

## **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 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

## KEY TERMS

Bar Graph

Class Mark

Class Midpoint

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

## 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.

<u>Age</u>	<u>f</u>
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 \_\_\_\_\_.

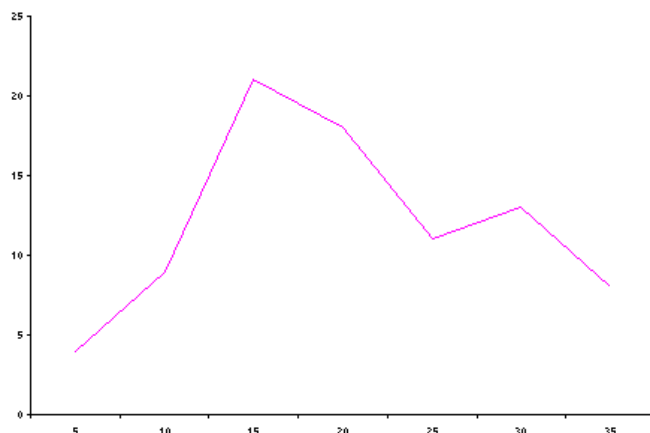
27 21 8 10 9 16 11 12 21 11 29 19 17 22 28 28 29 19 18 26 17 34 19 16 20

6. The class midpoint can be determined by \_\_\_\_\_.

7-9 Examine the frequency distribution below:

<u>class</u>	<u>frequency</u>
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:

<u>Category</u>	<u>Amount</u>
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 \_\_\_\_\_.

### ANSWERS TO STUDY QUESTIONS

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

**SOLUTIONS TO THE ODD-NUMBERED PROBLEMS IN CHAPTER 2**

2.1

- a) One possible 5 class frequency distribution:

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

- b) One possible 10 class frequency distribution:

<u>Class Interval</u>	<u>Frequency</u>
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

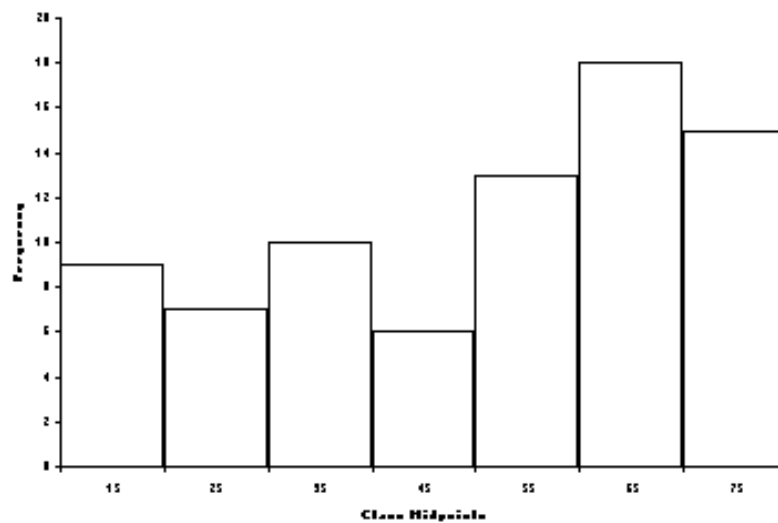
<u>Class</u> <u>Interval</u>	<u>Frequency</u>	<u>Class</u> <u>Midpoint</u>	<u>Relative</u> <u>Frequency</u>	<u>Cumulative</u> <u>Frequency</u>
0 - 5	6	2.5	$6/86 = .0698$	6
5 - 10	8	7.5	.0930	14
10 - 15	17	12.5	.1977	31
15 - 20	23	17.5	.2674	54
20 - 25	18	22.5	.2093	72
25 - 30	10	27.5	.1163	82
30 - 35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

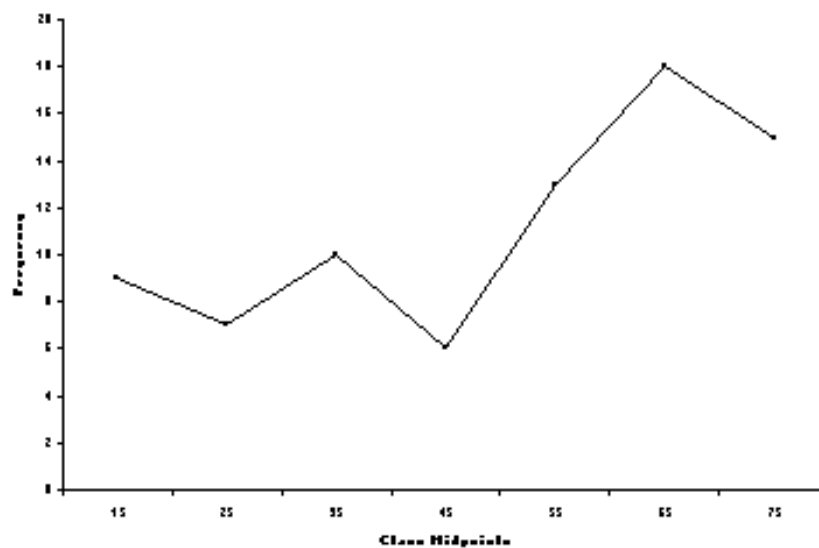
## 2.5 Some examples of cumulative frequencies in business:

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:



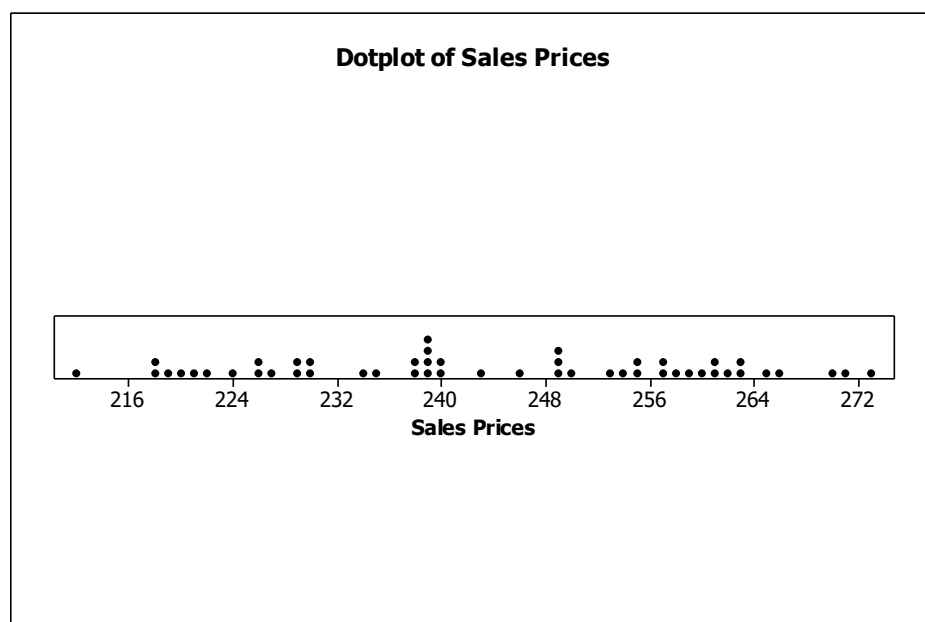
## Frequency Polygon:



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 The histogram shows that there are only three airports with more than 70 million passengers. 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. We can tell from the graph that there is one airport with between

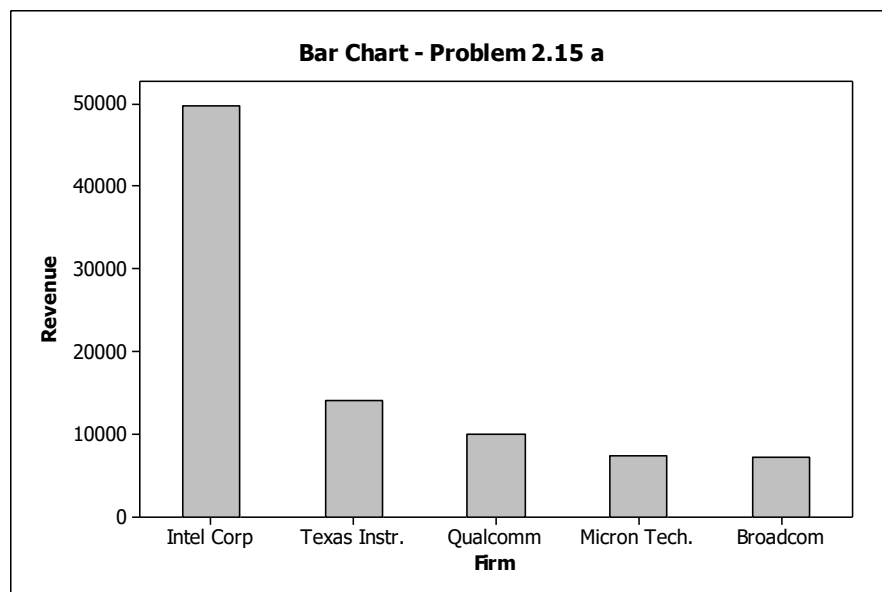


80 and 90 million passengers and another airport with between 70 and 80 million passengers. Four airports have between 60 and 70 million passengers. Eighteen of the top 30 airports have between 40 and 60 million passengers.

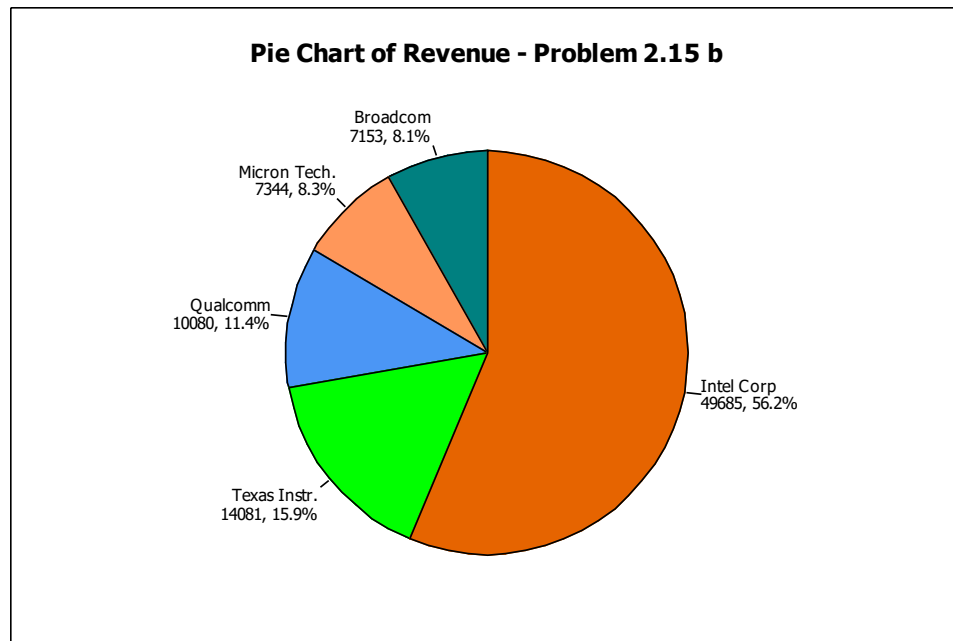
- 2.13 From the stem and leaf display, the original raw data can be obtained. For example, the fewest number of cars washed on any given day are 25, 29, 29, 33, etc. The most cars washed on any given day are 141, 144, 145, and 147. The modal stems are 3, 4, and 10 in which there are 6 days with each 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	<u>Firm</u>	<u>Proportion</u>	<u>Degrees</u>
	Intel Corp.	.5624	202.5
	Texas Instruments	.1594	57.4
	Qualcomm	.1141	41.1
	Micron Technology	.0831	29.9
	Broadcom	<u>.0810</u>	<u>29.2</u>
	TOTAL	1.0000	360.1

a.) Bar Graph:



b.) Pie Chart:

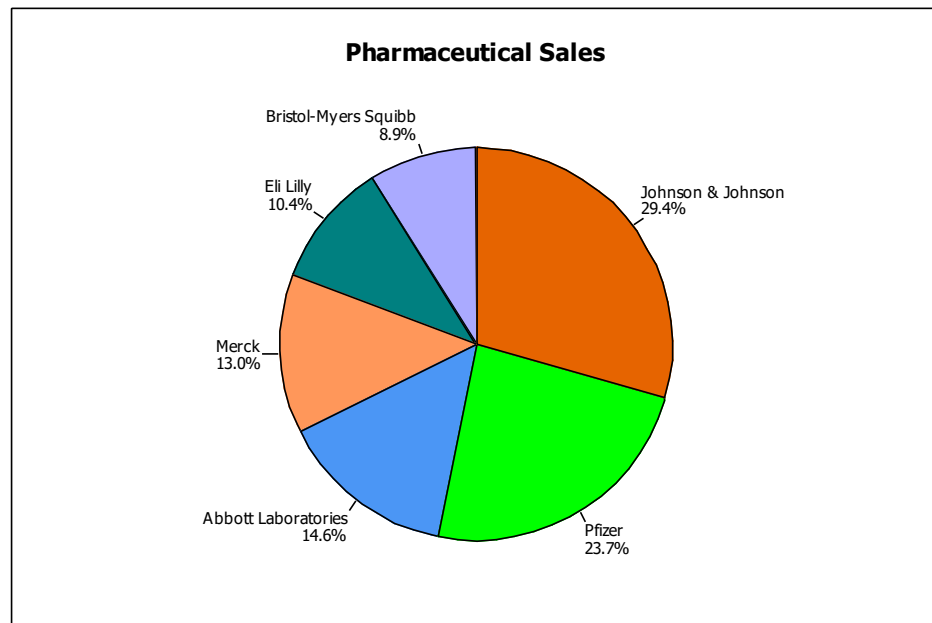


- c.) While pie charts are sometimes interesting and familiar to observe, in this problem it is virtually impossible from the pie chart to determine the relative difference between Micron Technology and Broadcom. In fact, it is somewhat difficult to judge the size of Qualcomm and Texas Instruments. From the bar chart, however, relative size is easier to judge, especially the difference between Qualcomm and Texas Instruments.

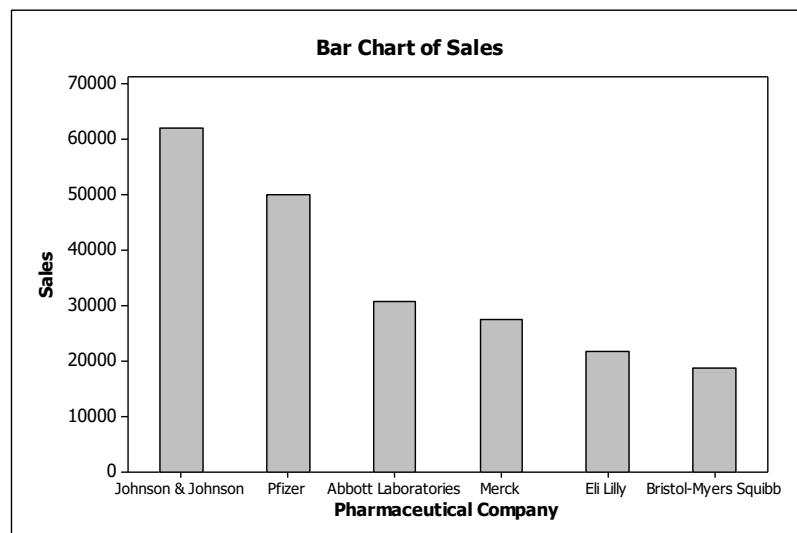
2.17

<u>Brand</u>	<u>Proportion</u>	<u>Degrees</u>
Johnson & Johnson	.294	106
Pfizer	.237	85
Abbott Laboratories	.146	53
Merck	.130	47
Eli Lilly	.104	37
Bristol-Myers Squibb	.089	32
TOTAL	1.000	360

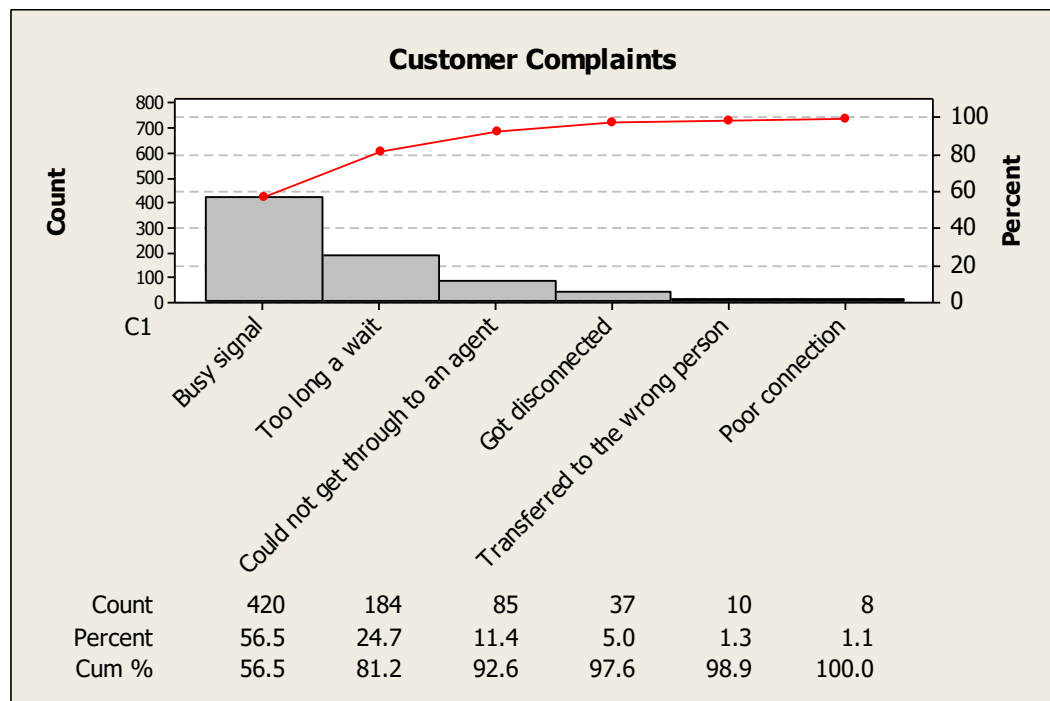
Pie Chart:



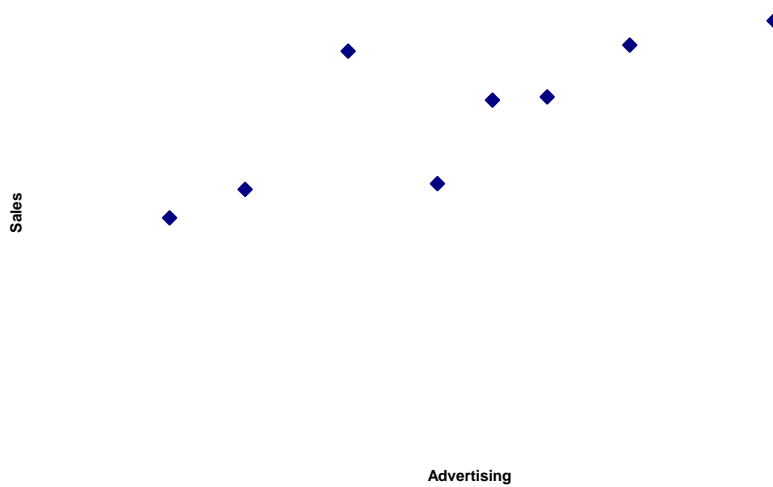
Bar Graph:



2.19	<u>Complaint</u>	<u>Number</u>	<u>% of Total</u>
	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



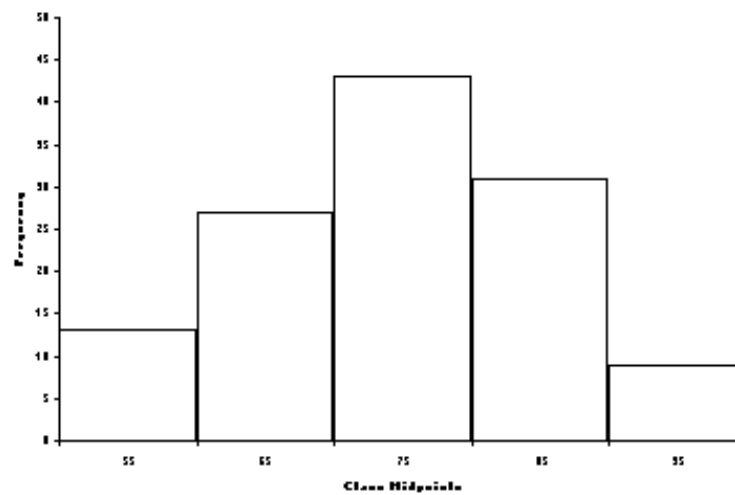
Generally, as advertising dollars increase, sales are increasing.

- 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.

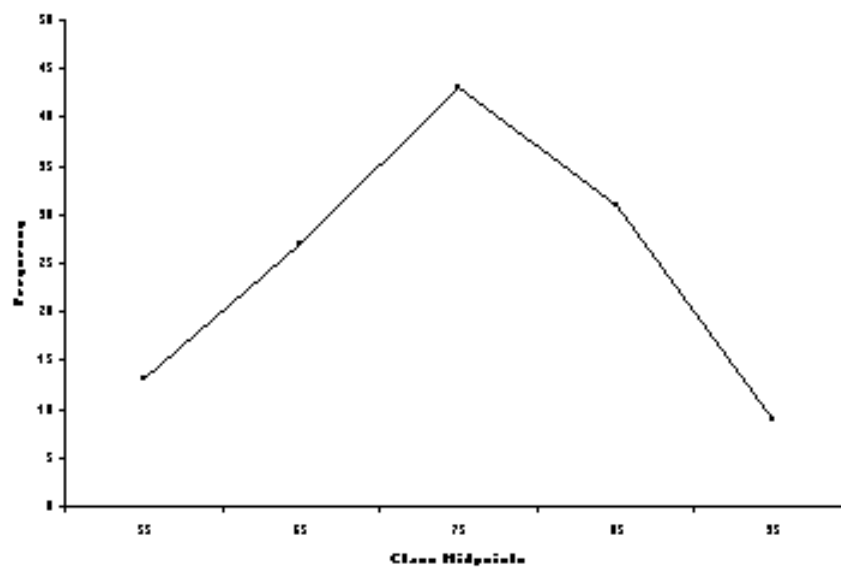
2.25	<u>Class Interval</u>	<u>Frequencies</u>
	16 - under 23	6
	23 - under 30	9
	30 - under 37	4
	37 - under 44	4
	44 - under 51	4
	51 - under 58	<u>3</u>
	TOTAL	30

2.27	<u>Class Interval</u>	<u>Frequencies</u>
	50 - under 60	13
	60 - under 70	27
	70 - under 80	43
	80 - under 90	31
	90 - under 100	9
	TOTAL	123

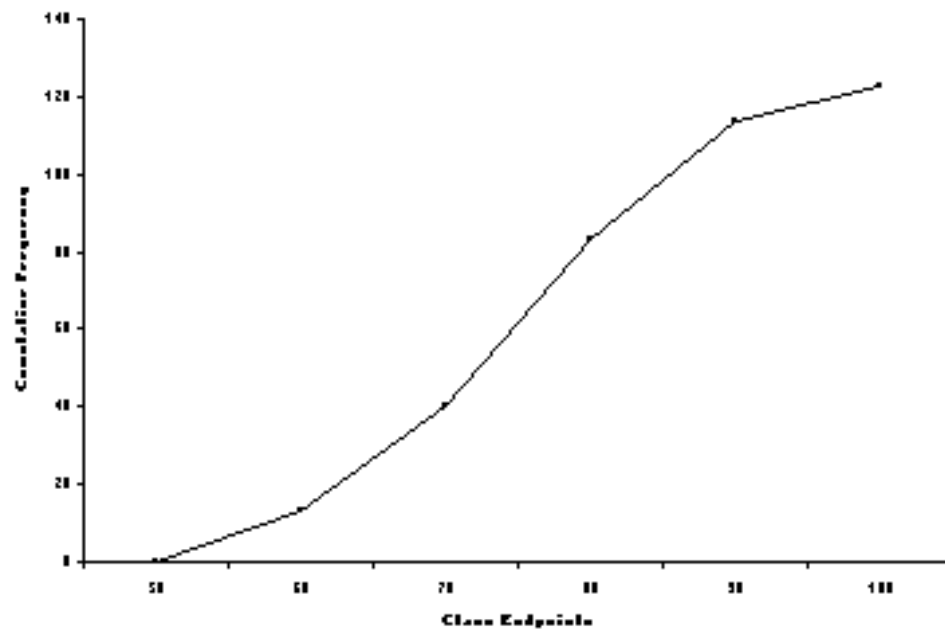
Histogram:



Frequency Polygon:



Ogive:



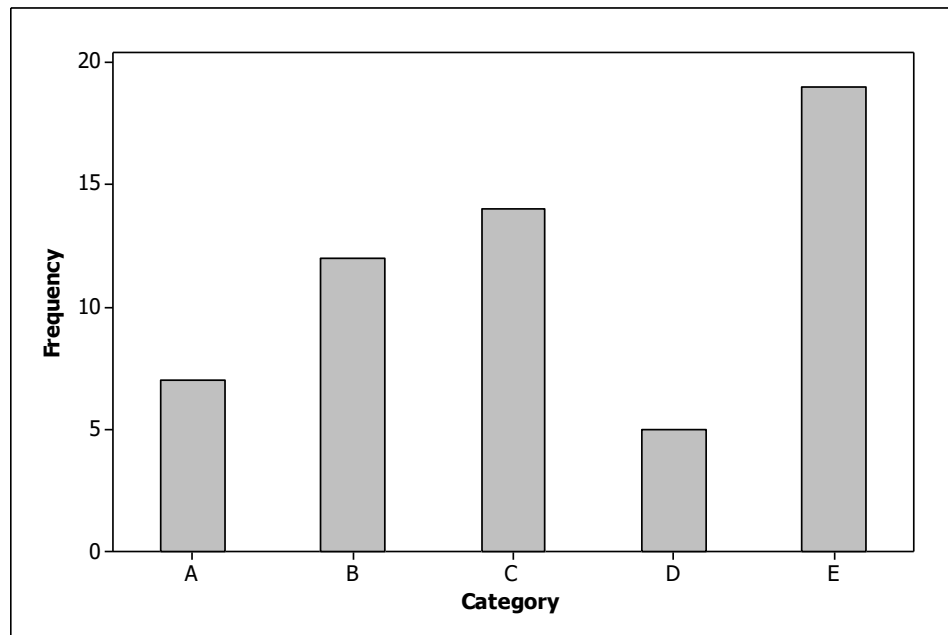
2.29

STEM LEAF

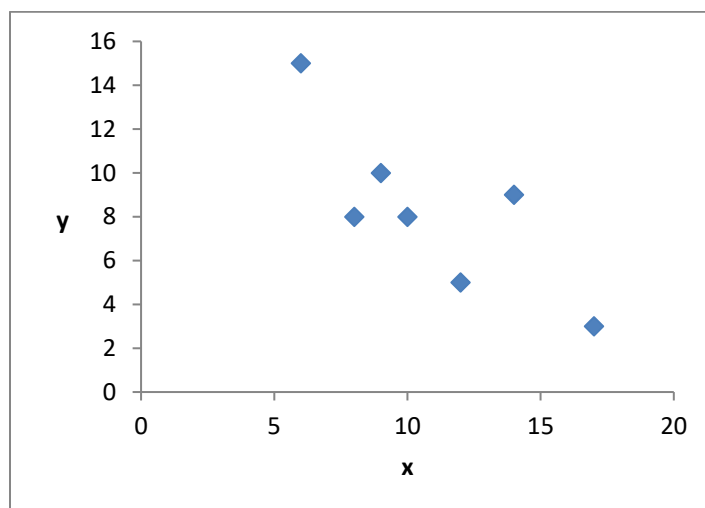
28	4 6 9
29	0 4 8
30	1 6 8 9
31	1 2 4 6 7 7
32	4 4 6
33	5

## 2.31 Bar Graph:

<u>Category</u>	<u>Frequency</u>
A	7
B	12
C	14
D	5
E	19



## 2.33 Scatter Plot





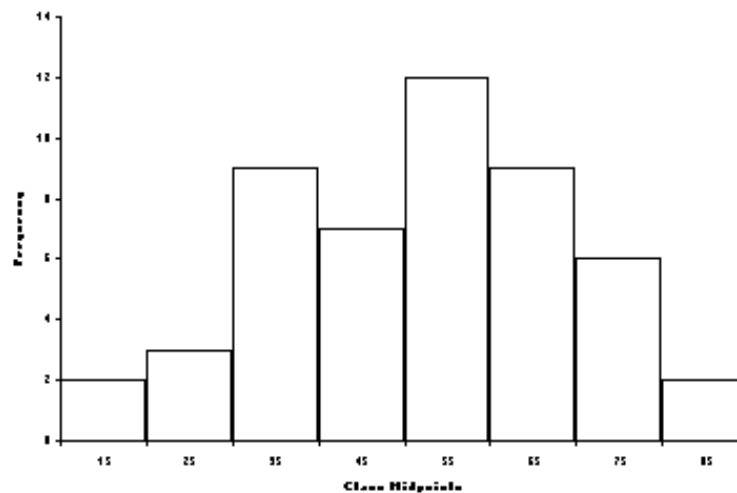
2.35

<u>Class Interval</u>	<u>Frequency</u>	<u>Class Midpoint</u>	<u>Relative Frequency</u>	<u>Cumulative Frequency</u>
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	<u>7</u>	47.5	<u>.1321</u>	53
TOTAL	53		.9999	

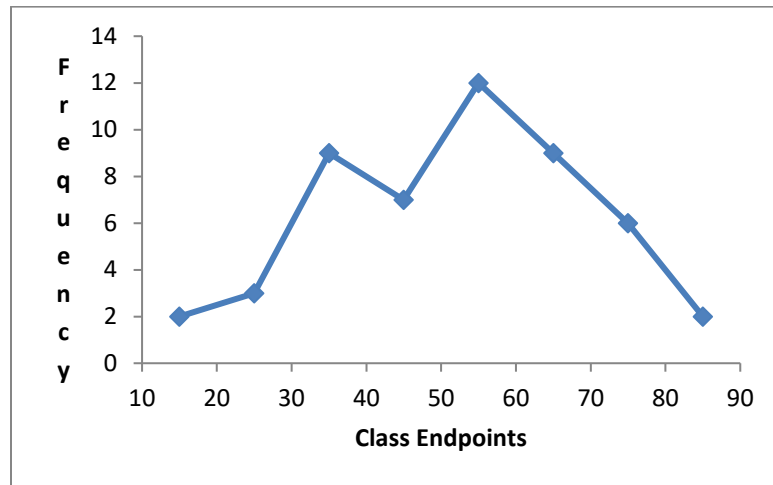
2.37 Frequency Distribution:

<u>Class Interval</u>	<u>Frequency</u>
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	<u>2</u>
	50

Histogram:



## Frequency Polygon:

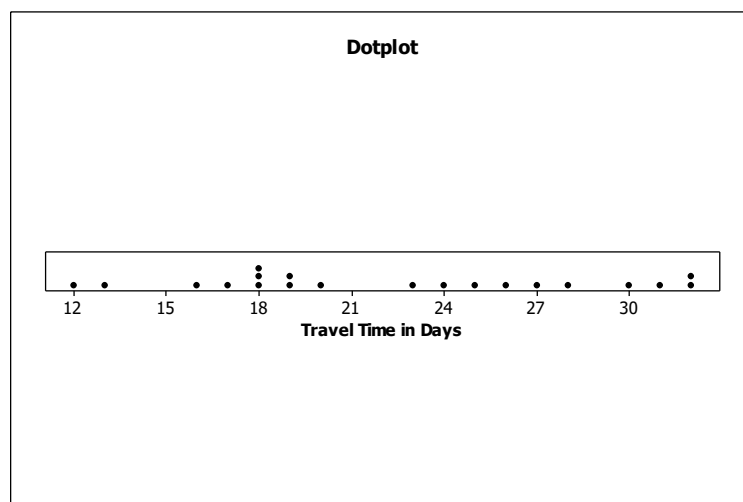


The normal distribution appears to peak near the center and diminish towards the end intervals.

## 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



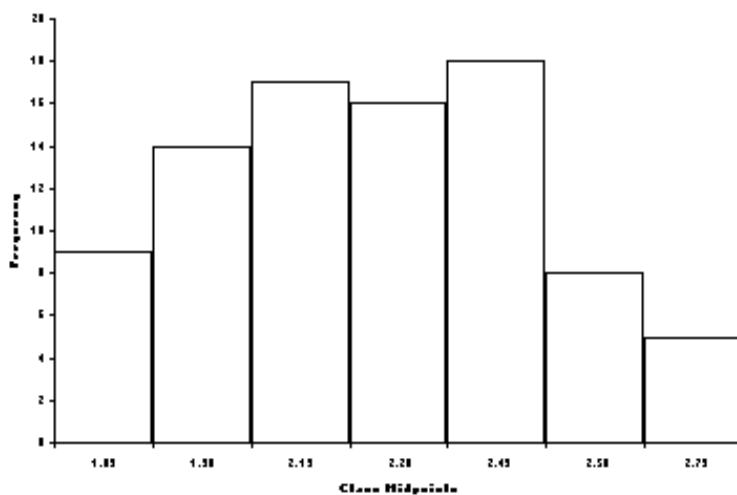
## c.) Comments:

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).

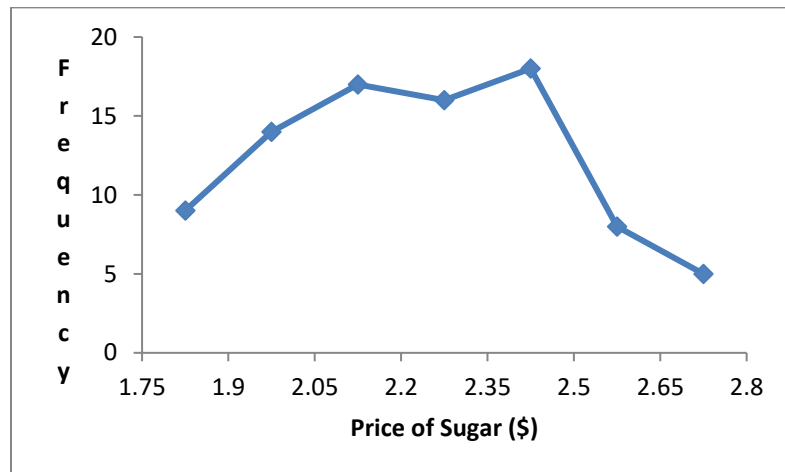
2.41

<u>Price</u>	<u>Frequency</u>	<u>Cumulative Frequency</u>
\$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	<u>5</u>	87
	87	

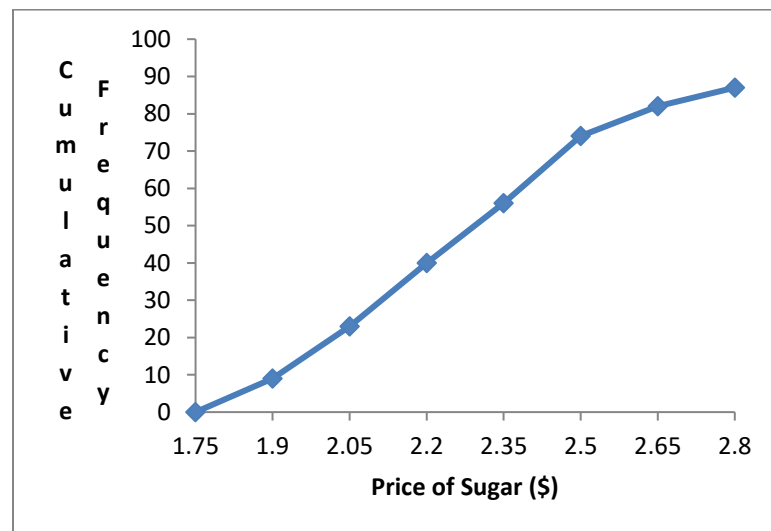
Histogram:



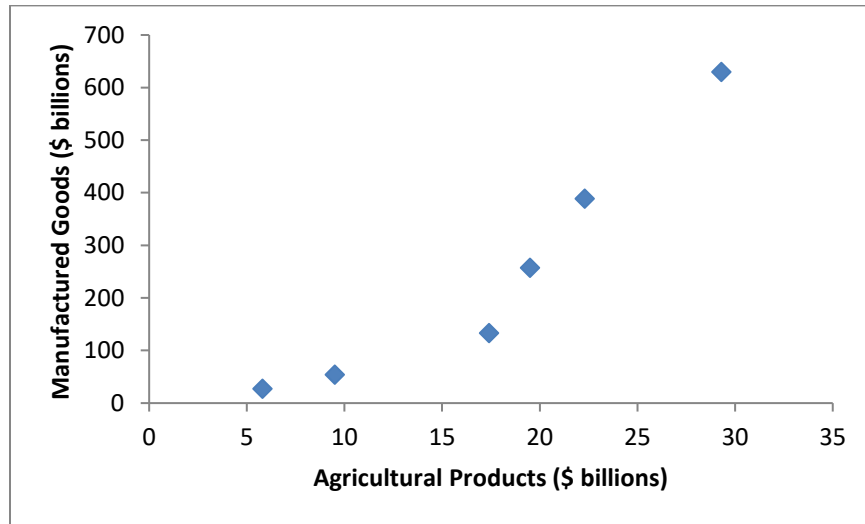
## Frequency Polygon:



## Ogive:

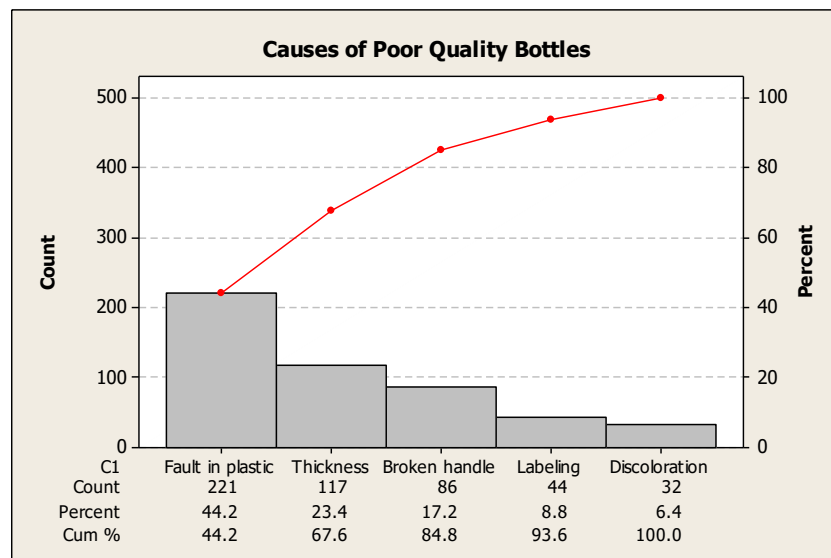


2.43



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.

2.45



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.

- 2.47 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.
- 2.49 Family practice is the most prevalent specialty with about 20% of physicians being in family practice and pediatrics next at slightly less than. A virtual tie exists between ob/gyn, general surgery, anesthesiology, and psychiatry at about 14% each.
- 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. The slope of the graph was steeper for lower values of the DJIA and for higher values of the DJIA. However, in the middle, when the DJIA was from about 8600 to about 10,500, the slope was considerably less indicating that over this interval as the DJIA rose, the NASDAQ-100 did not increase as fast as it did over other intervals.