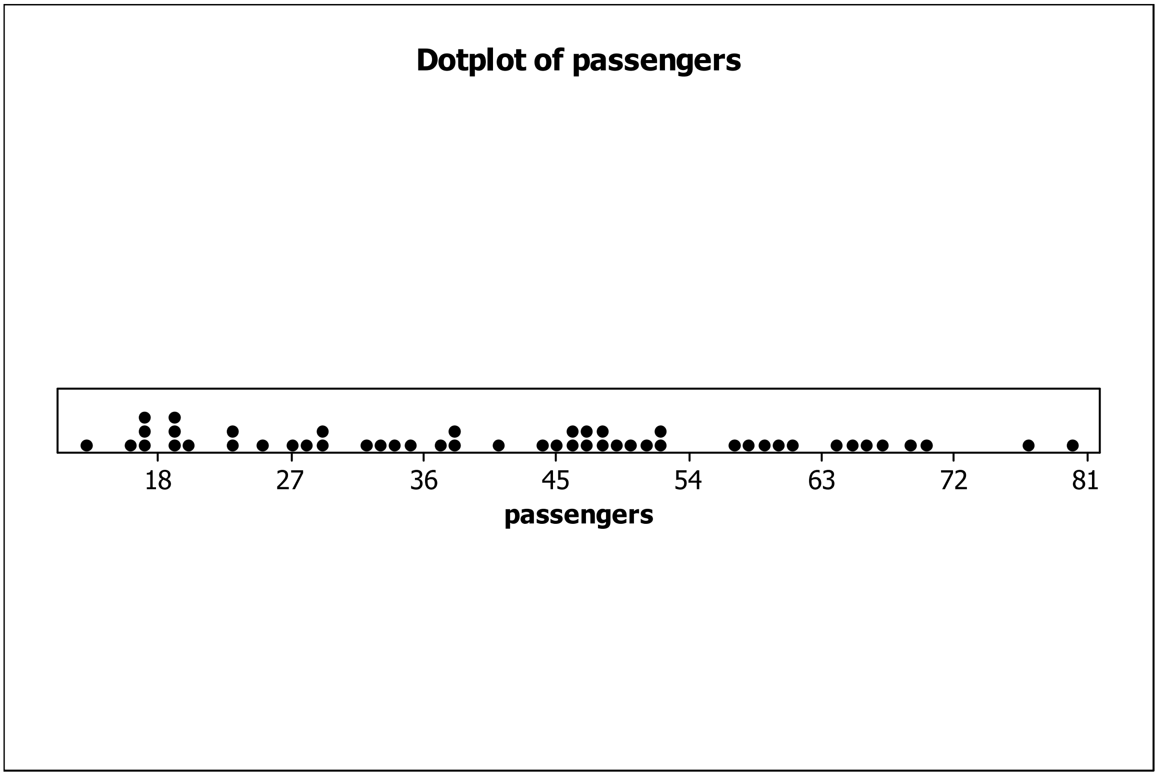
STEM	LEAF
21	2 8 8 9
22	0 1 2 4 6 6 7 9 9
23	0 0 4 5 8 8 9 9 9 9
24	0 0 3 6 9 9 9
25	0 3 4 5 5 7 7 8 9
26	0 1 1 2 3 3 5 6
27	0 1 3

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 is Atlanta's Hartsfield-Jackson International Airport which has over 95 million passengers. The histogram shows that there are only two airports with between 65 and 75 million passengers and one airport with between 55 and 65 million passengers. Three airports have between 45 and 55 million passengers. Sixteen of the top 30 airports have between 35 and 75 million passengers.

2.12 The dotplot shows that all but two states have less than 100,000 farms. There is only one state with more than 110,000 farms and that state has around 250,000 farms. Most of the states have between a few hundred and about 90,000 farms. There are five states that appear to have around 10,000 farms (modal number). Four states have about 5,000 farms, four states have about 30,000 farms, four states have about 50,000 farms, and four states have about 80,000 farms. The actual mean for the data (including the two states with the very large number of farms) is 44,060 (not shown in graph) and the actual median is 37,500. The difference between the mean and the median indicate a skewness toward the right which the dotplot shows.

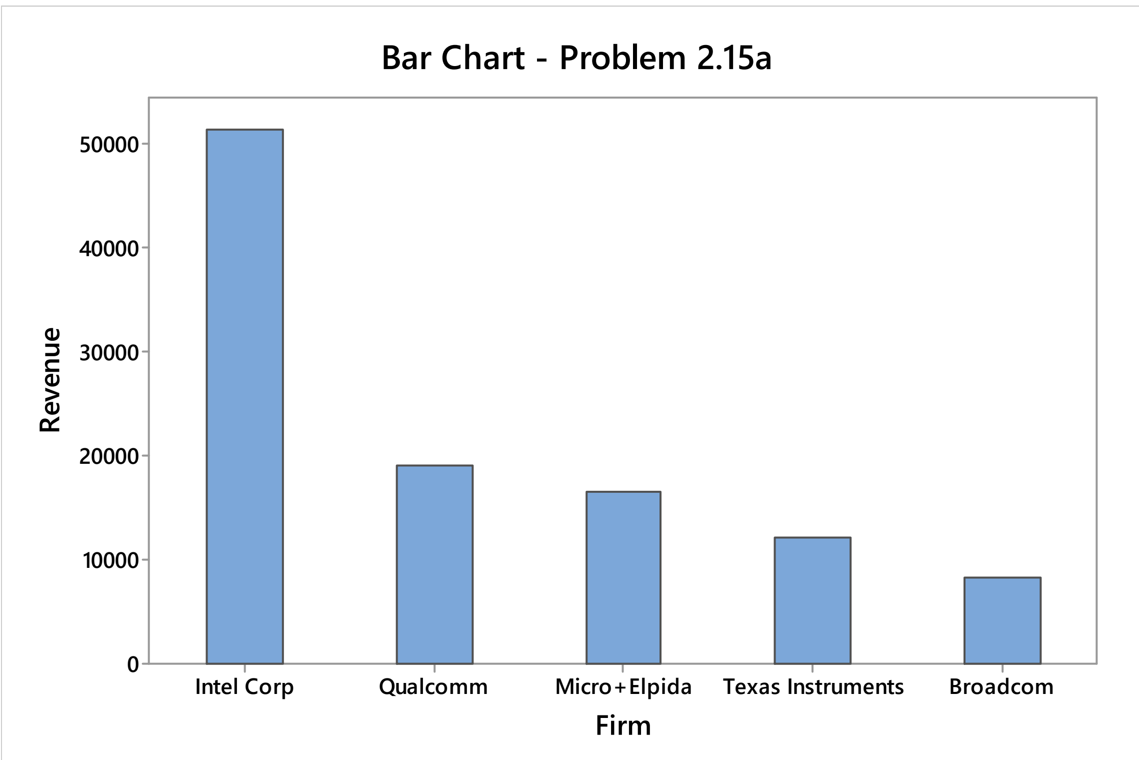
2.13 From the stem and leaf display, the original raw data can be obtained. For example, the fewest number of cars washed on any given day are 25, 29, 29, 33, etc. The most cars washed on any given day are 141, 144, 145, and 147. The modal stems are 3, 4, and 10 in which there are 6 days with each of 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 ogive tells us several things. Out of 200 pots, 50 of them contained only 10 King crabs. From the ogive, it is possible to see that nearly 100 (or about 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 (less than 10%) have as many as 100 crabs.

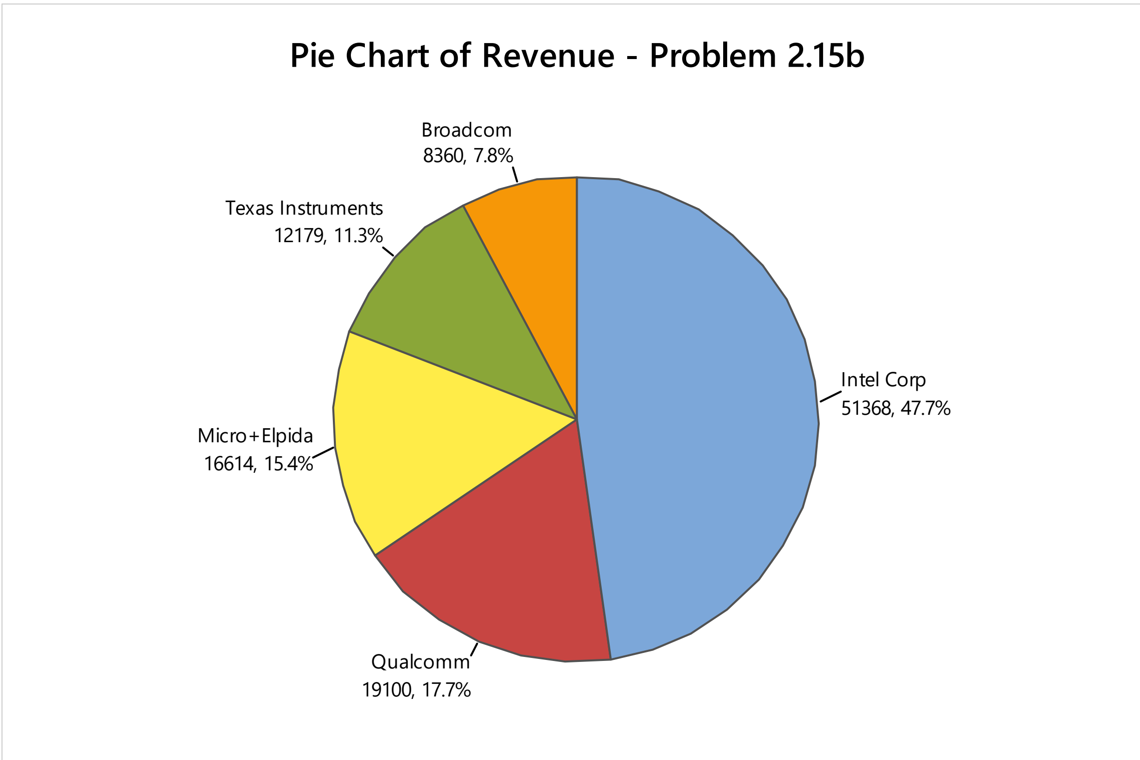
2.15

Firm	Proportion	Degrees
Intel Corp.	.477	171.7
Qualcomm	.178	64.1
Micro + Elpida	.154	55.4
Texas Instruments	.113	40.7
Broadcom	.078	28.1
TOTAL	1.000	360.0

a.) Bar Graph:



b.) Pie Chart:

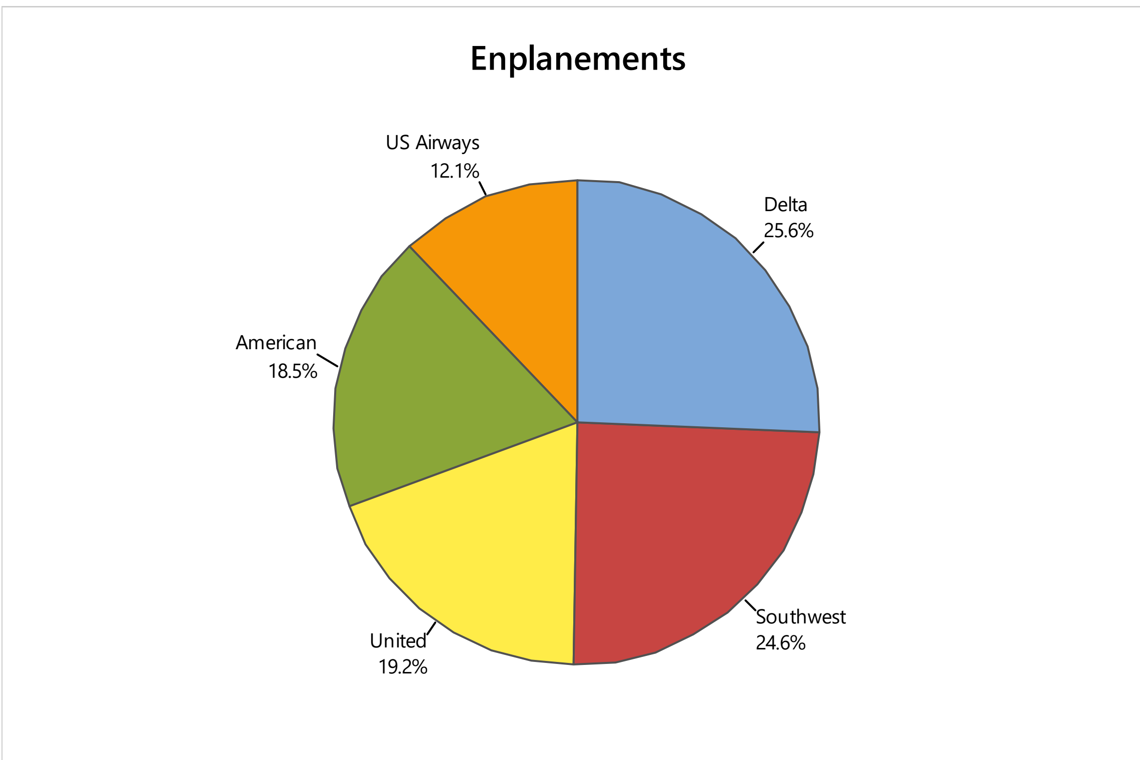


c.) While pie charts are sometimes interesting and familiar to observe, in this problem it is virtually impossible from the pie chart to determine the relative difference between Micro + Elpida and Qualcom without the data labels. From the bar chart, however, it is slightly easier to judge the difference between Micro + Elpida and Qualcomm.

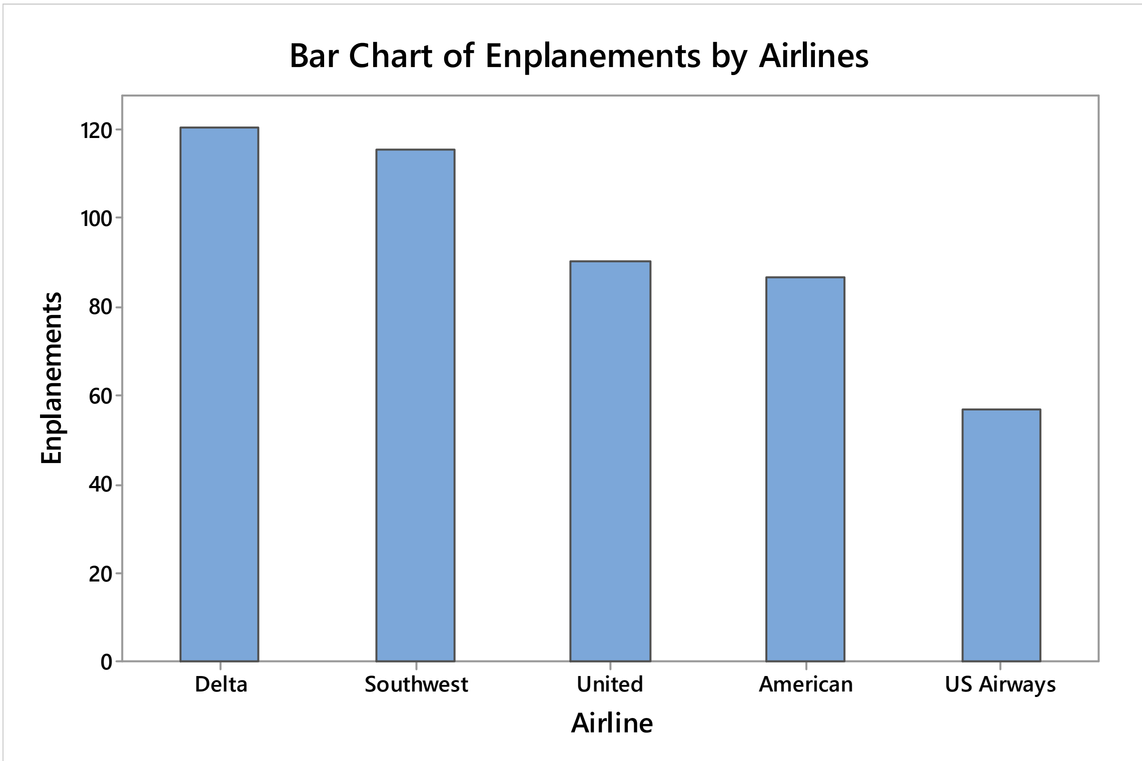
2. 16

Company	Proportion	Degrees
Delta	.256	92.2
Southwest Airlines	.246	88.6
United Airlines	.192	69.1
American Airlines	.185	66.6
US Airways	.121	43.6
TOTAL	1.000	360.1

Pie Chart:



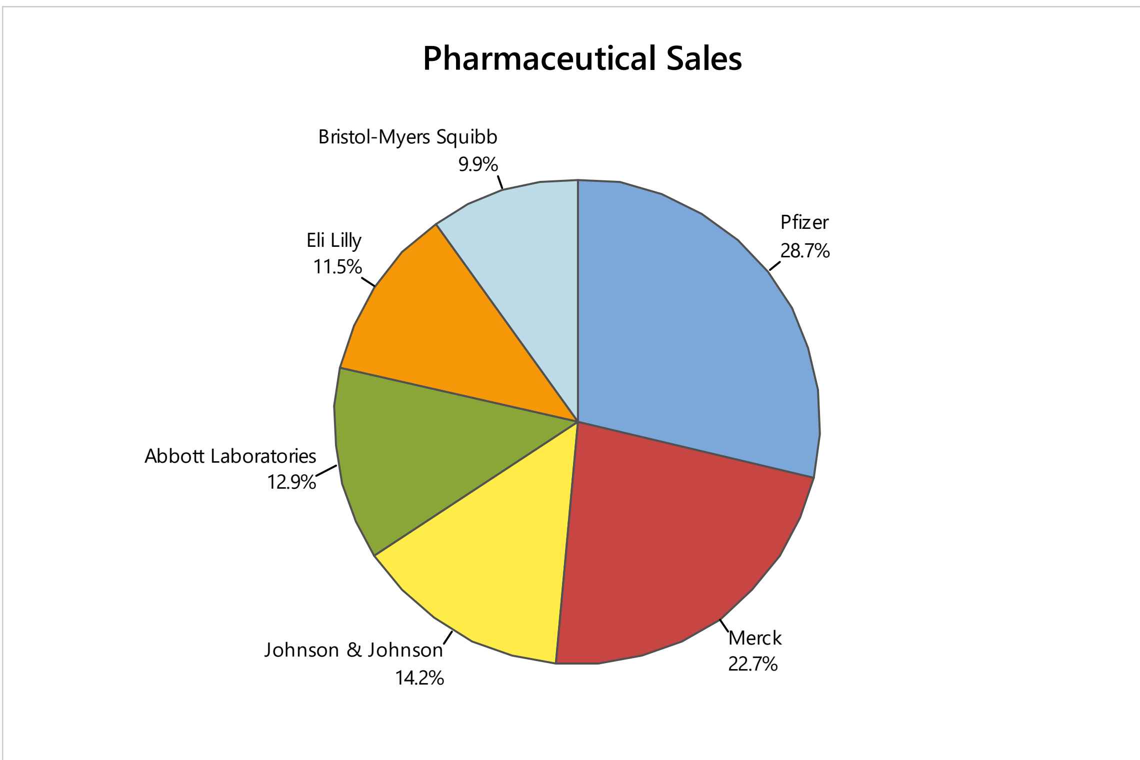
Bar Graph:


Bar Chart of Enplanements by Airlines
Enplanements
120
100
80
60
40
20
0
Delta
Southwest
United
American
US Airways
Airline

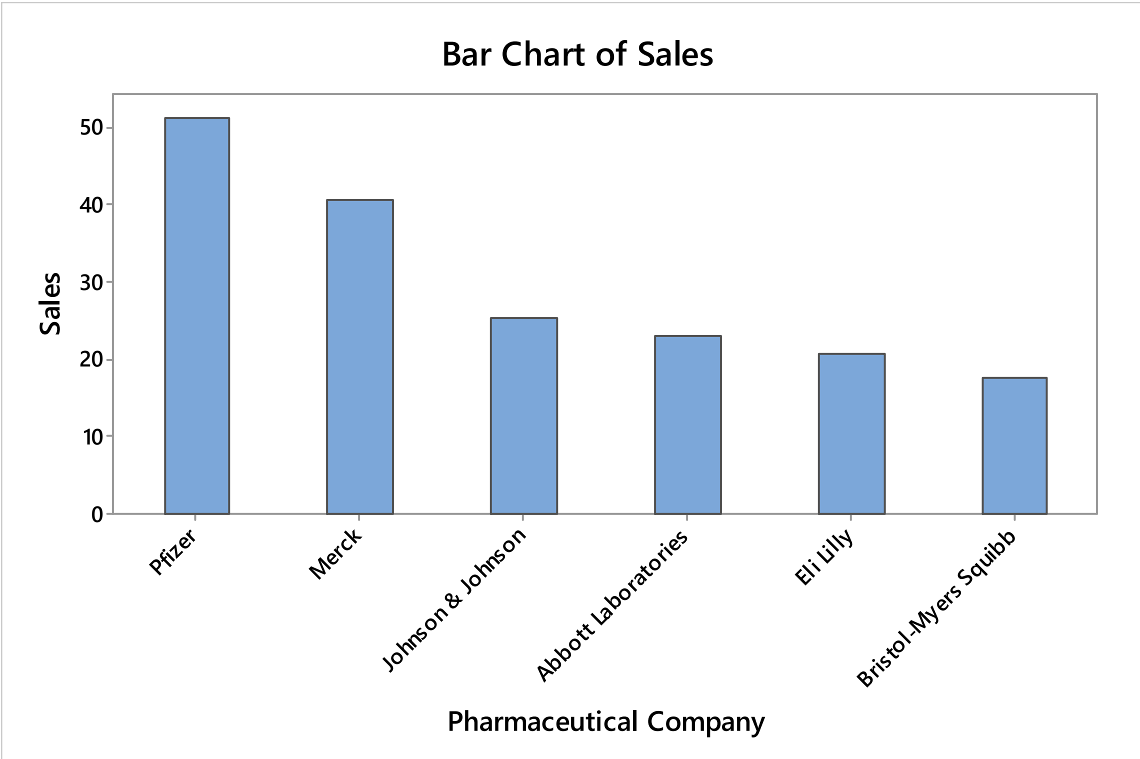
2.17

Brand	Proportion	Degrees
Pfizer	.287	103
Merck	.227	82
Johnson & Johnson	.142	51
Abbott Laboratories	.129	47
Eli Lilly	.115	41
Bristol-Myers Squibb	.099	36
TOTAL	0.999	360

Pie Chart:



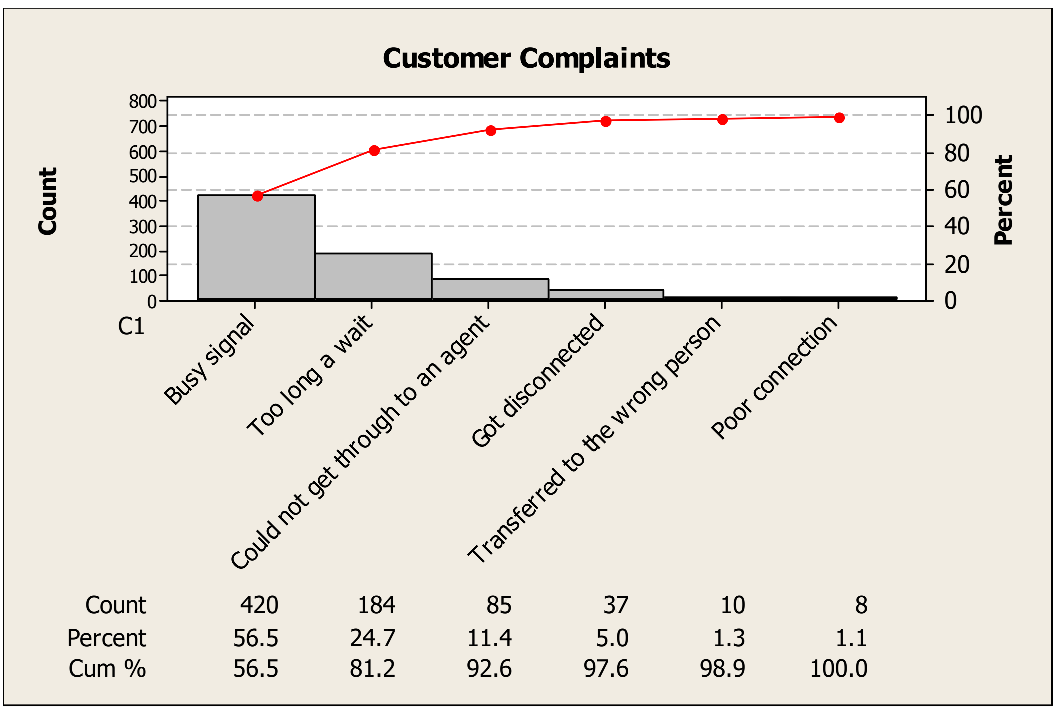
Bar Graph:


Bar Chart of Sales
Sales
50
40
30
20
10
0
Pfizer
Merck
Johnson & Johnson
Abbott Laboratories
Eli Lilly
Bristol-Myers Squibb
Pharmaceutical Company

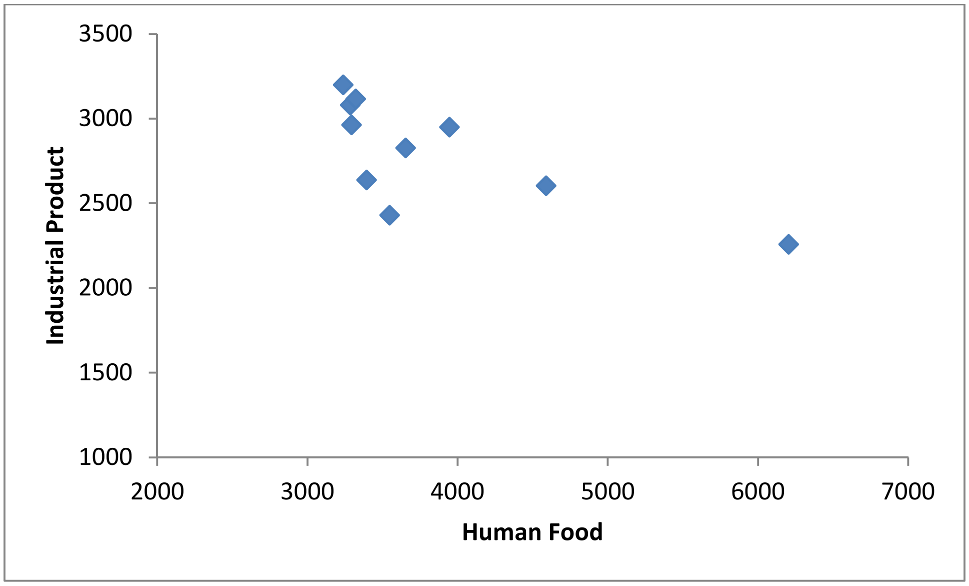
2.18 The bar chart shows that of all the currencies considered here, the Euro is strongest against the U.S. dollar (each Euro is worth $1.10). The India rupee is weakest against the U.S. dollar with a worth of only about two cents. The Canadian dollar is worth seventy-six cents compared to the U.S. dollar and the New Zealand dollar is worth about sixty-five U.S. cents. The Malaysia ringgit is worth twenty-five U.S. cents, the UAE dirham is worth about twenty-seven U.S. cents, the Chinese yuan is worth sixteen U.S. cents, and the Mexican peso is worth about six 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

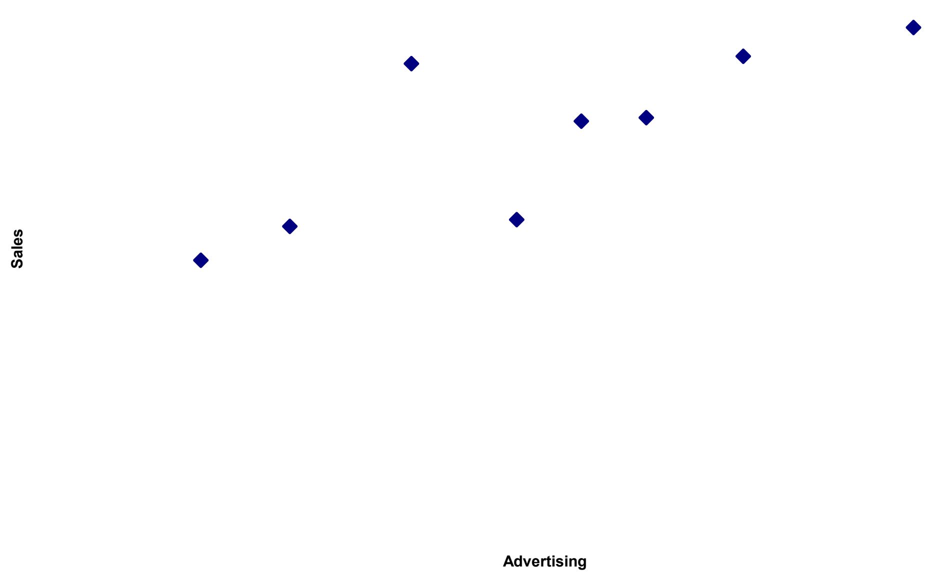


2.20



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

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

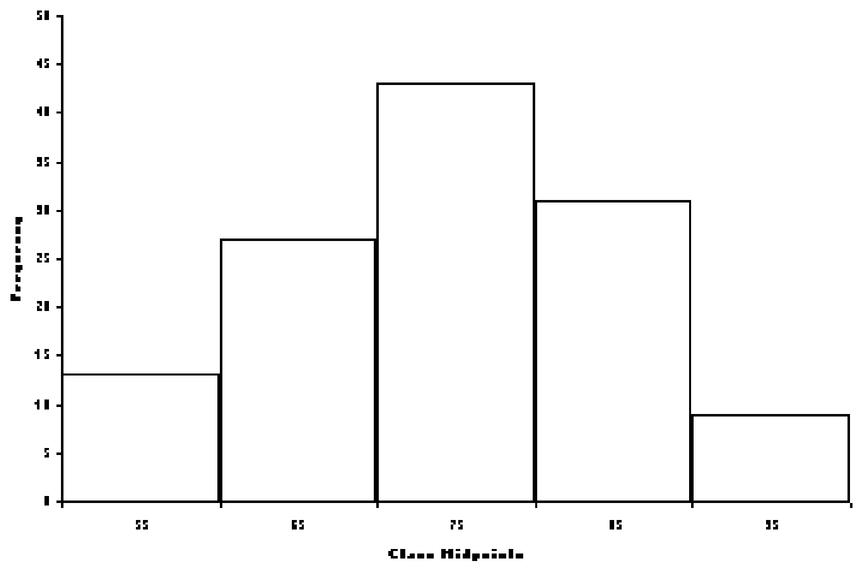
2.26

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

2.27

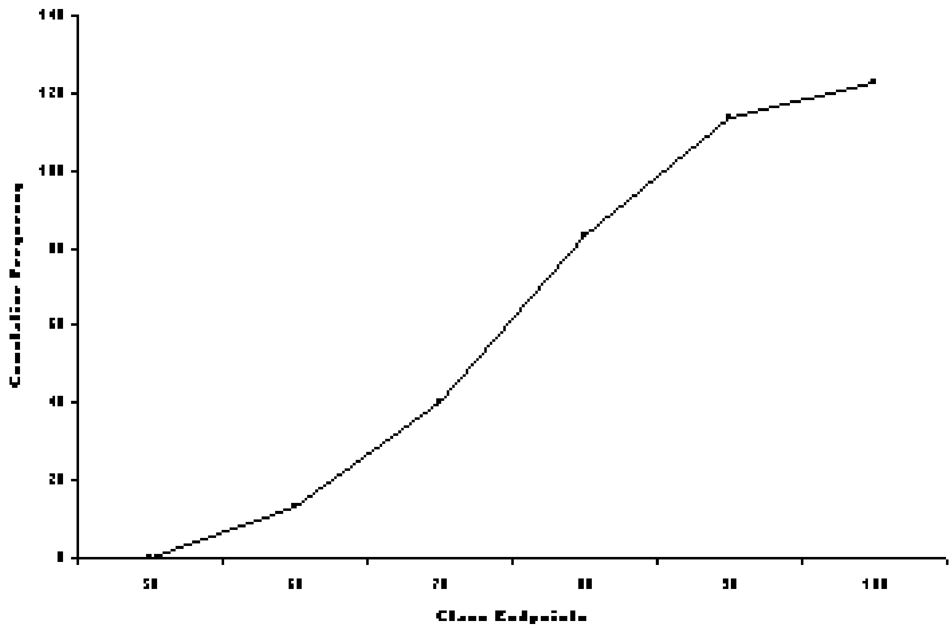
Class Interval	Frequencies
50 - under 60	13
60 - under 70	27
70 - under 80	43
80 - under 90	31
90 - under 100	9
TOTAL	123

Histogram:

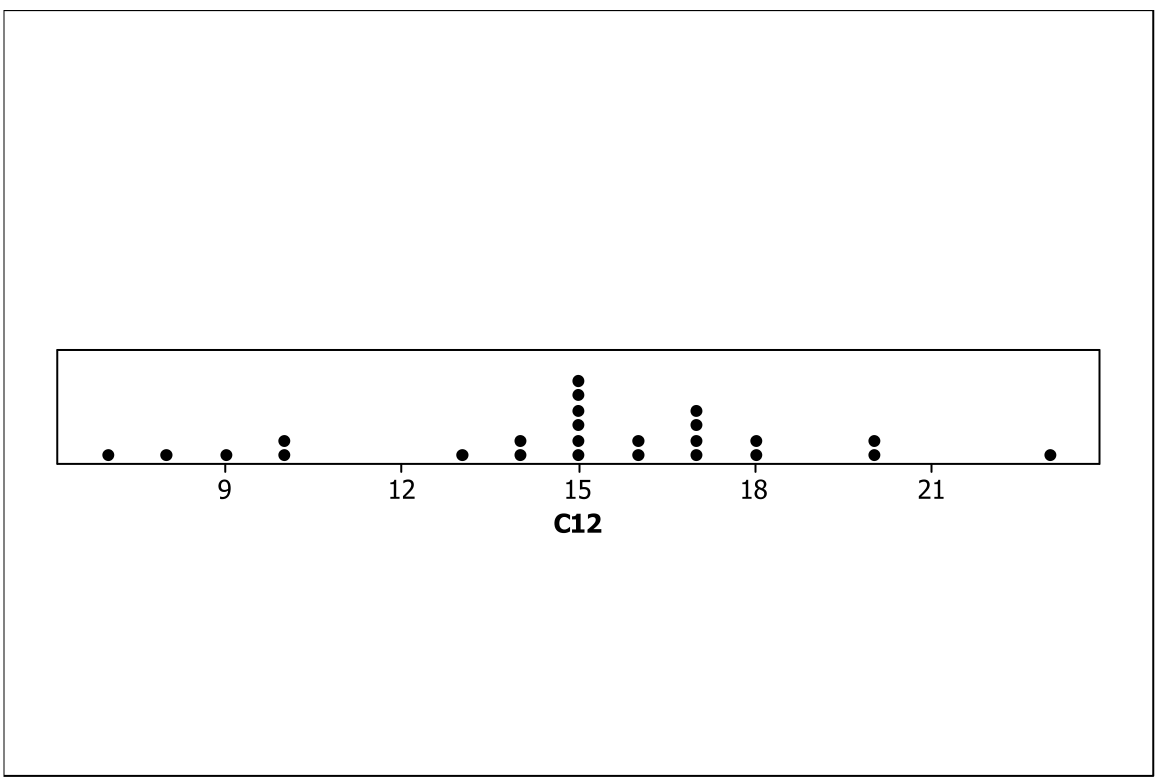


Frequency Polygon:

Ogive:



2.28 Dot Plot



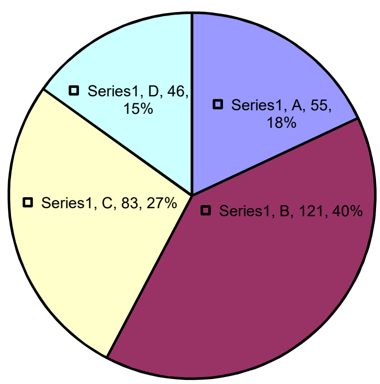
2.29

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

2.30

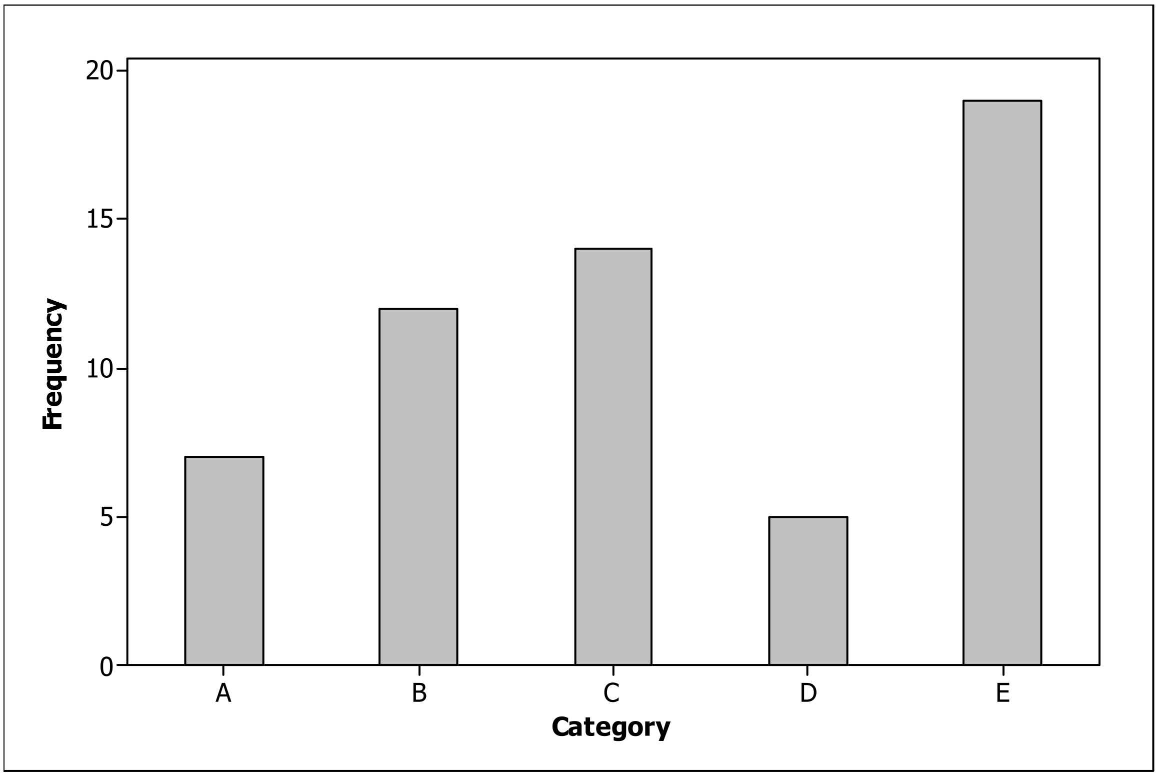
Label	Value	Proportion	Degrees
A	55	.180	65
B	121	.397	143
C	83	.272	98
D	46	.151	54
TOTAL	305	1.000	360

Pie Chart:



2.31 Bar Graph:

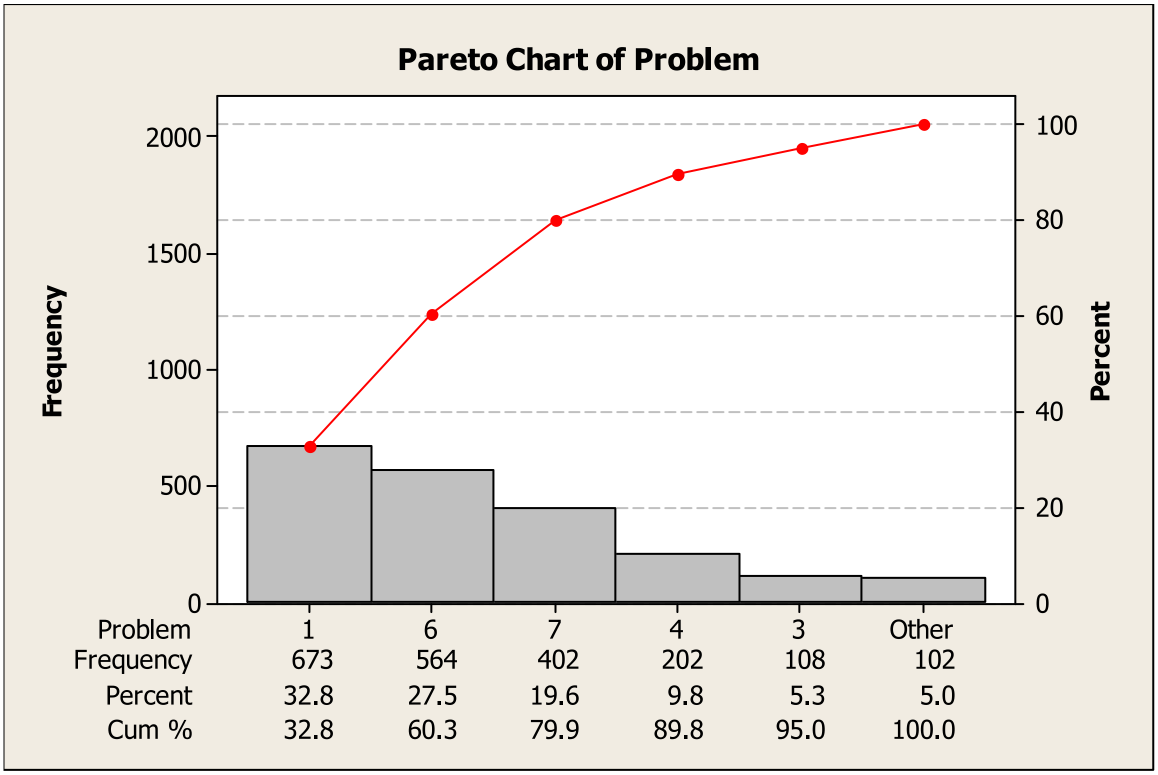
Category	Frequency
A	7
B	12
C	14
D	5
E	19



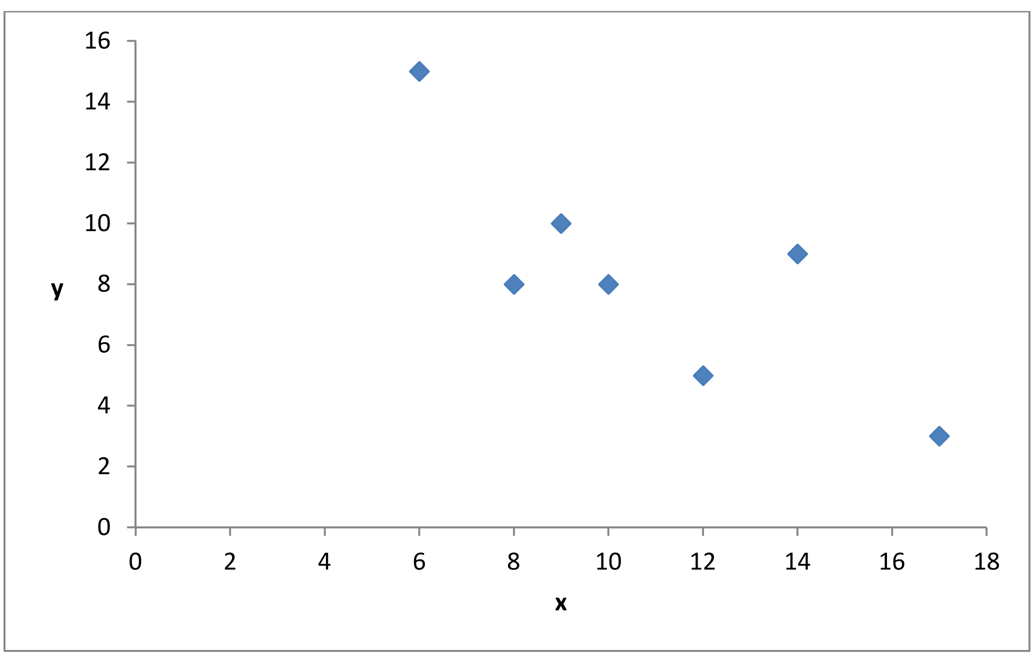
2.32

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:

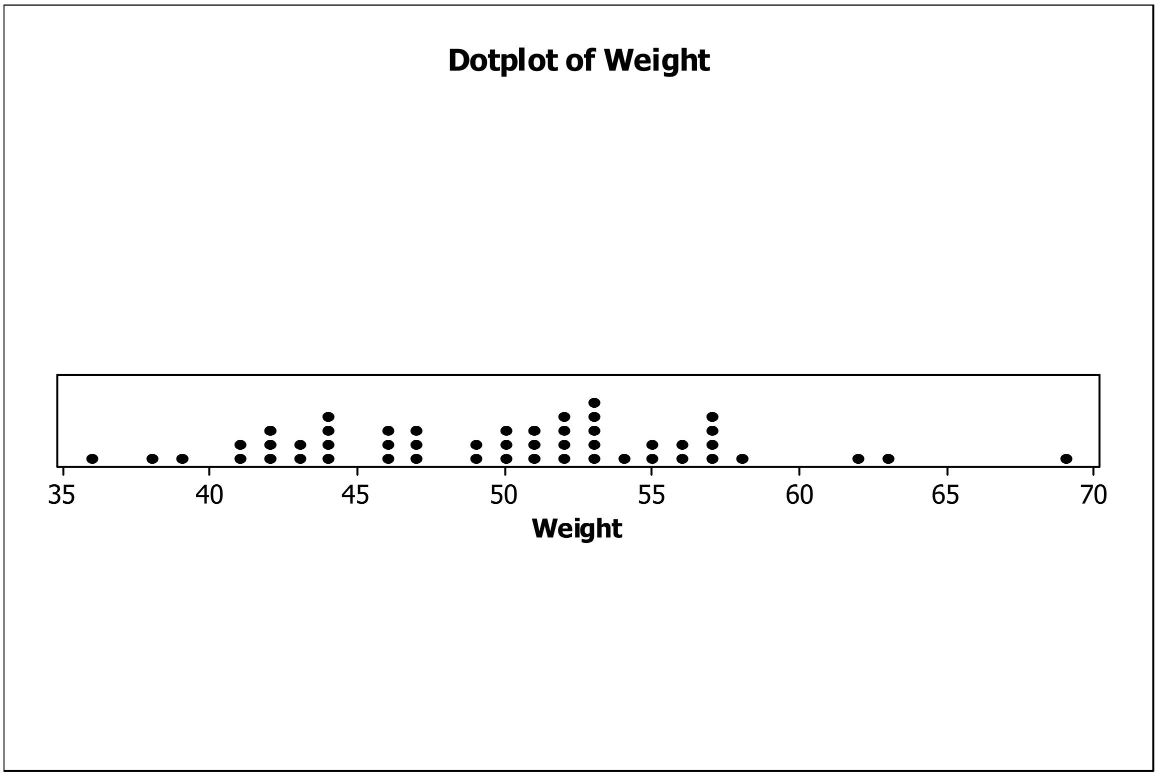


2.33 Scatter Plot



2.34 Whitcomb Company

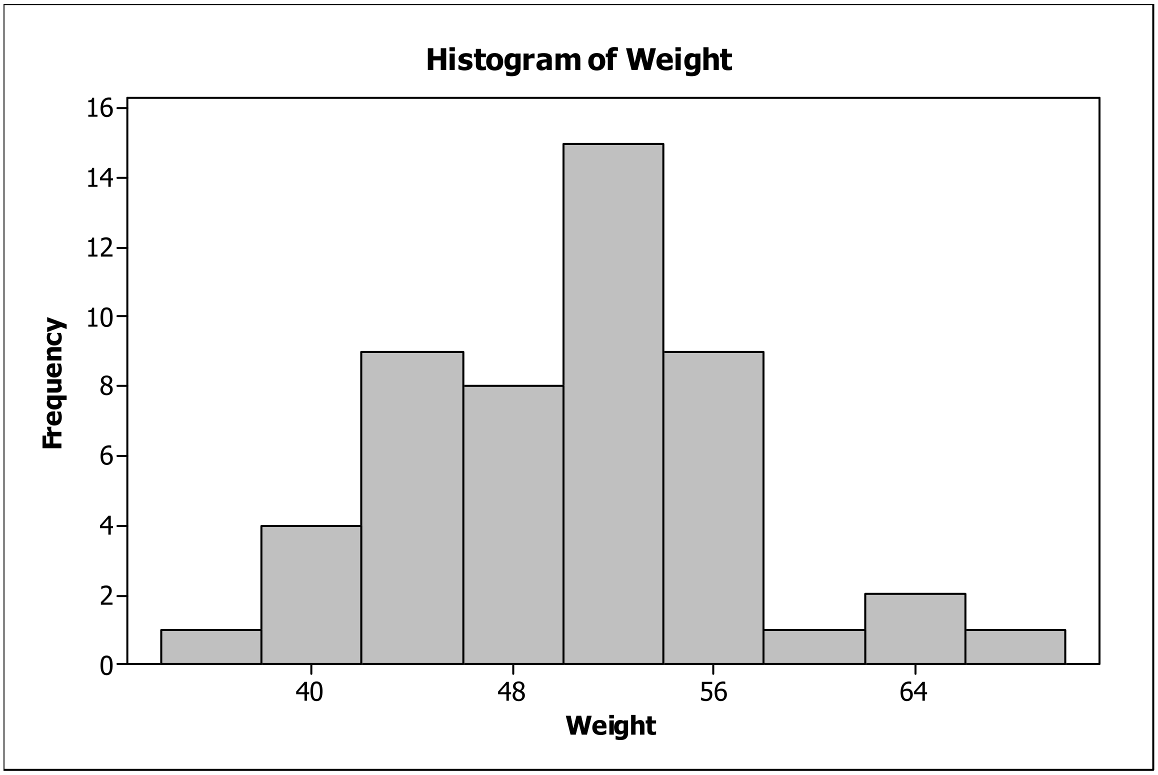
a.) Dot Plot. The metal ring is supposed to weigh around 50 ounces. The dot plot shows that most of the rings weigh between 41 and 58 ounces with the highest number being 53 (in Chapter 3 we will name this as the mode) and other weights piling up at 44, 52, and 57 ounces. There are some extreme values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

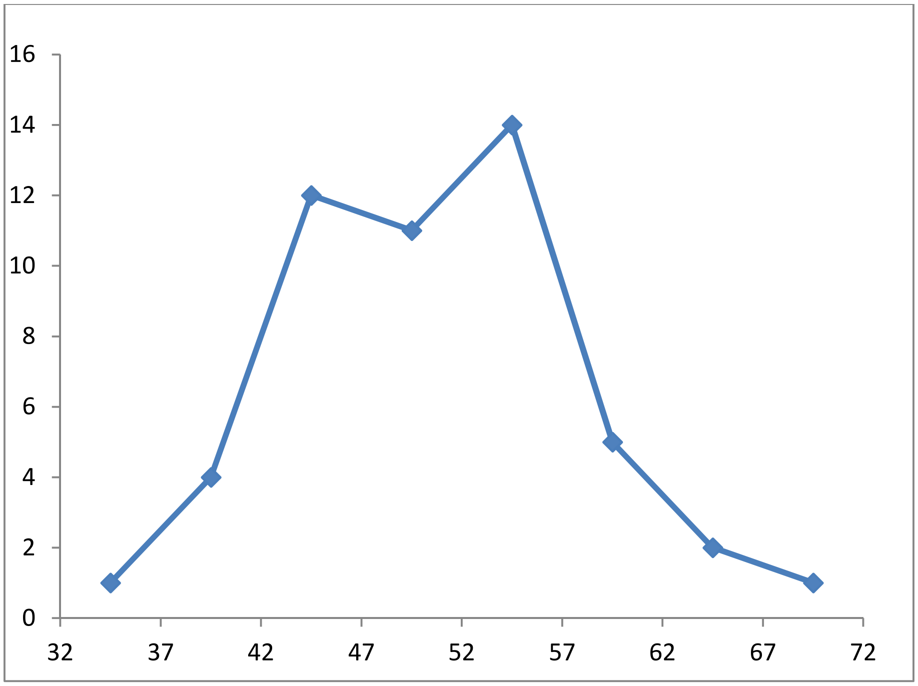
Shown below is a histogram of the weights.



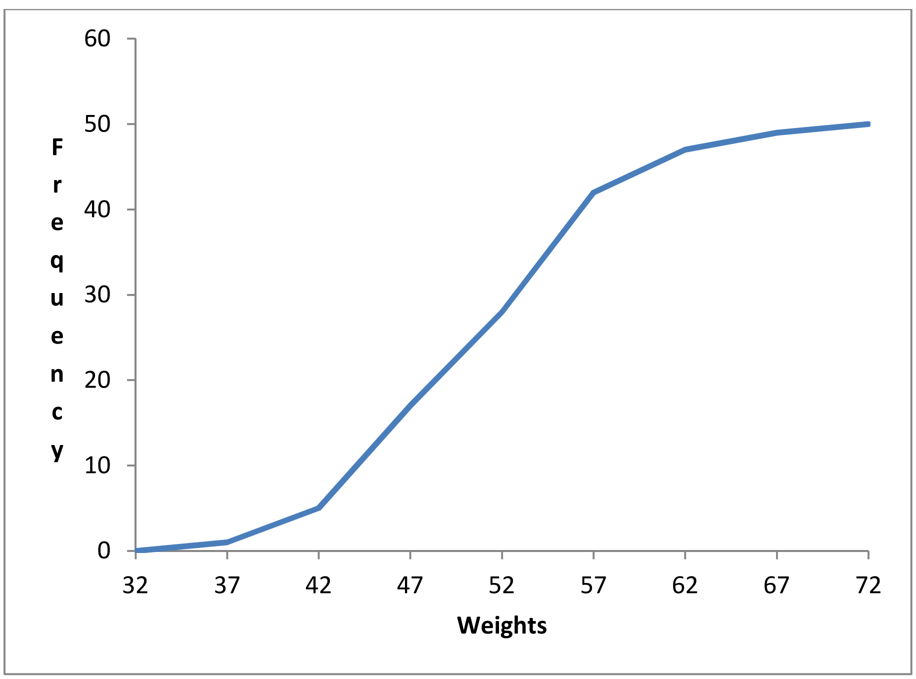
From the histogram, we can see that the bulk of the weights pile up between about 42 ounces and 58 ounces. The distribution appears to have one high point (in Chapter 3 we will call it the modal class – 50 to 54).

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.

2.35

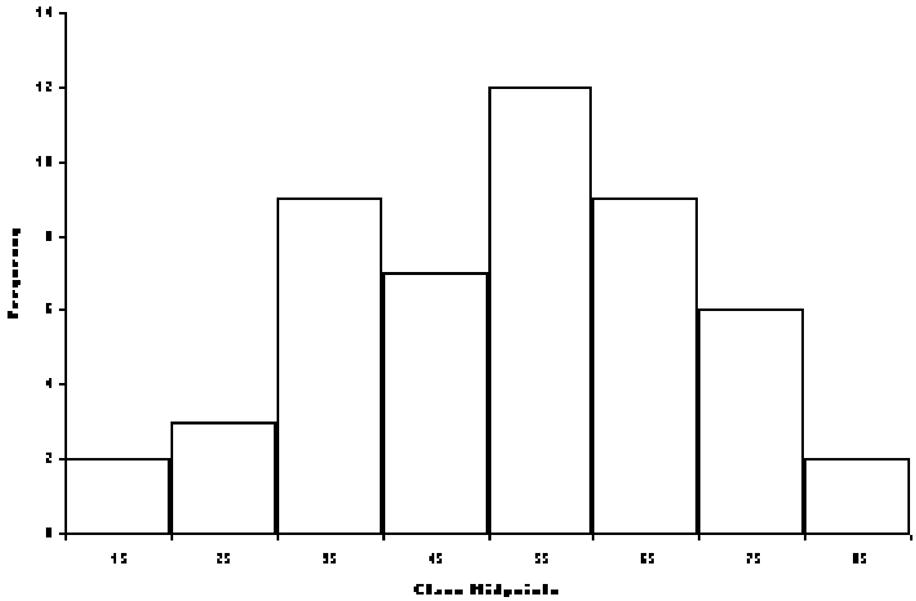
Class Interval	Frequency	Class Midpoint	Relative Frequency	Cumulative Frequency
20 – 25	8	22.5	8/53 = .1509	8
25 – 30	6	27.5	.1132	14
30 – 35	5	32.5	.0943	19
35 – 40	12	37.5	.2264	31
40 – 45	15	42.5	.2830	46
45 – 50	7	47.5	.1321	53
TOTAL	53		.9999	

2.36 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.

2.37 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:



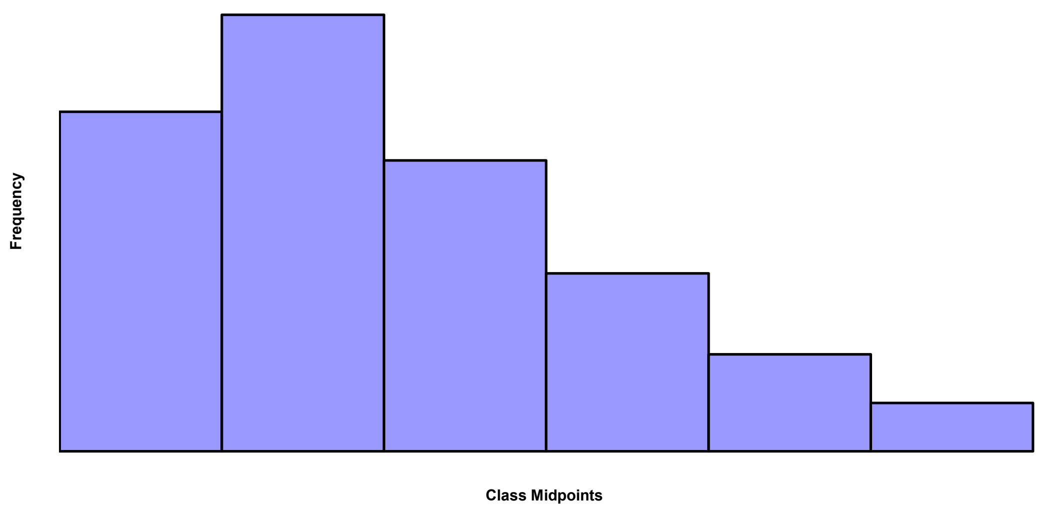
Frequency Polygon:

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

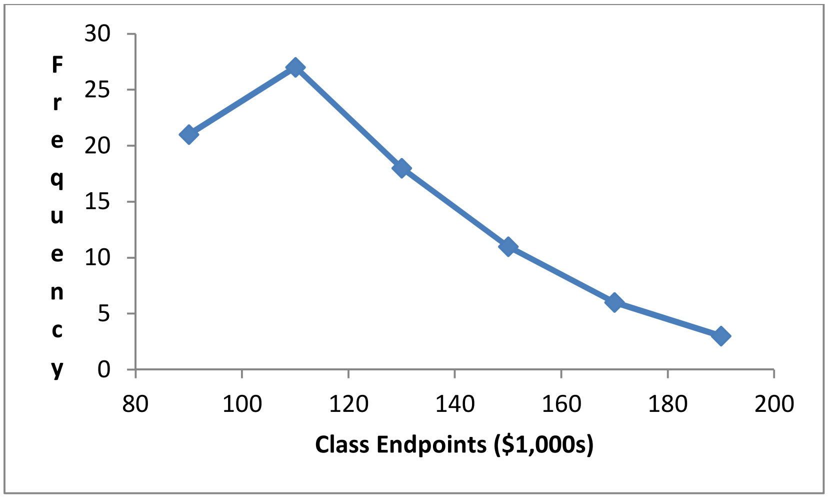
2.38

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	

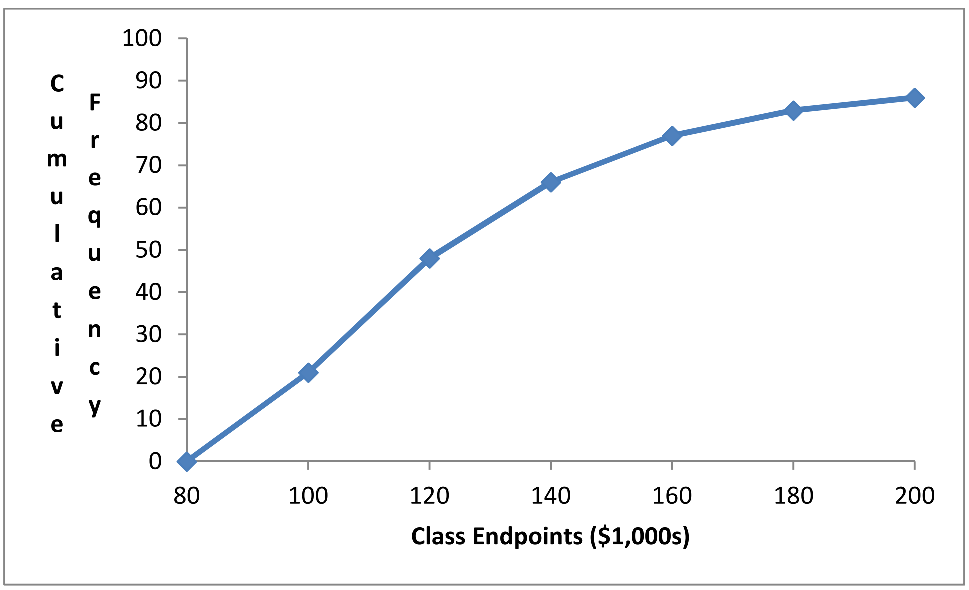
Histogram:



Frequency Polygon:



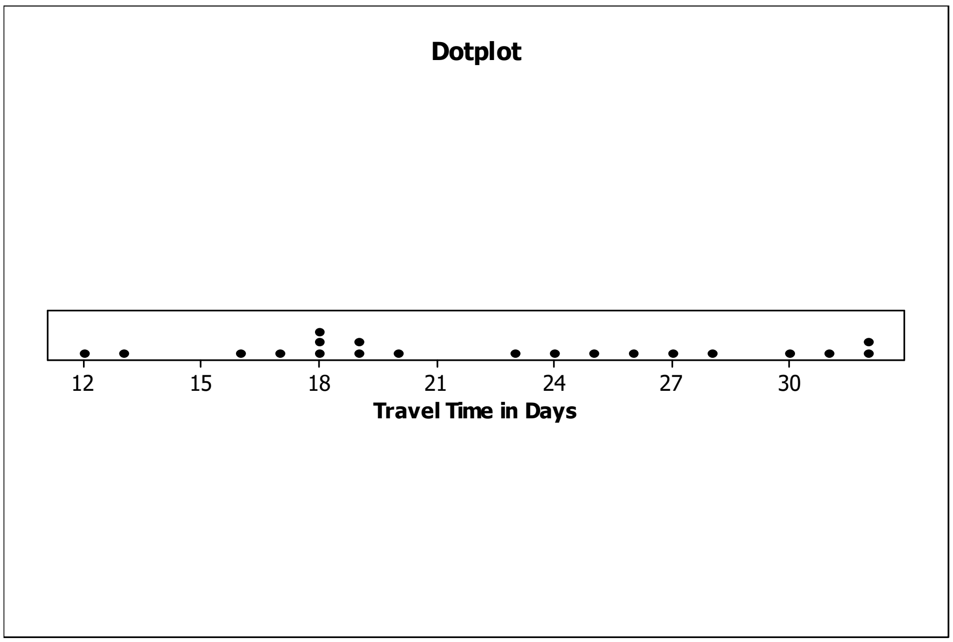
Ogive:



2.39 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



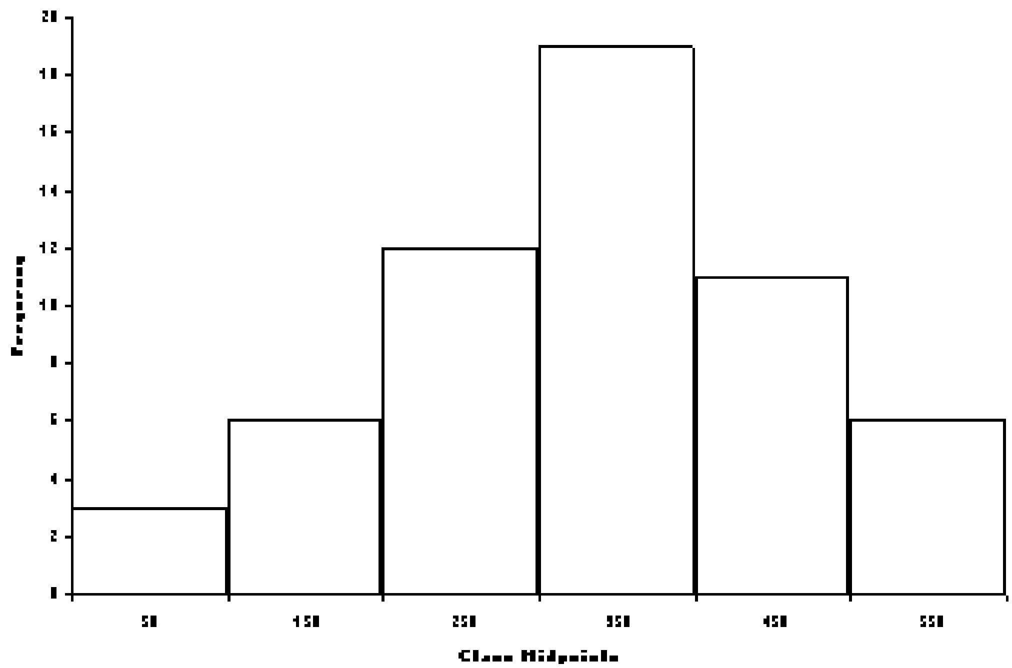
c.) Comments:

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).

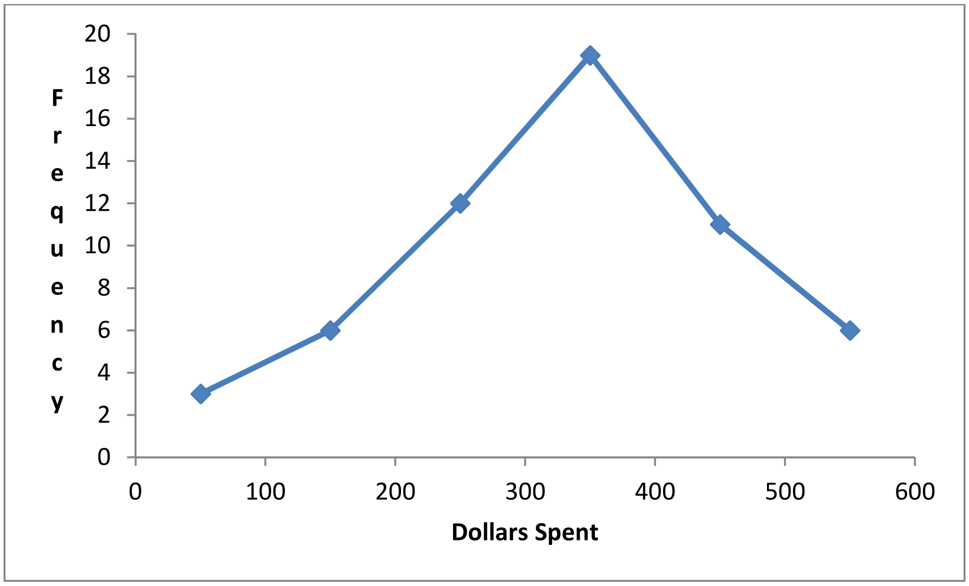
2.40

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	

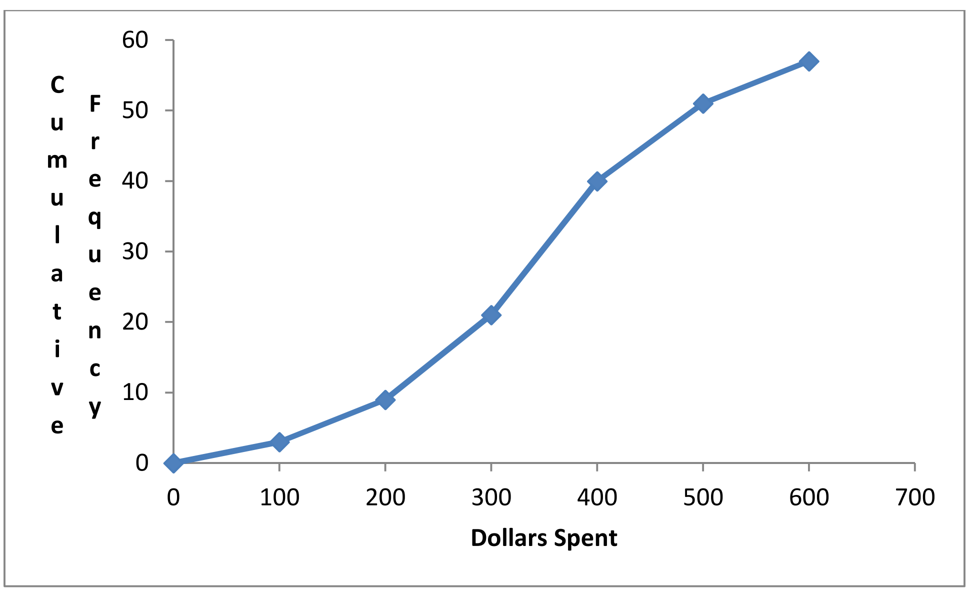
Histogram:



Frequency Polygon:



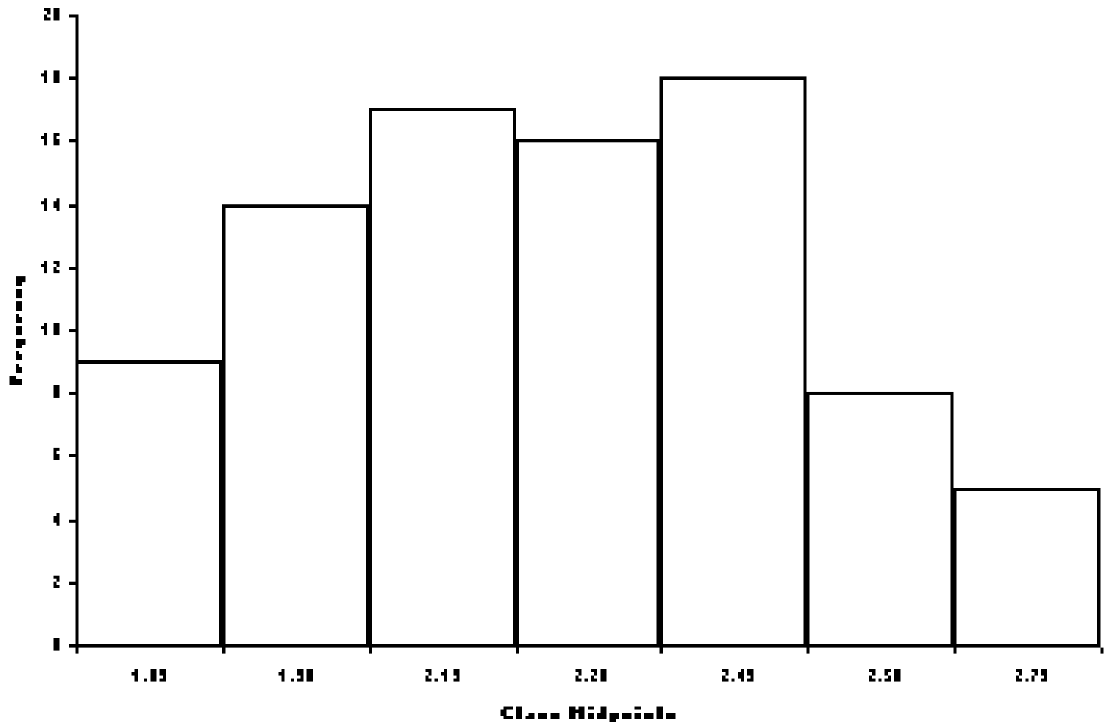
Ogive:



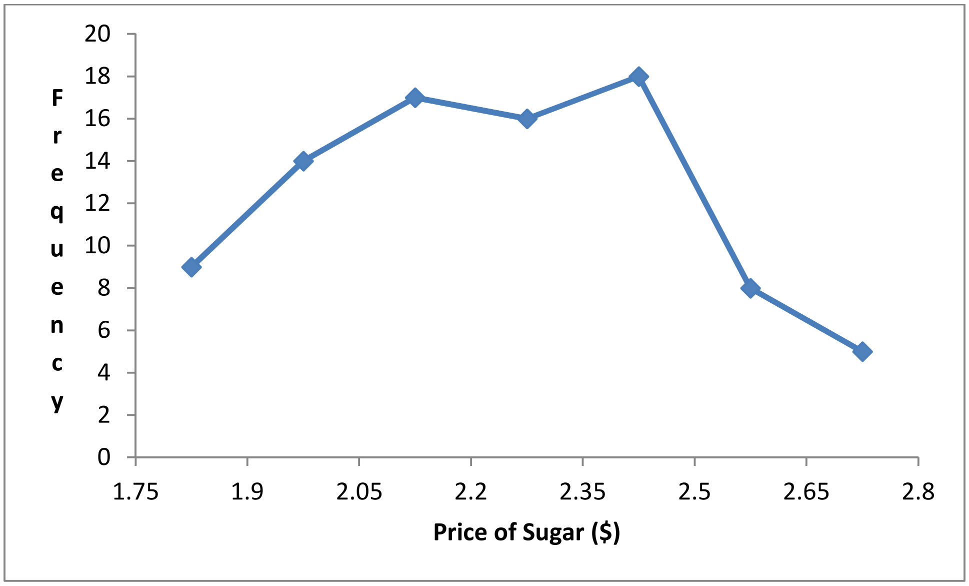
2.41

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	

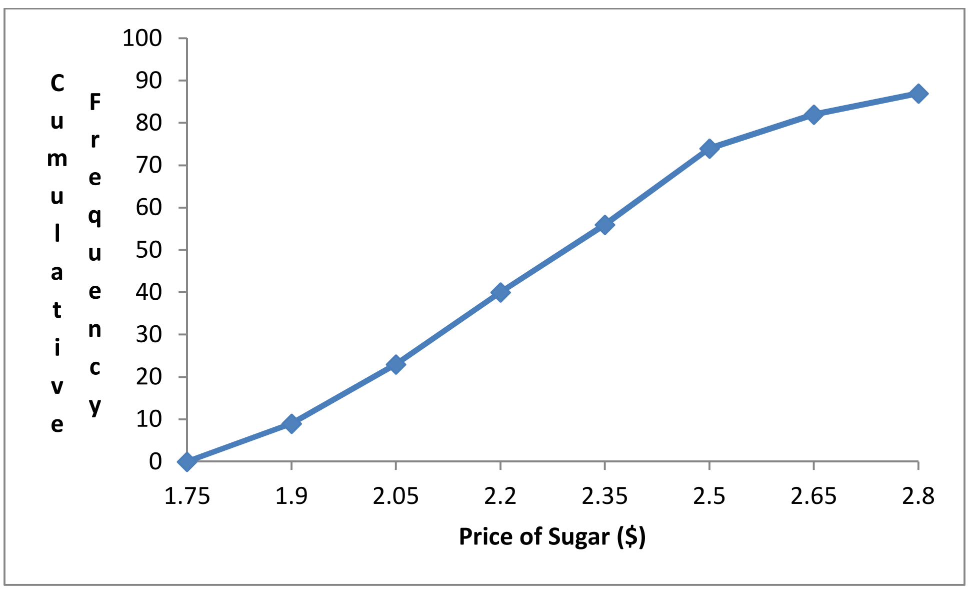
Histogram:



Frequency Polygon:



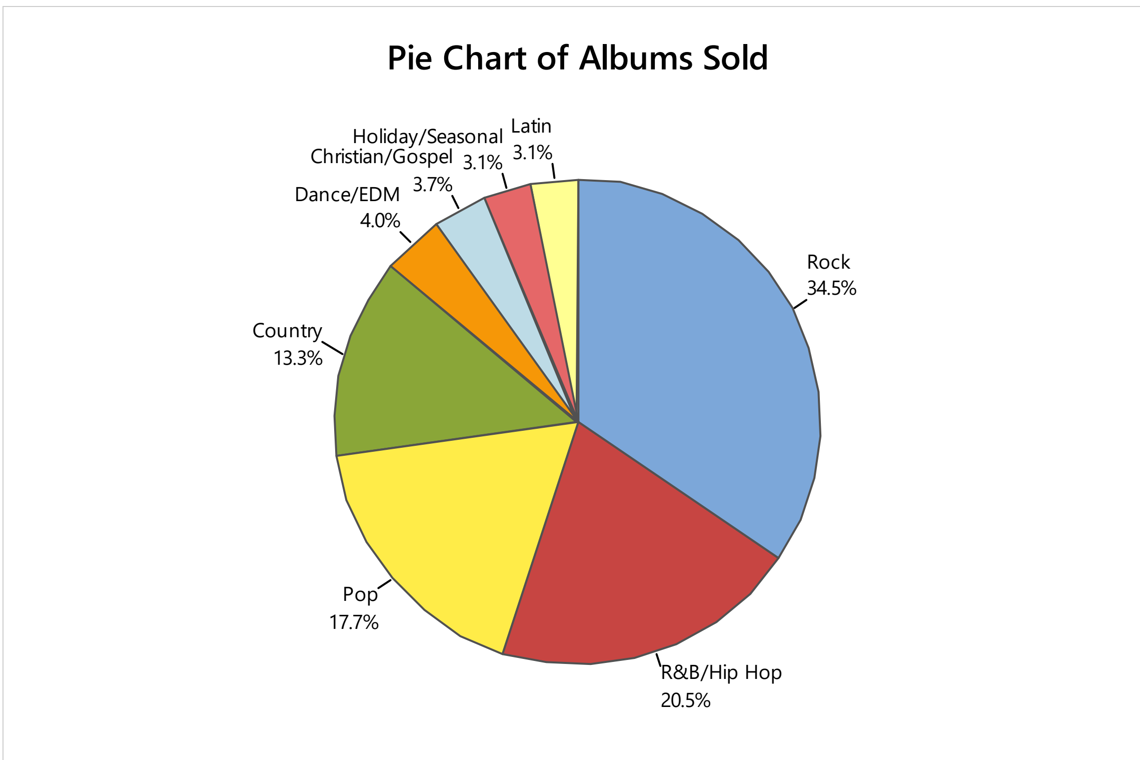
Ogive:



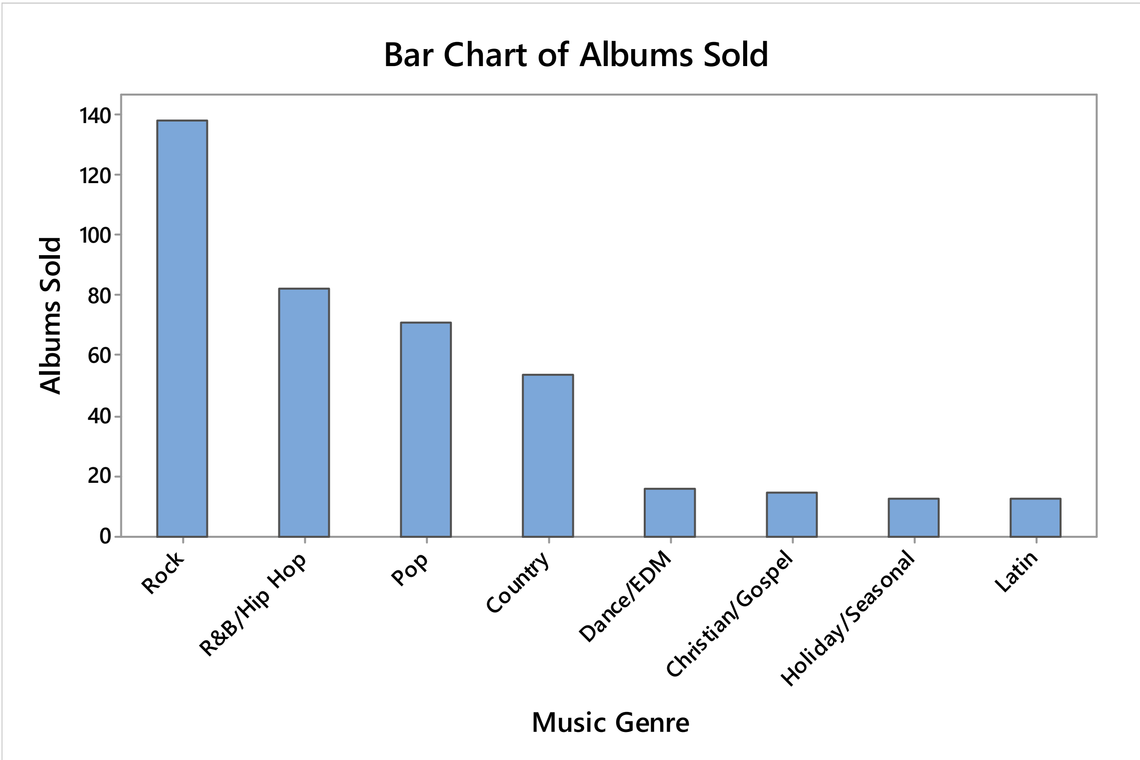
2.42

Genre	Albums Sold	Proportion	Degrees
Rock	138.2	.35	126
R&B/Hip Hop	82.0	.20	72
Pop	71.0	.18	65
Country	53.4	.13	47
Dance/EDM	16.2	.04	14
Christian/Gospel	14.8	.04	14
Holiday/Seasonal	12.4	.03	11
Latin	12.4	.03	11
TOTAL		1.00	360

a.) Pie Chart:

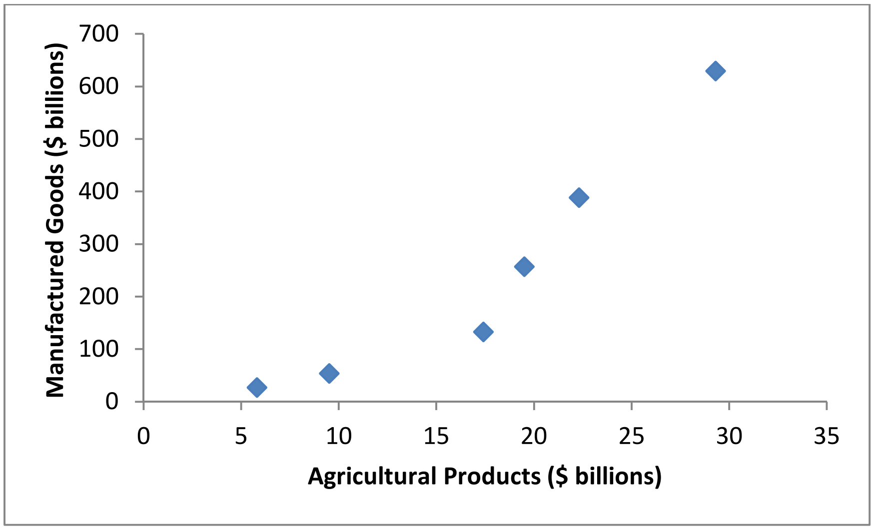


b.) Bar Chart:



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 R&B/Hip Hop sales are greater than the pop sales. In the pie chart, without the labels, differentiating between these two genres would be difficult.

2.43

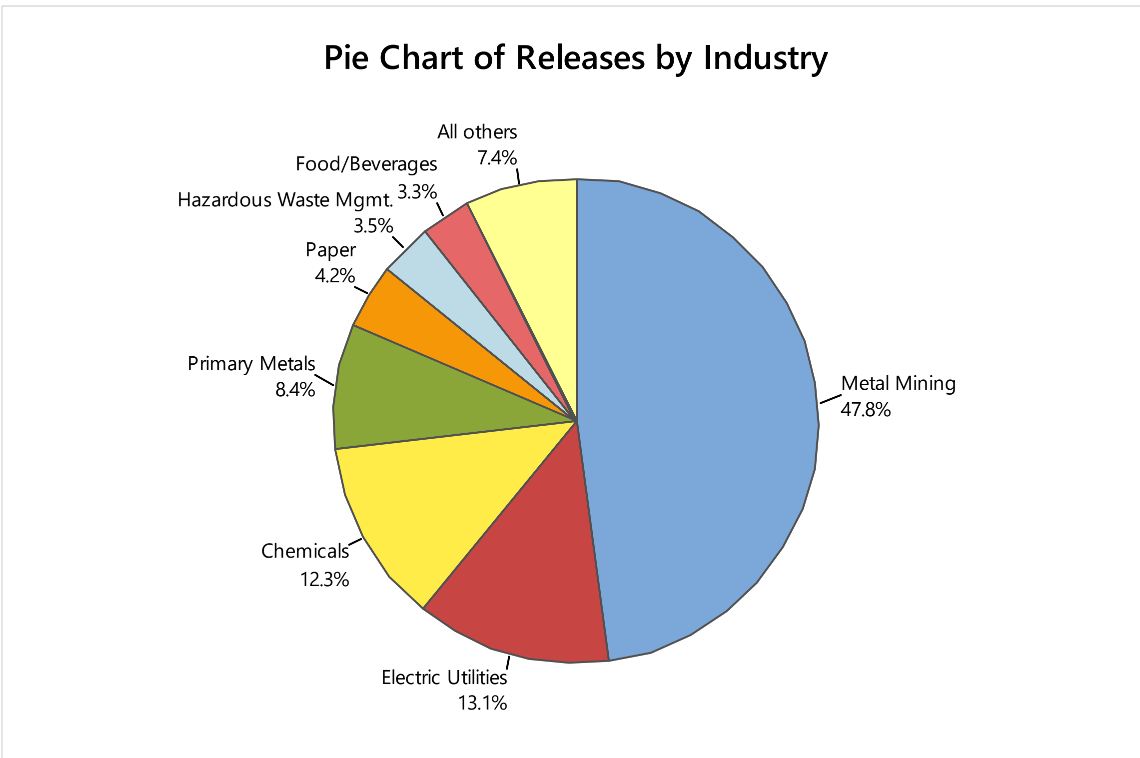


It can be observed that as the U.S. import of agricultural products increased, the U.S. import of manufactured goods also increased. As a matter of fact, a non-linear association may exist between the two variables.

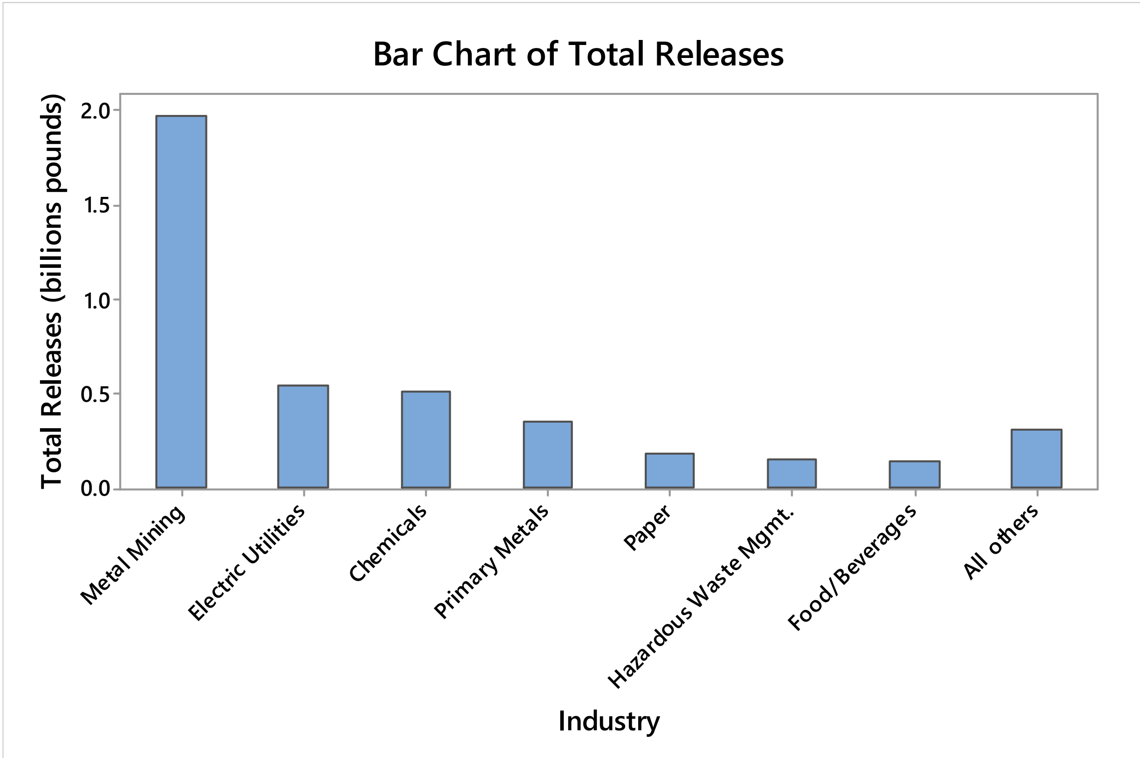
2.44

Industry	Total Release (billions lbs.)	Proportion	Degrees
Metal Mining	1.970	.478	172
Electric Utilities	0.539	.131	47
Chemicals	0.505	.123	44
Primary Metals	0.346	.084	30
Paper	0.175	.042	15
Hazardous Waste Mgmt.	0.146	.035	13
Food/Beverages	0.134	.033	12
All Others	0.306	.074	27
TOTAL	4.121	1.000	360

Pie Chart:

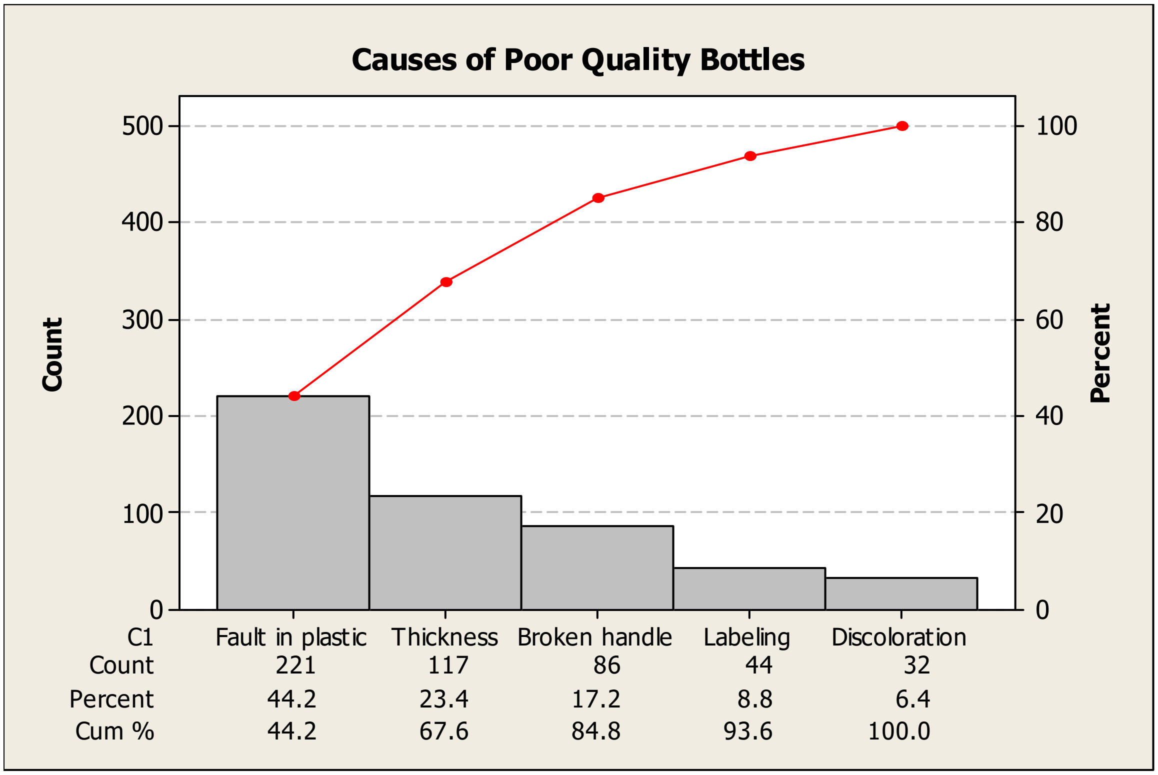


Bar Graph:



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 three categories from paper through hazardous waste mgmt. are close enough to be problematic in discipering differences. The same goes for electric utilities and chemicals and for primary metals and all others. As for the pie chart, were the categories not labeled, it would be difficult to determine the relative differences in these same categories in the pie chart. Perhaps any 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.

2.45



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.

2.46

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

2.47 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.

2.48 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.

2.49 This pie chart without numerical labels gives inconclusive information. Certainly we can conclude that of the six specialties graphed, cardiology is the smallest one. The other five have pie slices that are so similar, it is difficult to discern the differences. This points out the weakness of the pie chart without numerical information. That is, often it is difficult by just eyeing the chart to determine relative sizes of the slices.

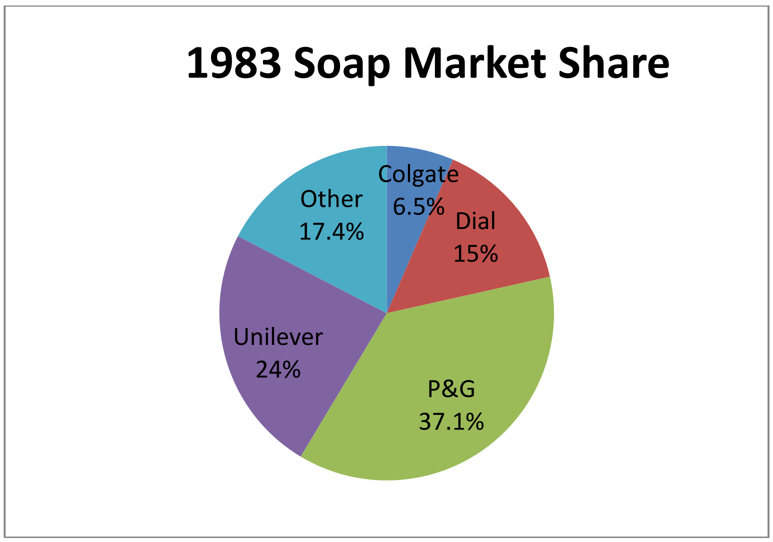
2.50 The fewest number of audits is 12 and the most is 42. More companies (8) performed 27 audits than any other number. Thirty-five companies performed between 12 and 19 audits. Only 7 companies performed 40 or more audits.

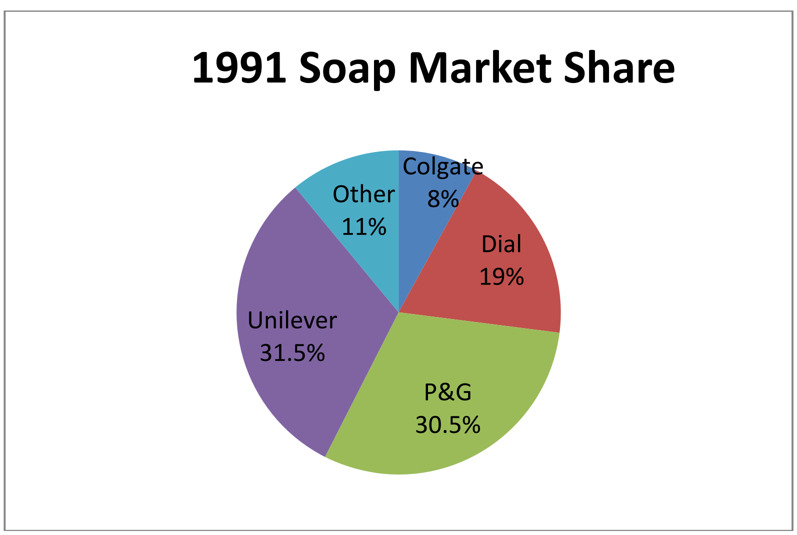
2.51 There appears to be a relatively strong positive relationship between the NASDAQ-100 and the DJIA. Note that as the DJIA became higher, the NASDAQ-100 tended to also get higher. Except for a few outliers, the slope of the graph was generally the same over time. The strong relationship between the two indices was especially good until the DJIA reached about 17,500 at which point there was more spread of the NASDAQ-100 points about the various values of the DJIA.

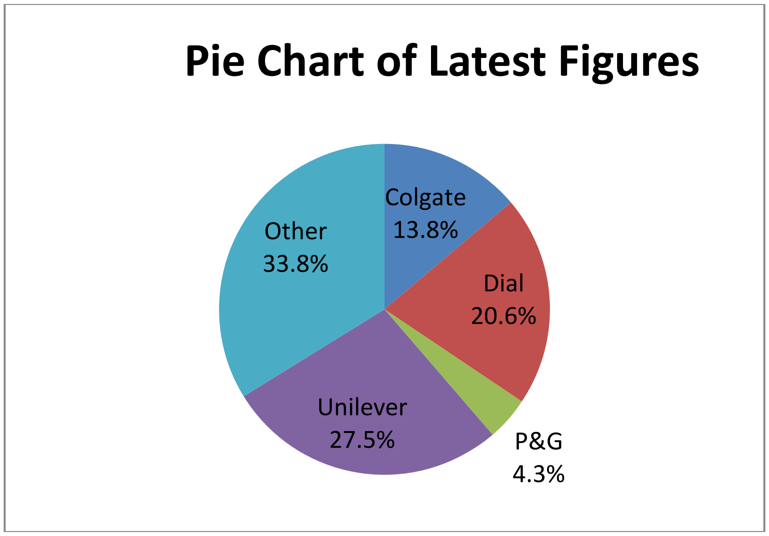
Chapter 2
Soap Companies Do Battle

The pie chart is useful in displaying the market shares in one device adjacent to each other. Many decision makers are used to viewing pie charts in connection which budgets and therefore might feel more at ease with a pie chart. On the other hand, when percentages are close such as with Dial and "Others" in 1983, it can be difficult to discern the difference using the pie chart slices. In this case, the bar chart shown above is more desirable.

1. Shown below are pie charts for the 1983, the 1991, and the latest market shares.



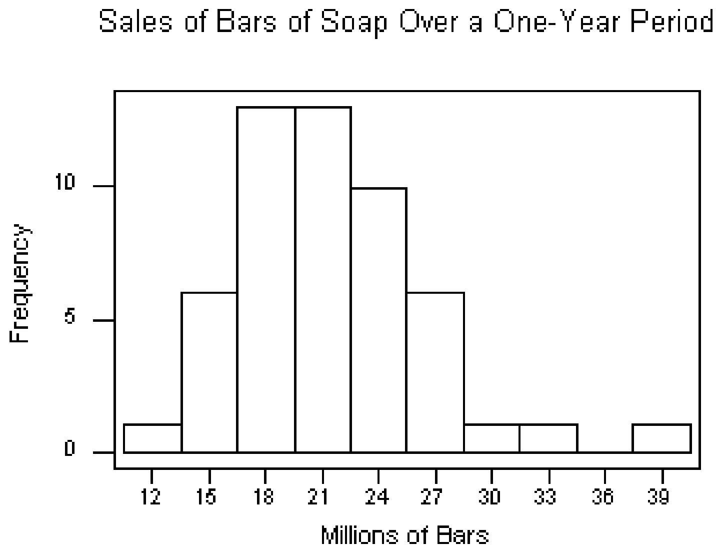




An examination of the pie charts from 1983 through the latest reveals that the slice sizes of Dial and Colgate-Palmolive have grown and the sizes of the Procter & Gamble slices have shrunk substantially since 1983. Shown below are the actual percentage figures for the three time periods so that you have the option of displaying the data in different ways:

Company	1983 Share	1991 Share	Latest Share
Procter & Gamble	37.1	30.5	4.3
Unilever	24.0	31.5	27.5
Dial	15.0	19.0	20.6
Colgate-Palmolive	6.5	8.0	13.8
Others	17.4	11.0	33.8

2. Shown below is a histogram of the weekly sales of bars of soaps over the year. The histogram was constructed using 10 classes. In Minitab, the student has the option of trying several different values for the number of intervals. In Excel, students can explore various bin options. The shape of the histogram will somewhat change according to the number of class intervals. Note the shape of this histogram is mound shaped with some skewness to the right. The center of the distribution appears to be near to 20 million as would be expected since Procter & Gamble sells about 20 million bars per week, for the sample year selected. Note, however that some weeks actually average as much as 39 million bars per week and others only 12 million bars. What inventory, production, and human resource implications might this have? How does a company "cope" with such fluctuations?

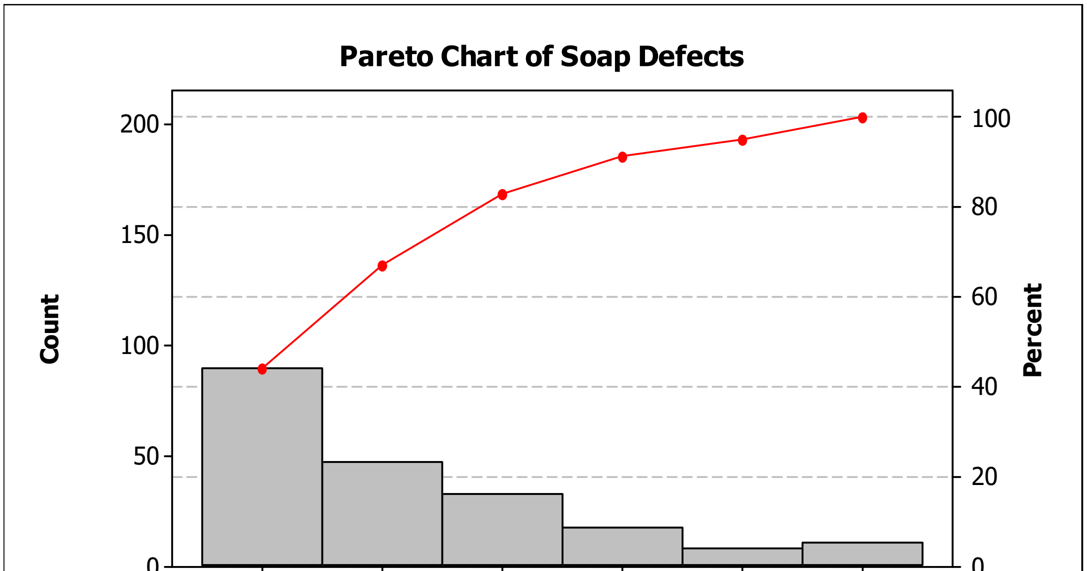


The stem and leaf plot for these data is shown below. The advantage of the stem and leaf over histograms, pie charts, bar charts, and others is that the stem and leaf retains the original data in case the researcher wants to calculate other statistics on the numbers. Production people would likely find the histogram the most interesting because it displays to them where the bulk of production occurs and the magnitude of the unusual size runs.

Stem	Leaf
12	2
13	6
14	7
15	0, 4, 4, 5
16	8
17	0, 1, 1, 4, 5
18	2, 3, 4, 5, 7
19	1, 3, 6, 9
20	0, 3, 3, 4, 4, 6, 7, 9
21	3, 4, 4
22	5, 8
23	1, 4, 8, 9
24	0, 3
25	2, 2
26	2, 2, 3, 3, 6, 9
27	
28	
29	
30	6
31	
32	8
33	

34	
35	
36	
37	
38	
39	8

3. Following is a Minitab-produced Pareto Chart for the frequency of problems associated with defective soap bars. Note that Bar Surface is by far the most frequently occurring problem causing nearly 44% of the defects. If the organization wanted to improve the quality of the packaged bar soap, they should identify ways to reduce marring of the bar surface. Improvement in this area will greatly reduce the overall number of defects. Similarly, improvements in the seal and/ or label will reduce the overall number of defects because 23% of the defects are due to a bad seal and almost 16% of the defects are due to poor labeling. Together, marred bar surface, bad seal, and poor labeling, account for approximately 80% of all defects (83%, to be exact).



C1	Bar surface	Seal	Label	Color	Shape	Other
Count	89	47	32	17	8	10
Percent	43.8	23.2	15.8	8.4	3.9	4.9
Cum %	43.8	67.0	82.8	91.1	95.1	100.0

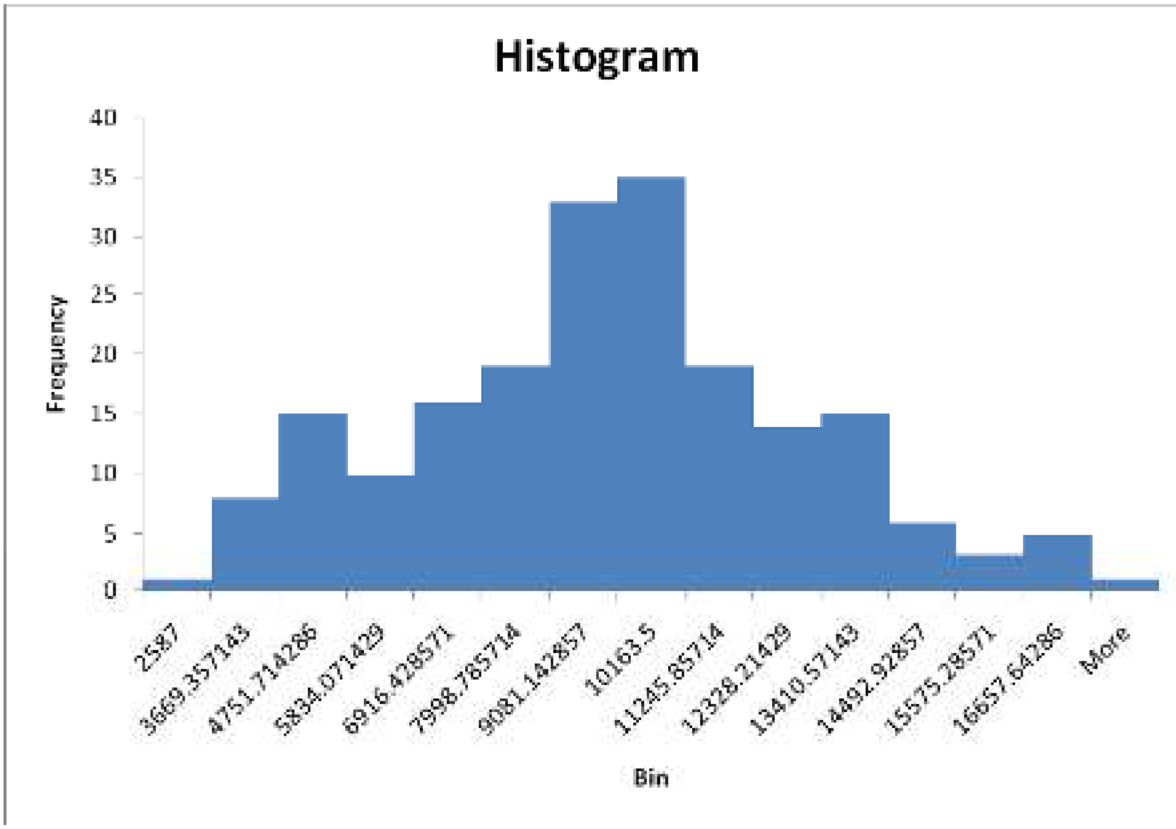
Answers to Analyzing the Databases Questions in Chapter 2

1. Frequency Distribution of Number of Production Workers using Excel. Note in Excel if you do not specify the interval classes, it arbitrarily selects the class intervals. The last interval says 'more'. To avoid this, the user can specify the 'bin'. See the second table.

Bin	*Frequency*
50	68
100	30
150	19
200	10
250	6
300	1
350	1
400	2
450	1
500	0
More	2

0 to 50	68
51 to 100	30
101 to 150	19
151 to 200	10
201 to 250	6
251 to 300	1
301 to 350	1
351 to 400	2
401 to 450	1
451 to 500	0
501 to 550	0
551 to 600	1
601 to 650	1

This tells us that the number of production workers "pile up" at the lower numbers. In chapter 3, we will learn that this distribution is skewed right. Number of production workers is in units of 1,000. Thus, 68 of the 140 companies in this database have between 1,000 and 50,000 workers. This represents almost 49% of the companies. An additional 30/140 or over 21% of the companies have between 50,001 and 100,000 production workers. Taken together, these two categories account for over 70% of the companies in this database.

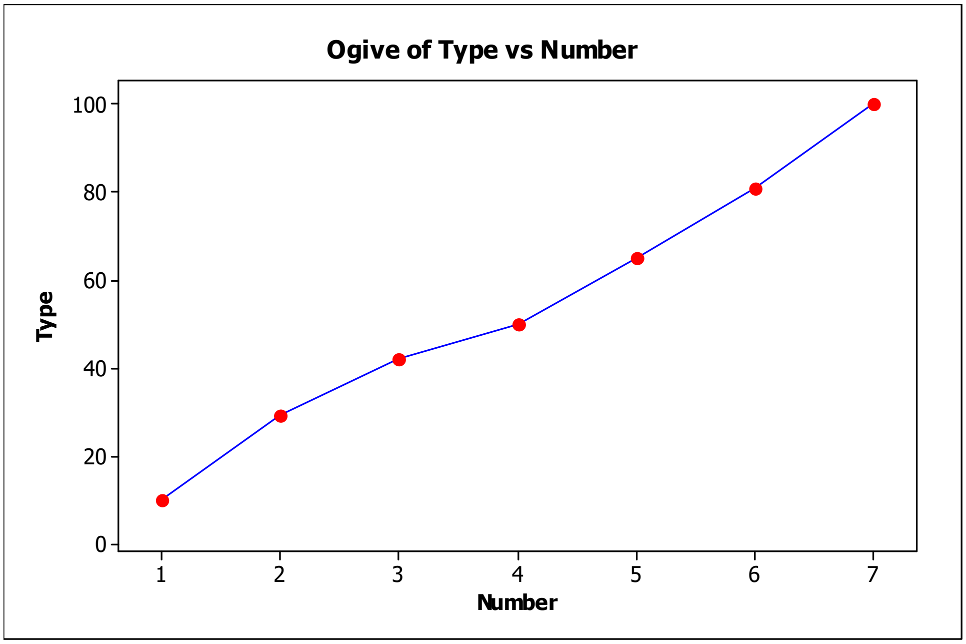


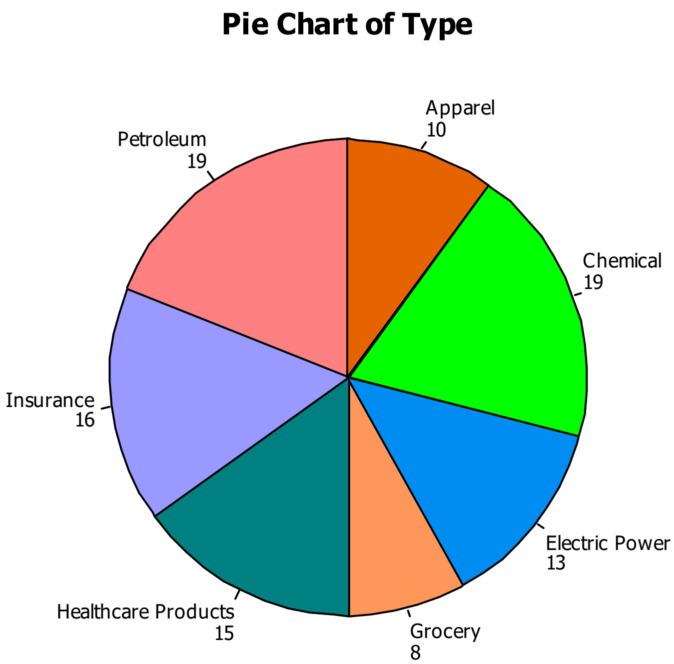
The histogram has a somewhat symmetrical shape. It is high in the middle and "thin" at the ends. It is not unlike the normal distribution (bell-shaped curve) that will be introduced later in the text. Annual food spending seems to be centered at the $10,000 amount. A large chunk of the histogram falls between $7,000 and $13,000.

3. Graphing discrete numbers does not exactly produce a true ogive. However, one can produce a cumulative line graph over the numbers progressing from 1 to 7. The cumulative distribution is:

Type	Cumulative Frequency
1	10
2	29
3	42
4	50
5	65
6	81
7	100

The ogive-like graph is shown below. The line in the graph is relatively straight indicating that most increases are fairly constant and there are few "jumps" or inordinate increases. The raw cumulative numbers shown above point this out. If you were to look at the actual frequency counts, you would see that Type 2 and Type 7 have the highest number with a frequency of 19 and Type 4 has the fewest with 8. Thus, as the ogive goes from Type 3 to Type 4, there is a slight dip in the increase of the line. As the ogive goes from Type 1 to Type 2 and from Type 6 to Type 7, you see the greatest slope or increase in the line.





The pie chart demonstrates that more companies are chemical and petroleum (19) than any other categories and these two categories are equal. It also shows that there are fewer grocery companies. Without the frequency labels, however, it would be more difficult to differentiate between the numbers of companies in the various categories.

4. Stem and Leaf Plot for Italy:

Stem-and-Leaf Display: Italy

Stem-and-leaf of Italy N = 40
Leaf Unit = 0.10

```
 2  478
 3  12234455577789
 4  114489
 5  499
 6  09
 7  035899
 8
 9
10  2
11  3
12  0133
```

The stem and leaf plot retains all of the original data so we know, for example, that some of the annual unemployment rates included 2.4%, 2.7%, 2.8%, 6.9%, 10.2%, and 12.3%. Studying this plot, we can tell that for most years (34 out of 40), the unemployment rate was below 8%. In fact, a high number of unemployment rates were in the 3% bracket. However, since these rates were gathered over a 40-year period, we cannot see from the stem and leaf plot, if there are any trends that would help predict ensuing years' unemployment rates and/or if the high unemployment rates were many years ago or more recent.