CHAPTER 2: ORGANIZING AND VISUALIZING VARIABLES

SCENARIO 2-1

An insurance company evaluates many numerical variables about a person before deciding on an appropriate rate for automobile insurance. A representative from a local insurance agency selected a random sample of insured drivers and recorded, *X,* the number of claims each made in the last 3 years, with the following results.

X	f
1	14
2	18
3	12
4	5
5	1

1. Referring to Scenario 2-1, how many drivers are represented in the sample?
   a) 5
   b) 15
   c) 18
   d) 50

ANSWER:
d
TYPE: MC DIFFICULTY: Easy
KEYWORDS: frequency distribution

2. Referring to Scenario 2-1, how many total claims are represented in the sample?
   a) 15
   b) 50
   c) 111
   d) 250

ANSWER:
c
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: interpretation, frequency distribution

3. A type of vertical bar chart in which the categories are plotted in the descending rank order of the magnitude of their frequencies is called a
   a) contingency table.
   b) Pareto chart.
   c) stem-and-leaf display.
   d) pie chart.

ANSWER:
b
TYPE: MC DIFFICULTY: Easy
KEYWORDS: Pareto chart

Copyright ©2017 Pearson Education, Inc.

SCENARIO 2-2

At a meeting of information systems officers for regional offices of a national company, a survey was taken to determine the number of employees the officers supervise in the operation of their departments, where X is the number of employees overseen by each information systems officer.

X	f
1	7
2	5
3	11
4	8
5	9

4. Referring to Scenario 2-2, how many regional offices are represented in the survey results?
    a) 5
    b) 11
    c) 15
    d) 40

ANSWER:
d
TYPE: MC DIFFICULTY: Easy
KEYWORDS: interpretation, frequency distribution

5. Referring to Scenario 2-2, across all of the regional offices, how many total employees were supervised by those surveyed?
    a) 15
    b) 40
    c) 127
    d) 200

ANSWER:
c
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: interpretation, frequency distribution

6. The width of each bar in a histogram corresponds to the
    a) differences between the boundaries of the class.
    b) number of observations in each class.
    c) midpoint of each class.
   d) percentage of observations in each class.

ANSWER:
a
TYPE: MC DIFFICULTY: Easy
KEYWORDS: histogram

SCENARIO 2-3

Every spring semester, the School of Business coordinates a luncheon with local business leaders for graduating seniors, their families, and friends. Corporate sponsorship pays for the lunches of each of the seniors, but students have to purchase tickets to cover the cost of lunches served to guests they bring with them. The following histogram represents the attendance at the senior luncheon, where X is the number of guests each graduating senior invited to the luncheon and f is the number of graduating seniors in each category.
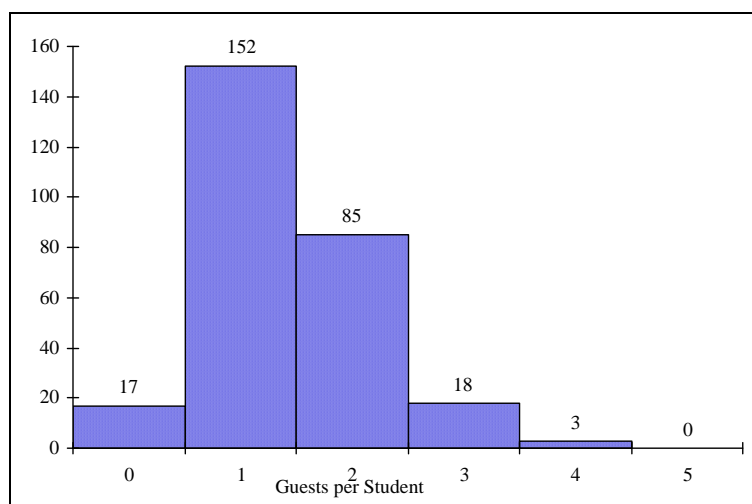
7. Referring to the histogram from Scenario 2-3, how many graduating seniors attended the luncheon?
    a) 4
    b) 152
    c) 275
    d) 388

ANSWER:
c
TYPE: MC DIFFICULTY: Difficult
EXPLANATION: The number of graduating seniors is the sum of all the frequencies, f.
KEYWORDS: interpretation, histogram

8. Referring to the histogram from Scenario 2-3, if all the tickets purchased were used, how many guests attended the luncheon?
    a) 4
   b) 152
    c) 275
    d) 388

ANSWER:
d
TYPE: MC DIFFICULTY: Difficult
EXPLANATION: The total number of guests is $\sum_{i=1}^{6} X_i f_i$
KEYWORDS: interpretation, histogram

9. A professor of economics at a small Texas university wanted to determine what year in school
    students were taking his tough economics course. Shown below is a pie chart of the results. What
    percentage of the class took the course prior to reaching their senior year?
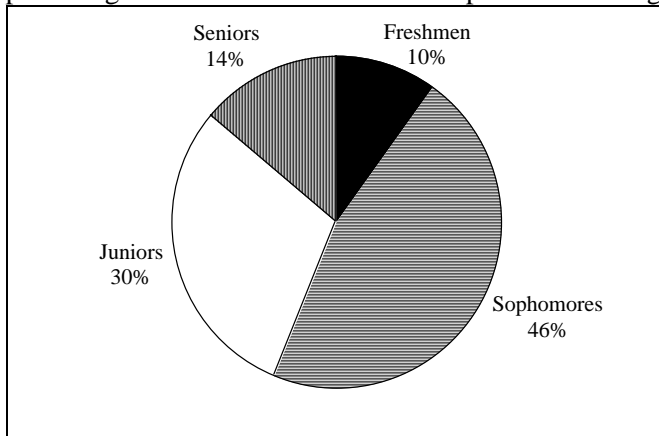
    a) 14%
    b) 44%
    c) 54%
    d) 86%

ANSWER:
d
TYPE: MC DIFFICULTY: Easy
KEYWORDS: interpretation, pie chart

10. When polygons or histograms are constructed, which axis must show the true zero or "origin"?
    a) The horizontal axis.
    b) The vertical axis.
    c) Both the horizontal and vertical axes.
    d) Neither the horizontal nor the vertical axis.

ANSWER:
b
TYPE: MC DIFFICULTY: Easy
KEYWORDS: polygon, histogram

11. When constructing charts, the following is plotted at the class midpoints:
    a) frequency histograms.
    b) percentage polygons.
    c) cumulative percentage polygon (ogives).
    d) All of the above.

ANSWER:
b
TYPE: MC DIFFICULTY: Easy
KEYWORDS: percentage polygon

SCENARIO 2-4

A survey was conducted to determine how people rated the quality of programming available on television. Respondents were asked to rate the overall quality from 0 (no quality at all) to 100 (extremely good quality). The stem-and-leaf display of the data is shown below.

Stem	Leaves
3	24
4	03478999
5	0112345
6	12566
7	01
8	
9	2

12. Referring to Scenario 2-4, what percentage of the respondents rated overall television quality with a rating of 80 or above?
    a) 0
    b) 4
    c) 96
    d) 100

ANSWER:
b
TYPE: MC DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display, interpretation

13. Referring to Scenario 2-4, what percentage of the respondents rated overall television quality with a rating of 50 or below?
    a) 11
    b) 40
    c) 44
    d) 56

ANSWER:
c
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, interpretation

14. Referring to Scenario 2-4, what percentage of the respondents rated overall television quality with a rating from 50 through 75?
    a) 11
    b) 40
    c) 44
    d) 56

ANSWER:
d
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, interpretation

SCENARIO 2-5
The following are the duration in minutes of a sample of long-distance phone calls made within the continental United States reported by one long-distance carrier.

Time (in Minutes)	Relative Frequency
0 but less than 5	0.37
5 but less than 10	0.22
10 but less than 15	0.15
15 but less than 20	0.10
20 but less than 25	0.07
25 but less than 30	0.07
30 or more	0.02

15. Referring to Scenario 2-5, what is the width of each class?
    a) 1 minute
    b) 5 minutes
    c) 2%
    d) 100%

ANSWER:
b
TYPE: MC DIFFICULTY: Easy
KEYWORDS: class interval, relative frequency distribution

16. Referring to Scenario 2-5, if 1,000 calls were randomly sampled, how many calls lasted under 10 minutes?
    a. 220
    b. 370
    c. 410
    d. 590

ANSWER:
d
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: relative frequency distribution, interpretation

17. Referring to Scenario 2-5, if 100 calls were randomly sampled, how many calls lasted 15 minutes or longer?
    a. 10
    b. 14
    c. 26
    d. 74

ANSWER:
c
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: relative frequency distribution, interpretation

18. Referring to Scenario 2-5, if 10 calls lasted 30 minutes or more, how many calls lasted less than 5 minutes?
    a) 10
    b) 185
    c) 295
    d) 500

ANSWER:
b
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: relative frequency distribution, interpretation

19. Referring to Scenario 2-5, what is the cumulative relative frequency for the percentage of calls that lasted under 20 minutes?
    a) 0.10
    b) 0.59
    c) 0.76
    d) 0.84

ANSWER:
d
TYPE: MC DIFFICULTY: Easy
KEYWORDS: cumulative relative frequency

20. Referring to Scenario 2-5, what is the cumulative relative frequency for the percentage of calls that lasted 10 minutes or more?
    a) 0.16
    b) 0.24
    c) 0.41
    d) 0.90

ANSWER:
c
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: cumulative relative frequency

21. Referring to Scenario 2-5, if 100 calls were randomly sampled, _____ of them would have lasted at least 15 minutes but less than 20 minutes
    a) 6
    b) 8
    c) 10
    d) 16

ANSWER:
c
TYPE: MC DIFFICULTY: Easy
KEYWORDS: relative frequency distribution, interpretation

22. Referring to Scenario 2-5, if 100 calls were sampled, _____ of them would have lasted less than 15 minutes.
    a)   26
    b)   74
    c)   10
    d)   None of the above.

ANSWER:
b
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: relative frequency distribution, interpretation

23. Referring to Scenario 2-5, if 100 calls were sampled, _____of them would have lasted 20 minutes or more.
    a)   26
    b)   16
    c)   74
    d)   None of the above.

ANSWER:
b
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: relative frequency distribution, interpretation

24. Referring to Scenario 2-5, if 100 calls were sampled, _____ of them would have lasted less than 5 minutes or at least 30 minutes or more.
    a)   35
    b)   37
    c)   39
    d)   None of the above.

ANSWER:
c
TYPE: MC DIFFICULTY: Difficult
KEYWORDS: relative frequency distribution, interpretation

25. Which of the following is appropriate for displaying data collected on the different brands of cars students at a major university drive?

    a)   A Pareto chart
    b)   A two-way classification table
    c)   A histogram
    d)   A scatter plot

ANSWER:
a
TYPE: MC DIFFICULTY: Easy
KEYWORDS: Pareto diagram

26. One of the developing countries is experiencing a baby boom, with the number of births rising for the fifth year in a row, according to a BBC News report. Which of the following is best for displaying this data?
   a) A Pareto chart
   b) A two-way classification table
   c) A histogram
   d) A time-series plot

ANSWER:
d
TYPE: MC DIFFICULTY: Easy
KEYWORDS: time-series plot

27. When studying the simultaneous responses to two categorical questions, you should set up a
   a) contingency table.
   b) frequency distribution table.
   c) cumulative percentage distribution table.
   d) histogram.

ANSWER:
a
TYPE: MC DIFFICULTY: Easy
KEYWORDS: contingency table

28. Data on 1,500 students' height were collected at a larger university in the East Coast. Which of the following is the best chart for presenting the information?
   a) A pie chart.
   b) A Pareto chart.
   c) A side-by-side bar chart.
   d) A histogram.

ANSWER:
d
TYPE: MC DIFFICULTY: Easy
KEYWORDS: choice of chart, histogram

29. Data on the number of part-time hours students at a public university worked in a week were collected. Which of the following is the best chart for presenting the information?
   a) A pie chart.
   b) A Pareto chart.
   c) A percentage table.
   d) A percentage polygon.

ANSWER:
d
TYPE: MC DIFFICULTY: Easy
KEYWORDS: choice of chart, percentage polygon

30. Data on the number of credit hours of 20,000 students at a public university enrolled in a Spring semester were collected. Which of the following is the best for presenting the information?
    a) A pie chart.
    b) A Pareto chart.
    c) A stem-and-leaf display.
    d) A contingency table.

ANSWER:
c
TYPE: MC DIFFICULTY: Easy
KEYWORDS: choice of chart, stem-and-leaf

31.   A survey of 150 executives were asked what they think is the most common mistake candidates make during job interviews. Six different mistakes were given. Which of the following is the best for presenting the information?
    a) A bar chart.
    b) A histogram
    c) A stem-and-leaf display.
    d) A contingency table.

ANSWER:
a
TYPE: MC DIFFICULTY: Easy
KEYWORDS: choice of chart, bar chart

32. You have collected information on the market share of 5 different search engines used by U.S. Internet users in a particular quarter. Which of the following is the best for presenting the information?
    a) A pie chart.
    b) A histogram
    c) A stem-and-leaf display.
    d) A contingency table.

ANSWER:
a
TYPE: MC DIFFICULTY: Easy
KEYWORDS: choice of chart, pie chart

33. You have collected information on the consumption by the 15 largest coffee-consuming nations. Which of the following is the best for presenting the shares of the consumption?
   a) A pie chart.
   b) A Pareto chart
   c) A side-by-side bar chart.
   d) A contingency table.

ANSWER:
b
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: choice of chart, Pareto chart
NOTE: Even though a pie chart can also be used, the Pareto chart is preferable for separating the "vital few" from the "trivial many".

34. You have collected data on the approximate retail price (in $) and the energy cost per year (in $) of 15 refrigerators. Which of the following is the best for presenting the data?
   a) A pie chart.
   b) A scatter plot
   c) A side-by-side bar chart.
   d) A contingency table.

ANSWER:
b
TYPE: MC DIFFICULTY: Easy
KEYWORDS: choice of chart, scatter plot

35. You have collected data on the number of U.S. households actively using online banking and/or online bill payment over a 10-year period. Which of the following is the best for presenting the data?
   a) A pie chart.
   b) A stem-and-leaf display
   c) A side-by-side bar chart.
   d) A time-series plot.

ANSWER:
d
TYPE: MC DIFFICULTY: Easy
KEYWORDS: choice of chart, time-series plot

36. You have collected data on the monthly seasonally adjusted civilian unemployment rate for the United States over a 10-year period. Which of the following is the best for presenting the data?
   a) A contingency table.
   b) A stem-and-leaf display
   c) A time-series plot.
   d) A side-by-side bar chart.

ANSWER:
c
TYPE: MC DIFFICULTY: Easy
KEYWORDS: choice of chart, time-series plot

2-12   Organizing and Visualizing Variables

37. You have collected data on the number of complaints for 6 different brands of automobiles sold in the US over a 10-year period. Which of the following is the best for presenting the data?
    a) A contingency table.
    b) A stem-and-leaf display
    c) A time-series plot.
    d) A side-by-side bar chart.

ANSWER:
d
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: choice of chart, side-by-side bar chart

38. You have collected data on the responses to two questions asked in a survey of 40 college students majoring in business—What is your gender (Male = M; Female = F) and What is your major (Accountancy = A; Computer Information Systems = C; Marketing = M). Which of the following is the best for presenting the data?
    a) A contingency table.
    b) A stem-and-leaf display
    c) A time-series plot.
    d) A Pareto chart.

ANSWER:
a
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: choice of chart, contingency table

SCENARIO 2-6
A sample of 200 students at a Big-Ten university was taken after the midterm to ask them whether they went bar hopping the weekend before the midterm or spent the weekend studying, and whether they did well or poorly on the midterm. The following table contains the result.

	Did Well in Midterm	Did Poorly in Midterm
Studying for Exam	80	20
Went Bar Hopping	30	70

39. Referring to Scenario 2-6, of those who went bar hopping the weekend before the midterm in the sample, _____ percent of them did well on the midterm.
    a) 15
    b) 27.27
    c) 30
    d) 55

ANSWER:
c
TYPE: MC DIFFICULTY: Easy
KEYWORDS: contingency table, interpretation

Copyright ©2017 Pearson Education, Inc.

40. Referring to Scenario 2-6, of those who did well on the midterm in the sample, _____ percent of them went bar hopping the weekend before the midterm.
   a) 15
   b) 27.27
   c) 30
   d) 50

ANSWER:
b
TYPE: MC DIFFICULTY: Easy
KEYWORDS: contingency table, interpretation

41. Referring to Scenario 2-6, _____ percent of the students in the sample went bar hopping the weekend before the midterm and did well on the midterm.
   a) 15
   b) 27.27
   c) 30
   d) 50

ANSWER:
a
TYPE: MC DIFFICULTY: Easy
KEYWORDS: contingency table, interpretation

42. Referring to Scenario 2-6, _____ percent of the students in the sample spent the weekend studying and did well on the midterm.
   a) 40
   b) 50
   c) 72.72
   d) 80

ANSWER:
a
TYPE: MC DIFFICULTY: Easy
KEYWORDS: contingency table, interpretation

43. Referring to Scenario 2-6, if the sample is a good representation of the population, we can expect _____ percent of the students in the population to spend the weekend studying and do poorly on the midterm.
   a) 10
   b) 20
   c) 45
   d) 50

ANSWER:
a
TYPE: MC DIFFICULTY: Easy
KEYWORDS: contingency table, interpretation

44. Referring to Scenario 2-6, if the sample is a good representation of the population, we can expect
_____ percent of those who spent the weekend studying to do poorly on the midterm.
   a) 10
   b) 20
   c) 45
   d) 50

ANSWER:
b
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: contingency table, interpretation

45. Referring to Scenario 2-6, if the sample is a good representation of the population, we can expect
_____ percent of those who did poorly on the midterm to have spent the weekend studying.
   a) 10
   b) 22.22
   c) 45
   d) 50

ANSWER:
b
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: contingency table, interpretation

46. In a contingency table, the number of rows and columns
   a) must always be the same.
   b) must always be 2.
   c) must add to 100%.
   d) None of the above.

ANSWER:
d
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: contingency table

47. Retailers are always interested in determining why a customer selected their store to make a purchase. A sporting goods retailer conducted a customer survey to determine why its customers shopped at the store. The results are shown in the bar chart below. What proportion of the customers responded that they shopped at the store because of the merchandise or the convenience?
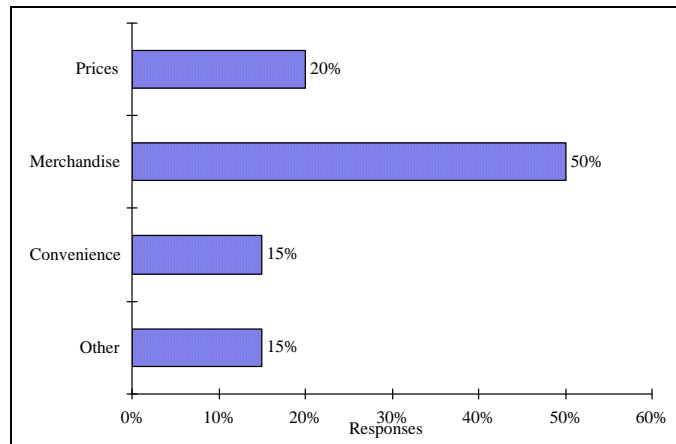
    a)   35%
    b)   50%
    c)   65%
    d)   85%

ANSWER:
c
TYPE: MC DIFFICULTY: Easy
KEYWORDS: bar chart, interpretation

SCENARIO 2-7

The Stem-and-Leaf display below contains data on the number of months between the date a civil suit is filed and when the case is actually adjudicated for 50 cases heard in superior court.

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

48. Referring to Scenario 2-7, locate the first leaf, i.e., the lowest valued leaf with the lowest valued stem. This represents a wait of _____ months.

ANSWER:
12
TYPE: FI DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display, interpretation

49. Referring to Scenario 2-7, the civil suit with the longest wait between when the suit was filed and when it was adjudicated had a wait of _____ months.

ANSWER:
68
TYPE: FI DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display, interpretation

50. Referring to Scenario 2-7, the civil suit with the fourth shortest waiting time between when the suit was filed and when it was adjudicated had a wait of _____ months.

ANSWER:
14
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, interpretation

51. Referring to Scenario 2-7, _____ percent of the cases were adjudicated within the first 2 years.

ANSWER:
30
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, interpretation

52. Referring to Scenario 2-7, _____ percent of the cases were not adjudicated within the first 4 years.

ANSWER:
20
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, interpretation

53. Referring to Scenario 2-7, if a frequency distribution with equal sized classes was made from this data, and the first class was "10 but less than 20," the frequency of that class would be _____.

ANSWER:
9
TYPE: FI DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display, interpretation

54.   Referring to Scenario 2-7, if a frequency distribution with equal sized classes was made from this data, and the first class was "10 but less than 20," the relative frequency of the third class would be _____.

ANSWER:
0.20 or 20% or 10/50
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: relative frequency distribution

55. Referring to Scenario 2-7, if a frequency distribution with equal sized classes was made from this data, and the first class was "10 but less than 20," the cumulative percentage of the second class would be _____.

ANSWER:
46% or 0.46 or 23/50
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: cumulative percentage distribution

SCENARIO 2-8

The Stem-and-Leaf display represents the number of times in a year that a random sample of 100 "lifetime" members of a health club actually visited the facility.

Stem	Leaves
0	0122222333333445666666667789999
1	1111222234444455669999
2	00011223455556889
3	0000446799
4	011345567
5	0077
6	8
7	67
8	3
9	0247

56. Referring to Scenario 2-8, the person who has the largest leaf associated with the smallest stem visited the facility _____ times.

ANSWER:
9
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, interpretation

57. Referring to Scenario 2-8, the person who visited the health club less than anyone else in the sample visited the facility _____ times.

ANSWER:
0 or no
TYPE: FI DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display, interpretation

58. Referring to Scenario 2-8, the person who visited the health club more than anyone else in the sample visited the facility _____ times.

ANSWER:
97
TYPE: FI DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display, interpretation

59. Referring to Scenario 2-8, _____ of the 100 members visited the health club at least 52 times in a year.

ANSWER:
10
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, interpretation

60. Referring to Scenario 2-8, _____ of the 100 members visited the health club no more than 12 times in a year.

ANSWER:
38
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, interpretation

61. Referring to Scenario 2-8, if a frequency distribution with equal sized classes was made from this data, and the first class was "0 but less than 10," the frequency of the fifth class would be _____.

ANSWER:
9
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, frequency distribution

62. Referring to Scenario 2-8, if a frequency distribution with equal sized classes was made from this data, and the first class was "0 but less than 10," the relative frequency of the last class would be _____.

ANSWER:
4% or 0.04 or 4/100
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, relative frequency distribution

63. Referring to Scenario 2-8, if a frequency distribution with equal sized classes was made from this data, and the first class was "0 but less than 10," the cumulative percentage of the next-to-last class would be _____.

ANSWER:
96% or 0.96 or 96/100
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, cumulative percentage distribution

64. Referring to Scenario 2-8, if a frequency distribution with equal sized classes was made from this data, and the first class was "0 but less than 10," the class midpoint of the third class would be _____.

ANSWER:
25 or (20+30)/2
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: stem-and-leaf display, class midpoint

SCENARIO 2-9

The frequency distribution below represents the rents of 250 randomly selected federally subsidized apartments in a small town.

Rent in $	Frequency
1,100 but less than 1,200	113
1,200 but less than 1,300	85
1,300 but less than 1,400	32
1,400 but less than 1,500	16
1,500 but less than 1,600	4

65. Referring to Scenario 2-9, _____ apartments rented for at least $1,200 but less than $1,400.

ANSWER:
117
TYPE: FI DIFFICULTY: Easy
KEYWORDS: frequency distribution

66. Referring to Scenario 2-9, _____ percent of the apartments rented for $1,400 or more.

ANSWER:
8% or 20/250
TYPE: FI DIFFICULTY: Easy
KEYWORDS: frequency distribution, cumulative percentage distribution

67. Referring to Scenario 2-9, _____ percent of the apartments rented for at least $1,300.

ANSWER:
20.8% or 52/250
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: frequency distribution, cumulative percentage distribution

68. Referring to Scenario 2-9, the class midpoint of the second class is _____.

ANSWER:
1,250
TYPE: FI DIFFICULTY: Easy
KEYWORDS: frequency distribution, class midpoint

69. Referring to Scenario 2-9, the relative frequency of the second class is _____.

ANSWER:
85/250 or 17/50 or 34% or 0.34
TYPE: FI DIFFICULTY: Easy
KEYWORDS: frequency distribution, relative frequency distribution

70. Referring to Scenario 2-9, the percentage of apartments renting for less than $1,400 is _____.

ANSWER:
230/250 or 23/25 or 92% or 0.92
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: frequency distribution, cumulative percentage distribution

SCENARIO 2-10

The histogram below represents scores achieved by 200 job applicants on a personality profile.
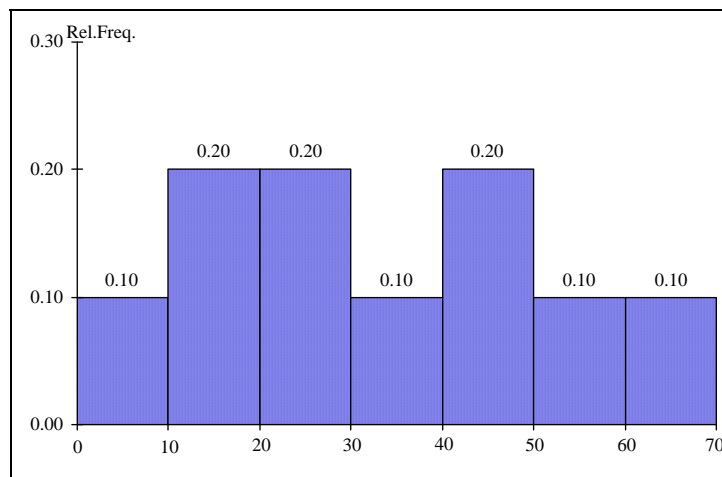
71. Referring to the histogram from Scenario 2-10, _____ percent of the job applicants scored between 10 and 20.

ANSWER:
20%
TYPE: FI DIFFICULTY: Easy
KEYWORDS: histogram, percentage distribution

72. Referring to the histogram from Scenario 2-10, _____ percent of the job applicants scored below 50.

ANSWER:
80%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: histogram, percentage distribution

73. Referring to the histogram from Scenario 2-10, the number of job applicants who scored between 30 and below 60 is _____.

ANSWER:
80
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: histogram

74. Referring to the histogram from Scenario 2-10, the number of job applicants who scored 50 or above is _____.

ANSWER:
40
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: histogram

75. Referring to the histogram from Scenario 2-10, 90% of the job applicants scored above or equal to _____.

ANSWER:
10
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: histogram, cumulative percentage distribution

76. Referring to the histogram from Scenario 2-10, half of the job applicants scored below _____.

ANSWER:
30
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: histogram, cumulative percentage distribution

77. Referring to the histogram from Scenario 2-10, _____ percent of the applicants scored below 20 or at least 50.

ANSWER:
50%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: histogram, cumulative percentage distribution

78. Referring to the histogram from Scenario 2-10, _____ percent of the applicants scored between 20 and below 50.

ANSWER:
50%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: histogram, cumulative percentage distribution

SCENARIO 2-11

The ordered array below resulted from selecting a sample of 25 batches of 500 computer chips and determining how many in each batch were defective.

Defects
1 2 4 4 5 5 6 7 9 9 12 12 15
17 20 21 23 23 25 26 27 27 28 29 29

79. Referring to Scenario 2-11, if a frequency distribution for the defects data is constructed, using "0 but less than 5" as the first class, the frequency of the "20 but less than 25" class would be _____.

ANSWER:
4
TYPE: FI DIFFICULTY: Easy
KEYWORDS: frequency distribution

80. Referring to Scenario 2-11, if a frequency distribution for the defects data is constructed, using "0 but less than 5" as the first class, the relative frequency of the "15 but less than 20" class would be _____.

ANSWER:
0.08 or 8% or 2/25
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: relative frequency distribution

81. Referring to Scenario 2-11, construct a frequency distribution for the defects data, using "0 but less than 5" as the first class.
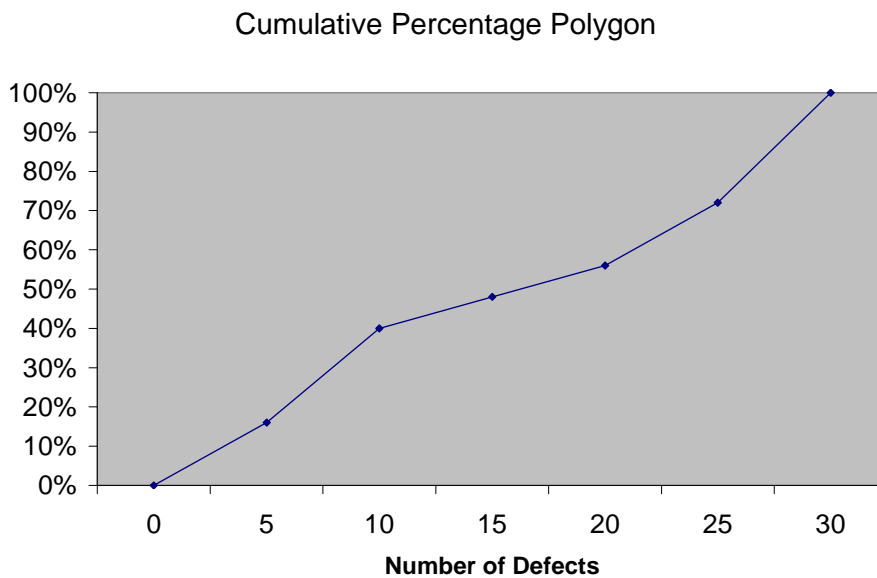
ANSWER:

Defects	Frequency
0 but less than 5	4
5 but less than 10	6
10 but less than 15	2
15 but less than 20	2
20 but less than 25	4
25 but less than 30	7

TYPE: PR DIFFICULTY: Easy
KEYWORDS: frequency distribution

82. Referring to Scenario 2-11, construct a relative frequency or percentage distribution for the defects data, using "0 but less than 5" as the first class.

ANSWER:

Defects	Percentage
0 but less than 5	16
5 but less than 10	24
10 but less than 15	8
15 but less than 20	8
20 but less than 25	16
25 but less than 30	28

TYPE: PR DIFFICULTY: Moderate
KEYWORDS: relative frequency distribution, percentage distribution

83. Referring to Scenario 2-11, construct a cumulative percentage distribution for the defects data if the corresponding frequency distribution uses "0 but less than 5" as the first class.

ANSWER:

Defects	CumPct
0	0
5	16
10	40
15	48
20	56
25	72
30	100

TYPE: PR DIFFICULTY: Moderate
KEYWORDS: cumulative percentage distribution

84. Referring to Scenario 2-11, construct a histogram for the defects data, using "0 but less than 5" as the first class.

ANSWER:
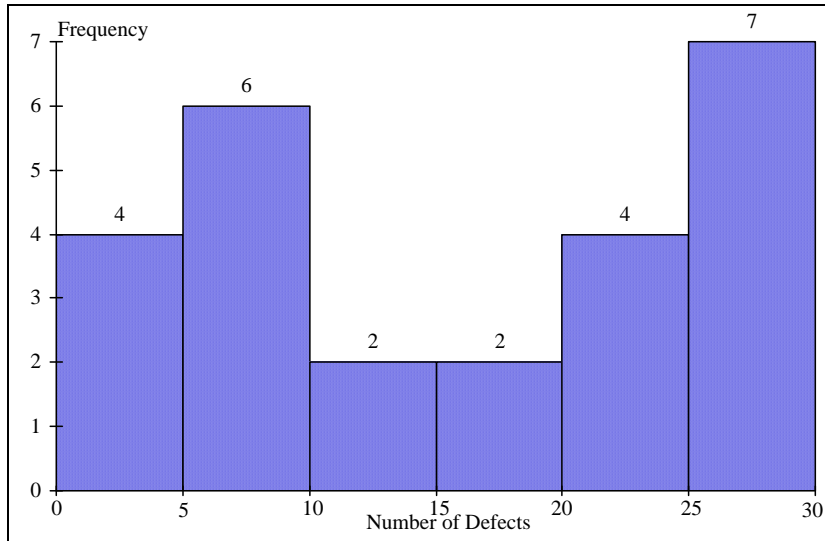
TYPE: PR DIFFICULTY: Easy
KEYWORDS: histogram, frequency distribution

85. Referring to Scenario 2-11, construct a cumulative percentage polygon for the defects data if the corresponding frequency distribution uses "0 but less than 5" as the first class.

ANSWER:

TYPE: PR DIFFICULTY: Moderate
KEYWORDS: cumulative percentage polygon

86. The point halfway between the boundaries of each class interval in a grouped frequency distribution is called the _____.

ANSWER:
class midpoint
TYPE: FI DIFFICULTY: Easy
KEYWORDS: cumulative percentage polygon, frequency distribution

87. A _____ is a vertical bar chart in which the rectangular bars are constructed at the boundaries of each class interval.

ANSWER:
histogram
TYPE: FI DIFFICULTY: Easy
KEYWORDS: histogram

88. It is essential that each class grouping or interval in a frequency distribution be _____ and _____.

ANSWER:
non-overlapping and of equal width
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: frequency distribution, class interval

89. In order to compare one large set of numerical data to another, a _____ distribution must be developed from the frequency distribution.

ANSWER:
relative frequency or percentage
TYPE: FI DIFFICULTY: Easy
KEYWORDS: relative frequency distribution, percentage distribution

90. When comparing two or more large sets of numerical data, the distributions being developed should use the same _____.

ANSWER:
class boundaries.
TYPE: FI DIFFICULTY: Easy
KEYWORDS: class boundaries

91. The width of each class grouping or interval in a frequency distribution should be _____.

ANSWER:
the same or equal
TYPE: FI DIFFICULTY: Easy
KEYWORDS: class interval, frequency distribution

92. In constructing a polygon, each class grouping is represented by its _____ and then these are consecutively connected to one another.

ANSWER:
midpoint
TYPE: FI DIFFICULTY: Easy
KEYWORDS: polygon, class interval, midpoint

93. A _____ is a summary table in which numerical data are tallied into class intervals or categories.

ANSWER:
frequency distribution
TYPE: FI DIFFICULTY: Easy
KEYWORDS: frequency distribution, class interval

94. True or False: In general, grouped frequency distributions should have between 5 and 15 class intervals.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: frequency distribution, number of classes

95. True or False: The sum of relative frequencies in a distribution always equals 1.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: relative frequency

96. True or False: The sum of cumulative frequencies in a distribution always equals 1.

ANSWER:
False
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: cumulative distribution

97. True or False: In graphing two categorical data, the side-by-side bar chart is best suited when comparing joint responses.

ANSWER:
True
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: side-by-side bar chart

98. True or False: When constructing a frequency distribution, classes should be selected so that they are of equal width.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: frequency distribution

99. True or False: A research analyst was directed to arrange raw data collected on the yield of wheat, ranging from 40 to 93 bushels per acre, in a frequency distribution. He should choose 30 as the class interval width.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: frequency distribution, class interval

100.   True or False: If the values of the seventh and eighth class in a cumulative percentage distribution are the same, we know that there are no observations in the eighth class.

ANSWER:
True
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: cumulative percentage distribution

101.   True or False: One of the advantages of a pie chart is that it clearly shows that the total of all the categories of the pie adds to 100%.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: pie chart

102.   True or False: The larger the number of observations in a numerical data set, the larger the number of class intervals needed for a grouped frequency distribution.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: class interval, frequency distribution

103.   True or False: Determining the class boundaries of a frequency distribution is highly subjective.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: class boundaries, frequency distribution

104.   True or False: The original data values cannot be determined once they are grouped into a frequency distribution table.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: frequency distribution

105.   True or False: The percentage distribution cannot be constructed from the frequency distribution directly.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: percentage distribution, frequency distribution

106.   True or False: The stem-and-leaf display is often superior to the frequency distribution in that it maintains the original values for further analysis.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display, frequency distribution

107.   True or False: The relative frequency is the frequency in each class divided by the total number of observations.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: relative frequency distribution

108.   True or False: Ogives are plotted at the midpoints of the class groupings.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: ogives, midpoint

109.   True or False: Percentage polygons are plotted at the boundaries of the class groupings.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: percentage polygons

110. True or False: The main principle behind the Pareto chart is the ability to separate the "vital few" from the "trivial many."

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: Pareto chart

111. True or False: A histogram can have gaps between the bars, whereas bar charts cannot have gaps.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: histogram, bar chart

112. True or False: Histograms are used for numerical data while bar charts are suitable for categorical data.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: histogram, bar chart

113. True or False: A Walmart store in a small town monitors customer complaints and organizes these complaints into six distinct categories. Over the past year, suppose the company has received 534 complaints. One possible graphical method for representing these data would be a Pareto chart.

ANSWER:
True
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: Pareto chart

114. True or False: Apple Computer, Inc. collected information on the age of their customers. Suppose the youngest customer was 12 and the oldest was 72. To study the distribution of the age among its customers, it can use a Pareto chart.

ANSWER:
False
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: Pareto chart

115.   True or False: Apple Computer, Inc. collected information on the age of their customers. Suppose the youngest customer was 12 and the oldest was 72. To study the distribution of the age among its customers, it is best to use a pie chart.

ANSWER:
False
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: pie chart

116.   True or False: Apple Computer, Inc. collected information on the age of their customers. Suppose the youngest customer was 12 and the oldest was 72. To study the distribution of the age among its customers, it can use a percentage polygon.

ANSWER:
True
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: percentage polygon

117.   True or False: Apple Computer, Inc. collected information on the age of their customers. Suppose the youngest customer was 12 and the oldest was 72. To study the percentage of their customers who are below a certain age, it can use an ogive.

ANSWER:
True
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: ogive

118.   True or False: If you wish to construct a graph of a relative frequency distribution, you would most likely construct an ogive first.

ANSWER:
False
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: Ogive

119.   True or False: An ogive is a cumulative percentage polygon.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: Ogive, cumulative percentage polygon

120.   True or False: A side-by-side bar chart is two histograms plotted side-by-side.

ANSWER:
False
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: side-by-side bar chart

121. True or False: A good choice for the number of class groups to use in constructing frequency distribution is to have at least 5 but no more than 15 class groups.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: number of classes

122. True or False: In general, a frequency distribution should have at least 8 class groups but no more than 20.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: number of classes

123. True of False: To determine the width of class interval, divide the number of class groups by the range of the data.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: class interval

124. True or False: The percentage polygon is formed by having the lower boundary of each class represent the data in that class and then connecting the sequence of lower boundaries at their respective class percentages.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: percentage polygon

125. True or False: A polygon can be constructed from a bar chart.

ANSWER:
False
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: polygon

126. To evaluate two categorical variables at the same time, a _____ could be developed.

ANSWER:
contingency or cross-classification table or side-by-side bar chart
TYPE: FI DIFFICULTY: Easy
KEYWORDS: contingency table, cross-classification table

127. Relationships in a contingency table can be examined more fully if the frequencies are converted into _____ .

ANSWER:
percentages or proportions
TYPE: FI DIFFICULTY: Easy
KEYWORDS: contingency table

SCENARIO 2-12

The table below contains the opinions of a sample of 200 people broken down by gender about the latest congressional plan to eliminate anti-trust exemptions for professional baseball.

	For	Neutral	Against	Totals
Female	38	54	12	104
Male	12	36	48	96
Totals	50	90	60	200

128. Referring to Scenario 2-12, construct a table of row percentages.

ANSWER:

	For	Neutral	Against	Totals
Female	36.54	51.92	11.54	100.00
Male	12.50	37.50	50.00	100.00
Totals	25.00	45.00	30.00	100.00

TYPE: PR DIFFICULTY: Easy
KEYWORDS: row percentages

129. Referring to Scenario 2-12, construct a table of column percentages.

ANSWER:

	For	Neutral	Against	Totals
Female	76.00	60.00	20.00	52.00
Male	24.00	40.00	80.00	48.00
Totals	100.00	100.00	100.00	100.00

TYPE: PR DIFFICULTY: Easy
KEYWORDS: column percentages

130. Referring to Scenario 2-12, construct a table of total percentages.

ANSWER:

	For	Neutral	Against	Totals
Female	19.00	27.00	6.00	52.00
Male	6.00	18.00	24.00	48.00
Totals	25.00	45.00	30.00	100.00

TYPE: PR DIFFICULTY: Easy
KEYWORDS: total percentages

131. Referring to Scenario 2-12, of those for the plan in the sample, _____ percent were females.

ANSWER:
76%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

132. Referring to Scenario 2-12, of those neutral in the sample, _____ percent were males.

ANSWER:
40%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

133. Referring to Scenario 2-12, of the males in the sample, _____ percent were for the plan.

ANSWER:
12.50%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table

134. Referring to Scenario 2-12, of the females in the sample, _____ percent were against the plan.

ANSWER:
11.54%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table

135. Referring to Scenario 2-12, of the females in the sample, _____ percent were either neutral or against the plan.

ANSWER:
63.46% or (51.92+11.54)%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table

136. Referring to Scenario 2-12, _____ percent of the 200 were females who were against the plan.

ANSWER:
6%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table

137.   Referring to Scenario 2-12, _____ percent of the 200 were males who were neutral.

ANSWER:
18%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table

138.   Referring to Scenario 2-12, _____ percent of the 200 were females who were either neutral
       or against the plan.

ANSWER:
33%
TYPE: FI DIFFICULTY: Difficult
KEYWORDS: contingency table

139.   Referring to Scenario 2-12, _____ percent of the 200 were males who were not against the
       plan.

ANSWER:
24%
TYPE: FI DIFFICULTY: Difficult
KEYWORDS: contingency table

140.   Referring to Scenario 2-12, _____ percent of the 200 were not neutral.

ANSWER:
55%
TYPE: FI DIFFICULTY: Difficult
KEYWORDS: contingency table, row percentages

141.   Referring to Scenario 2-12, _____ percent of the 200 were against the plan.

ANSWER:
30%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, row percentages

142.   Referring to Scenario 2-12, _____ percent of the 200 were males.

ANSWER:
48%
TYPE: FI DIFFICULTY: Easy
KEYWORDS: contingency table, column percentages

143. Referring to Scenario 2-12, if the sample is a good representation of the population, we can expect _____ percent of the population will be for the plan.

ANSWER:
25%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, row percentages

144. Referring to Scenario 2-12, if the sample is a good representation of the population, we can expect _____ percent of the population will be males.

ANSWER:
48%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: column percentages, contingency table

145. Referring to Scenario 2-12, if the sample is a good representation of the population, we can expect _____ percent of those for the plan in the population will be males.

ANSWER:
24%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table

146. Referring to Scenario 2-12, if the sample is a good representation of the population, we can expect _____ percent of the males in the population will be against the plan.

ANSWER:
50%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table

147. Referring to Scenario 2-12, if the sample is a good representation of the population, we can expect _____ percent of the females in the population will not be against the plan.

ANSWER:
88.46% or (36.54+51.92)
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table

SCENARIO 2-13

Given below is the stem-and-leaf display representing the amount of detergent used in gallons (with leaves in 10ths of gallons) in a day by 25 drive-through car wash operations in Phoenix.

```
 9 | 147
10 | 02238
11 | 135566777
12 | 223489
13 | 02
```

148.   Referring to Scenario 2-13, if a frequency distribution for the amount of detergent used is constructed, using "9.0 but less than 10.0 gallons" as the first class, the frequency of the "11.0 but less than 12.0 gallons" class would be _____.

ANSWER:
9
TYPE: FI DIFFICULTY: Easy
KEYWORDS: frequency distribution

149.   Referring to Scenario 2-13, if a percentage histogram for the detergent data is constructed, using "9.0 but less than 10.0 gallons" as the first class, the percentage of drive-through car wash operations that use "12.0 but less than 13.0 gallons" of detergent would be _____.

ANSWER:
24%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: relative frequency distribution, percentage distribution

150.   Referring to Scenario 2-13, if a percentage histogram for the detergent data is constructed, using "9.0 but less than 10.0 gallons" as the first class, what percentage of drive-through car wash operations use less than 12 gallons of detergent in a day?

ANSWER:
68%
TYPE: FI DIFFICULTY: Easy
KEYWORDS: percentage distribution, cumulative relative frequency

151.   Referring to Scenario 2-13, if a relative frequency or percentage distribution for the detergent data is constructed, using "9.0 but less than 10.0 gallons" as the first class, what percentage of drive-through car wash operations use at least 10 gallons of detergent in a day?

ANSWER:
88%
TYPE: FI DIFFICULTY: Easy
KEYWORDS: relative frequency distribution, percentage distribution

152.   Referring to Scenario 2-13, if a relative frequency or percentage distribution for the detergent data is constructed, using "9.0 but less than 10.0 gallons" as the first class, what percentage of drive-through car wash operations use at least 10 gallons but less than 13 gallons of detergent in a day?

ANSWER:
80%
TYPE: FI DIFFICULTY: Easy
KEYWORDS: relative frequency distribution, percentage distribution

153. Referring to Scenario 2-13, construct a frequency distribution for the detergent data, using "9.0 but less than 10.0 gallons" as the first class.

ANSWER:

Purchases (gals)	Frequency
9.0 but less than 10.0	3
10.0 but less than 11.0	5
11.0 but less than 12.0	9
12.0 but less than 13.0	6
13.0 but less than 14.0	2

TYPE: PR DIFFICULTY: Moderate
KEYWORDS: frequency distribution

154. Referring to Scenario 2-13, construct a relative frequency or percentage distribution for the detergent data, using "9.0 but less than 10.0" as the first class.

ANSWER:
Gasoline

Purchases (gals)	Percentage
9.0 but less than 10.0	12%
10.0 but less than 11.0	20
11.0 but less than 12.0	36
12.0 but less than 13.0	24
13.0 but less than 14.0	8

TYPE: PR DIFFICULTY: Moderate
KEYWORDS: relative frequency distribution, percentage distribution

155. Referring to Scenario 2-13, construct a cumulative percentage distribution for the detergent data if the corresponding frequency distribution uses "9.0 but less than 10.0" as the first class.

ANSWER:

Gasoline Purchases (gals)	Frequency Less Than	Percentage Less Than
9.0 but less than 10.0	3	12
10.0 but less than 11.0	8	32
11.0 but less than 12.0	17	68
12.0 but less than 13.0	23	92
13.0 but less than 14.0	25	100

TYPE: PR DIFFICULTY: Moderate
KEYWORDS: cumulative percentage distribution

156. Referring to Scenario 2-13, construct a percentage histogram for the detergent data, using "9.0 but less than 10.0" as the first class.

ANSWER:
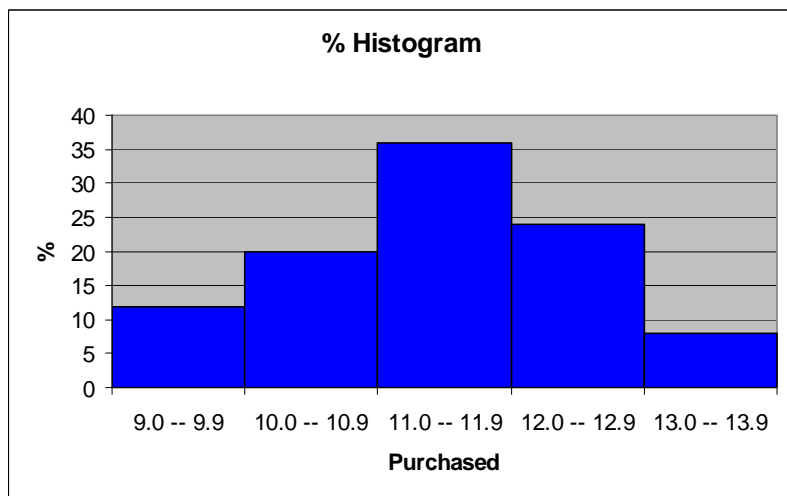
% Histogram

TYPE: PR DIFFICULTY: Moderate
KEYWORDS: histogram, frequency distribution

157. Referring to Scenario 2-13, construct a cumulative percentage polygon for the detergent data if the corresponding frequency distribution uses "9.0 but less than 10.0" as the first class.

ANSWER:

Ogive
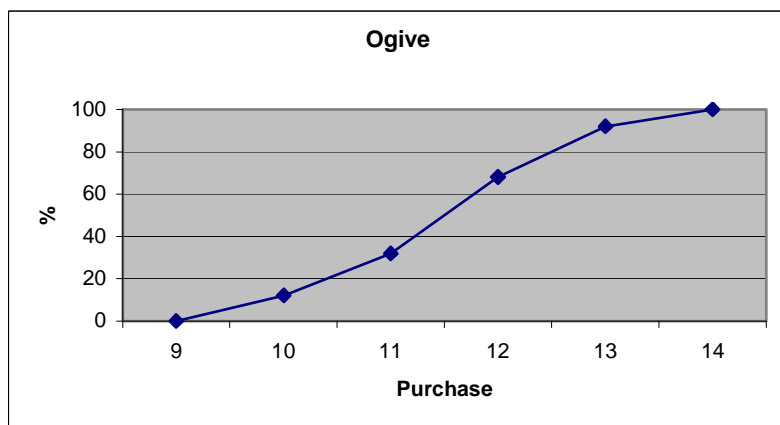
TYPE: PR DIFFICULTY: Moderate
KEYWORDS: cumulative percentage polygon

158. Referring to Scenario 2-13, construct a percentage polygon for the detergent data if the corresponding frequency distribution uses "9.0 but less than 10.0" as the first class.

ANSWER:
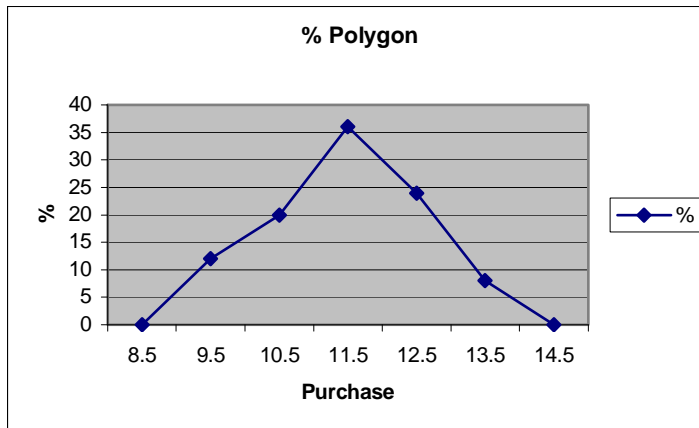
TYPE: PR DIFFICULTY: Moderate
KEYWORDS: percentage distribution, percentage polygon

SCENARIO 2-14

The table below contains the number of people who own a portable Blu-ray player in a sample of 600 broken down by gender.

Own a Portable Blu-ray player	Male	Female
Yes	96	40
No	224	240

159. Referring to Scenario 2-14, construct a table of row percentages.

ANSWER:

Own	Male	Female	Total
Yes	70.59%	29.41%	100.00%
No	48.28%	51.72%	100.00%
Total	53.33%	46.67%	100.00%

TYPE: PR DIFFICULTY: Easy
KEYWORDS: row percentages

160. Referring to Scenario 2-14, construct a table of column percentages.

ANSWER:

Own	Male	Female	Total
Yes	30.00%	14.29%	22.67%
No	70.00%	85.71%	77.33%
Total	100.00%	100.00%	100.00%

TYPE: PR DIFFICULTY: Easy
KEYWORDS: column percentages

161.   Referring to Scenario 2-14, construct a table of total percentages.

ANSWER:

Own	Male	Female	Total
Yes	16.00%	6.67%	22.67%
No	37.33%	40.00%	77.33%
Total	53.33%	46.67%	100.00%

TYPE: PR DIFFICULTY: Easy
KEYWORDS: total percentages

162.   Referring to Scenario 2-14, of those who owned a portable Blu-ray player in the sample, _____ percent were females.

ANSWER:
29.41%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, row percentages

163.   Referring to Scenario 2-14, of those who did not own a portable Blu-ray player in the sample, _____ percent were males.

ANSWER:
48.28%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, row percentages

164.   Referring to Scenario 2-14, of the males in the sample, _____ percent owned a portable Blu-ray player.

ANSWER:
30%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

165.   Referring to Scenario 2-14, of the females in the sample, _____ percent did not own a portable Blu-ray player.

ANSWER:
85.71%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

166.   Referring to Scenario 2-14 of the females in the sample, _____ percent owned a portable Blu-ray player.

ANSWER:
14.29%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

167. Referring to Scenario 2-14, _____ percent of the 600 were females who owned a portable Blu-ray player.

ANSWER:
6.67%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, total percentage

168. Referring to Scenario 2-14, _____ percent of the 600 were males who owned a portable Blu-ray player.

ANSWER:
16%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, total percentage

169. Referring to Scenario 2-14, _____ percent of the 600 were females who either owned or did not own a portable Blu-ray player.

ANSWER:
46.67%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, total percentage

170. Referring to Scenario 2-14, _____ percent of the 600 were males who did not own a portable Blu-ray player.

ANSWER:
37.33%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, total percentage

171. Referring to Scenario 2-14, _____ percent of the 600 owned a portable Blu-ray player.

ANSWER:
22.67%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

172. Referring to Scenario 2-14, _____ percent of the 600 did not own a portable Blu-ray player.

ANSWER:
77.33%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

173.   Referring to Scenario 2-14, _____ percent of the 600 were females.

ANSWER:
46.67%
TYPE: FI DIFFICULTY: Easy
KEYWORDS: contingency table, row percentages

174.   Referring to Scenario 2-14, if the sample is a good representation of the population, we can expect _____ percent of the population will own a portable Blu-ray player.

ANSWER:
22.67%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

175.   Referring to Scenario 2-14, if the sample is a good representation of the population, we can expect _____ percent of the population will be males.

ANSWER:
53.33%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

176.   Referring to Scenario 2-14, if the sample is a good representation of the population, we can expect _____ percent of those who own a portable Blu-ray player in the population will be males.

ANSWER:
70.59%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, row percentages

177.   Referring to Scenario 2-14, if the sample is a good representation of the population, we can expect _____ percent of the males in the population will own a portable Blu-ray player.

ANSWER:
30%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

178.   Referring to Scenario 2-14, if the sample is a good representation of the population, we can expect _____ percent of the females in the population will not own a portable Blu-ray player.

ANSWER:
85.71%
TYPE: FI DIFFICULTY: Moderate
KEYWORDS: contingency table, column percentages

SCENARIO 2-15

The figure below is the ogive for the amount of fat (in grams) for a sample of 36 pizza products where the upper boundaries of the intervals are: 5, 10, 15, 20, 25, and 30.

Cumulative Percentage Polygon for Fat
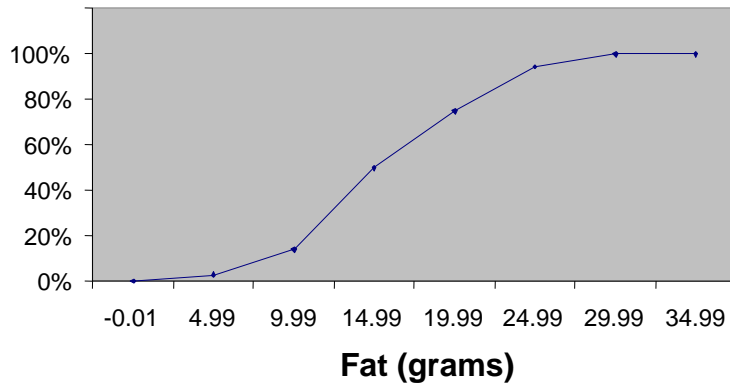
179. Referring to Scenario 2-15, roughly what percentage of pizza products contains less than 10 grams of fat?
    a)   3%
    b)   14%
    c)   50%
    d)   75%

ANSWER:
b
TYPE: MC DIFFICULTY: Easy
KEYWORDS: cumulative percentage polygon, ogive, interpretation

180. Referring to Scenario 2-15, what percentage of pizza products contains at least 20 grams of fat?
    a)   5%
    b)   25%
    c)   75%
    d)   96%

ANSWER:
b
TYPE: MC DIFFICULTY: Easy
KEYWORDS: cumulative percentage polygon, ogive, interpretation

181. Referring to Scenario 2-15, what percentage of pizza products contains between 10 and 25 grams of fat?
    a) 14%
    b) 44%
    c) 62%
    d) 81%

ANSWER:
d
TYPE: MC DIFFICULTY: Easy
KEYWORDS: cumulative percentage polygon, ogive, interpretation

SCENARIO 2-16

The figure below is the percentage polygon for the amount of calories for a sample of 36 pizzas products where the upper limits of the intervals are: 310, 340, 370, 400 and 430.

Percentage Polygon for Calories
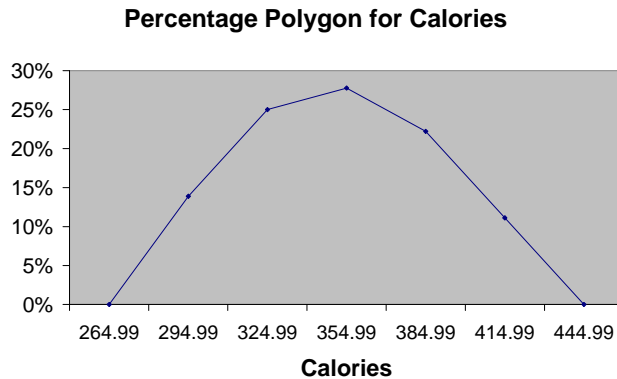
182. Referring to Scenario 2-16, roughly what percentage of pizza products contains between 400 and 430 calories?
    a)   0%
    b)   11%
    c)   89%
    d) 100%

ANSWER:
b
TYPE: MC DIFFICULTY: Easy
KEYWORDS: percentage polygon, interpretation

183. Referring to Scenario 2-16, roughly what percentage of pizza products contains between 340 and 400 calories?
a) 22%
b) 25%
c) 28%
d) 50%

ANSWER:
d
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: percentage polygon, interpretation

184. Referring to Scenario 2-16, roughly what percentage of pizza products contains at least 340 calories?
a) 25%
b) 28%
c) 39%
d) 61%

ANSWER:
d
TYPE: MC DIFFICULTY: Moderate
KEYWORDS: percentage polygon, interpretation

SCENARIO 2-17

The following table presents total retail sales in millions of dollars for the leading apparel companies over a two-year period in the past.

APPAREL COMPANY	Year 1	Year 2
Gap	1,159.0	962.0
TJX	781.7	899.0
Limited	596.5	620.4
Kohl's	544.9	678.9
Nordstrom	402.6	418.3
Talbots	139.9	130.1
AnnTaylor	114.2	124.8

185. Referring to Scenario 2-17, construct a table of column percentages.

ANSWER:

Apparel Company	Year 1	Year 2
Gap	31.00%	25.09%
TJX	20.91%	23.45%
Limited	15.95%	16.18%
Kohl's	14.57%	17.71%
Nordstrom	10.77%	10.91%
Talbots	3.74%	3.39%
AnnTaylor	3.05%	3.26%
Total	100.00%	100.00%

TYPE: PR DIFFICULTY: Moderate
KEYWORDS: column percentages

186. Referring to Scenario 2-17, construct a side-by-side bar chart.
ANSWER:
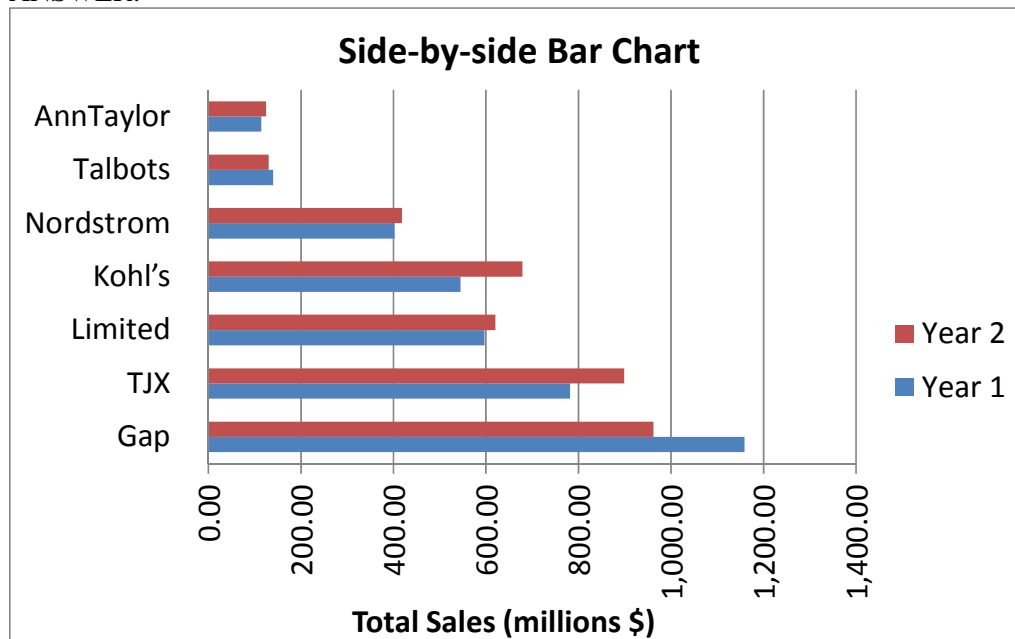
TYPE: PR DIFFICULTY: Moderate
KEYWORDS: column percentages, side-by-side bar chart

187. True or False: Referring to Scenario 2-17, in general, retail sales for the apparel industry have seen a modest growth between Year 1 and Year 2.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: column percentages, side-by-side bar chart, interpretation

188. Referring to Scenario 2-17, among the 8 stores, _____ saw a sales decline.

ANSWER:
Gap and Talbots
TYPE: FI DIFFICULTY: Easy
KEYWORDS: column percentages, side-by-side bar chart, interpretation

SCENARIO 2-18

The stem-and-leaf display below shows the result of a survey on 50 students on their satisfaction with their school with the higher scores represent higher level of satisfaction.

			Stem-and-Leaf Display	
			Stem unit 10	
Statistics			4	1 3 6 6 7
Sample Size	50		5	0 0 3 8 9
Mean	71.06		6	0 1 1 4 4 5 7 7 9 9
Median	73.5		7	0 0 0 1 3 4 4 5 5 6 6 6 7 8 8
Std. Deviation	14.13695		8	0 1 1 3 4 4 5 7 7 8 9
Minimum	41		9	0 2 2 7
Maximum	97			

189. Referring to Scenario 2-18, what was the highest level of satisfaction?

ANSWER:
97
TYPE: PR DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display

190. Referring to Scenario 2-18, what was the lowest level of satisfaction?

ANSWER:
41
TYPE: PR DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display

191. Referring to Scenario 2-18, how many students have a satisfaction level in the 50s?

ANSWER:
5
TYPE: PR DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display

192. Referring to Scenario 2-18, how many students have a satisfaction level below 60?

ANSWER:
10
TYPE: PR DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display

193.   Referring to Scenario 2-18, how many students have a satisfaction level of at least 80?

ANSWER:
15
TYPE: PR DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display

194.   True or False: Referring to Scenario 2-18,   the level of satisfaction is concentrated around 75.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display

195.   True or False: Referring to Scenario 2-18, if a student is randomly selected, his/her most likely
       level of satisfaction will be in the 70s among the 40s, 50s, 60s, 70s, 80s and 90s.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display

196.   True or False: Referring to Scenario 2-18, if a student is randomly selected, his/her most likely
       level of satisfaction will be in the 60s among the 40s, 50s, 60s, 70s, 80s and 90s.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: stem-and-leaf display

197. True or False: Given below is the scatter plot of the price/earnings ratio versus earnings per share of 20 U.S. companies. There appears to be a negative relationship between price/earnings ratio and earnings per share.
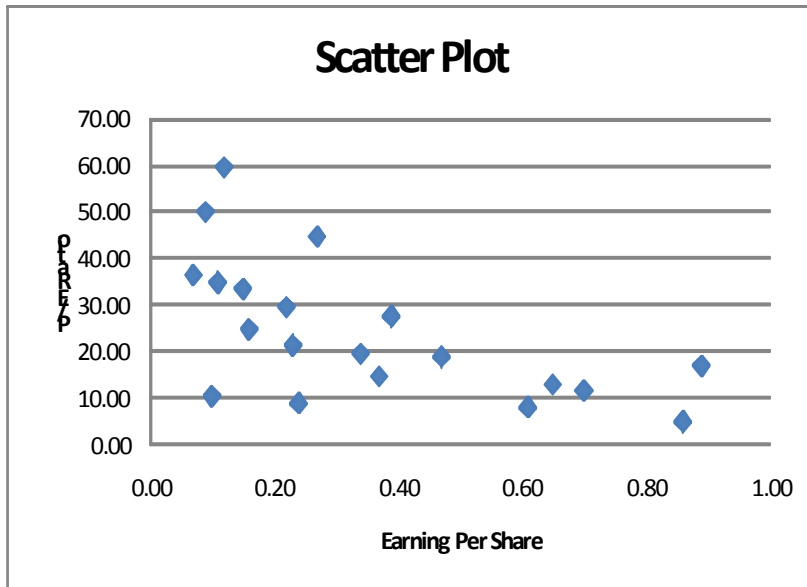
ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: scatter plot

198.   True or False: Given below is the scatter plot of the price/earnings ratio versus earnings per share of 20 U.S. companies. There appear to be a positive relationship between price/earnings ratio and earnings per share.
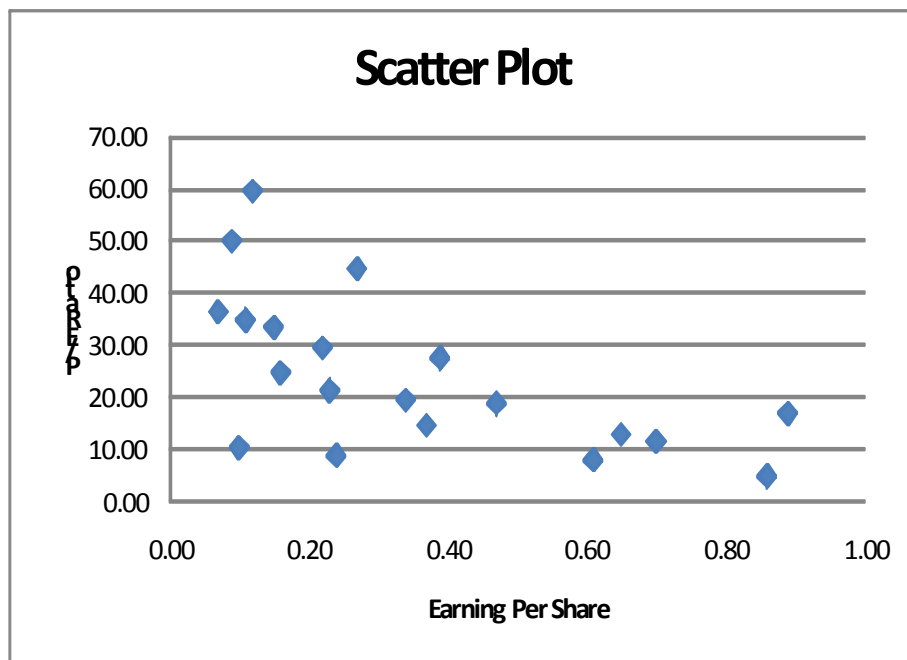
Scatter Plot

ANSWER:
False
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: scatter plot

199. True or False: Given below is the scatter plot of the market value (thousands$) and profit (thousands$) of 50 U.S. companies. Higher market values appear to be associated with higher profits.
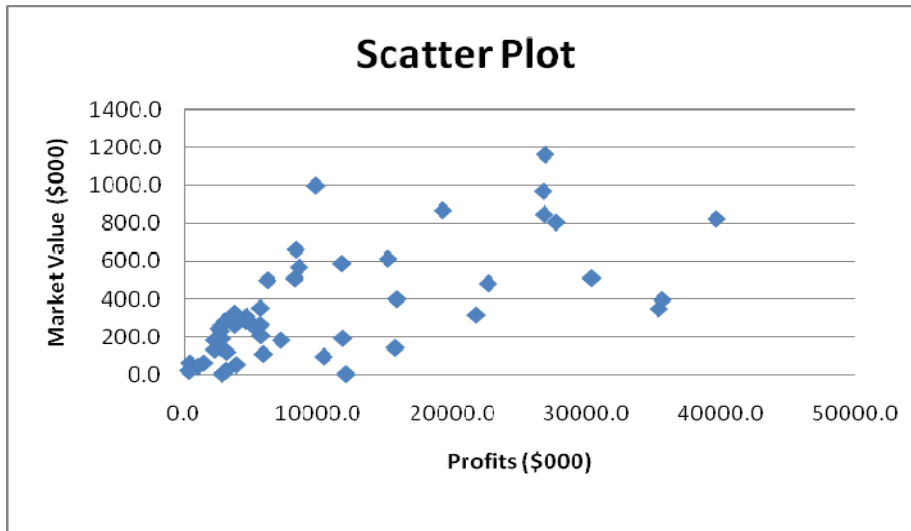
ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: scatter plot

200. True or False: Given below is the scatter plot of the market value (thousands$) and profit (thousands$) of 50 U.S. companies. There appears to be a negative relationship between market value and profit.
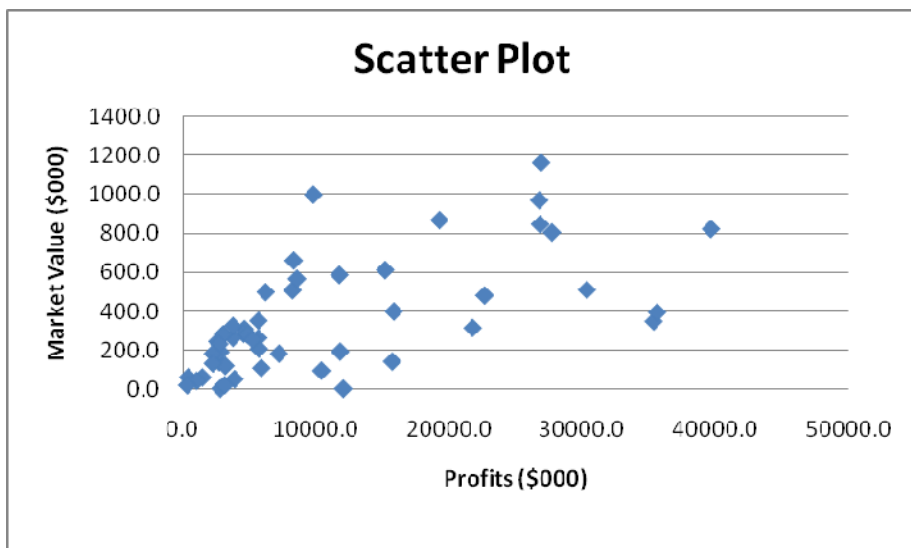
ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: scatter plot

201. True or False: Given below is the scatter plot of the number of employees and the total revenue ($millions) of 20 U.S. companies. There appears to be a positive relationship between total revenue and the number of employees.
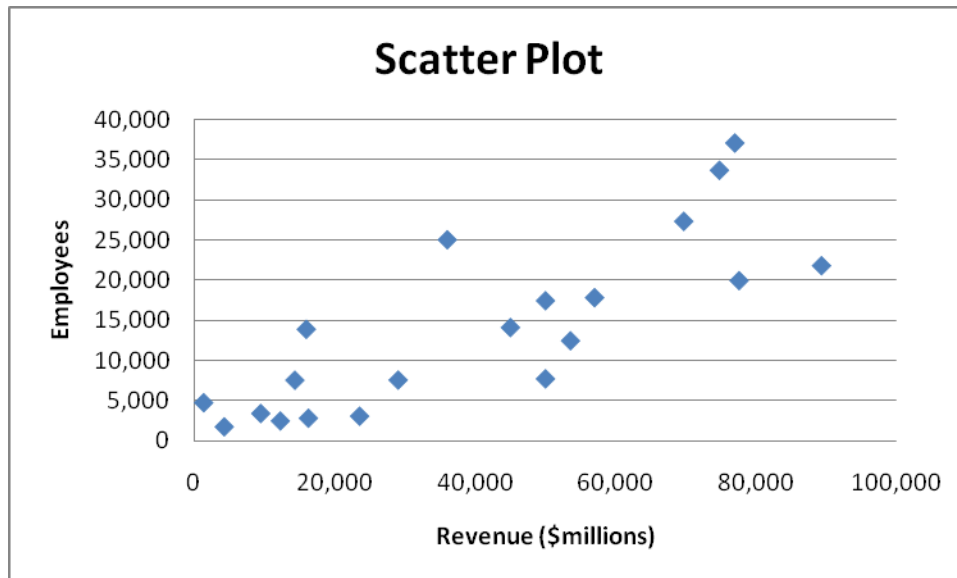
ANSWER:
True
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: scatter plot

202. True or False: Given below is the scatter plot of the number of employees and the total revenue ($millions) of 20 U.S. companies. Companies that have higher numbers of employees appear to also have higher total revenue.
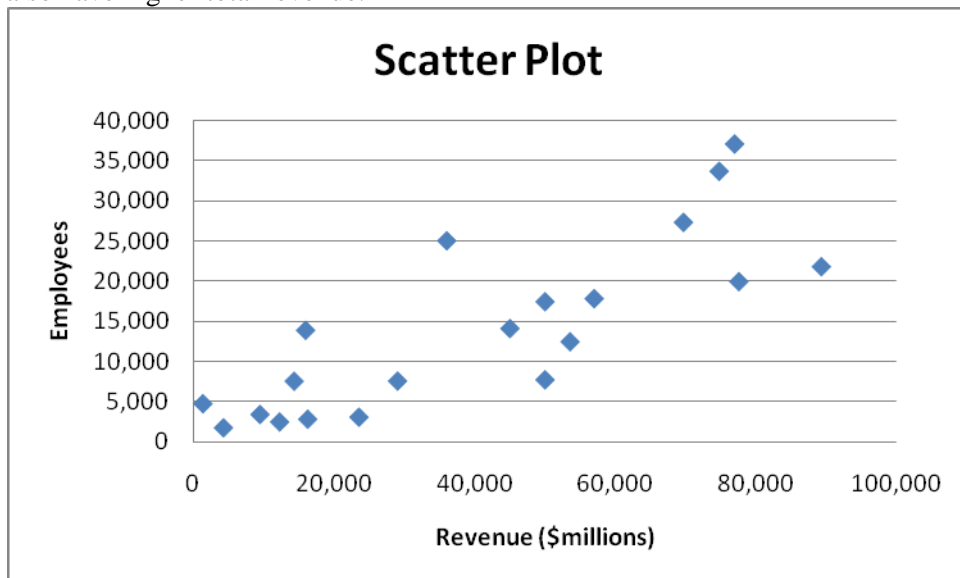
ANSWER: True
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: scatter plot

203. The addition of visual elements that either fail to convey any useful information or that obscure important points about the data in an attempt to enhance the visualization of data is called _____.

ANSWER:
chart junk
TYPE: FI DIFFICULTY: Easy
KEYWORDS: challenges in visualizing data

204. True or False: The Guidelines for Developing Visualizations recommend avoiding uncommon chart type such as doughnut, radar, cone and pyramid charts.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: challenges in visualizing data

205. True or False: The Guidelines for Developing Visualizations recommend using the simplest possible visualization.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: challenges in visualizing data

206. True or False: The Guidelines for Developing Visualizations recommend labeling all axes only when it is possible.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: challenges in visualizing data

207. True or False: The Guidelines for Developing Visualizations recommend using varying scale to conserve precious space whenever possible.

ANSWER:
False
TYPE: TF DIFFICULTY: Easy
KEYWORDS: challenges in visualizing data

208. True or False: The Guidelines for Developing Visualizations recommend always starting the scale for a vertical axis at zero.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: challenges in visualizing data

209.   True or False: The Guidelines for Developing Visualizations recommend always including a
       scale for each axis if the chart contains axes.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: challenges in visualizing data

210.   True or False: When you work with many variables, you must be mindful of the limits of the
       information technology as well as the limits of the ability of your readers to perceive and comprehend
       your results.

ANSWER:
True
TYPE: TF DIFFICULTY: Easy
KEYWORDS: organizing and visualizing many variables

211.   True or False: A multidimensional contingency table allows you to tally the responses of more than
       two continuous variables.

ANSWER:
False
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: multidimensional contingency table, organizing and visualizing many variables

212.   True or False: A multidimensional contingency table allows you to tally the responses of more than
       two categorical variables.

ANSWER:
True
TYPE: TF DIFFICULTY: Moderate
KEYWORDS: multidimensional contingency table, organizing and visualizing many variables

SCENARIO 2-19-A

You are the quality control manager of a water bottles company. One of the biggest complaints in the past years has been the breakage and, hence, the concern on the durability of the connector between the lid and the bottle which many users use as a handle for the bottles. To collect evidence before implementing any modification to the production process, your department has subjected 50 water bottles to a durability test and the following data on the number of times the handles have been used to lift the bottles before they break are contained in the file Scenario2-19-DataA.XLSX.

495	499	502	500	491	498	498	495	488	516
513	486	504	503	493	504	489	500	495	499
501	507	511	496	486	497	510	504	493	482
511	502	520	514	486	514	500	505	512	500
504	498	503	514	474	489	488	506	517	490

213. Referring to Scenario 2-19-A, construct a frequency distribution using "473 but less than 480" as the first class.

ANSWER:

Number of lifts	Frequency
473 but less than 480	1
480 but less than 487	4
487 but less than 494	8
494 but less than 501	14
501 but less than 508	12
508 but less than 515	8
515 but less than 522	3

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: frequency distribution

214. Referring to Scenario 2-19-A, construct a relative frequency or percentage distribution if the corresponding frequency distribution uses "473 but less than 480" as the first class.

ANSWER:

Number of lifts	Percentage
473 but less than 480	2%
480 but less than 487	8%
487 but less than 494	16%
494 but less than 501	28%
501 but less than 508	24%
508 but less than 515	16%
515 but less than 522	6%

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: relative frequency distribution, percentage distribution

215. Referring to Scenario 2-19-A, construct a cumulative percentage distribution if the corresponding frequency distribution uses "473 but less than 480" as the first class.

ANSWER:

Number of lifts	Cumulative Percentage
473 but less than 480	2%
480 but less than 487	10%
487 but less than 494	26%
494 but less than 501	54%
501 but less than 508	78%
508 but less than 515	94%
515 but less than 522	100%

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage distribution

216. Referring to Scenario 2-19-A, construct a histogram using "473 but less than 480" as the first class.
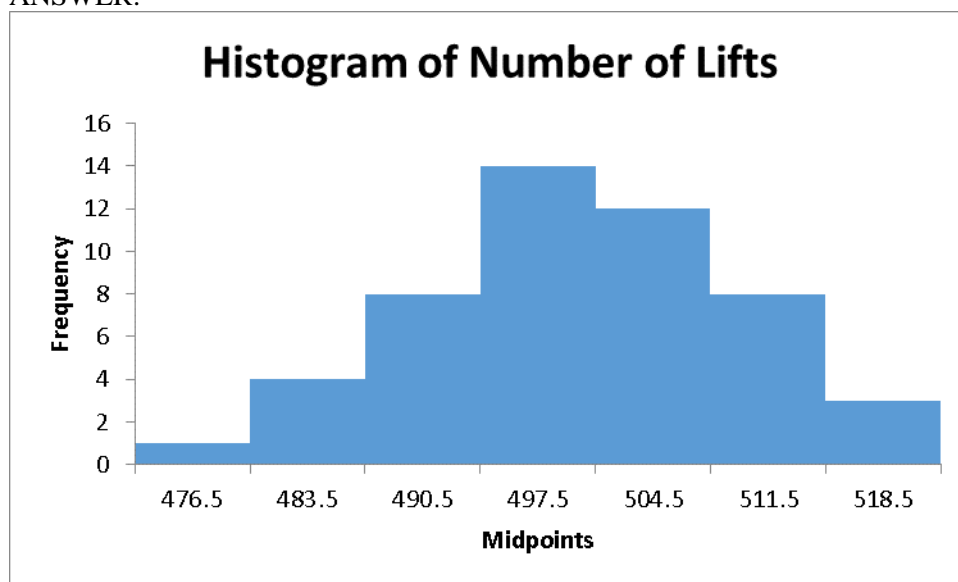
ANSWER:

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: histogram, frequency distribution

217. Referring to Scenario 2-19-A, construct a cumulative percentage polygon using "473 but less than 480" as the first class.
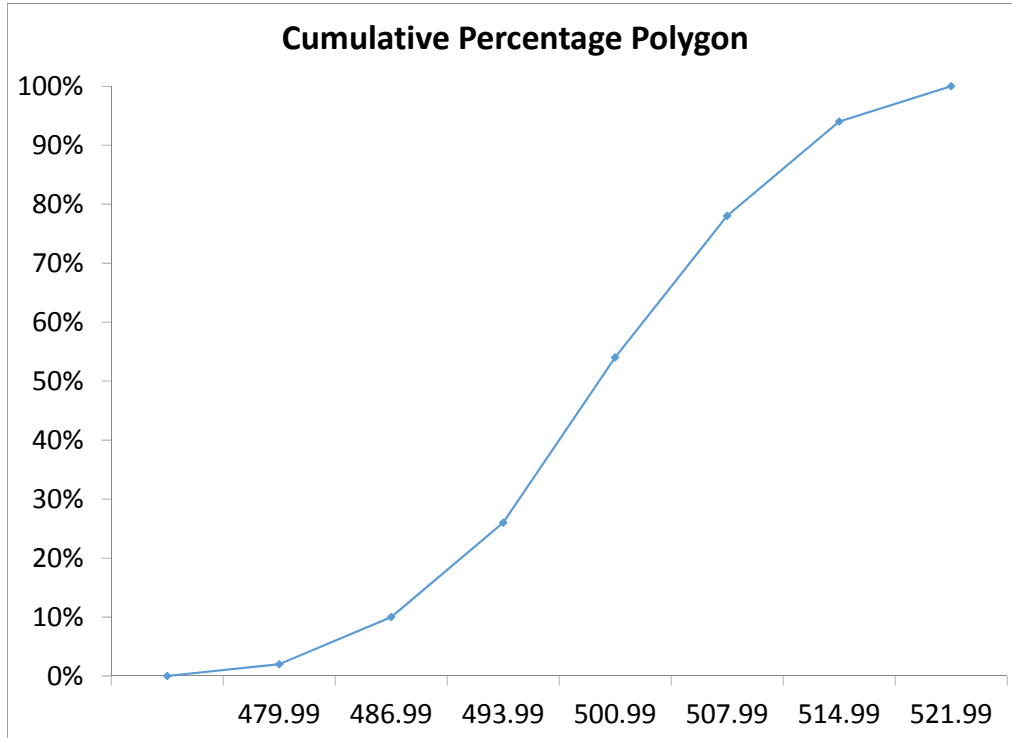
ANSWER:

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage polygon

218.   Referring to Scenario 2-19-A, construct a frequency polygon using "473 but less than 480" as the first class.

ANSWER:
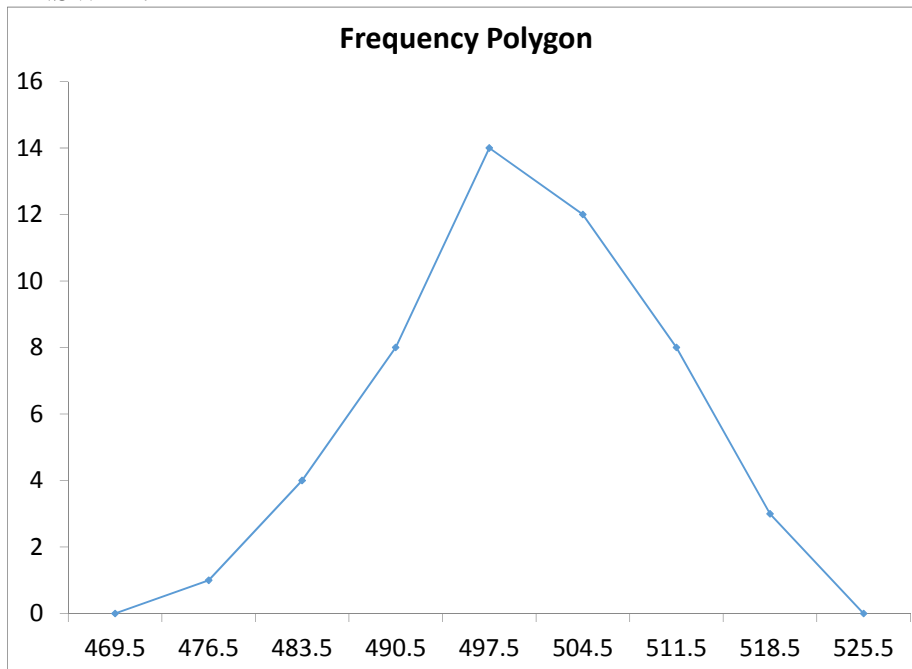
Frequency Polygon

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: frequency distribution, frequency polygon

219. Referring to Scenario 2-19-A, construct a percentage polygon using "473 but less than 480" as the first class.
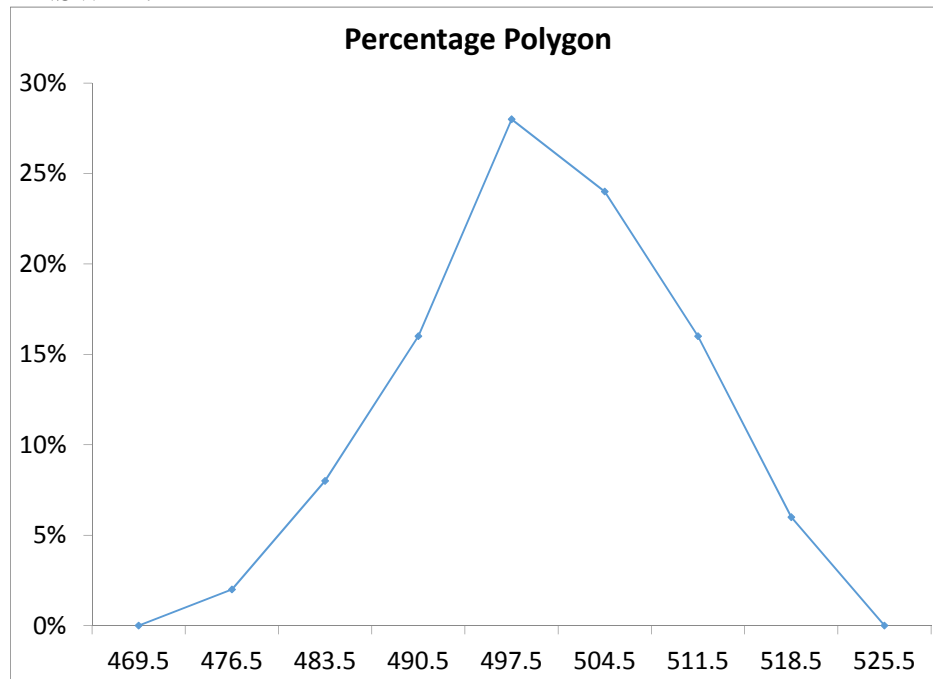
ANSWER:

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: percentage distribution, percentage polygon

220. Referring to Scenario 2-19-A, based on the cumulative percentage polygon or cumulative percentage distribution constructed using "473 but less than 480" as the first class, the percentage of bottles with handles that broke after being used for lifting fewer than 494 times would be _____.

ANSWER:
26%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage distribution, cumulative percentage polygon

221. Referring to Scenario 2-19-A, based on the cumulative percentage polygon or cumulative percentage distribution constructed using "473 but less than 480" as the first class, the percentage of bottles with handles that broke after being used for lifting fewer than 508 times would be _____.

ANSWER:
78%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage distribution, cumulative percentage polygon

222. Referring to Scenario 2-19-A, based on the cumulative percentage polygon or cumulative percentage distribution constructed using "473 but less than 480" as the first class, the percentage of bottles with handles that broke after being used for lifting at least 487 times would be _____.

ANSWER:
90%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage distribution, cumulative percentage polygon

223. Referring to Scenario 2-19-A, based on the cumulative percentage polygon or cumulative percentage distribution constructed using "473 but less than 480" as the first class, the percentage of bottles with handles that broke after being used for lifting at least 515 times would be _____.

ANSWER:
6%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage distribution, cumulative percentage polygon

224. Referring to Scenario 2-19-A, if a percentage histogram was constructed using "473 but less than 480" as the first class, the percentage of bottles with handles that broke after being used for lifting at least 494 but less than 501 times would be _____.

ANSWER:
28%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: relative frequency distribution, percentage distribution

225. Referring to Scenario 2-19-A, if a percentage histogram or percentage distribution was constructed using "473 but less than 480" as the first class, the percentage of bottles with handles that broke after being used for lifting at least 494 but less than 515 times would be _____.

ANSWER:
68%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: relative frequency distribution, percentage distribution

226. Referring to Scenario 2-19-A, if a percentage histogram or percentage distribution was constructed using "473 but less than 480" as the first class, the percentage of bottles with handles that broke after being used for lifting at least 480 but less than 508 times would be _____.

ANSWER:
76%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: relative frequency distribution, percentage distribution

SCENARIO 2-19-B

You are the quality control manager of a water bottles company. One of the biggest complaints in the past years has been the breakage and, hence, the concern on the durability of the connector between the lid and the bottle which many users use as a handle for the bottles. To collect evidence before implementing any modification to the production process, your department has subjected 50 water bottles to a durability test and the following data on the number of times the handles have been used to lift the bottles before they break are contained in the file Scenario2-19-DataB.XLSX.

493	506	515	491	500	505	517	510	506	503
503	491	495	496	496	505	493	486	504	483
514	494	497	501	493	490	510	494	494	495
494	486	495	506	506	507	502	498	510	501
500	505	492	486	501	496	501	521	510	498

227. Referring to Scenario 2-19-B, construct a frequency distribution using "480 but less than 487" as the first class.

ANSWER:

Number of lifts	Frequency
480 but less than 487	4
487 but less than 494	7
494 but less than 501	15
501 but less than 508	16
508 but less than 515	5
515 but less than 522	3

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: frequency distribution

228. Referring to Scenario 2-19-B, construct a relative frequency or percentage distribution if the corresponding frequency distribution uses "480 but less than 487" as the first class.

ANSWER:

Number of lifts	Percentage
480 but less than 487	8%
487 but less than 494	14%
494 but less than 501	30%
501 but less than 508	32%
508 but less than 515	10%
515 but less than 522	6%

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: relative frequency distribution, percentage distribution

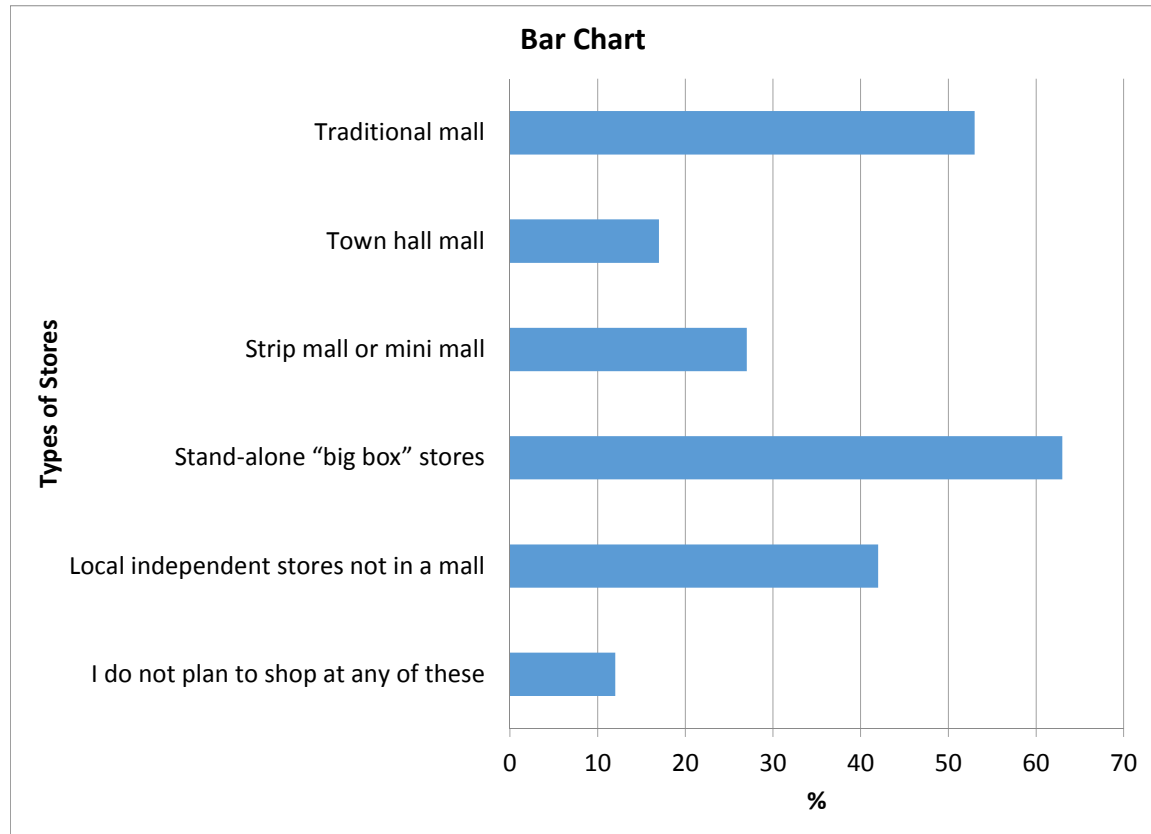229. Referring to Scenario 2-19-B, construct a cumulative percentage distribution if the
     corresponding frequency distribution uses "480 but less than 487" as the first class.

ANSWER:

Number of lifts	Cumulative Percentage
480 but less than 487	8%
487 but less than 494	22%
494 but less than 501	52%
501 but less than 508	84%
508 but less than 515	94%
515 but less than 522	100%

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage distribution

230. Referring to Scenario 2-19-B, construct a histogram using "480 but less than 487" as the first
     class.
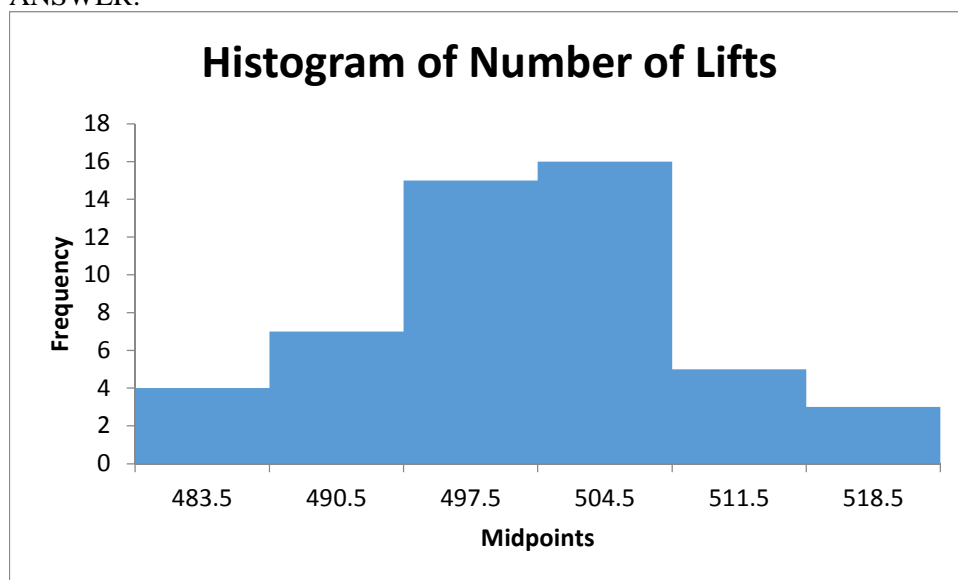
ANSWER:

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: histogram, frequency distribution

231. Referring to Scenario 2-19-B, construct a cumulative percentage polygon using "480 but less than 487" as the first class.

ANSWER:

Cumulative Percentage Polygon
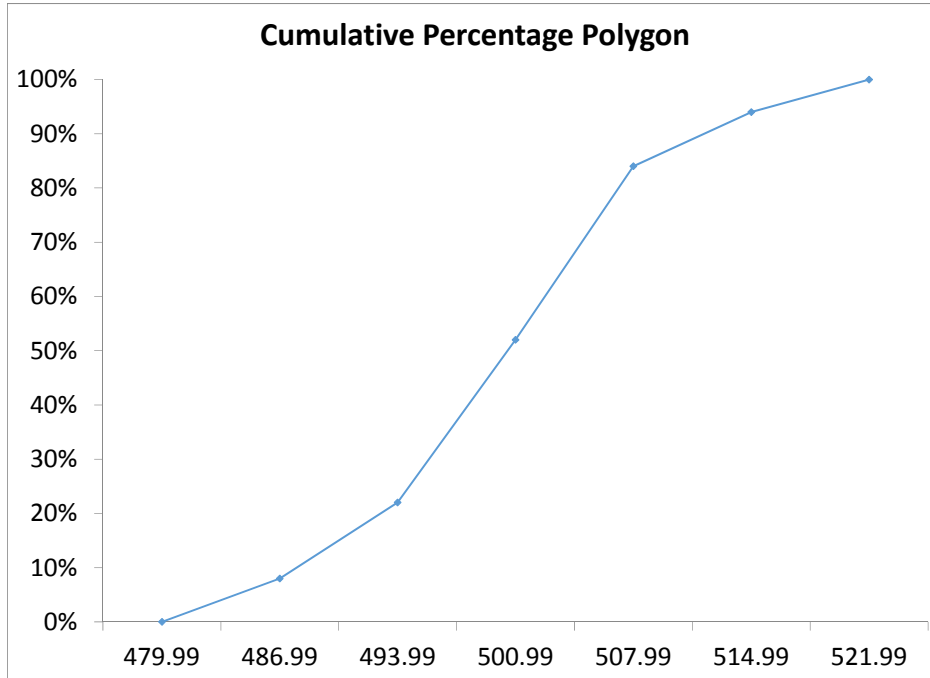
TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage polygon

232.   Referring to Scenario 2-19-B, construct a frequency polygon using "473 but less than 480" as the first class.

ANSWER:
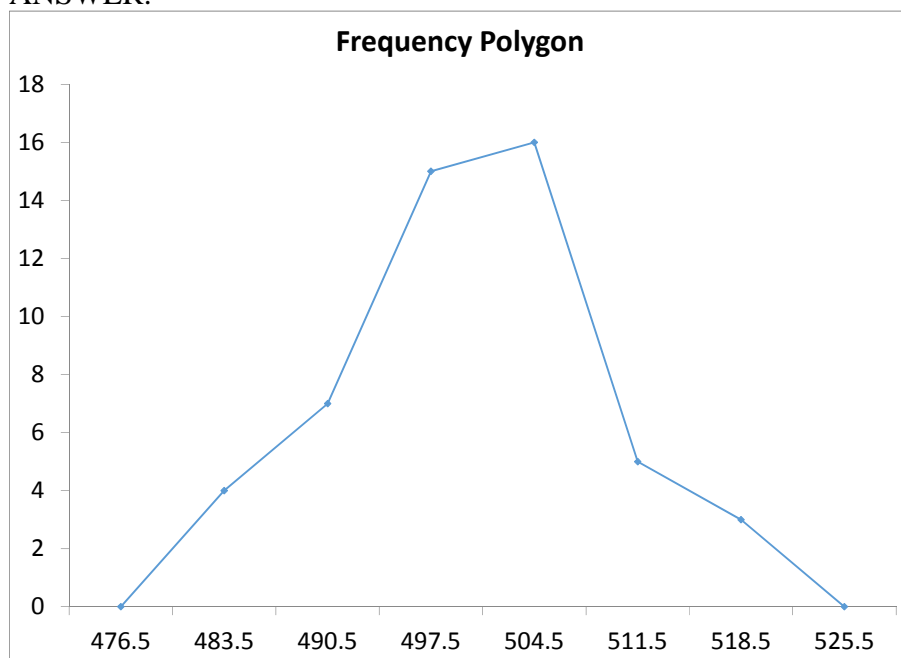
Frequency Polygon

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: frequency distribution, frequency polygon

233.   Referring to Scenario 2-19-B, construct a percentage polygon using "473 but less than 480" as the first class.
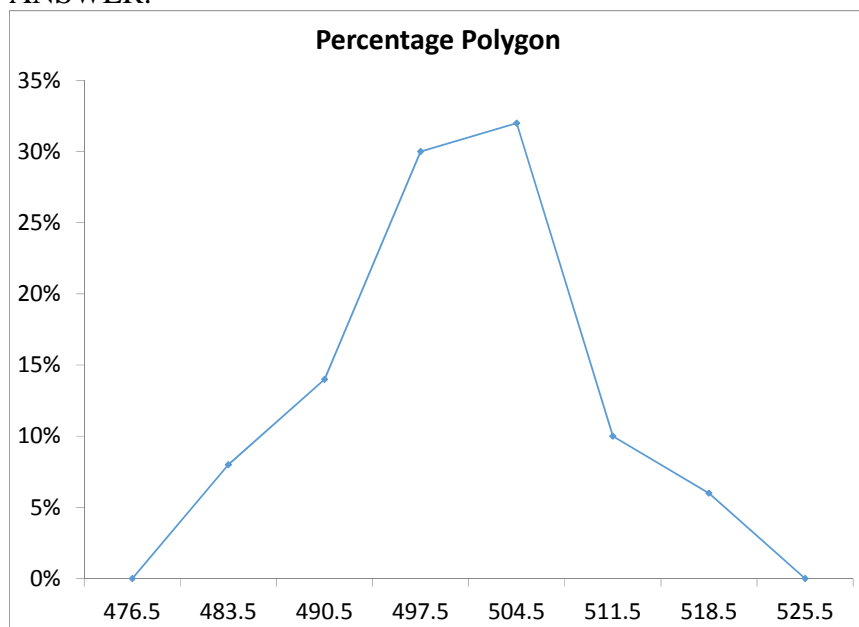
ANSWER:

Percentage Polygon

TYPE: PR DIFFICULTY: Difficult (using PHStat)
KEYWORDS: percentage distribution, percentage polygon

234. Referring to Scenario 2-19-B, based on the cumulative percentage polygon or cumulative percentage distribution constructed using "480 but less than 487" as the first class, the percentage of bottles with handles that broke after being used for lifting fewer than 494 times would be _____.

ANSWER:
22%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage distribution, cumulative percentage polygon

235. Referring to Scenario 2-19-B, based on the cumulative percentage polygon or cumulative percentage distribution constructed using "480 but less than 487" as the first class, the percentage of bottles with handles that broke after being used for lifting fewer than 508 times would be _____.

ANSWER:
84%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage distribution, cumulative percentage polygon

236. Referring to Scenario 2-19-B, based on the cumulative percentage polygon or cumulative percentage distribution constructed using "480 but less than 487" as the first class, the percentage of bottles with handles that broke after being used for lifting at least 487 times would be _____.

ANSWER:
92%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage distribution, cumulative percentage polygon

237. Referring to Scenario 2-19-B, based on the cumulative percentage polygon or cumulative percentage distribution constructed using "480 but less than 487" as the first class, the percentage of bottles with handles that broke after being used for lifting at least 515 times would be _____.

ANSWER:
6%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: cumulative percentage distribution, cumulative percentage polygon

238. Referring to Scenario 2-19-B, if a percentage histogram was constructed using "473 but less than 480" as the first class, the percentage of bottles with handles that broke after being used for lifting at least 494 but less than 501 times would be _____.

ANSWER:
30%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: relative frequency distribution, percentage distribution

239. Referring to Scenario 2-19-B, if a percentage histogram or percentage distribution was constructed using "473 but less than 480" as the first class, the percentage of bottles with handles that broke after being used for lifting at least 494 but less than 515 times would be _____.

ANSWER:
72%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: relative frequency distribution, percentage distribution

240. Referring to Scenario 2-19-B, if a percentage histogram or percentage distribution was constructed using "473 but less than 480" as the first class, the percentage of bottles with handles that broke after being used for lifting at least 487 but less than 508 times would be _____.

ANSWER:
76%
TYPE: FI DIFFICULTY: Difficult (using PHStat)
KEYWORDS: relative frequency distribution, percentage distribution

SCENARIO 2-20-A

A recent consumer survey on holiday shopping reveals the following information on the types of stores at which consumers plan to shop (The table is also available in the file Scenario2-20-DataA.xlsx:

Types of Stores	% of Customers
Stand-alone "big box" stores	63
Traditional mall	53
Local independent stores not in a mall	42
Strip mall or mini mall	27
Town hall mall	17
I do not plan to shop at any of these	12

241. Referring to Scenario 2-20-A, construct a bar chart for the types of stores customers plan to shop at.

ANSWER:

TYPE: PR DIFFICULTY: Easy (using PHStat)
KEYWORDS: bar chart

242. Referring to Scenario 2-20-A, construct a pie chart for the types of stores customers plan to shop at.
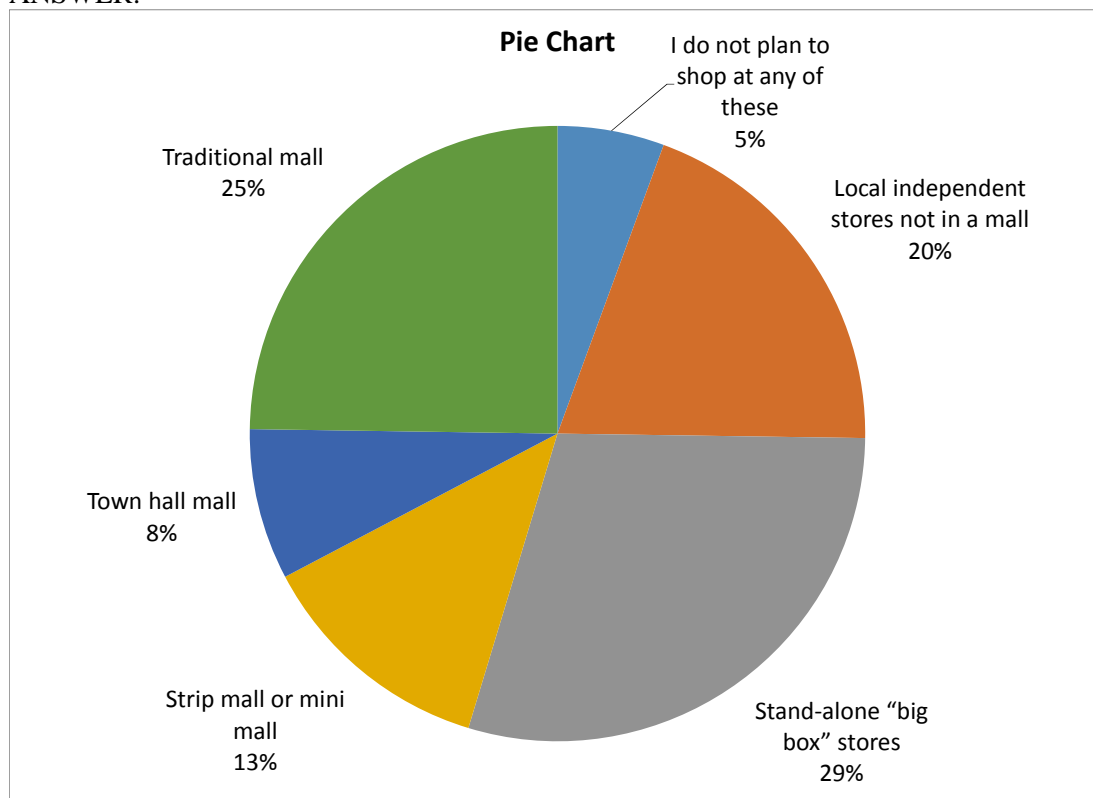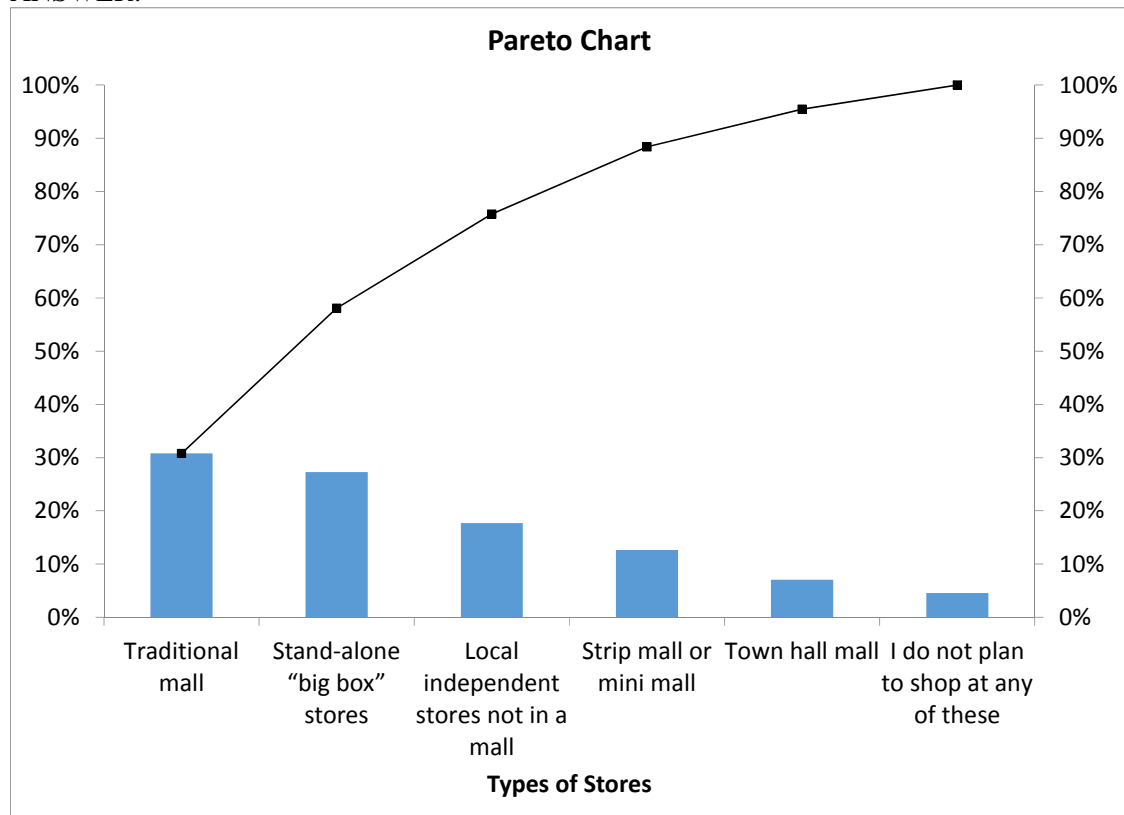
ANSWER:

Pie Chart

TYPE: PR DIFFICULTY: Easy (using PHStat)
KEYWORDS: pie chart

243. Referring to Scenario 2-20-A, construct a Pareto chart for the types of stores customers plan to shop at.

ANSWER:
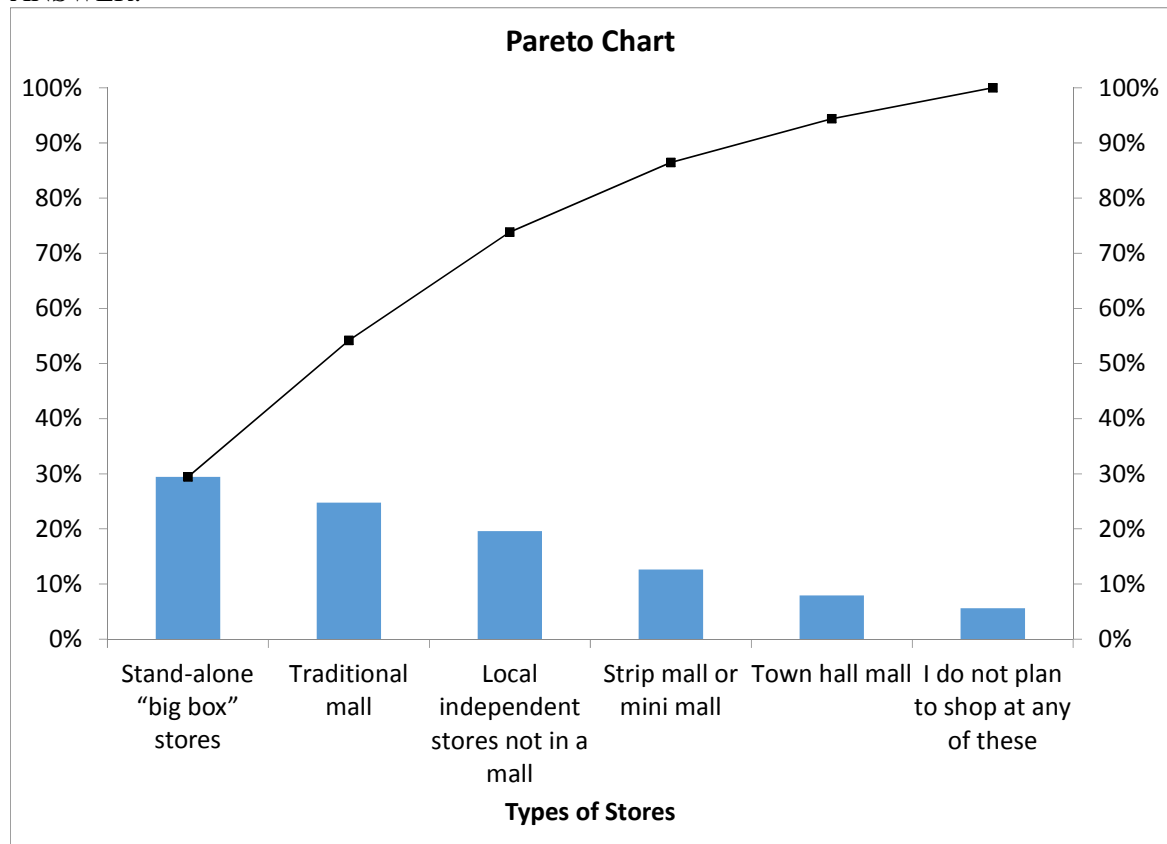
Pareto Chart

TYPE: PR DIFFICULTY: Easy (using PHStat)
KEYWORDS: Pareto chart

244. Referring to Scenario 2-20-A, the type of stores that the most customers plan to shop at is _____.

ANSWER: Stand-alone "big box" stores
TYPE: FI DIFFICULTY: Easy (using PHStat)
KEYWORDS: Pareto chart, pie chart, bar chart, interpretation

245. Referring to Scenario 2-20-A, the top 2 categories of stores that customers plan to shop at make up ____% of the 6 categories of shopping preferences.

ANSWER: 54%
TYPE: FI DIFFICULTY: Moderate (using PHStat)
KEYWORDS: Pareto chart, pie chart, bar chart, interpretation

246.   Referring to Scenario 2-20-A, the category "I do not plan to shop at any of these" makes up _____ % of the 6 categories of shopping preferences.

ANSWER: 5%
TYPE: FI DIFFICULTY: Moderate (using PHStat)
KEYWORDS: Pareto chart, pie chart, bar chart, interpretation

247.   Referring to Scenario 2-20-A, _____ % of the customers surveyed mentioned that they did not plan to shop at any of these stores.

ANSWER: 12%
TYPE: FI DIFFICULTY: Moderate (using PHStat)
KEYWORDS: Pareto chart, pie chart, bar chart, interpretation

248.   Referring to Scenario 2-20-A, what are the top 3 "vital few" types of stores that customers plan to shop at.

ANSWER: "Stand-alone "big box" stores", "Traditional mall" and "Local independent stores not in a mall".
TYPE: FI DIFFICULTY: Moderate (using PHStat)
KEYWORDS: Pareto chart, pie chart, bar chart, interpretation

SCENARIO 2-20-B

A recent consumer survey on holiday shopping reveals the following information on the types of stores at which consumers plan to shop (The table is also available in the file Scenario2-20-DataB.xlsx:

Types of Stores	% of Customers
Stand-alone "big box" stores	54
Traditional mall	61
Local independent stores not in a mall	35
Strip mall or mini mall	25
Town hall mall	14
I do not plan to shop at any of these	9

249. Referring to Scenario 20-20-B, construct a bar chart for the types of stores customers plan to shop at.
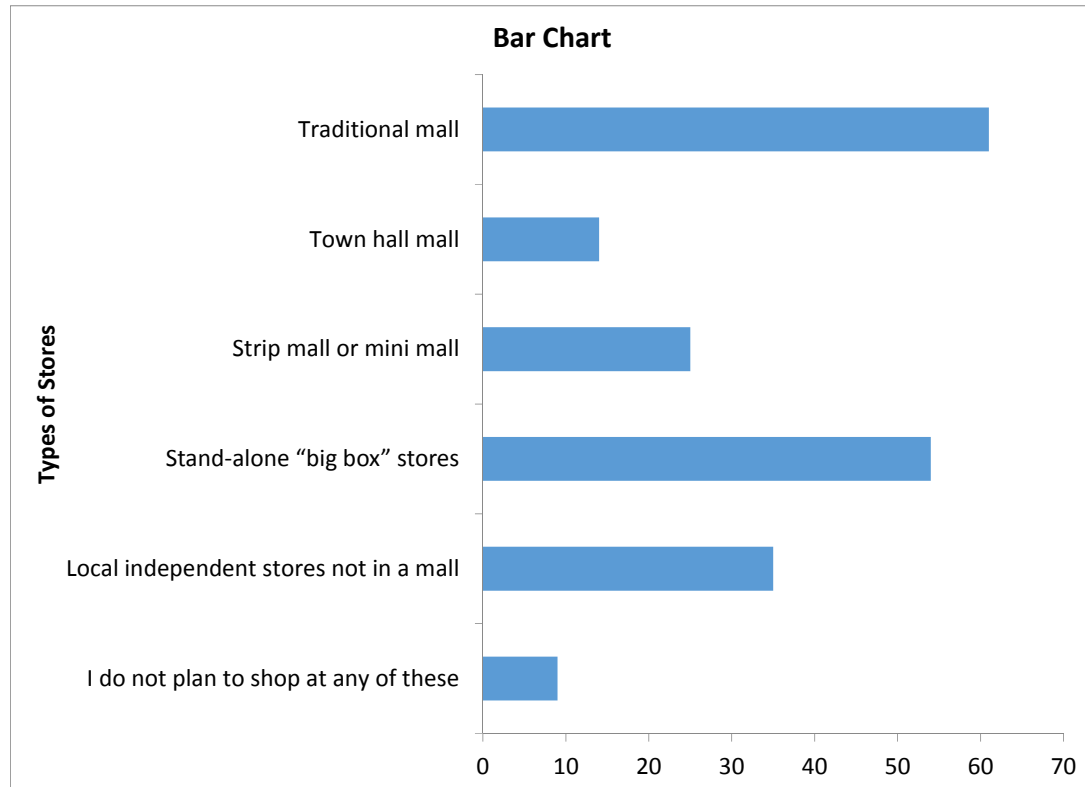
ANSWER:

TYPE: PR DIFFICULTY: Easy (using PHStat)
KEYWORDS: bar chart

250.   Referring to Scenario 20-20-B, construct a pie chart for the types of stores customers plan to shop at.
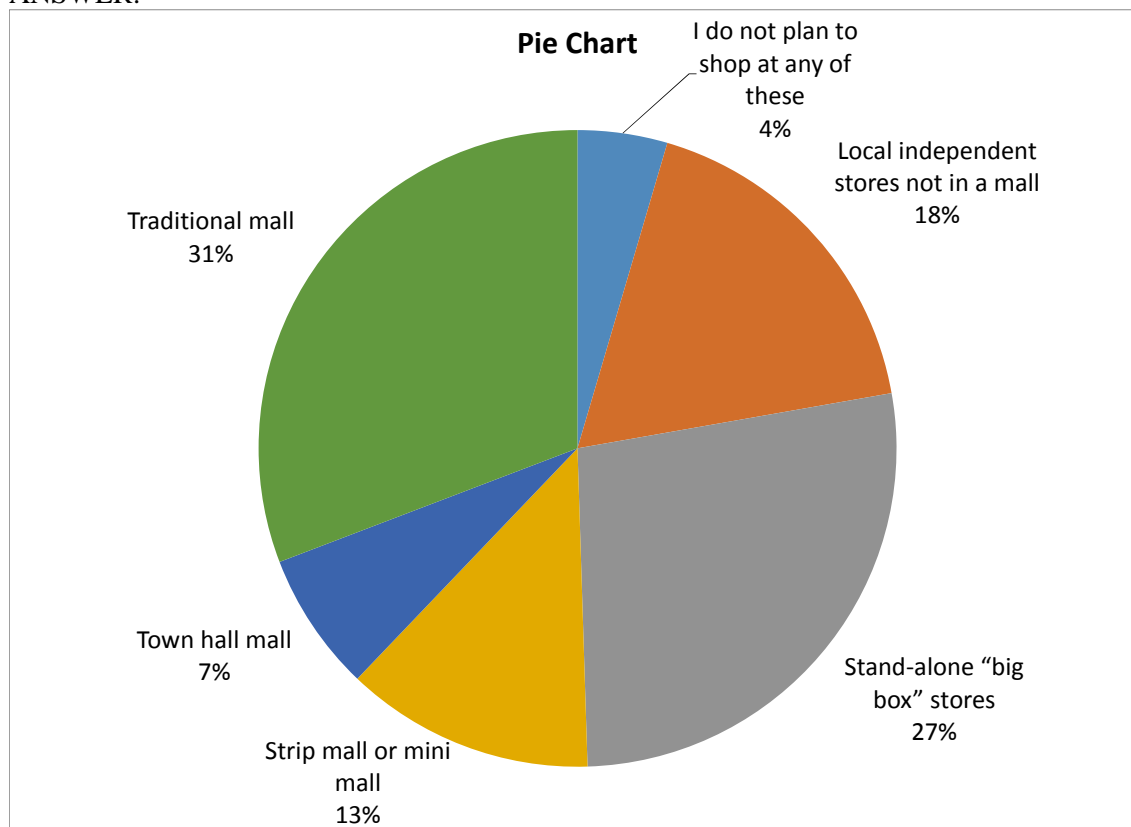
ANSWER:

TYPE: PR DIFFICULTY: Easy (using PHStat)
KEYWORDS: pie chart

251. Referring to Scenario 20-20-B, construct a Pareto chart for the types of stores customers plan to shop at.

ANSWER:

Pareto Chart

TYPE: PR DIFFICULTY: Easy (using PHStat)
KEYWORDS: Pareto chart

252. Referring to Scenario 20-20-B, the type of stores that the most customers plan to shop at is _____.

ANSWER: Traditional mall
TYPE: FI DIFFICULTY: Easy (using PHStat)
KEYWORDS: Pareto chart, pie chart, bar chart, interpretation

253. Referring to Scenario 20-20-B, the top 2 categories of stores that customers plan to shop at make up ____% of the 6 categories of shopping preferences.

ANSWER: 58%
TYPE: FI DIFFICULTY: Moderate (using PHStat)
KEYWORDS: Pareto chart, pie chart, bar chart, interpretation

254. Referring to Scenario 20-20-B, the category "I do not plan to shop at any of these" makes up
      ____ % of the 6 categories of shopping preferences.

ANSWER: 5%
TYPE: FI DIFFICULTY: Moderate (using PHStat)
KEYWORDS: Pareto chart, pie chart, bar chart, interpretation

255. Referring to Scenario 20-20-B, ____ % of the customers surveyed mentioned that they did not
      plan to shop at any of these stores.

ANSWER: 9%
TYPE: FI DIFFICULTY: Moderate (using PHStat)
KEYWORDS: Pareto chart, pie chart, bar chart, interpretation

256. Referring to Scenario 20-20-B, what are the top 3 "vital few" types of stores that customers
      plan to shop at.

ANSWER: "Traditional mall", "Stand-alone "big box" stores" and "Local independent stores not in a
mall".
TYPE: FI DIFFICULTY: Moderate (using PHStat)
KEYWORDS: Pareto chart, pie chart, bar chart, interpretation

CHAPTER 2

2.1   (a)

Category	Frequency	Percentage
A	13	26%
B	28	56
C	9	18

   (b)   Category "B" is the majority.

2.2   (a)   Table frequencies for all student responses

Student Major Categories

Gender	A	C	M	Totals
Male	14	9	2	25
Female	6	6	3	15
Totals	20	15	5	40

   (b)   Table percentages based on overall student responses

Student Major Categories

Gender	A	C	M	Totals
Male	35.0%	22.5%	5.0%	62.5%
Female	15.0%	15.0%	7.5%	37.5%
Totals	50.0%	37.5%	12.5%	100.0%

Table based on row percentages

Student Major Categories

Gender	A	C	M	Totals
Male	56.0%	36.0%	8.0%	100.0%
Female	40.0%	40.0%	20.0%	100.0%
Totals	50.0%	37.5%	12.5%	100.0%

Table based on column percentages

Student Major Categories

Gender	A	C	M	Totals
Male	70.0%	60.0%	40.0%	62.5%
Female	30.0%	40.0%	60.0%	37.5%
Totals	100.0%	100.0%	100.0%	100.0%

2.3   Answers will vary.
   (a)   You can conclude that Android smartphones have seen steady increase in market shares while Blackberry and Other OS smartphones have seen steady decrease in market shares since 2011. Android smartphones dominated the market in all those three years.
   (b)   The iOS smartphones have overtaken Other OS smartphones and owned the second largest market share since 2012. The Microsoft smartphones have arisen to the third place in terms of market share in 2013 from the fifth place position in 2011 while the Other OS smartphones have dropped from the second place in 2011 to the last place in 2013 and last but one in 2014.

2.4 (a) The percentage of complaints for each automaker:

Automaker	Frequency	Percentage	Cumulative Pct.
General Motors	551	18.91%	18.91%
Other	516	17.71%	36.62%
Nissan Motors Corporation	467	16.03%	52.64%
Ford Motor Company	440	15.10%	67.74%
Chrysler LLC	439	15.07%	82.81%
Toyota Motor Sales	332	11.39%	94.20%
American Honda	169	5.80%	100.00%

(b) General Motors has the most complaints, followed by Other, Nissan Motors Corporation, Ford Motor Company, Chryler LLC, Toyota Motor Sales and American Honda.

(c) The percentage of complaints for each category:

Category	Frequency	Percentage	Cumulative Pct.
Powertrain	1148	42.82%	42.82%
Steering	397	14.81%	57.63%
Interior Electronics/Hardware	279	10.41%	68.03%
Fuel/Emission/Exhaust System	240	8.95%	76.99%
Airbags and Seatbelts	201	7.50%	84.48%
Body and Glass	182	6.79%	91.27%
Brakes	163	6.08%	97.35%
Tires and Wheels	71	2.65%	100.00%

(d) Powertrain has the most complaints, followed by steering, interior electronics/hardware, fuel/emission/exhaust system, airbags and seatbelts, body and glass, brakes, and, finally, tires and wheels.

2.5 Answers will vary.
"High pay" has the highest percentage at 23%, followed closely by "good work-life balance" at 22%.

2.6 (a)

Region	Oil Production	Percentage
	(millions of barrels a day)	
Iran	2.69	3.27%
Saudi Arabia	9.58	11.66%
Other OPEC countries	17.93	21.82%
Non-OPEC countries	51.99	63.26%
Total	82.19	100.00%

(b) More than half the oil produced is from non-OPEC countries. About 22% is produced by OPEC countries other than Iran and Saudi Arabia.

2.7 (a) The percentage of values for each response need:

Barriers	Frequency	%
Data must be integrated from multiple sources	68	22.67%
Lack of automation/repeatable process	51	17.00%
Metrics need to be identified or defined	45	15.00%
Production is cumbersome	42	14.00%
Data quality is not reliable	36	12.00%
Sharing findings is challenging	21	7.00%
Analytic tools are too complex	17	5.67%
Ensuring security and integrity of workforce data	17	5.67%
Other	3	1.00%
Total	300	100.00%

(b) Answer will vary. "Data must be integrated from multiple sources" is the most frequently mentioned need, followed by "Lack of automation/repeatable process", "Metrics need to be identified or defined", "Production is cumbersome" and "Data quality is not reliable".

2.8 (a) Table of total percentages

	Gender		
Influenced	Male	Female	Total
Yes	5%	10%	15%
No	45%	40%	85%
Total	50%	50%	100%

Table of row percentages

	Gender		
Influenced	Male	Female	Total
Yes	34%	66%	100%
No	53%	47%	100%
Total	50%	50%	100%

Table of column percentages

	Gender		
Influenced	Male	Female	Total
Yes	10%	20%	15%
No	90%	80%	85%
Total	100%	100%	100%

(b) Answer will vary. A higher percentage of females are influenced by social media.

2.9   (a)

Table of total percentages:

Category	Outcome		Total
	Successful	Not Successful	
Film & Video	16%	25%	41%
Games	5%	9%	14%
Music	18%	16%	34%
Technology	2%	8%	11%
Total	41%	59%	100%

Note: The numbers in the Total column may not appear to be the sum of the different outcomes due to rounding.

Table of row percentages:

Category	Outcome		Total
	Successful	Not Successful	
Film & Video	39%	61%	100%
Games	34%	66%	100%
Music	53%	47%	100%
Technology	23%	77%	100%
Total	41%	59%	100%

Table of column percentages:

Category	Outcome		Total
	Successful	Not Successful	
Film & Video	39%	43%	41%
Games	12%	16%	14%
Music	44%	27%	34%
Technology	6%	14%	11%
Total	100%	100%	100%

   (b)   The row percentages is most informative for these data as they show that among the different categories, music is the most successful at 53% while technology is most unsuccessful at only 23%.

   (c)   Answer may vary. Music is the most successful at 53% followed by film & video at 39%, games at 34% and finally technology at 23%.

2.10   Social recommendations had very little impact on correct recall. Those who arrived at the link from a recommendation had a correct recall of 73.07% as compared to those who arrived at the link from browsing who had a correct recall of 67.96%.

2.11   Ordered array: 63 64 68 71 75 88 94

2.12   Ordered array: 73 78 78 78 85 88 91

2.13

Average Time to Resolve Cyberattacks	Frequency	%	Cumulative %
Less than 1 day	101	29.97%	29.97%
Between 1 and less than 3 days	115	34.12%	64.09%
Between 3 and less than 7 days	47	13.95%	78.04%
Between 7 and less than 14 days	30	8.90%	86.94%
14 days or more	44	13.06%	100.00%
Total	337	100.00%	

(a)   64.09% of small businesses took less than 3 days, on average, to resolve cyberattacks.
(b)   56.97% of small businesses took between 1 and less than 14 days, on average, to resolve cyberattacks.
(c)   35.91% of small businesses took 3 or more days, on average, to resolve cyberattacks.

2.14   (a)   0 but less than 5 million, 5 million but less than 10 million, 10 million but less than 15 million, 15 million but less than 20 million, 20 million but less than 25 million, 25 million but less than 30 million.
(b)   5 million
(c)   2.5 million, 7.5 million, 12.5 million, 17.5 million, 22.5 million, and 27.5 million.

2.15   (a)   Ordered array:

Cost ($)	212.40	221.80	223.92	232.44	245.39	258.78	261.20	263.10
	271.74	275.74	278.90	280.28	281.06	289.71	295.40	312.20
	317.08	322.50	325.85	336.52	340.60	341.90	369.86	404.60
	435.72	444.16	468.20	477.32	541.00	676.42		

(b)   PHStat output:

	Cost		Frequency	Percentage	Cumulative Pctage.	Midpts.
200	but less than	260	6	20.00%	20.00%	230
260	but less than	320	11	36.67%	56.67%	290
320	but less than	380	6	20.00%	76.67%	350
380	but less than	440	2	6.67%	83.33%	410
440	but less than	500	3	10.00%	93.33%	470
500	but less than	560	1	3.33%	96.67%	530
560	but less than	620	0	0.00%	96.67%	590
620	but less than	680	1	3.33%	100.00%	650

(c)   The costs of attending a basketball game is concentrating between $200 and $380.

2.16   (a)
Electricity Costs	Frequency	Percentage
$80 to $99	4	8%
$100 to $119	7	14
$120 to $139	9	18
$140 to $159	13	26
$160 to $179	9	18
$180 to $199	5	10
$200 to $219	3	6

(b)

Electricity Costs	Frequency	Percentage	Cumulative %
$99	4	8%	8%
$119	7	14%	22%
$139	9	18%	40%
$159	13	26%	66%
$179	9	18%	84%
$199	5	10%	94%
$219	3	6%	100%

(c)   The majority of utility charges are clustered between $120 and $180.

2.17   (a), (b) Annual Time Sitting in Traffic (hours)

Bin Cell	Frequency	Percentage	Cumulative Pctage.
15 but less than 20	1	3.23%	3.23%
20 but less than 25	4	12.90%	16.13%
25 but less than 30	4	12.90%	29.03%
30 but less than 35	2	6.45%	35.48%
35 but less than 40	7	22.58%	58.06%
40 but less than 45	3	9.68%	67.74%
45 but less than 50	4	12.90%	80.65%
50 but less than 55	2	6.45%	87.10%
55 but less than 60	1	3.23%	90.32%
60 but less than 65	1	3.23%	93.55%
65 but less than 70	0	0.00%	93.55%
70 but less than 75	2	6.45%	100.00%

Cost of Sitting in Traffic($)

Bin Cell	Frequency	Percentage	Cumulative Pctage.
300 but less than 450	4	12.90%	12.90%
450 but less than 600	6	19.35%	32.26%
600 but less than 750	6	19.35%	51.61%
750 but less than 900	5	16.13%	67.74%
900 but less than 1050	6	19.35%	87.10%
1050 but less than 1200	2	6.45%	93.55%
1200 but less than 1350	1	3.23%	96.77%
1350 but less than 1550	0	0.00%	96.77%
1550 but less than 1650	1	3.23%	100.00%

2.17   (c)   The annual time sitting in traffic is concentrated around 37.5 hours with a few spending
cont.   as much as around 72.5 hours.
       (d)   The cost of sitting in traffic per year is concentrated around $675 with one costing as
             much as $1,575.

2.18   (a), (b)

Credit Score			Frequency	Percentage	Cumulative Pctage.	Midpts.
600	but less than	610	1	0.70%	0.70%	615
610	but less than	620	0	0.00%	0.70%	625
620	but less than	630	2	1.40%	2.10%	635
630	but less than	640	15	10.49%	12.59%	645
640	but less than	650	18	12.59%	25.17%	655
650	but less than	660	24	16.78%	41.96%	665
660	but less than	670	22	15.38%	57.34%	675
670	but less than	680	28	19.58%	76.92%	685
680	but less than	690	21	14.69%	91.61%	695
690	but less than	700	11	7.69%	99.30%	705
700	but less than	710	1	0.70%	100.00%	715

Note: Due to rounding, some of the numbers in the table may not add up.
       (c)   The average credit scores are concentrated around 630 and 690.

2.19   (a), (b)

Bin	Frequency	Percentage	Cumulative %
-0.00350 but less than -0.00201	13	13.00%	13.00%
-0.00200 but less than -0.00051	26	26.00%	39.00%
-0.00050 but less than 0.00099	32	32.00%	71.00%
0.00100 but less than 0.00249	20	20.00%	91.00%
0.00250 but less than 0.00399	8	8.00%	99.00%
0.004 but less than 0.00549	1	1.00%	100.00%

       (c)   Yes, the steel mill is doing a good job at meeting the requirement as there is only one
             steel part out of a sample of 100 that is as much as 0.005 inches longer than the specified
             requirement.

2.20   (a), (b)

Bin	Frequency	Percentage	Cumulative %
8.310 -- 8.329	3	6.12%	6.12%
8.330 -- 8.349	2	4.08%	10.20%
8.350 -- 8.369	1	2.04%	12.24%
8.370 -- 8.389	4	8.16%	20.41%
8.390 -- 8.409	4	8.16%	28.57%
8.410 -- 8.429	15	30.61%	59.18%
8.430 -- 8.449	7	14.29%	73.47%
8.450 -- 8.469	5	10.20%	83.67%
8.470 -- 8.489	5	10.20%	93.88%
8.490 -- 8.509	3	6.12%	100.00%

       (c)   All the troughs will meet the company's requirements of between 8.31 and 8.61 inches
             wide.

2.21   (a),(b)

Strength	Frequency	Percentage	Cumulative Percentage
1500 -- 1549	1	3.33%	3.33%
1550 -- 1599	2	6.67%	10.00%
1600 -- 1649	2	6.67%	16.67%
1650 -- 1699	7	23.33%	40.00%
1700 -- 1749	5	16.67%	56.67%
1750 -- 1799	7	23.33%	80.00%
1800 -- 1849	3	10.00%	90.00%
1850 -- 1899	3	10.00%	100.00%

(c)   The strength of all the insulators meets the company's requirement of at least 1500 lbs.

2.22   (a), (b) Manufacturer A:

Bin Cell	Frequency	Percentage	Cumulative Pctage.
6,500 but less than 7,500	3	7.50%	7.50%
7,500 but less than 8,500	5	12.50%	20.00%
8,500 but less than 9,500	20	50.00%	70.00%
9,500 but less than 10,500	9	22.50%	92.50%
10,500 but less than 11,500	3	7.50%	100.00%

Manufacturer B:

Bin Cell	Frequency	Percentage	Cumulative Pctage.
7,500 but less than 8,500	2	5.00%	5.00%
9,500 but less than 9,500	8	20.00%	25.00%
9,500 but less than 10,500	16	40.00%	65.00%
10,500 but less than 11,500	9	22.50%	87.50%
11,500 but less than 12,500	5	12.50%	100.00%

(c)   Manufacturer B produces bulbs with longer lives than Manufacturer A. The cumulative percentage for Manufacturer B shows 65% of its bulbs lasted less than 10,500 hours, contrasted with 70% of Manufacturer A's bulbs, which lasted less than 9,500 hours. None of Manufacturer A's bulbs lasted more than 11,499 hours, but 12.5% of Manufacturer B's bulbs lasted between 11,500 and 12,499 hours. At the same time, 7.5% of Manufacturer A's bulbs lasted less than 7,500 hours, whereas all of Manufacturer B's bulbs lasted at least 7,500 hours

2.23   (a)

Amount of Soft Drink	Frequency	Percentage
1.850 – 1.899	1	2%
1.900 – 1.949	5	10
1.950 – 1.999	18	36
2.000 – 2.049	19	38
2.050 – 2.099	6	12
2.100 – 2.149	1	2

Amount of Soft Drink	Frequency Less Than	Percentage Less Than
1.899	1	2%
1.949	6	12
1.999	24	48
2.049	43	86
2.099	49	98
2.149	50	100

   (b)   The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.24   (a)

Percentages in decimals as proportions

2.24 (a)
cont.

Pie Chart
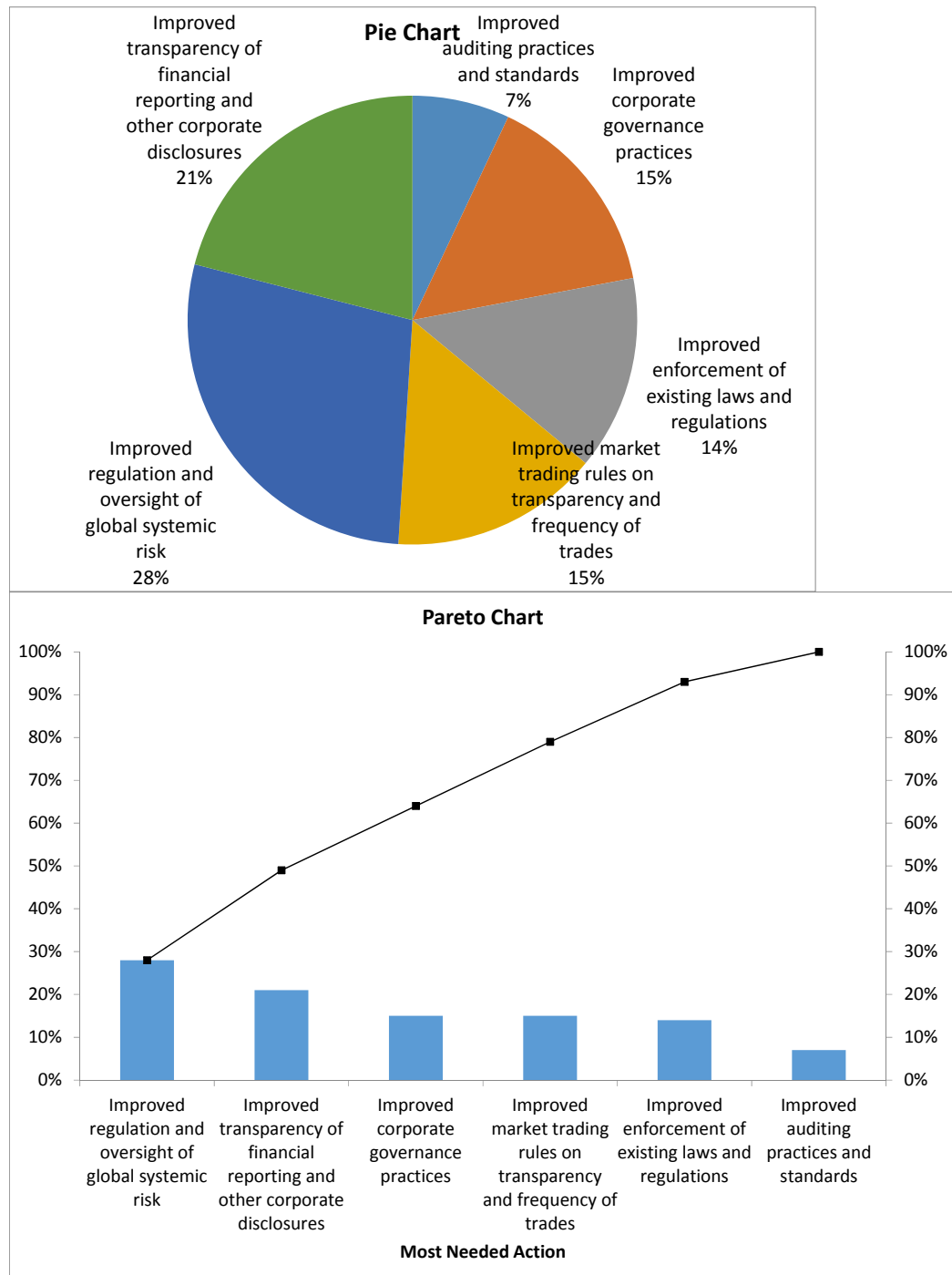
Pareto Chart

(b) The Pareto diagram is better than the pie chart to portray these data because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

(c) You can conclude that "improved regulation and oversight of global systemic risk" accounts for the largest percentage (28%) of the most needed action to improve investor trust and market integrity.

2.25   (a)
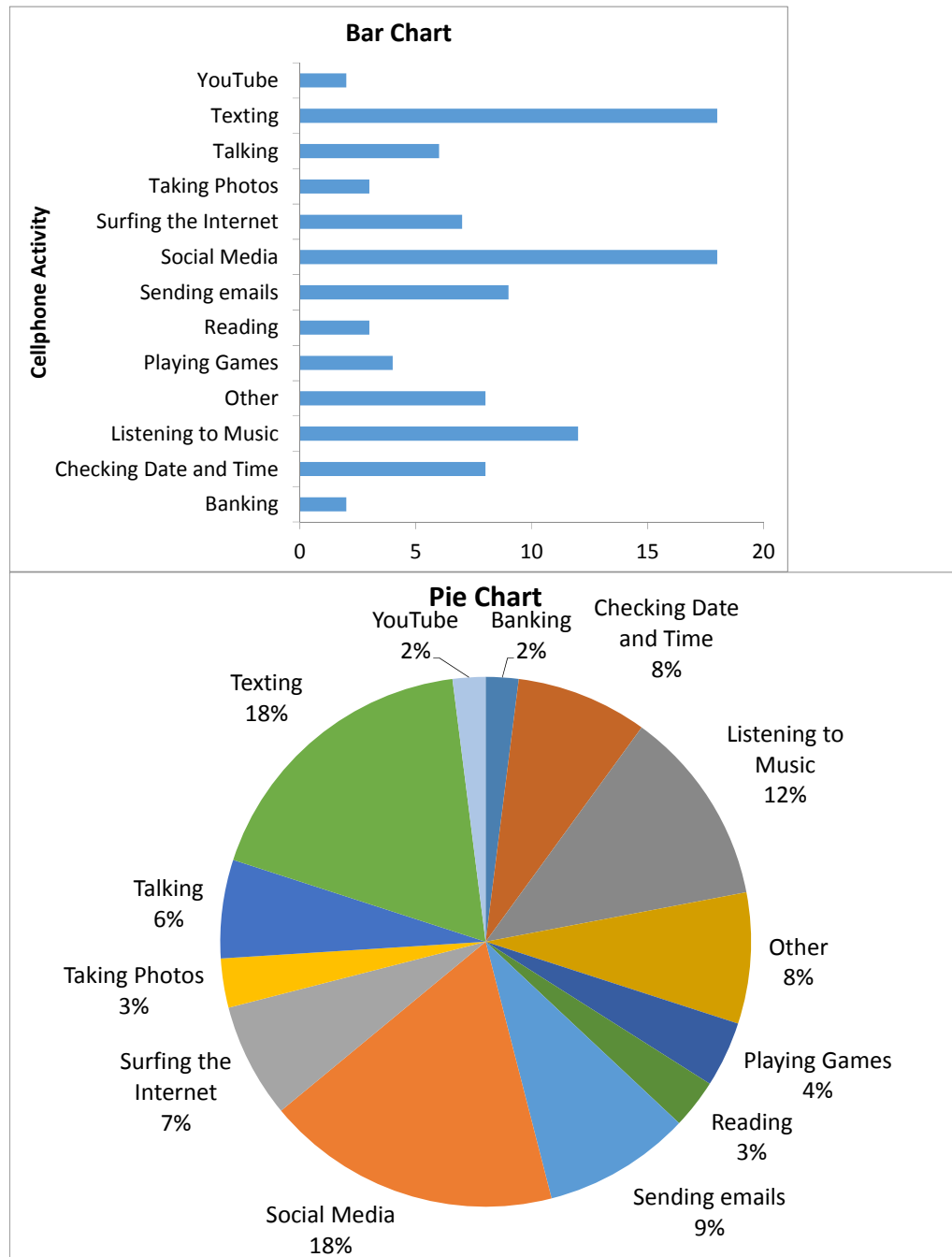
2.25 (a)
cont.
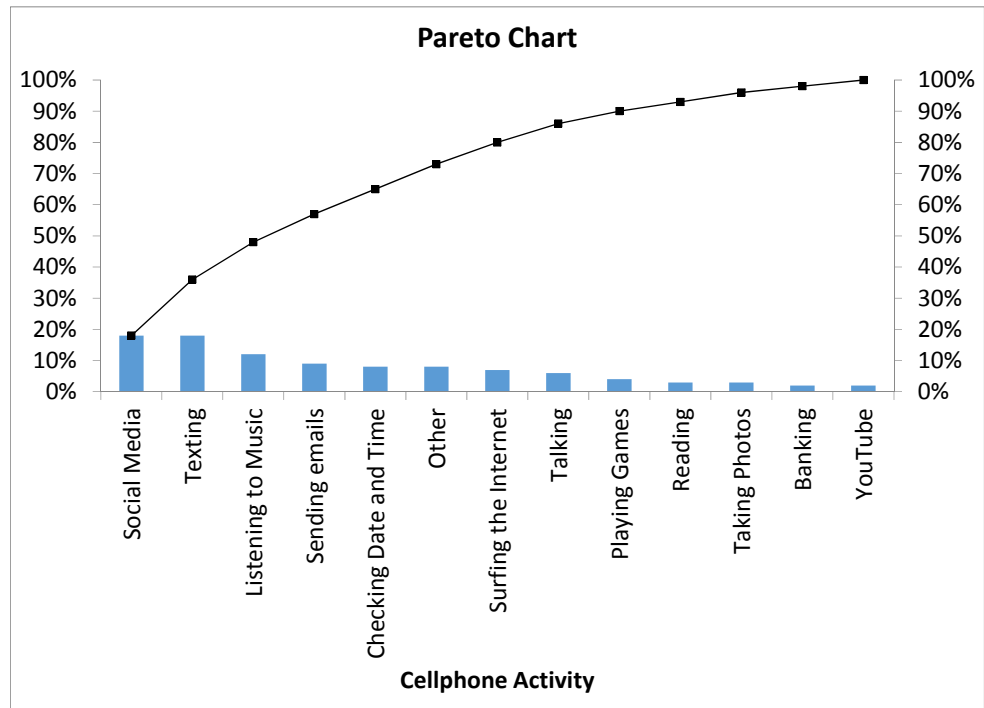
Pareto Chart

(b) The Pareto diagram is better than the pie chart or the bar chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

(c) From the Pareto diagram, it is obvious that slightly around 35% of them use their cell phones for social media, texting.

2.26 (a)
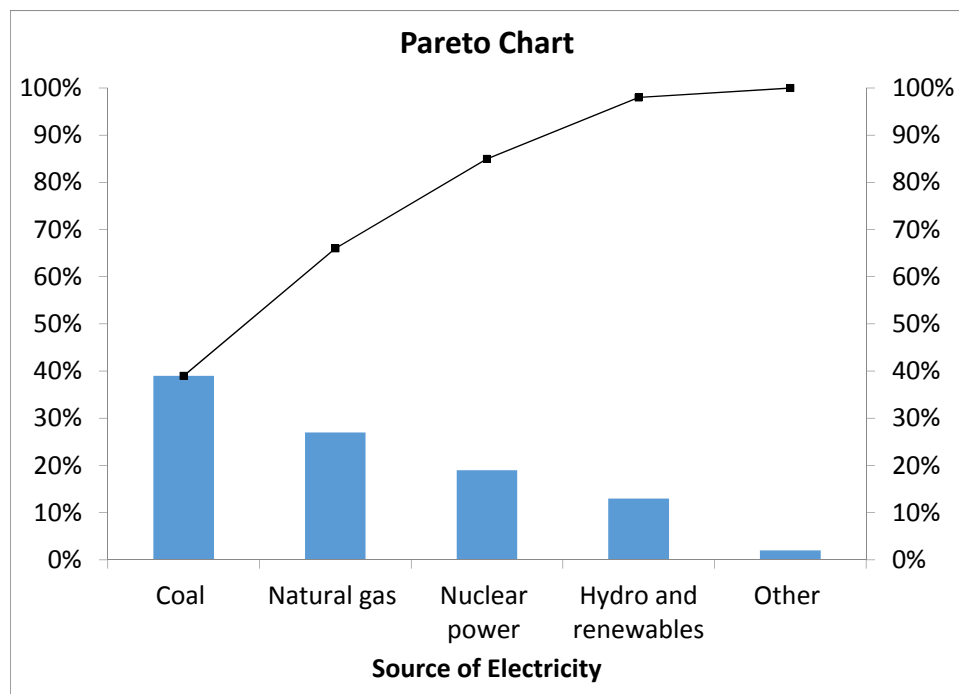
Pareto Chart

2.26   (b)   Eighty-five percent of power is derived from coal, natural gas, or nuclear power.
cont.   (c)
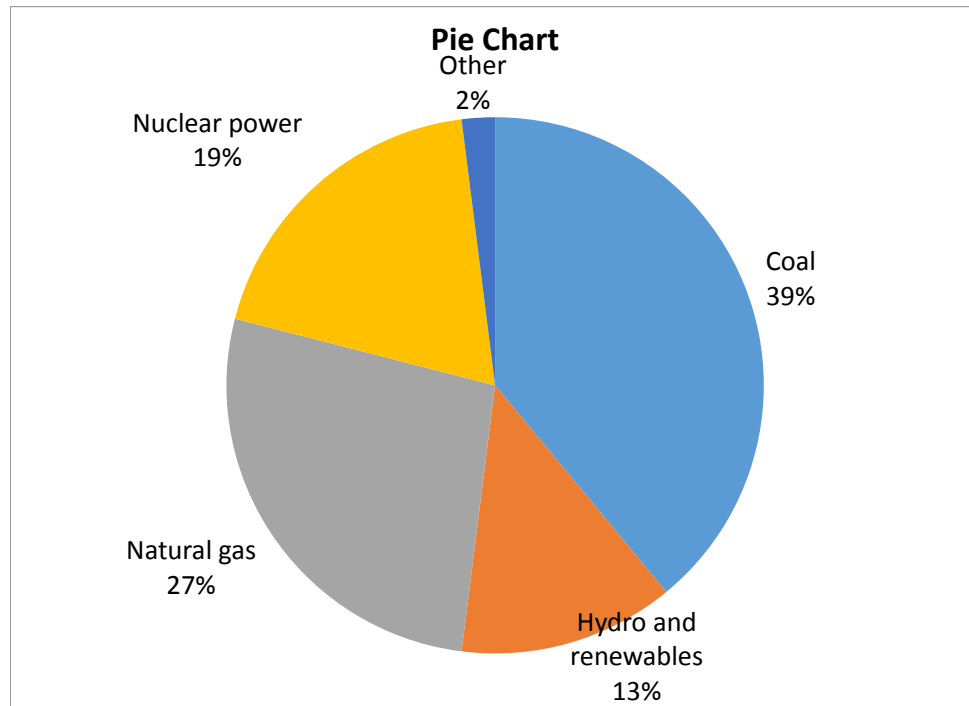
(d)   The Pareto diagram is better than the pie chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

2.27   (a)

2.27   (a)
cont.

Pie Chart

Toyota Motor Sales
11%

American Honda
6%

Chrysler LLC
15%

Other
18%

Ford Motor Company
15%

Nissan Motors Corporation
16%

General Motors
19%
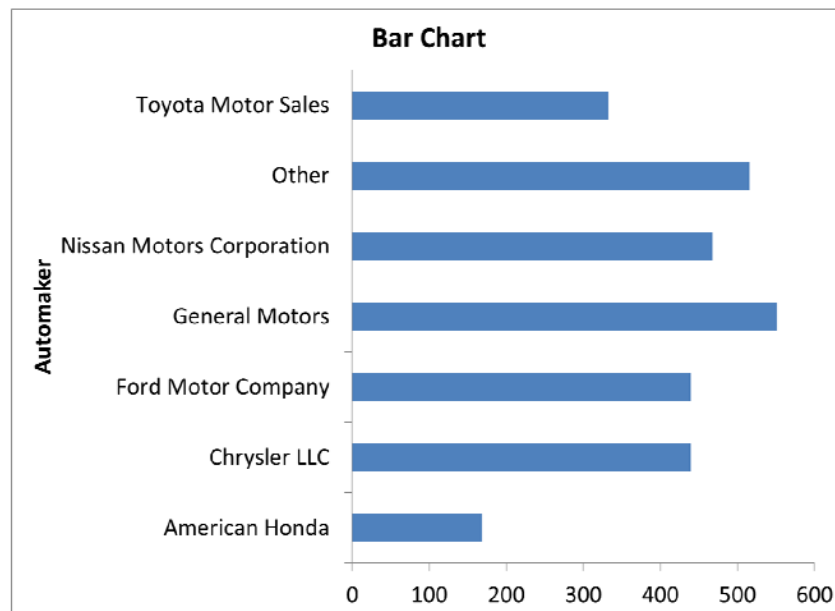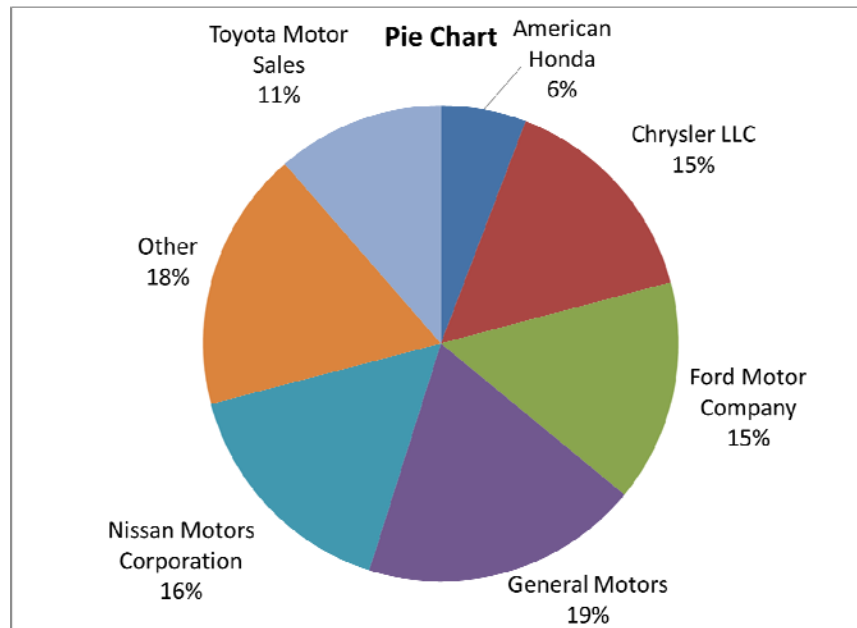
(b)   The bar chart is more suitable if the purpose is to compare the categories. The pie chart is more suitable if the main objective is to investigate the portion of the whole that is in a particular category. *

* Note: This is one of the many possible solutions for the question.

(c)

Pareto Chart

| | 100% | 90% | 80% | 70% | 60% | 50% | 40% | 30% | 20% | 10% | 0% |

Category: Powertrain, Steering, Interior Electronics/Hardware, Fuel/Emission/Exhaust System, Airbags and Seatbelts, Body and Glass, Brakes, Tires and Wheels
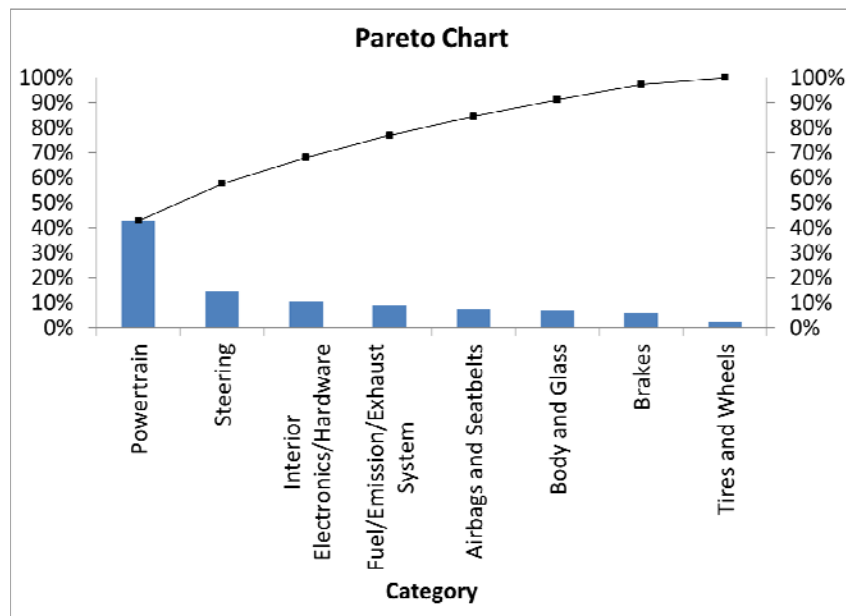
(d)   The "vital few" reasons for the categories of complaints are "powertrain", "steering", and "interior electronics/hardware" which account for more than 68% of the complaints. The remaining reasons are the "trivial many" which make up less than 32% of the complaints.

2.28   (a)

Bar Chart

Wet cleaning	
Water heating	
Refrigeration	
Other	
Lighting	
Heating	
Electronics	
Cooling	
Cooking	

Type of Use Percentage

0   0.05   0.1   0.15   0.2   0.25

Pie Chart

Cooking 2%
Cooling 15%
Electronics 9%
Heating 15%
Lighting 13%
Other 23%
Refrigeration 10%
Water heating 10%
Wet cleaning 3%
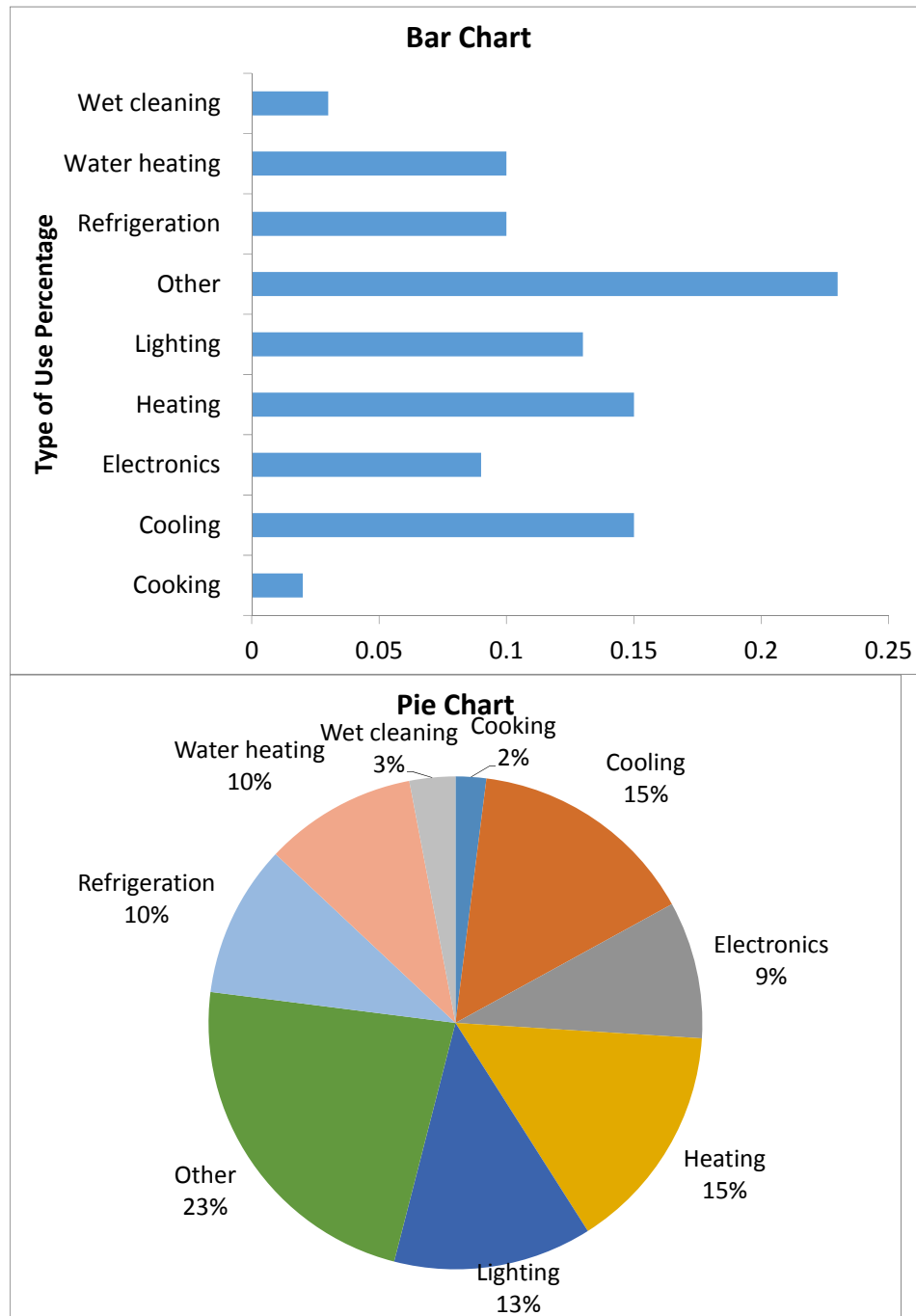
2.28   (a)
cont.
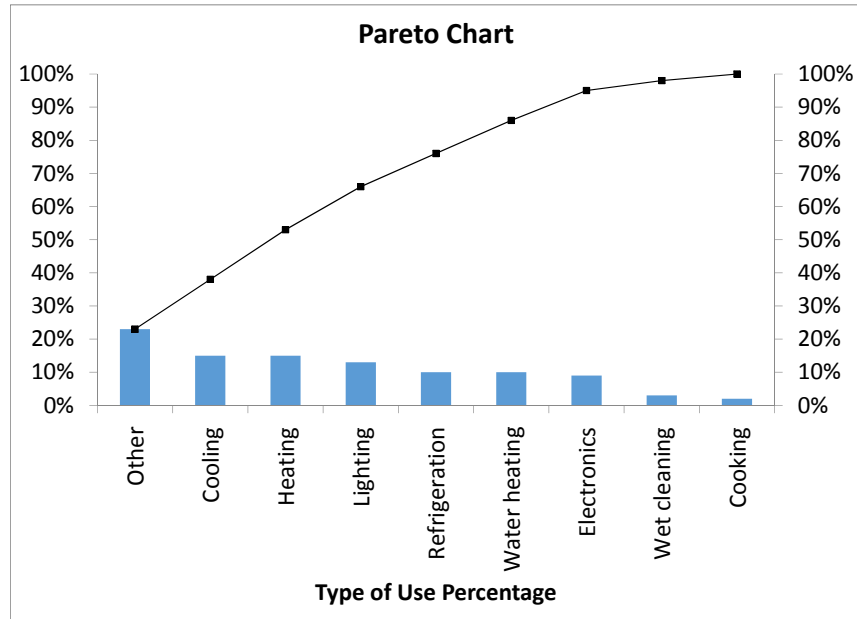
(b)   The Pareto diagram is better than the pie chart and bar chart because it not only sorts the frequencies in descending order; it also provides the cumulative polygon on the same scale.

(c)   Other, cooling, heating and lighting accounted for 66% of the residential electricity consumption in the United States.

2.29   (a)
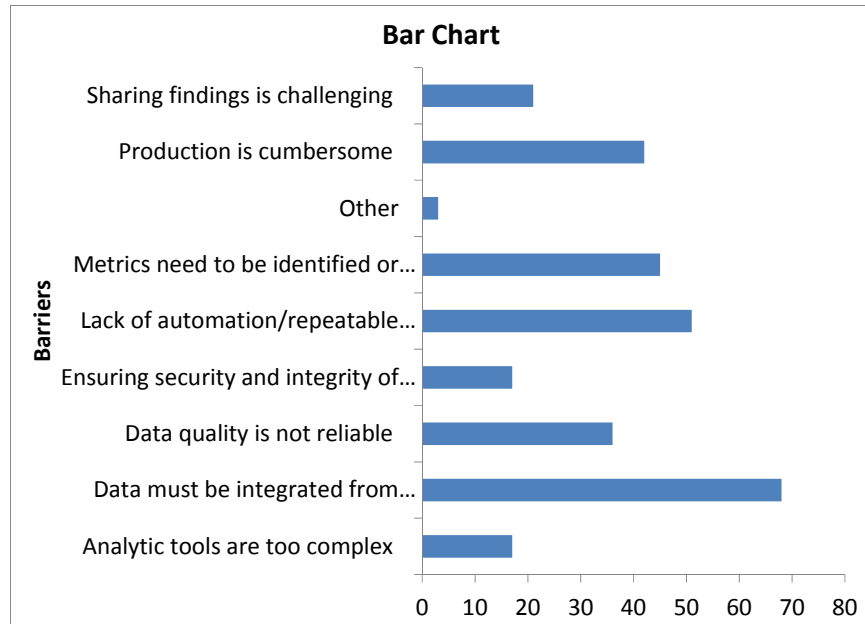
2.29   (a)
cont.
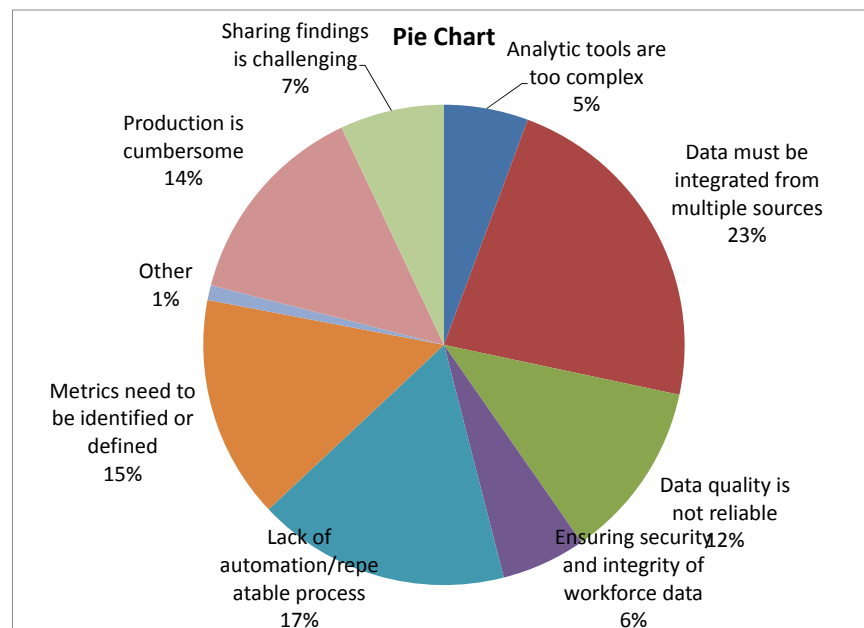
(b)   The highest percentage of technical barriers to workforce analytics is "data must be integrated from multiple sources" at 23% followed by "lack of automation/repeatable process" at 17%, "metrics need to be identified or defined" at 15% and "production is cumbersome" at 14%.

2.30   Whether you are a corporate affairs officer at the company or one of the pension fund managers seeking to ensure an independent board of directors, you should choose Chart 1 because it provides a more direct visual comparison in composition before and after the reform.

2.31   (a)
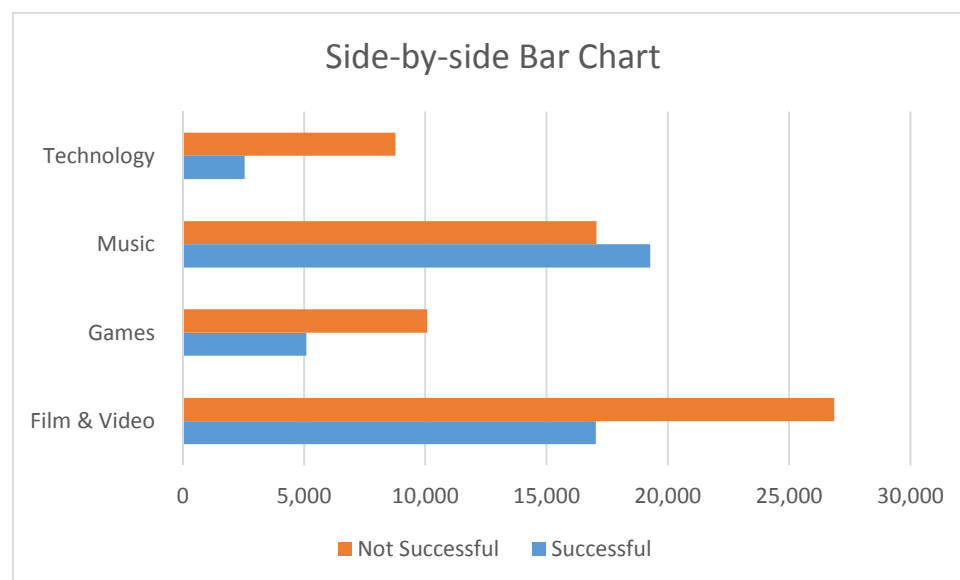
(b)   Music projects have the highest percentage of success while technology projects have the lowest.

2.32   (a)

Side-by-side Bar Chart
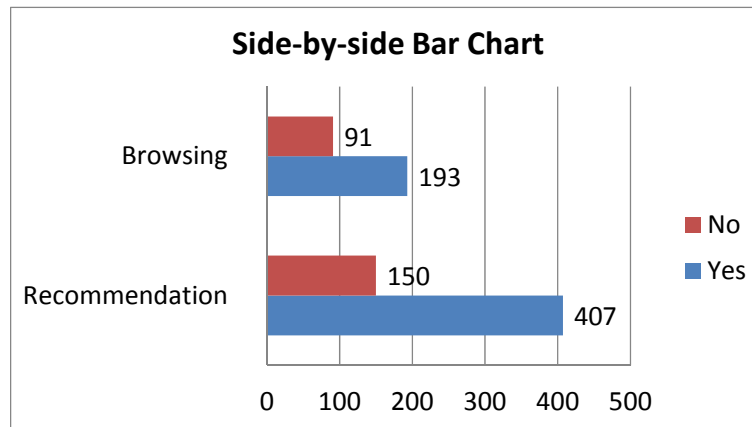
(b)   Social recommendations had very little impact on correct recall.

2.33   Stem-and-leaf of Finance Scores

        5   34
        6   9
        7   4
        9   38

2.34   Ordered array: 50 74 74 76 81 89 92

2.35   (a)   Ordered array: 9.1 9.4 9.7 10.0 10.2 10.2 10.3 10.8 11.1 11.2
                        11.5 11.5 11.6 11.6 11.7 11.7 11.7 12.2 12.2 12.3
                        12.4 12.8 12.9 13.0 13.2
        (b)   The stem-and-leaf display conveys more information than the ordered array. We can
               more readily determine the arrangement of the data from the stem-and-leaf display than
               we can from the ordered array. We can also obtain a sense of the distribution of the data
               from the stem-and-leaf display.
        (c)   The most likely gasoline purchase is between 11 and 11.7 gallons.
        (d)   Yes, the third row is the most frequently occurring stem in the display and it is located in the
               center of the distribution.

2.36   (a)

		Stem-and-Leaf Display	
		Stem unit 100	
Statistics		2	1 2 2 3 5 6 6 6 7 8 8 8 8 9
Sample Size	30	3	0 1 2 2 3 4 4 4 7
Mean	333.5263	4	0 4 4 7 8
Median	303.8000	5	4
Std. Deviation	105.3008	6	8
Minimum	212.4000		
Maximum	676.4200		

        (b)   The costs are concentrated around $200 and $370.

2.37   (a)   Ordered array:

Minimum Order for Free Shipping ($)		0	25	35	45	50
		75	99	150	175	195

(b)   Stem-and-leaf plot

Statistics			0	0
Sample Size	10		1	
Mean	84.9000		2	5
Median	62.5000		3	5
Std. Deviation	67.3885		4	5
Minimum	0.0000		5	0
Maximum	195.0000		6	
			7	5
			8	
			9	9
			10	
			11	
			12	
			13	
			14	
			15	0
			16	
			17	5
			18	
			19	5

(c)   The stem-and-leaf display usually conveys more information than the ordered array. We can more readily determine the arrangement of the data from the stem-and-leaf display than we can from the ordered array. We can also obtain a sense of the distribution of the data from the stem-and-leaf display. However, with just 10 data points, the stem-and-leaf display does not show its relative advantages over the ordered array.

(d)   The minimum online order required to receive free shipping is not concentrated around any value.

2.38   (a)
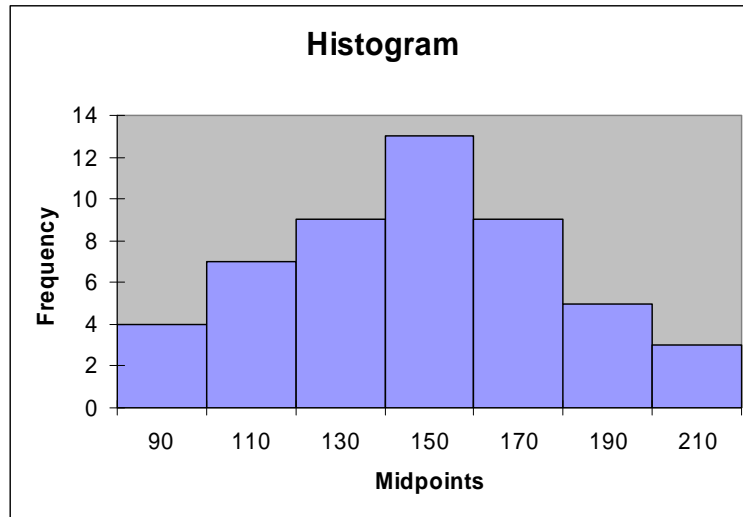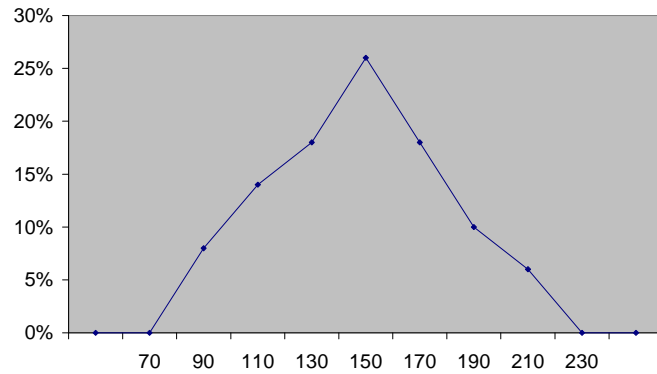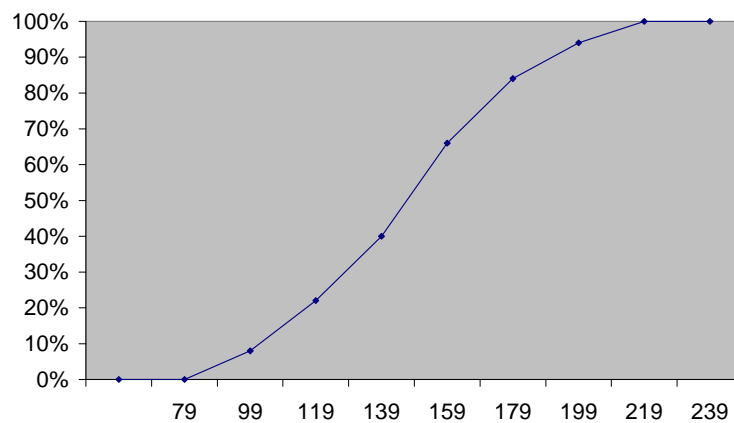
(b)

(c)   The majority of utility charges are clustered between $120 and $180.

2.39   The costs of attending a baseball game is concentrating between $160 and $240. There are a few
        outliers in the right tail with two teams having a cost higher than $300.

2.40   Property taxes seem concentrated between $1,000 and $1,500 and also between $500 and $1,000
        per capita. There were more states with property taxes per capita below $1,500 than above $1,500.

2.41   (a)
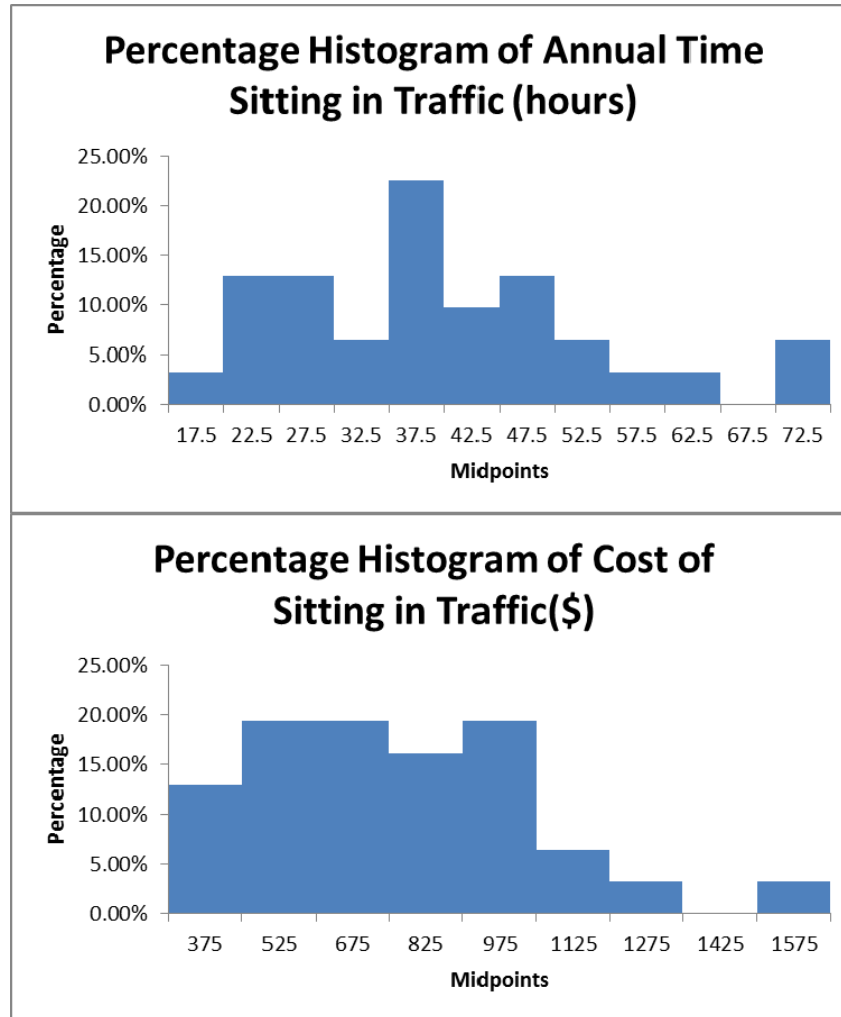
Percentage Histogram of Annual Time Sitting in Traffic (hours)

Percentage Histogram of Cost of Sitting in Traffic($)

2.41   (b)
cont.

Cumulative Percentage Polygon

Cumulative Percentage Polygon
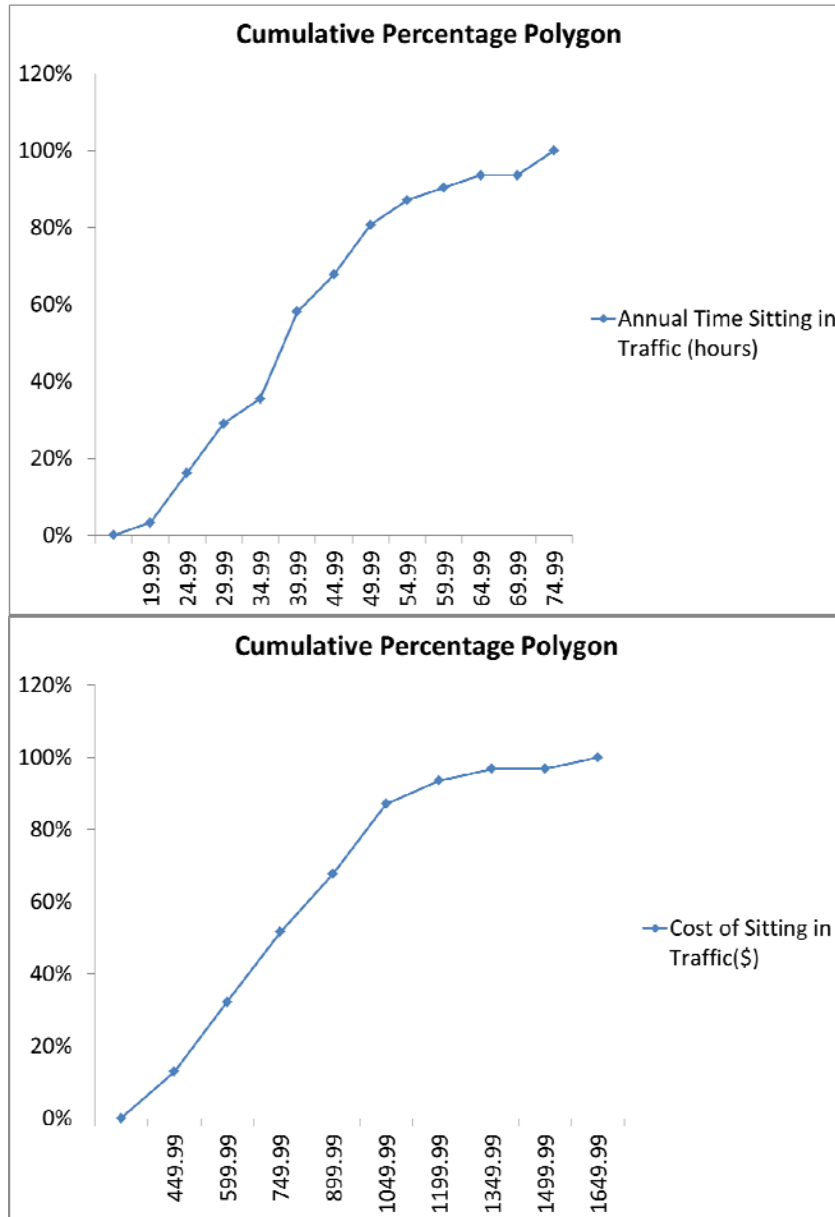
(c)   The annual time sitting in traffic is concentrated around 37.5 hours with a few spending as much as around 72.5 hours.

(d)   The cost of sitting in traffic per year is concentrated around $675 with one costing as much as $1,575.

2.42   (a)

Percentage Histogram of Average Credit Score

(b)

Cumulative Percentage Polygon
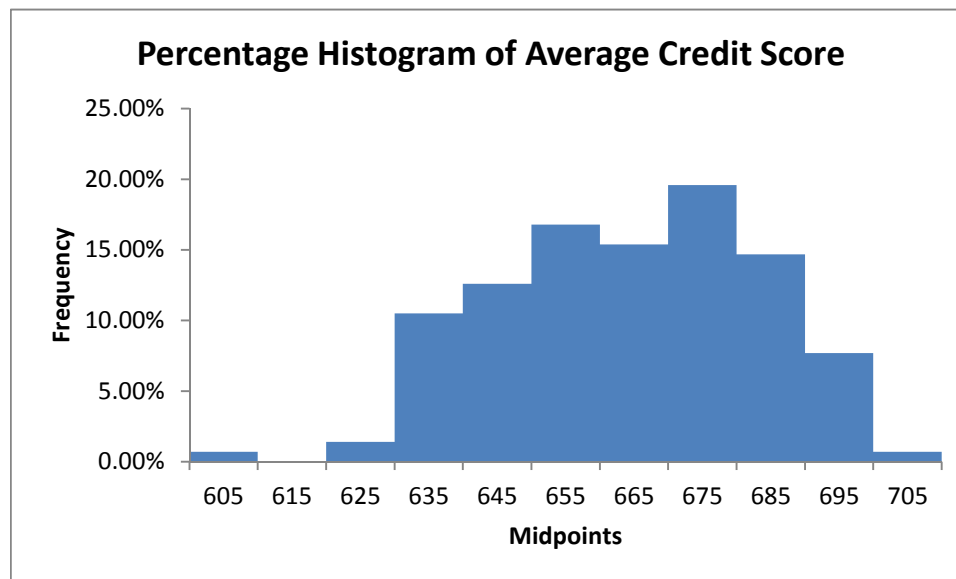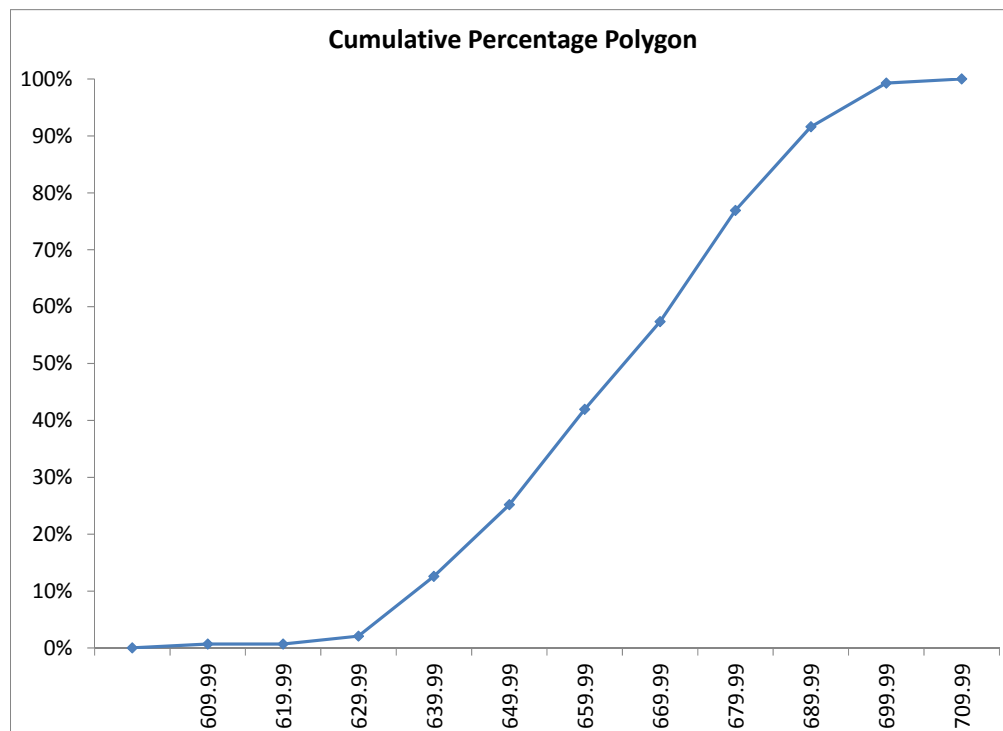
(c)   The average credit scores are concentrated between 630 and 690.

2.43   (a)
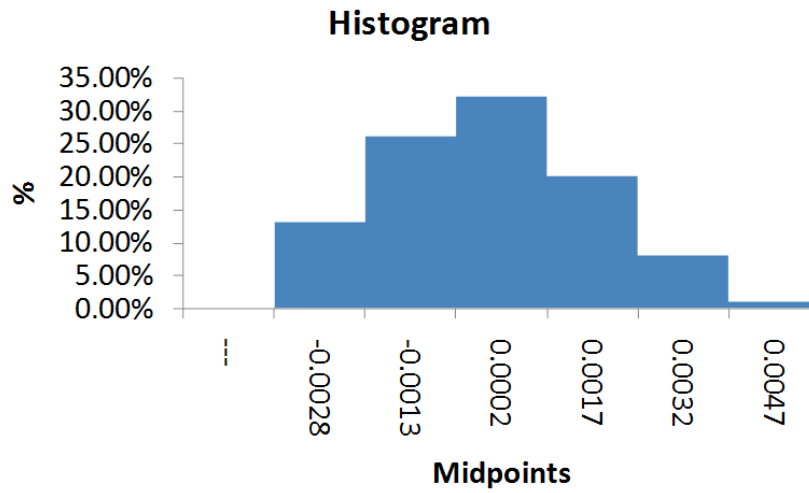
(b)   Yes, the steel mill is doing a good job at meeting the requirement as there is only one steel part out of a sample of 100 that is as much as 0.005 inches longer than the specified requirement.

2.44   (a)
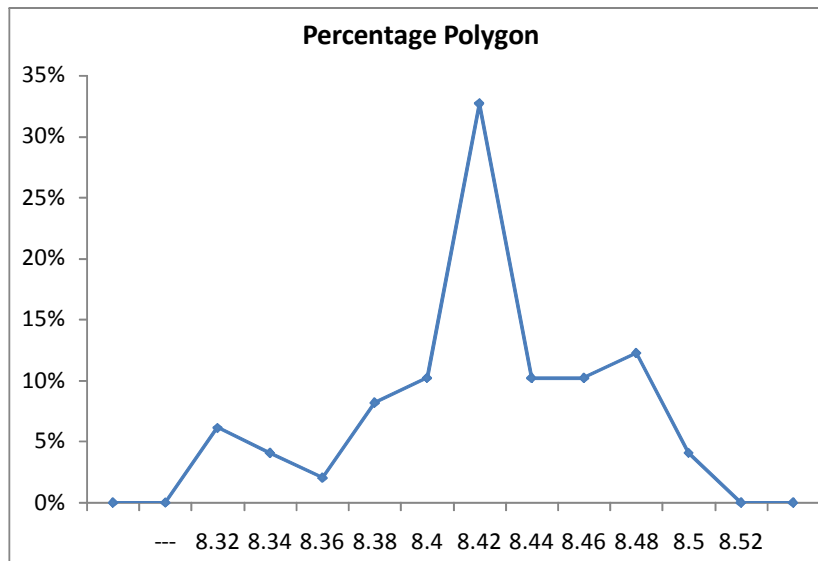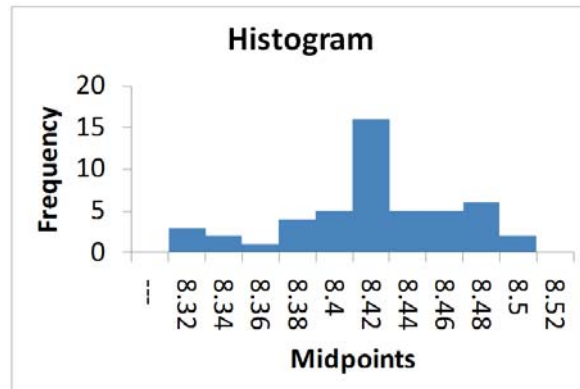
2.44   (b)
cont.

Cumulative Percentage Polygon
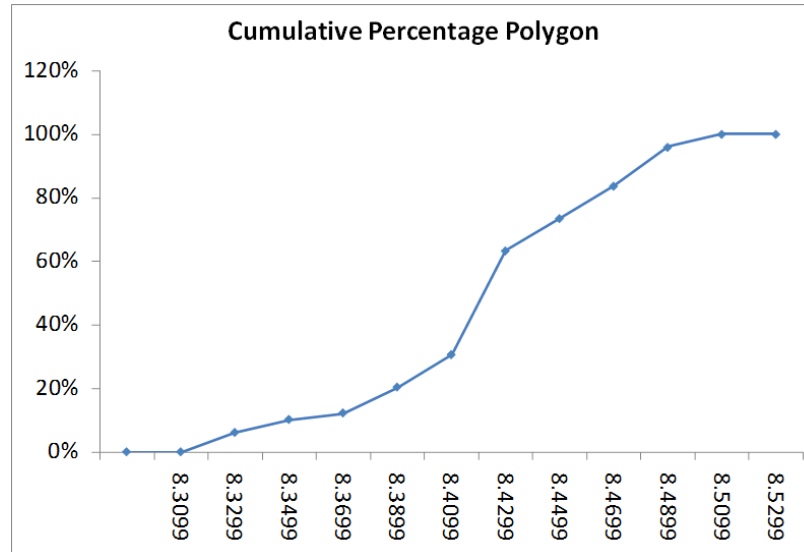
(c)   All the troughs will meet the company's requirements of between 8.31 and 8.61 inches wide.

2.45   (a)

%Histogram
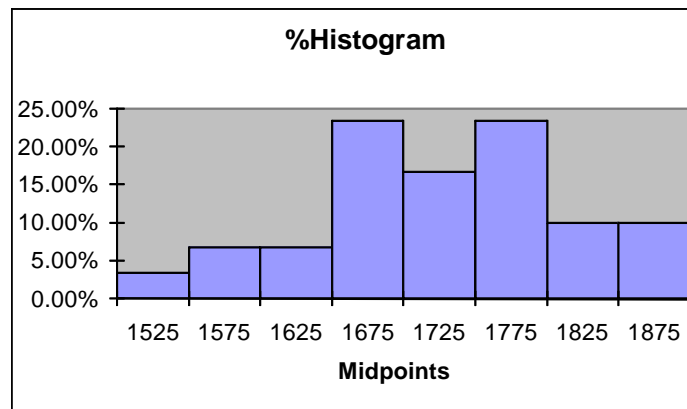
Percentage Polygon
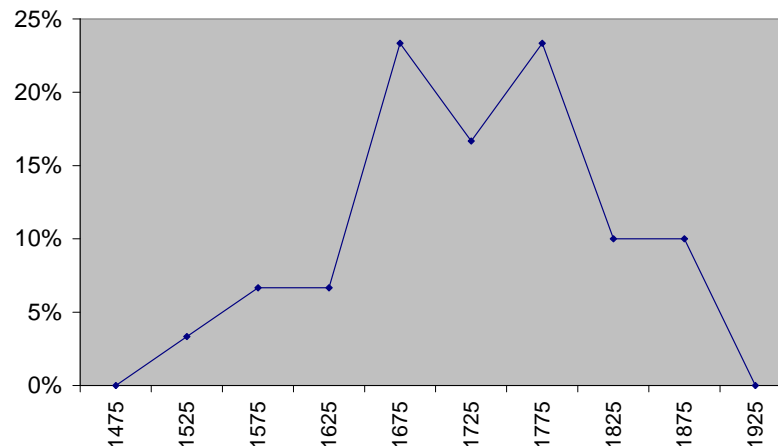
2.45   (b)
cont.

Cumulative Percentage Polygon
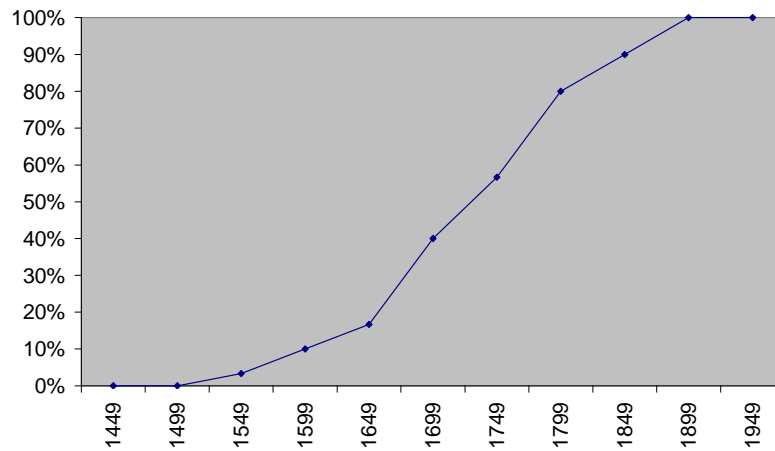
(c)   The strength of all the insulators meets the company's requirement of at least 1500 lbs.

2.46   (a)
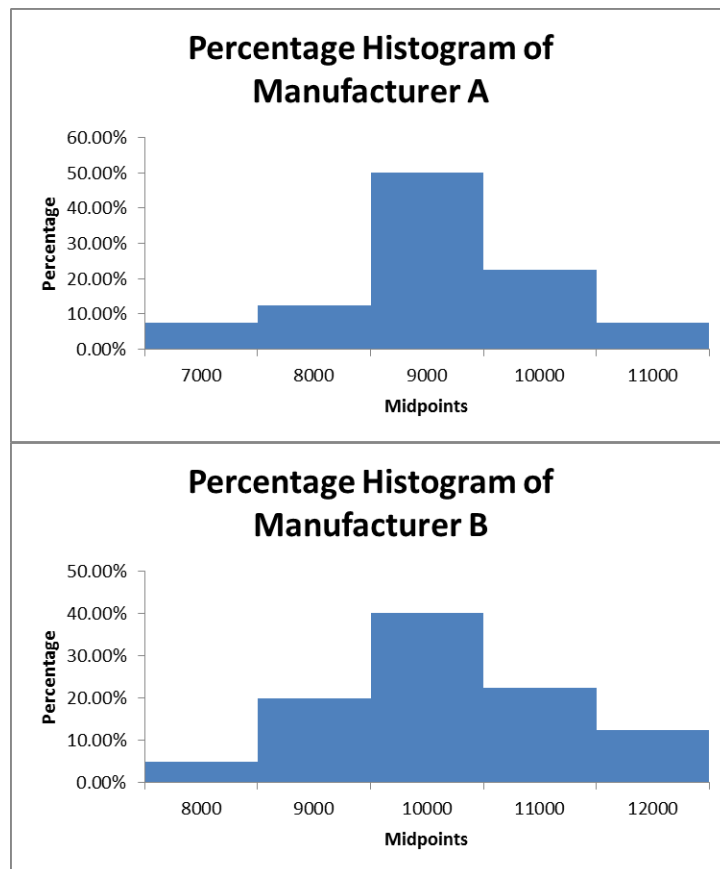
2.46   (a)
cont.
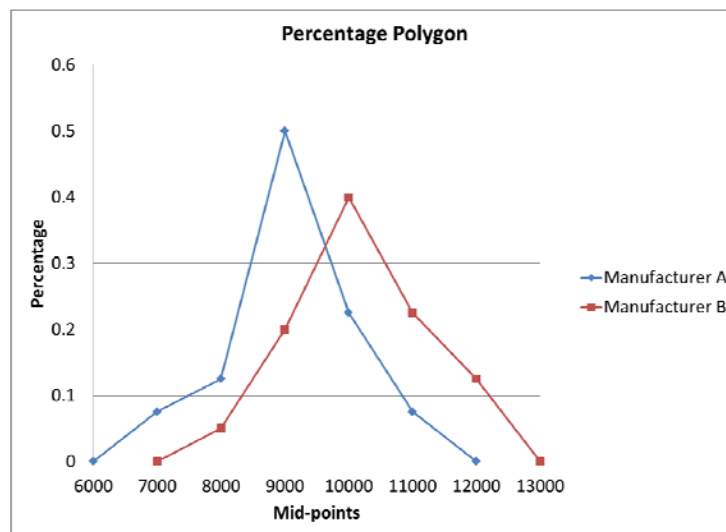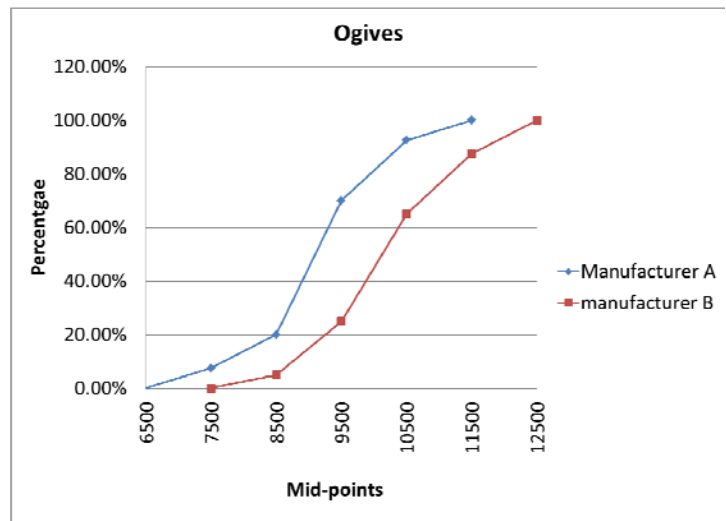
Percentage Polygon

(b)

Ogives

(c)   Manufacturer B produces bulbs with longer lives than Manufacturer A. The cumulative percentage for Manufacturer B shows 65% of their bulbs lasted 10499 hours or less contrasted with 70% of Manufacturer A's bulbs which lasted 9499 hours or less. None of Manufacturer A's bulbs lasted more than 11499 hours, but 12.5% of Manufacturer B's bulbs lasted between 11500 and 12499 hours. At the same time, 7.5% of Manufacturer A's bulbs lasted less than 7500 hours, while all of Manufacturer B's bulbs lasted at least 7500 hours.

2.47   (a)

Histogram

Percentage Polygon

(b)

Amount of Soft Drink	Frequency Less Than	Percentage Less Than
1.899	1	2%
1.949	6	12
1.999	24	48
2.049	43	86
2.099	49	98
2.149	50	100

Cumulative Percentage Polygon
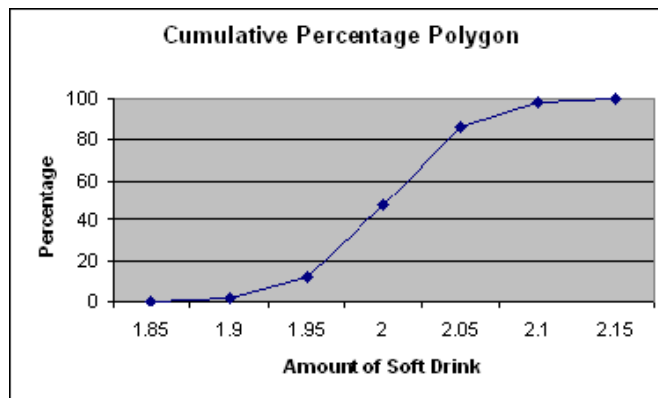
(c)   The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.48   (a)

    (b)   There is no relationship between *X* and *Y*.

2.49   (a)
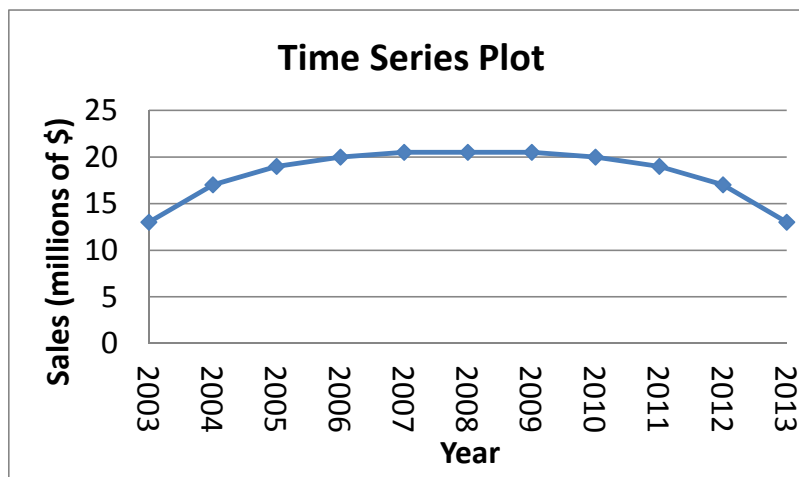
    (b)   Annual sales appear to be increasing in the earlier years before 2006 but start to decline after 2008.

2.50   (a)

2.50   (b)
cont.

Scatter Plot
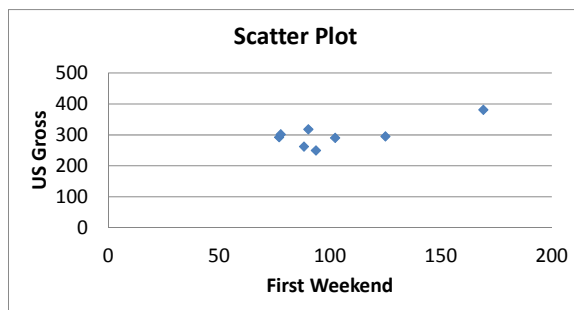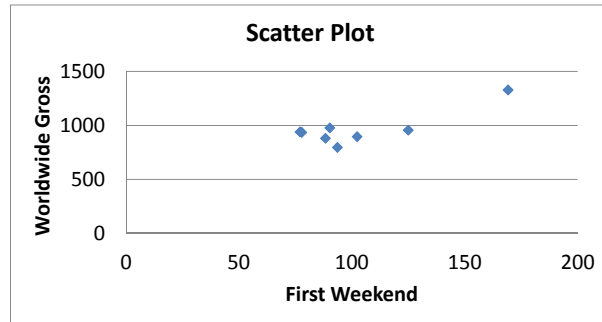
(c)   There appears to be a linear relationship between the first weekend gross and either the U.S. gross or the worldwide gross of Harry Potter movies. However, this relationship is greatly affected by the results of the last movie, *Deathly Hallows*, *Part II*.

2.51   (a)
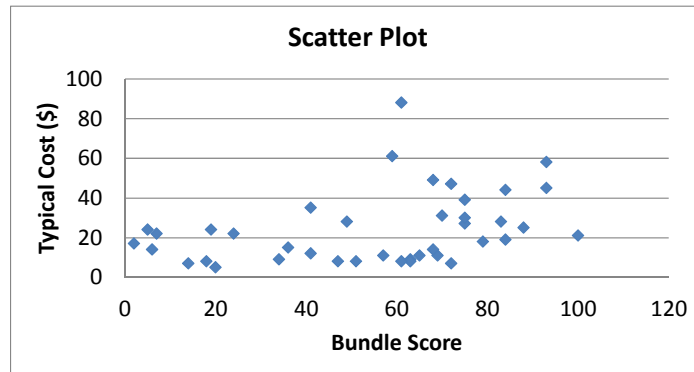
Scatter Plot

(b)   There appears to be a positive relationship between Bundle score and typical cost.

2.52   (a)   Yes, schools with higher revenues will also have higher coaches' total pay.
        (b)
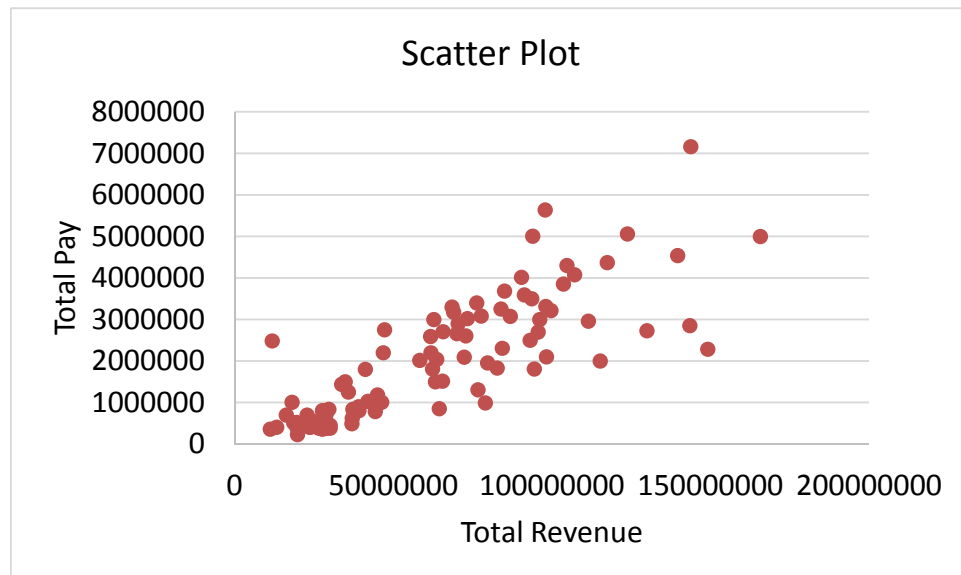
Scatter Plot

(c)   The scatter plot confirms your answer to (a).

2.53   (a)
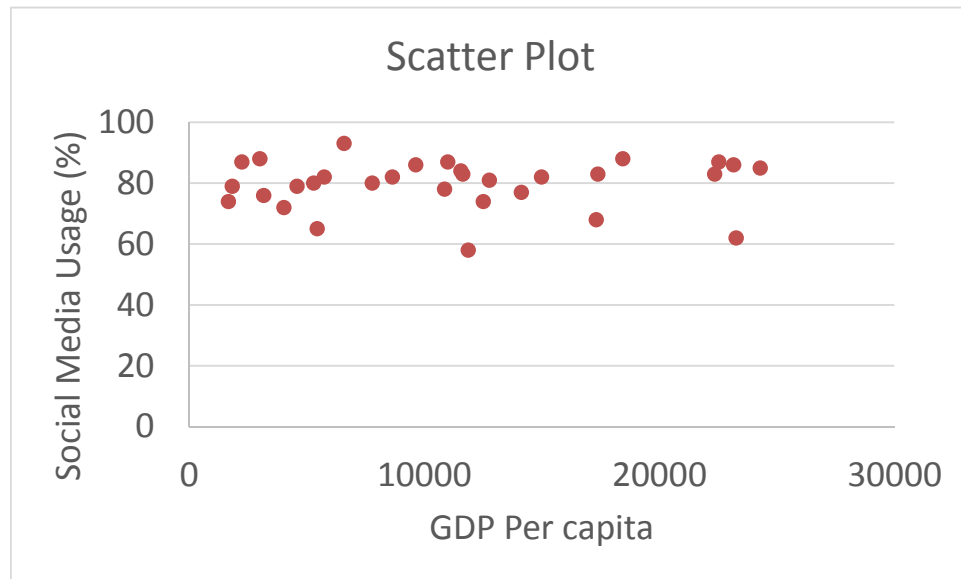
(b)   There does not appear to be any relationship between GDP and social media usage.

(c)
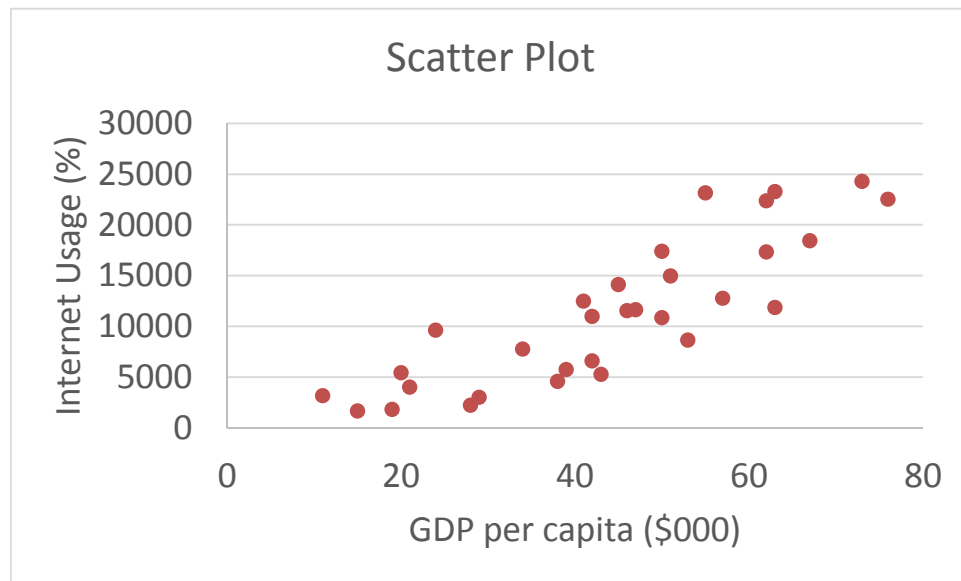
(d)   There is a positive relationship between GDP and internet usage.

2.54   (a)   Excel output:

Time-series Plot
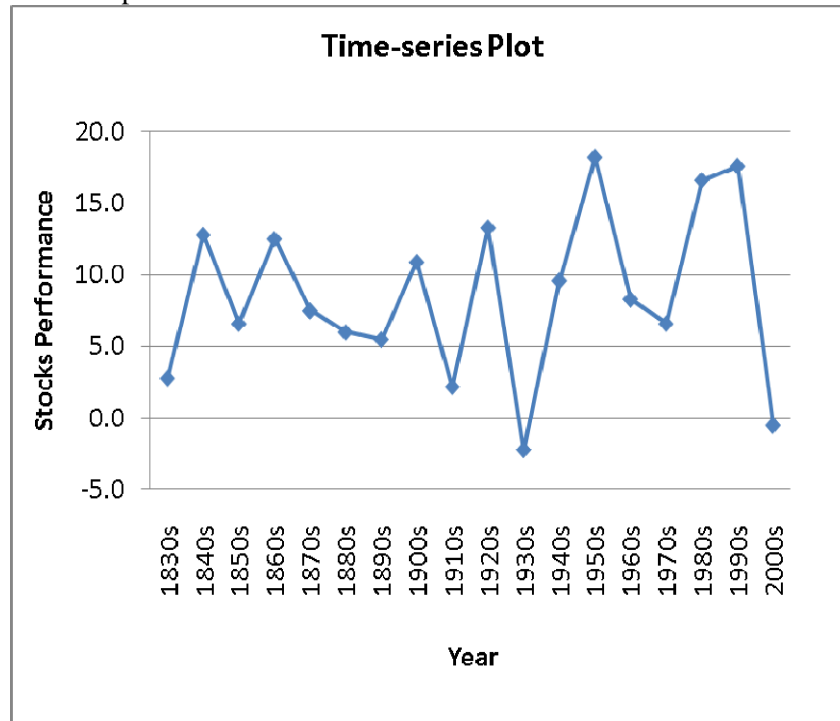
(b)   There is a great deal of variation in the returns from decade to decade. Most of the returns are between 5% and 15%. The 1950s, 1980s, and 1990s had exceptionally high returns, and only the 1930s and 2000s had negative returns.

2.55   (a)

Time-Series Plot
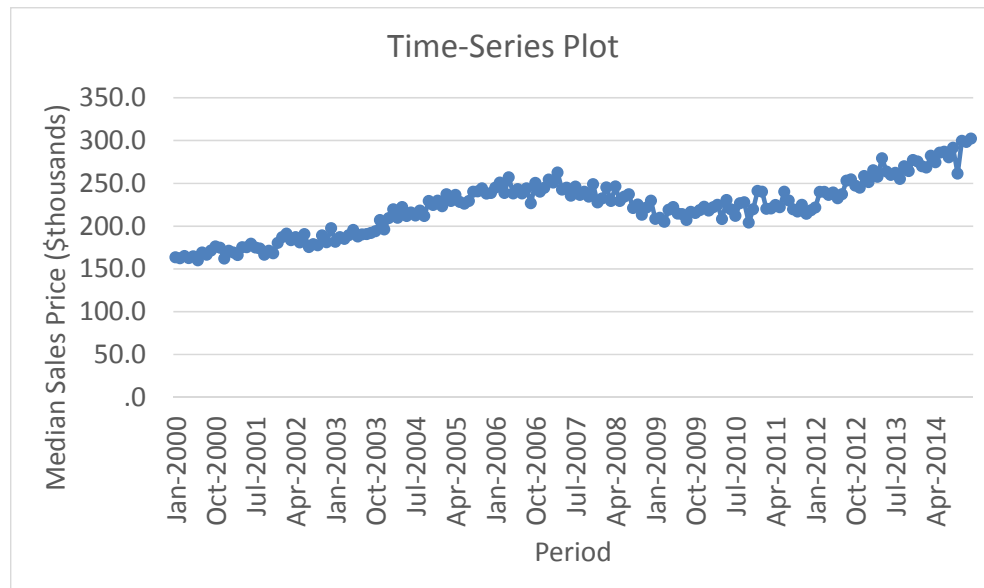
(b)   There is an upward trend on the median home sales price till 2007 and the sales price started a downward trend from then on till 2009 when it started to trend up again.

2.56   (a)
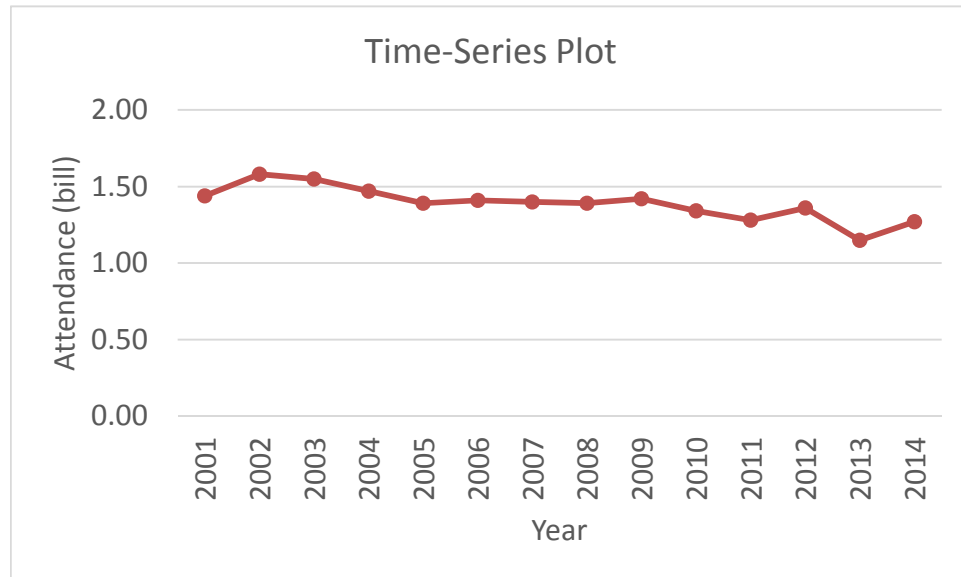
  (b)   There was a slight decline in movie attendance between 2001 and 2014. During that time, movie attendance increased from 2001 to 2002 but then after 2004 began decreasing to levels below that in 2001.

2.57   (a)
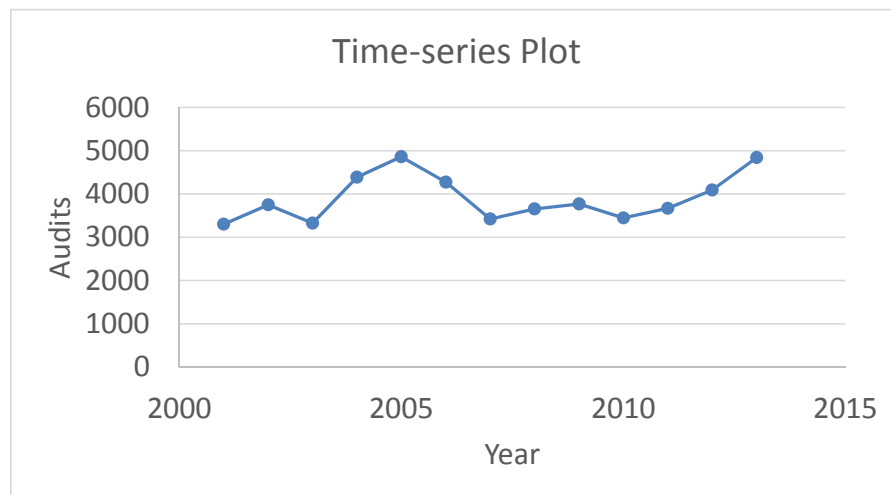
  (b)   The number of audits increased from 2001 to 2005, then declined back to the 2001 level in 2007, hover around the same level from then on until 2010 and increased again after 2010.

2.58   (a)   Pivotal table of tallies in terms of counts:

Count of 3YrReturn%	Column Labels					
Row Labels	Five	Four	One	Three	Two	Grand Total
Growth	9	45	19	118	78	269
Large	7	21	9	56	34	127
Mid-Cap	1	17	6	39	26	89
Small	1	7	4	23	18	53
Value	2	27	9	60	40	138
Large	1	16	5	39	22	83
Mid-Cap	1	6	3	10	10	30
Small		5	1	11	8	25
Grand Total	11	72	28	178	118	407

Pivotal table of tallies in terms of % of grand total:

Count of 3YrReturn%	Column Labels					
Row Labels	Five	Four	One	Three	Two	Grand Total
Growth	2.21%	11.06%	4.67%	28.99%	19.16%	66.09%
Large	1.72%	5.16%	2.21%	13.76%	8.35%	31.20%
Mid-Cap	0.25%	4.18%	1.47%	9.58%	6.39%	21.87%
Small	0.25%	1.72%	0.98%	5.65%	4.42%	13.02%
Value	0.49%	6.63%	2.21%	14.74%	9.83%	33.91%
Large	0.25%	3.93%	1.23%	9.58%	5.41%	20.39%
Mid-Cap	0.25%	1.47%	0.74%	2.46%	2.46%	7.37%
Small	0.00%	1.23%	0.25%	2.70%	1.97%	6.14%
Grand Total	2.70%	17.69%	6.88%	43.73%	28.99%	100.00%

(b)   Patterns of star rating conditioned on market cap:
For the growth funds as a group, most are rated as three-star, followed by two-star, four-star, one-star and five-star. The pattern of star rating is similar across the different market cap within the growth funds.
For the value funds as a group, most are rated as three-star, followed by two-star, four-star, one-star and five-star. Within the value funds, the large-cap and small-cap funds follow the same pattern as the value funds as a group. Most of the mid-cap funds are rated as three-star and two-star, followed by four-star, one-star and five-star.
Patterns of market cap conditioned on star rating:
Most of the growth funds are large-cap, followed by mid-cap and small-cap. The pattern is similar among the four-star, three-star, two-star and one-start growth funds but among the five-star growth funds, most are large-cap, followed by equal portions of small-cap and mid-cap.
The largest share of the value funds is large-cap, followed by mid-cap and small-cap. The pattern is similar among the four-star, three-star, two-star and one-star value funds. Among the five-star value funds, there are equal portions of mid-cap and large-cap with no small-cap funds.

2.58   (c)   The average three-year return for each type, market cap, and rating.
cont.

Average of 3YrReturn%	Column Labels					
Row Labels	Five	Four	One	Three	Two	Grand Total
Growth	**19.1144**	**17.0080**	**12.0311**	**15.4837**	**13.8681**	**15.1478**
Large	19.3957	17.9414	13.4389	16.3645	15.0629	16.2365
Mid-Cap	18.1000	16.1865	11.1133	14.7990	13.5696	14.4935
Small	18.1600	16.2029	10.2400	14.5004	12.0422	13.6379
Value	**19.4600**	**16.7463**	**9.9322**	**15.5308**	**14.4315**	**15.1418**
Large	17.3200	16.7813	8.7980	15.1287	14.8109	15.0081
Mid-Cap	21.6000	18.8950	10.9767	18.8030	15.3200	16.9710
Small		14.0560	12.4700	13.9818	12.2775	13.3908
Grand Total	**19.1773**	**16.9099**	**11.3564**	**15.4996**	**14.0591**	**15.1458**

(d)   There are 56 large cap growth funds with a rating of three. Below are the summary statistics for the three-year return:

	Assets	Turnover Ratio(%)	SD	Sharpe Ratio	Expense Ratio	1YrReturn%	3YrReturn%	5YrReturn%	10YrReturn%
Mean	1569.913393	55.77178571	10.76696	1.483928571	1.086785714	11.01410714	16.36446429	14.69053571	7.881607143
Standard Error	310.8859493	5.770010607	0.176072	0.025782153	0.029372338	0.479575961	0.211876728	0.178673034	0.214470017
Median	494.945	45.67	10.59	1.485	1.1	11.52	16.455	14.645	7.59
Mode	#N/A	52	11.88	1.41	0.97	#N/A	17.4	14.59	6.8
Standard Deviation	2326.457418	43.17880562	1.317601	0.192935963	0.21980245	3.588817874	1.585540252	1.337066558	1.604946645
Sample Variance	5412404.116	1864.409255	1.736072	0.037224286	0.048313117	12.87961373	2.51393789	1.787746981	2.575853734
Kurtosis	3.48732685	5.005942351	8.467439	-0.399428532	1.175285073	0.352492297	0.300157815	1.402229633	19.64901039
Skewness	1.996776548	1.890223	2.274308	-0.175188156	0.167532557	-0.476041631	-0.140678531	-0.330129068	3.668370541
Range	9832.5	226.51	7.91	0.81	1.25	18.35	7.63	7.69	11.41
Minimum	28.75	7.49	9.03	1.01	0.53	0.53	12.4	10.43	5.7
Maximum	9861.25	234	16.94	1.82	1.78	18.88	20.03	18.12	17.11
Sum	87915.15	3123.22	602.95	83.1	60.86	616.79	916.41	822.67	441.37
Count	56	56	56	56	56	56	56	56	56

2.59   (a)   Pivotal table of tallies in terms of counts:

Count of 3YrReturn%	Column Labels					
Row Labels	Five	Four	One	Three	Two	Grand Total
⊟Large	8	37	14	95	56	210
Average	7	29	13	72	46	167
High	1	1	1	2	2	7
Low		7		21	8	36
⊟Mid-Cap	2	23	9	49	36	119
Average	1	21	3	46	27	98
High		1	4	2	9	16
Low	1	1	2	1		5
⊟Small	1	12	5	34	26	78
Average	1	5		18	7	31
High		7	5	16	19	47
Grand Total	11	72	28	178	118	407

Pivotal table of tallies in terms of % of grand total:

Count of 3YrReturn%	Column Labels					
Row Labels	Five	Four	One	Three	Two	Grand Total
⊟Large	1.97%	9.09%	3.44%	23.34%	13.76%	51.60%
Average	1.72%	7.13%	3.19%	17.69%	11.30%	41.03%
High	0.25%	0.25%	0.25%	0.49%	0.49%	1.72%
Low	0.00%	1.72%	0.00%	5.16%	1.97%	8.85%
⊟Mid-Cap	0.49%	5.65%	2.21%	12.04%	8.85%	29.24%
Average	0.25%	5.16%	0.74%	11.30%	6.63%	24.08%
High	0.00%	0.25%	0.98%	0.49%	2.21%	3.93%
Low	0.25%	0.25%	0.49%	0.25%	0.00%	1.23%
⊟Small	0.25%	2.95%	1.23%	8.35%	6.39%	19.16%
Average	0.25%	1.23%	0.00%	4.42%	1.72%	7.62%
High	0.00%	1.72%	1.23%	3.93%	4.67%	11.55%
Grand Total	2.70%	17.69%	6.88%	43.73%	28.99%	100.00%

2.59   (b)   Patterns of star rating conditioned on risk:
cont.   For the large-cap funds as a group, most are rated as three-star, followed by four-star, two-star, five-star and then one-star. The pattern of star rating is the same among the low-risk large-cap funds. The pattern is different among the high-risk and average-risk large-cap funds. Among the high-risk large-cap funds, most are rated as two-star, followed by one three-star with no three-star, four-star or five-star rating. Among the average-risk large-cap funds, most are two-star and three-star, followed by one-star, four-star and five-star rating.

For the mid-cap funds as a group, most are rated as four-star, followed by three-star, two-star, five-star and then one-star. The pattern of star rating is different among the average-risk mid-cap funds with the largest portion of two-star, followed by three-star, four-star, one-star and five-star. Among the low-risk mid-cap funds, most are rated as four-star, followed by three-star, five-star, two-star and one-star.

For the small-cap funds as a group, most are rated as three-star, followed by four-star, two-star, one-star and then five-star. Among the average-risk small-cap funds, most are three-star, followed by two-star, four-star, one-star and five-star. Among the high-risk small-cap funds, most are rated as one-star, followed by equal portions of two-star, three-star and four-star and no five-star. Among the low-risk small-cap funds, most are four-star, followed by three-star and equal portions of two-star and five-star with none rated as one-star.

Patterns of risk conditioned on star rating:
Among the large-cap funds, most are low-risk, followed by average-risk and finally high-risk. The pattern is the same among the one-star, two-star, three-star, four-star and five-star large-cap funds. Among the mid-cap funds, most are low-risk, followed by average-risk with no high-risk. The pattern is the same among the five-star, four-star and three-star mid-cap funds.

Among the small-cap funds, most are average-risk, followed by low-risk and finally high-risk. The pattern is the same for the two-star and three-star small-cap funds. Among the one-star small-cap funds, most are high-risk, followed by average-risk with no low-risk. Among the four-star and five-star small-cap funds, most are low-risk, followed by average-risk and high-risk.

        (c)   The average three-year return for each market cap, risk, and rating.

Average of 3YrReturn%	Column Labels					
Row Labels	Five	Four	One	Three	Two	Grand Total
Large	19.1363	17.4397	11.7814	15.8572	14.9639	15.7510
Average	18.8986	17.8693	12.5423	16.2635	15.3628	16.1150
High	20.8000	19.0200	1.8900	16.7050	16.1700	15.3514
Low		15.4343		14.3833	12.3688	14.1400
Mid-Cap	19.8500	16.8930	11.0678	15.6161	14.0558	15.1181
Average	21.6000	16.5819	10.6400	15.6330	14.1159	15.3264
High		21.4100	9.7925	15.0500	13.8756	13.4725
Low	18.1000	18.9100	14.2600	15.9700		16.3000
Small	18.1600	15.3083	10.6860	14.3326	12.1146	13.5587
Average	18.1600	14.5280		14.1700	10.5371	13.5361
High		15.8657	10.6860	14.5156	12.6958	13.5736
Grand Total	19.1773	16.9099	11.3564	15.4996	14.0591	15.1458

2.59   (d)   There are 2 large cap funds that are high risk with a rating of three. Below are the
cont.   summary statistics for the three-year return:

	Assets	Turnover Ratio(%)	SD	Sharpe Ratio	1YrReturn%	3YrReturn%	5YrReturn%	10YrReturn%	Expense Ratio
Mean	201.675	44	15.725	1.09	1.715	16.705	16.09	10.315	1.285
Standard Error	165.705	5	1.215	0.08	1.185	0.525	0.93	0.815	0.015
Median	201.675	44	15.725	1.09	1.715	16.705	16.09	10.315	1.285
Mode	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Standard Deviation	234.3422584	7.071067812	1.718269478	0.113137085	1.675843071	0.74246212	1.315218613	1.152584053	0.021213203
Sample Variance	54916.29405	50	2.95245	0.0128	2.80845	0.55125	1.7298	1.32845	0.00045
Kurtosis	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
Skewness	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
Range	331.41	10	2.43	0.16	2.37	1.05	1.86	1.63	0.03
Minimum	35.97	39	14.51	1.01	0.53	16.18	15.16	9.5	1.27
Maximum	367.38	49	16.94	1.17	2.9	17.23	17.02	11.13	1.3
Sum	403.35	88	31.45	2.18	3.43	33.41	32.18	20.63	2.57
Count	2	2	2	2	2	2	2	2	2

2.60   (a)   Pivotal table of tallies in terms of counts:

Count of 3YrReturn%	Column Labels					
Row Labels	Five	Four	One	Three	Two	Grand Total
Growth	9	45	19	118	78	269
Average	7	35	11	95	51	199
High	1	7	7	16	23	54
Low	1	3	1	7	4	16
Value	2	27	9	60	40	138
Average	2	20	5	41	29	97
High		2	3	4	7	16
Low		5	1	15	4	25
Grand Total	11	72	28	178	118	407

Pivotal table of tallies in terms of % of grand total:

Count of 3YrReturn%	Column Labels					
Row Labels	Five	Four	One	Three	Two	Grand Total
Growth	2.21%	11.06%	4.67%	28.99%	19.16%	66.09%
Average	1.72%	8.60%	2.70%	23.34%	12.53%	48.89%
High	0.25%	1.72%	1.72%	3.93%	5.65%	13.27%
Low	0.25%	0.74%	0.25%	1.72%	0.98%	3.93%
Value	0.49%	6.63%	2.21%	14.74%	9.83%	33.91%
Average	0.49%	4.91%	1.23%	10.07%	7.13%	23.83%
High	0.00%	0.49%	0.74%	0.98%	1.72%	3.93%
Low	0.00%	1.23%	0.25%	3.69%	0.98%	6.14%
Grand Total	2.70%	17.69%	6.88%	43.73%	28.99%	100.00%

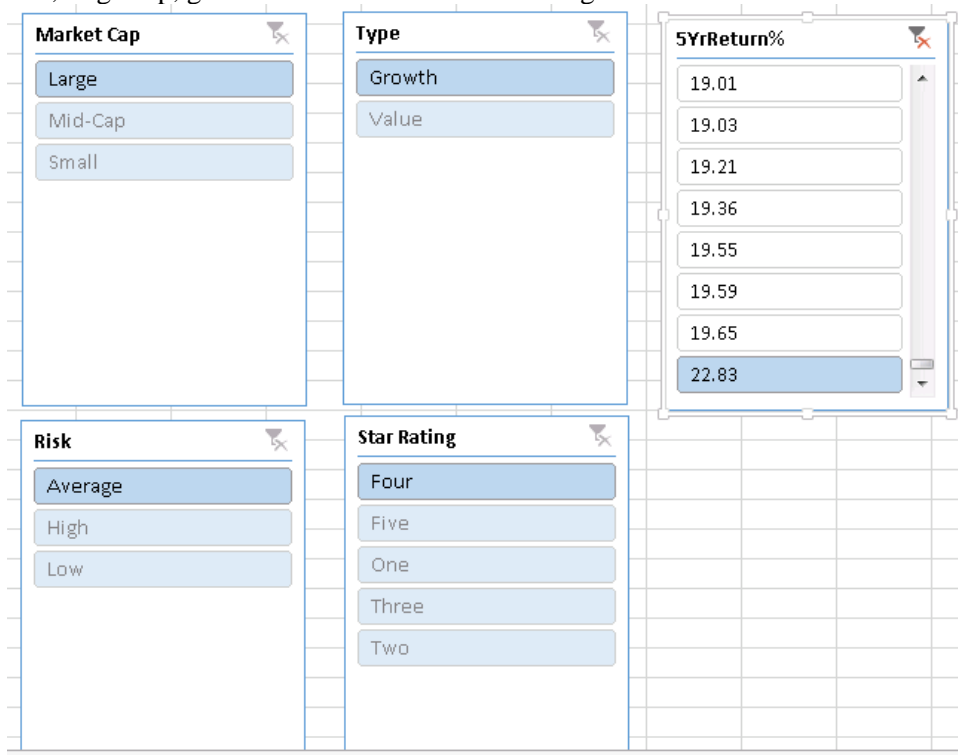2.60   (b)   Patterns of star rating conditioned on risk:
cont.   For the growth funds as a group, most are rated as three-star, followed by two-star, four-star, one-star and five-star. The pattern of star rating is the same among the low-risk and average-risk growth funds. The pattern is different among the high-risk growth funds. Among the high-risk growth funds, most are rated as two-star, followed by three-star, then equal portions of one-star and four-star and finally five-star.
               For the value funds as a group, most are rated as three-star, followed by two-star, four-star, one-star and five-star. The average-risk value funds follow the same pattern. Among the high-risk value funds, most are two-star, followed by three-star, one-star, four-star with no five-star. Among the low-risk value funds, most are three-star, followed by four-star, two-star, one-star with no one-star.
               Patterns of risk conditioned on star rating:
               Most of the growth funds are rated as average-risk, followed by high-risk and then low-risk. The pattern is the same among the one-star, two-star, three-star and four-star growth funds. Among the five-star growth funds, most are average-risk, followed by equal portions of high-risk and low-risk.
                Most of the value funds are rated as average-risk, followed by low-risk and then high-risk. The pattern is the same among the three-star and four-star value funds. Among the one-star and two-star value funds, most are average-risk, followed by high-risk and low-risk. Among the five-star value funds, all are average-risk with no low-risk or high-risk.

(c)

Average of 3YrReturn%	Column Labels					
Row Labels	Five	Four	One	Three	Two	Grand Total
⊟Growth		19.1144	17.0080	12.0311	15.4837	13.8681	15.1478
Average		19.0186	16.9457	12.9064	15.6177	14.0775	15.4263
High		20.8000	17.4457	10.3371	14.8094	13.3452	14.0587
Low		18.1000	16.7133	14.2600	15.2071	14.2050	15.3606
⊟Value		19.4600	16.7463	9.9322	15.5308	14.4315	15.1418
Average		19.4600	17.2985	10.6000	16.1334	15.2976	15.9071
High			14.6850	7.3767	14.7025	13.0714	12.6131
Low			15.3620	14.2600	14.1047	10.5325	13.7908
Grand Total		19.1773	16.9099	11.3564	15.4996	14.0591	15.1458

(d)   There are 16 growth funds with high risk with a rating of three. Below are the summary statistics for the three-year return:

	Assets	Turnover Ratio(%)	SD	Sharpe Ratio	1YrReturn%	3YrReturn%	5YrReturn%	10YrReturn%	Expense Ratio
Mean	288.66625	116.2125	14.45625	1.095625	0.1375	14.809375	15.6625	8.463125	1.394375
Standard Error	63.7015223	38.50097604	0.237966866	0.062769212	1.084409786	0.575337197	0.367889816	0.438094091	0.047019223
Median	257.64	77	14.385	1.02	-0.445	13.84	15.375	8.175	1.38
Mode	#N/A	#N/A	#N/A	1.29	#N/A	#N/A	#N/A	#N/A	1.3
Standard Deviation	254.8060892	154.0039042	0.951867463	0.251076847	4.337639143	2.301348789	1.471559264	1.752376363	0.18807689
Sample Variance	64926.14309	23717.2025	0.906051667	0.063039583	18.81511333	5.29620625	2.165486667	3.070822917	0.035372917
Kurtosis	-0.717209396	13.85191441	1.691247726	6.096226271	0.375691283	-0.366090479	-0.601262563	1.716192124	2.579478964
Skewness	0.690555804	3.615932824	1.129564954	2.179561503	0.686588585	0.846169605	-0.128329303	1.265339908	1.016917946
Range	756.93	659	3.62	1.03	16.48	7.58	4.94	6.48	0.8
Minimum	35.97	18	13.32	0.85	-6.51	12.26	13.02	6.5	1.1
Maximum	792.9	677	16.94	1.88	9.97	19.84	17.96	12.98	1.9
Sum	4618.66	1859.4	231.3	17.53	2.2	236.95	250.6	135.41	22.31
Count	16	16	16	16	16	16	16	16	16

2.61   (a)

Count of 3YrReturn% Column Labels	Five Average	High	Low	Five Total	Four Average	High	Low	Four Total	One Average	High	Low	One Total	Three Average	High	Low	Three Total	Two Average	High	Low	Two Total	Grand Total
Growth	7	1	1	9	35	7	3	45	11	7	1	19	95	16	7	118	51	23	4	78	269
Large	6	1		7	18	1	2	21	9			9	47	2	7	56	28	2	4	34	127
Mid-Cap			1	1	15	1	1	17	2	3	1	6	37	2		39	18	8		26	89
Small	1			1	2		5	7			4	4	11	12		23	5	13		18	53
Value	2			2	20	2	5	27	5	3	1	9	41	4	15	60	29	7	4	40	138
Large	1			1	11		5	16	4	1		5	25		14	39	18		4	22	83
Mid-Cap	1			1	6			6	1	1	1	3	9		1	10	9	1		10	30
Small					3	2		5			1	1	7	4		11	2	6		8	25
Grand Total	9	1	1	11	55	9	8	72	16	10	2	28	136	20	22	178	80	30	8	118	407

(b)   Patterns of star rating conditioned on type, market cap and risk:

From Problem 2.58 (b), we know that the growth funds as a group, most are rated as three-star, followed by two-star, four-star, one-star and five-star. The pattern of star rating is the same across the different market cap within the growth funds with most of the funds receiving a three-star rating, followed by two-star, four-star, one-star and five-star. If we want to bore further down into the subsets of star-rating among the large-cap growth funds, we see that similar pattern does not hold for the various risk ratings. For example, among the large-cap growth funds with an high-risk rating, most are rated as three-star and two-star each with equal portion, followed by equal shares of five-star and four-star with no one-star.

For the value funds as a group, most are rated as three-star, followed by two-star, four-star, one-star and five-star. Within the value funds, the large-cap and small-cap funds follow the same pattern as the value funds as a group. If we want to bore further down into the subsets of star-rating among the large-cap value funds, we see that similar pattern does not hold through for the various risk ratings. For example, among the large-cap value funds with an average-risk rating, the pattern is the same as the large-cap value funds as a group. However, among the large-cap value funds with a high-risk, they are all one-star funds.

Patterns of market cap conditioned on type, risk and star-rating:

Again, from Problem 2.58 (b), we know that most of the growth funds are large-cap, followed by mid-cap and small-cap. The pattern is similar among the four-star, three-star, two-star and one-start growth funds but among the five-star growth funds, most are large-cap, followed by equal portions of small-cap and mid-cap. If we bore further down into the subsets of risk-rating, we see that similar pattern does not hold across the different risk levels. For example, among the high-risk, four-star, growth funds, most are small-cap followed by equal portions of large-cap and mid-cap.

2.61   (c)   The tables for problems 2.58 through 2.60 are easier to interpret. With too many
cont.   dimensions in the contingency table, it becomes very difficult to discern any pattern.
        (d)   There are many empty cells with no observation in the table in this problem as a result of
                   increasing the dimension. Collapsing the table back to those in problems 2.58 through
                   2.60 though can result in potential loss of detailed information or pattern in the data.

2.62   With the help of the slicers, the fund with the highest five-year return at 22.83% is an average
        risk, large cap, growth fund with a four-star rating.

Market Cap	Type	5YrReturn%
Large	Growth	19.01
Mid-Cap	Value	19.03
Small		19.21
		19.36
		19.55
		19.59
		19.65
		22.83

Risk	Star Rating
Average	Four
High	Five
Low	One
	Three
	Two

2.63   There is only one small cap fund with a five-star rating and its five-year return is 19.01%.

Market Cap	Type	5YrReturn%
Large	Growth	19.01
Mid-Cap	Value	5.06
Small		5.12
		6.67
		7.13
		7.17
		7.27
		7.30

Risk	Star Rating
Average	Five
High	Four
Low	One
	Three
	Two

2.64   The fund with a fund number RF206 has the lowest five-year return at 5.06% and is a large cap, high risk, value fund with a one-star rating.

2.65   The five-star fund with the highest five-year return at 19.65% is a large cap, average risk, growth fund.

Market Cap	Type	5YrReturn%
Large	Growth	17.81
Mid-Cap	Value	18.16
Small		18.54
		18.84
		19.01
		19.36
		19.65
		5.06

Risk	Star Rating
Average	Five
High	Four
Low	One
	Three
	Two

2.66   The funds that have the lowest five-year return at 5.06% is a large cap, high risk, value fund with a one-star rating.

Market Cap
Large
Mid-Cap
Small

Type
Value
Growth

5YrReturn%
5.06
5.12
6.67
7.13
7.17
7.27
7.30
9.33

Risk
High
Average
Low

Star Rating
One
Five
Four
Three
Two

2.67   (a)

Type	2010	2011	2012	2013	2014	
Android	23.3	49.2	69.0	78.8	81.5	
iOS	15.6	18.8	18.7	15.1	14.8	
Microsoft	4.9	1.8	2.5	3.3	2.7	
Blackberry	16.0	10.3	4.5	1.9	0.4	
OtherOS	40.2	19.8	5.4	1.0	0.6	

(b)   The Android smartphone sales have been increasing since 2010 while those of the Blackberry and OtherOS have been decreasing since 2010. The iOS smartphone sales had been increasing since 2010 and reached the peak in 2011 and started a downward trend since. The Microsoft smartphone sales had been decreasing since 2010 and reached the trough in 2012 but had seen a comeback since.

2.68   (a)

Index	2009	2010	2011	2012	2013	
Dow Transportation	15.9	24.6	-1.7	5.7	39.5	
NASDAQ Composite	43.9	16.9	-1.8	15.9	38.3	
Russell 2000	25.2	25.3	-5.5	14.6	37.0	
NSADAQ 100	53.5	19.2	2.7	16.8	35.0	
S&P Midcap	35.0	24.9	-3.1	16.1	31.6	
Wilshire 5000	27.1	15.7	-1.0	13.7	31.4	
S&P 500	23.5	12.8	0.0	13.4	29.6	
Dow Industrials	18.8	11.0	5.5	7.3	26.5	
Dow Utilities	7.3	1.8	14.7	-2.5	8.3	

(b)   All indices reached their trough in 2011 and have been on the upward trend since with the exception of the Dow Utilities which reach its peak in 2011.

2.71   (a)   There is a title.
        (b)   None of the axes are labeled.
        (c)
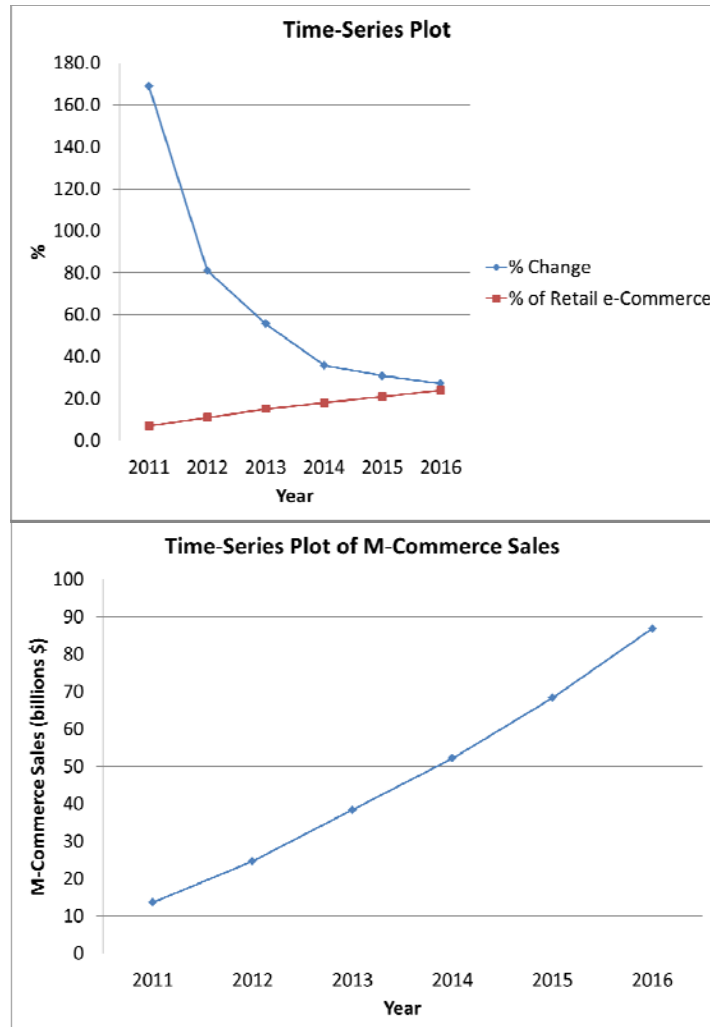
2.72   (a)   There is a title.
        (b)   The simplest possible visualization is not used.
        (c)
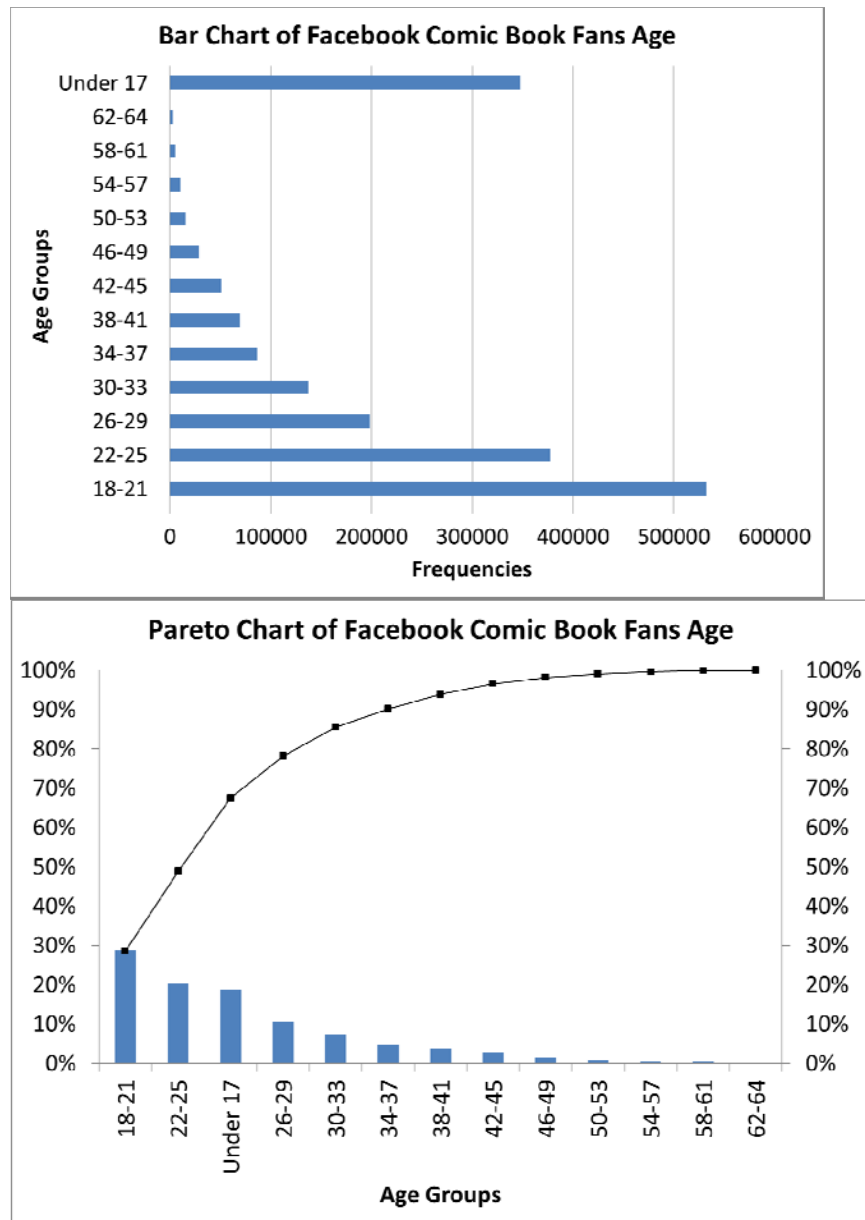
2.73   (a)   None.
       (b)   The use of chartjunk.
       (c)
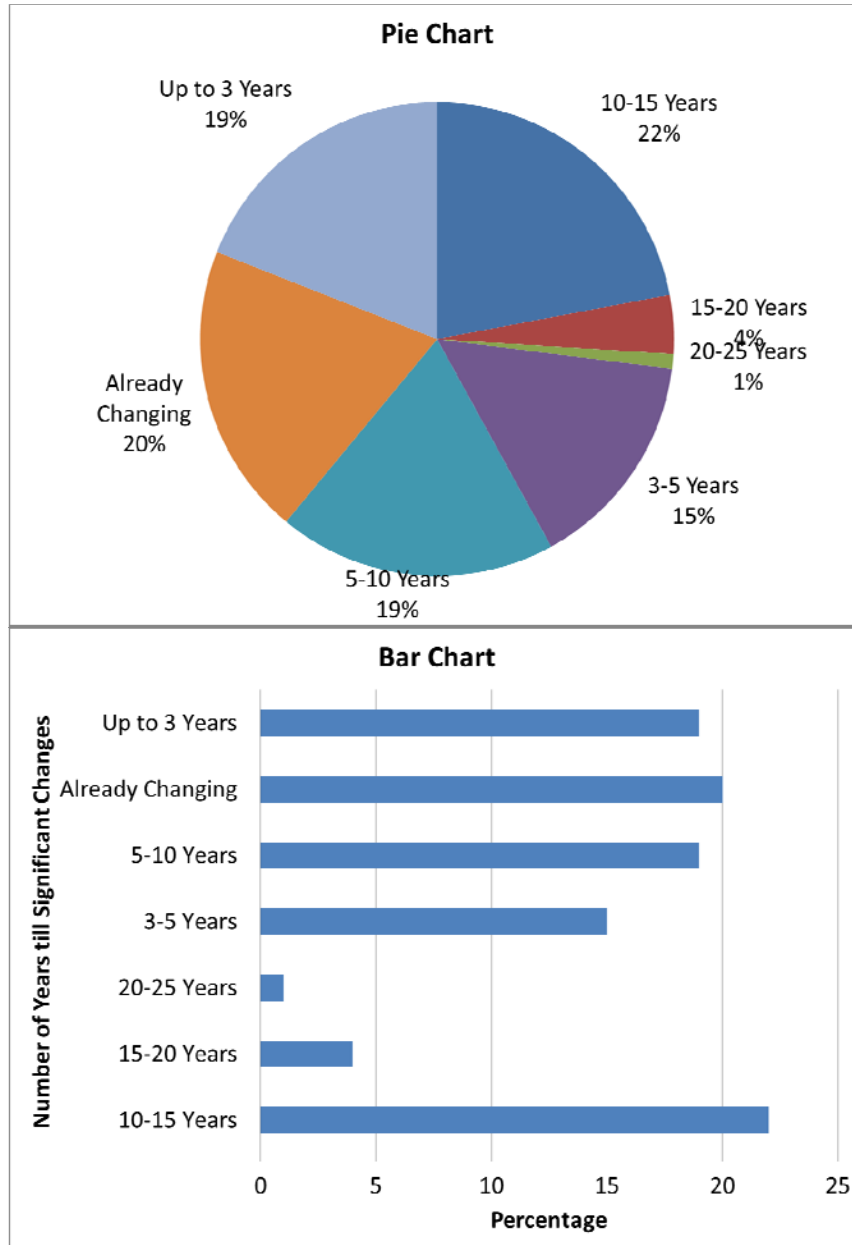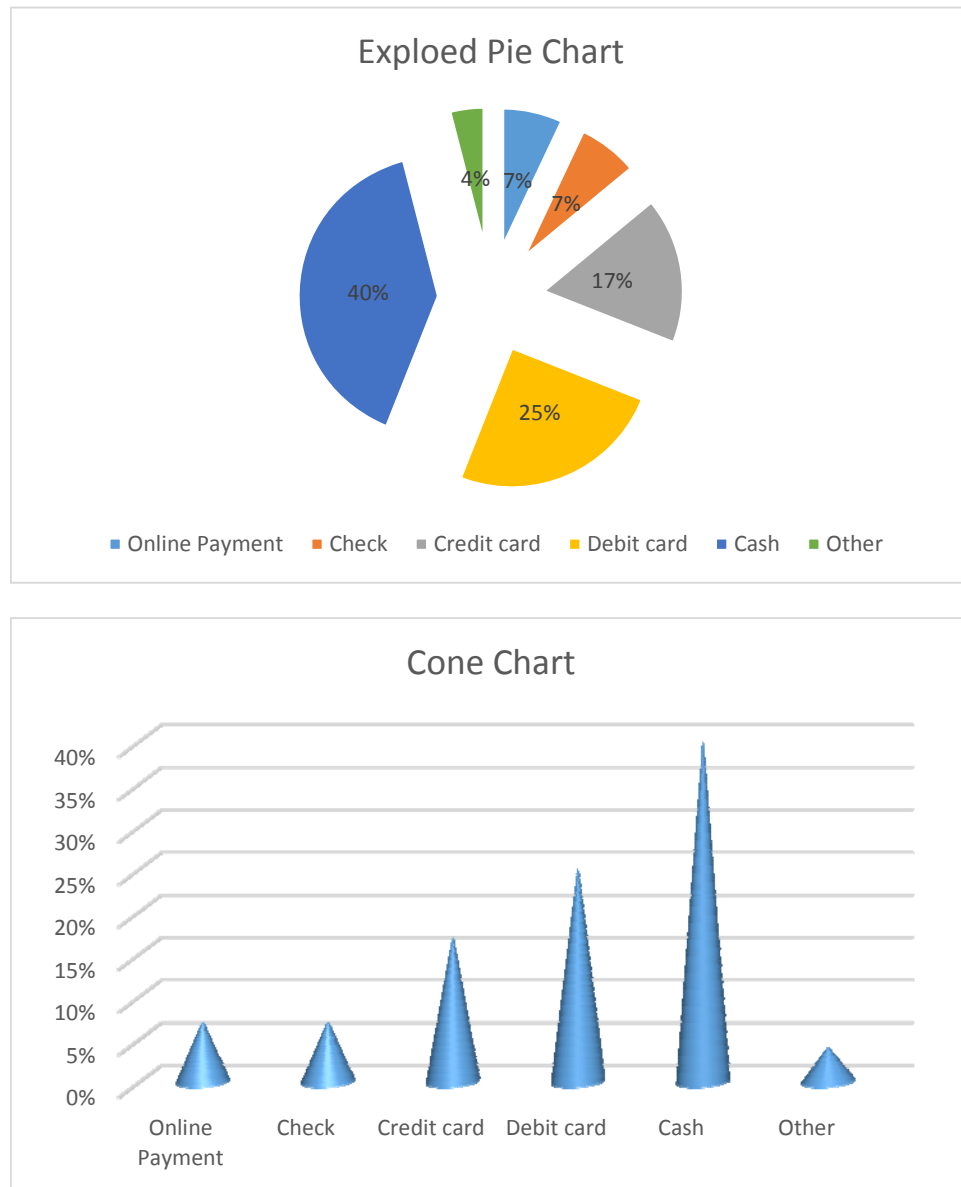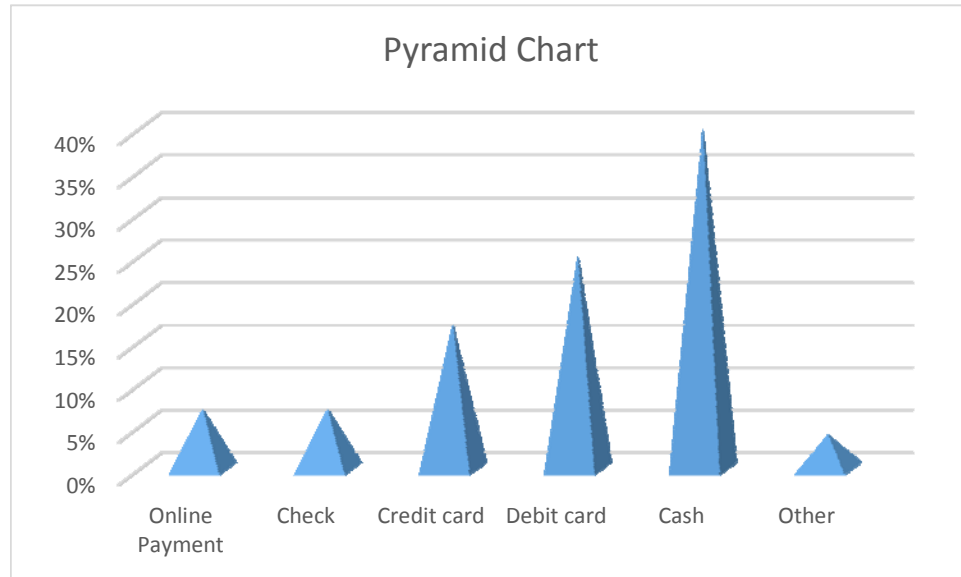
Pie Chart

Up to 3 Years
19%

10-15 Years
22%

15-20 Years
4%

20-25 Years
1%

3-5 Years
15%

5-10 Years
19%

Already
Changing
20%

Bar Chart

Number of Years till Significant Changes

Up to 3 Years

Already Changing

5-10 Years

3-5 Years

20-25 Years

15-20 Years

10-15 Years

0   5   10   15   20   25

Percentage

2.75   (a)

2.75   (a)
cont.

Pyramid Chart

(b)   The bar chart and the pie chart should be preferred over the exploded pie chart, doughnut chart, the cone chart and the pyramid chart since the former set is simpler and easier to interpret.

2.76   (a)

Exploed Pie Chart

2.76   (a)
cont.

Cone Chart

80%			
70%			
60%			
50%			
40%			
30%			
20%			
10%			
0%			
	Low risk	Average risk	High risk

Pyramid Chart

80%			
70%			
60%			
50%			
40%			
30%			
20%			
10%			
0%			
	Low risk	Average risk	High risk
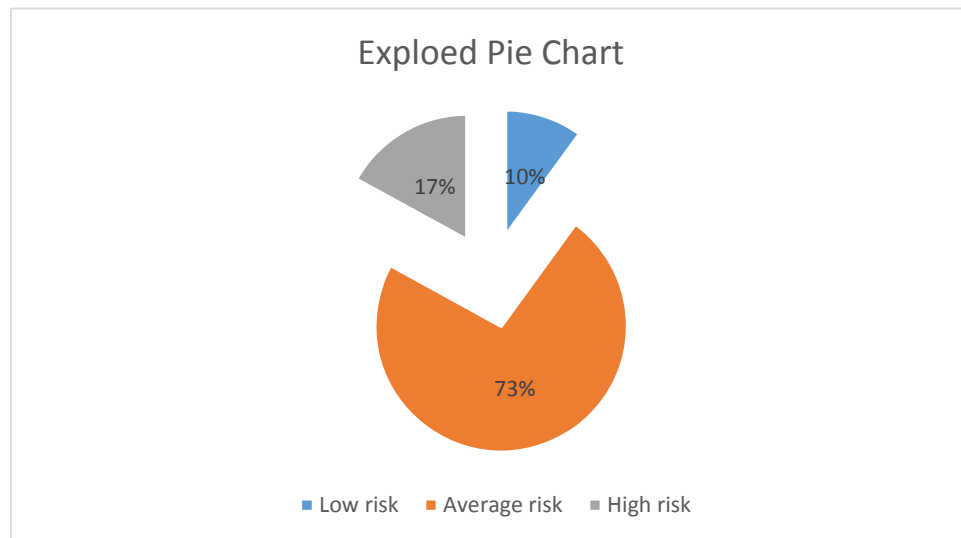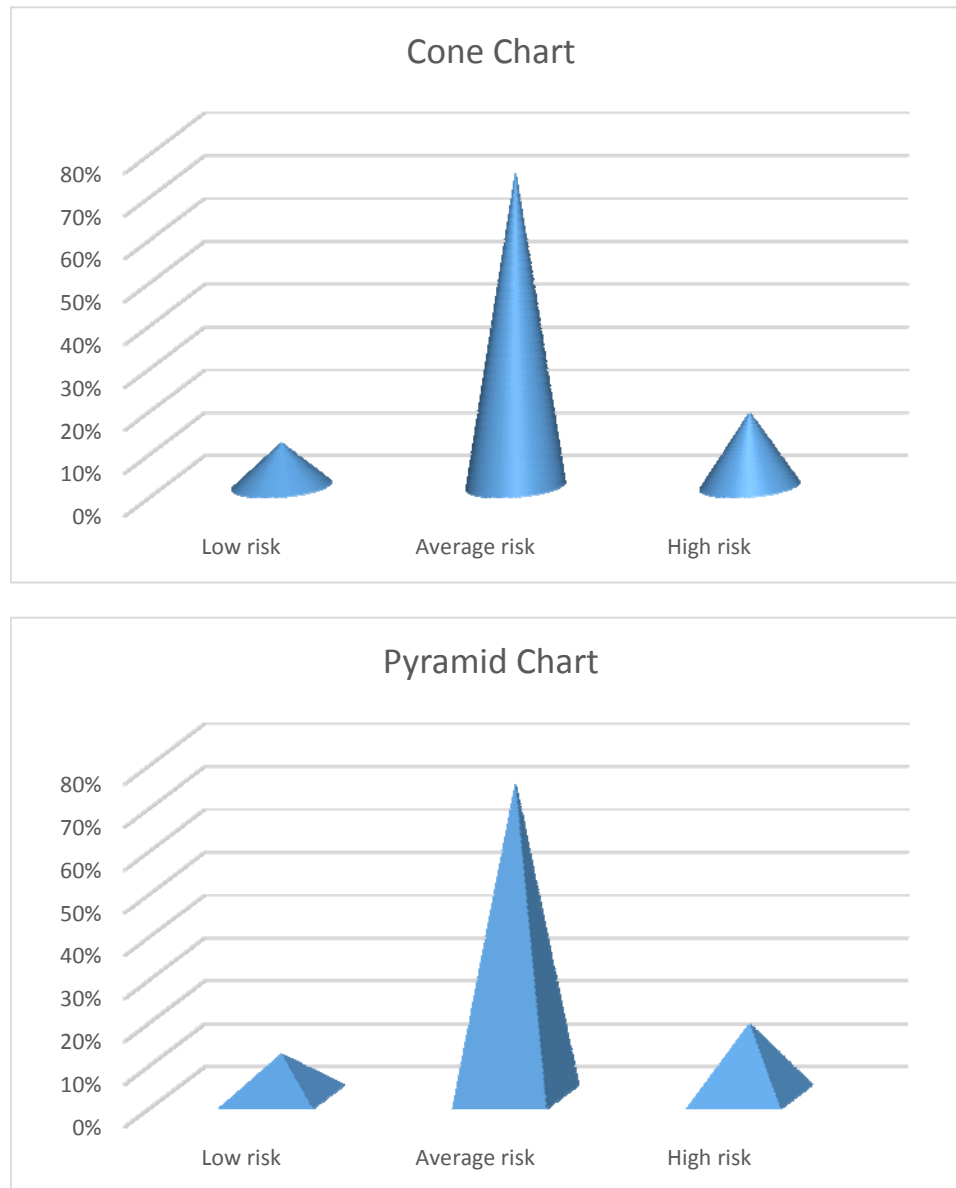
(b)   The bar chart and the pie chart should be preferred over the exploded pie chart, doughnut chart, the cone chart and the pyramid chart since the former set is simpler and easier to interpret.

2.77   A histogram uses bars to represent each class while a polygon uses a single point. The histogram should be used for only one group, while several polygons can be plotted on a single graph.

2.78   A summary table allows one to determine the frequency or percentage of occurrences in each category.

2.79   A bar chart is useful for comparing categories. A pie chart is useful when examining the portion of the whole that is in each category. A Pareto diagram is useful in focusing on the categories that make up most of the frequencies or percentages.

2.80   The bar chart for categorical data is plotted with the categories on the vertical axis and the frequencies or percentages on the horizontal axis. In addition, there is a separation between categories. The histogram is plotted with the class grouping on the horizontal axis and the frequencies or percentages on the vertical axis. This allows one to more easily determine the distribution of the data. In addition, there are no gaps between classes in the histogram.

2.81   A time-series plot is a type of scatter diagram with time on the x-axis.

2.82   Because the categories are arranged according to frequency or importance, it allows the user to focus attention on the categories that have the greatest frequency or importance.

2.83   Percentage breakdowns according to the total percentage, the row percentage, and/or the column percentage allow the interpretation of data in a two-way contingency table from several different perspectives.

2.84   A contingency table contains information on two categorical variables whereas a multidimensional table can display information on more than two categorical variables.

2.85   The multidimensional PivotTable can reveal additional patterns that cannot be seen in the contingency table. One can also change the statistic displayed and compute descriptive statistics which can add insight into the data.

2.86   In a PivotTable in Excel, double-clicking a cell drills down and causes Excel to display the underlying data in a new worksheet to enable you to then observe the data for patterns. In Excel, a slicer is a panel of clickable buttons that appears superimposed over a worksheet to enable you to work with many variables at once in a way that avoids creating an overly complex multidimensional contingency table that would be hard to comprehend and interpret.

2.87   Sparklines are compact time-series visualizations of numerical variables. Sparklines can also be used to plot time-series data using smaller time units than a time-series plot to reveal patterns that the time-series plot may not.

2.88   (a)

Bar Chart

Pie Chart

2.88 (a)
cont.

Pareto Diagram

(b)

Pareto Diagram
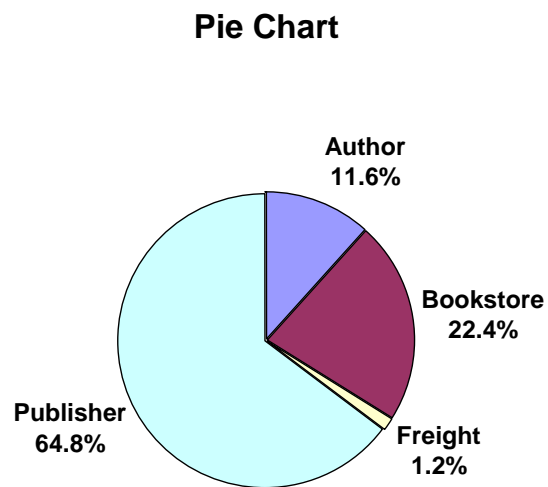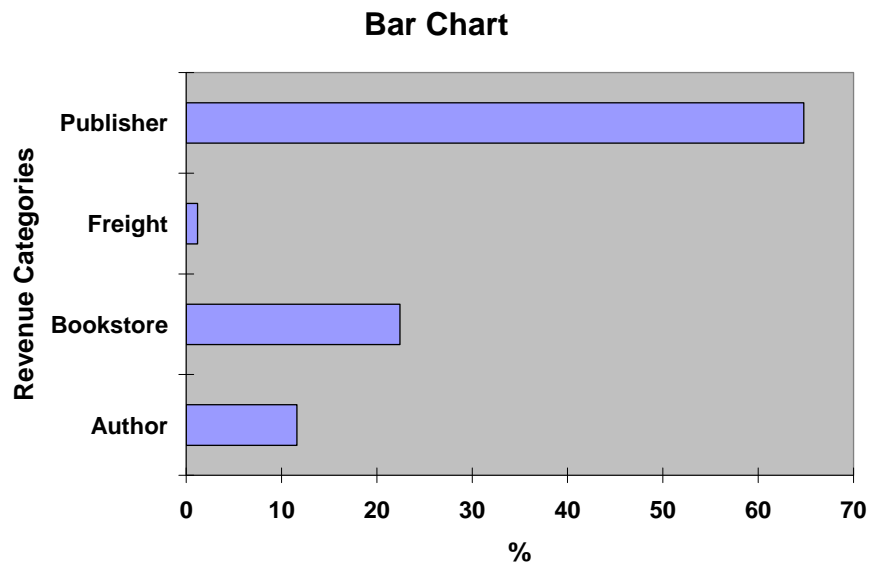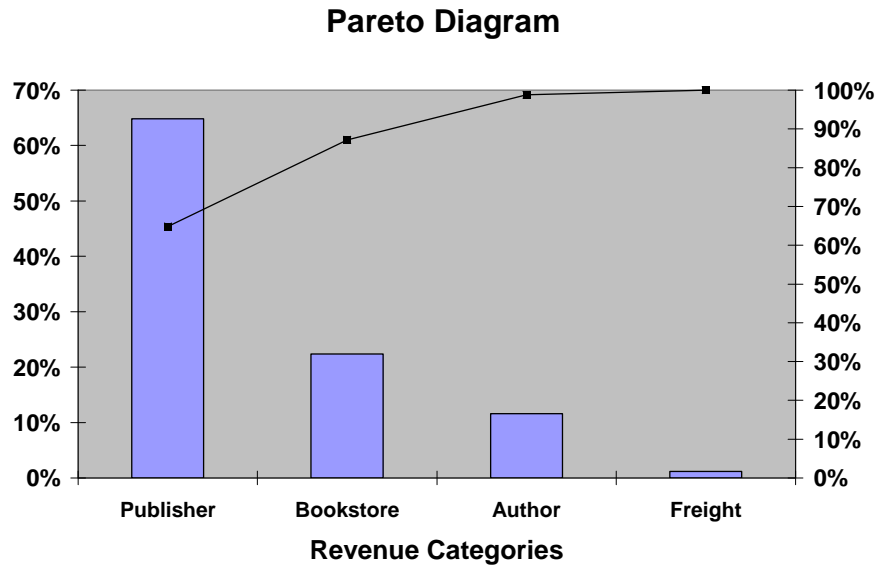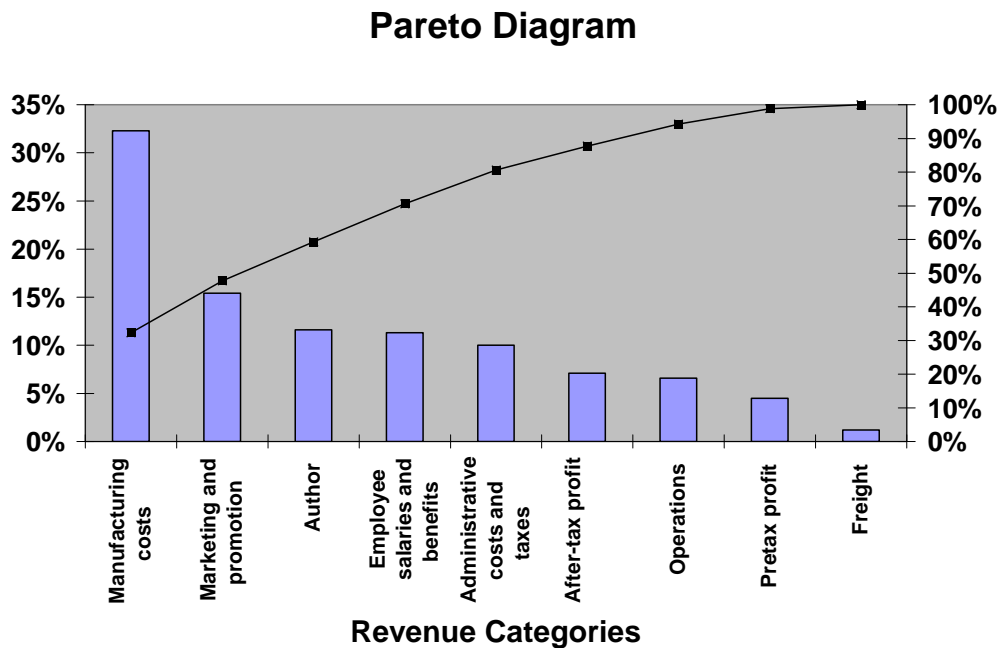
(c)   The publisher gets the largest portion (64.8%) of the revenue. About half (32.3%) of the revenue received by the publisher covers manufacturing costs. The publisher's marketing and promotion account for the next largest share of the revenue, at 15.4%. Author, bookstore employee salaries and benefits, and publisher administrative costs and taxes each account for around 10% of the revenue, whereas the publisher after-tax profit, bookstore operations, bookstore pretax profit, and freight constitute the "trivial few" allocations of the revenue. Yes, the bookstore gets twice the revenue of the authors.

2.89   (a)   **Number of Movies:**

2.89   (a)
cont.

Pareto Chart

Gross (in $millions):

Bar Chart

2.89   (a)
cont.

Pie Chart
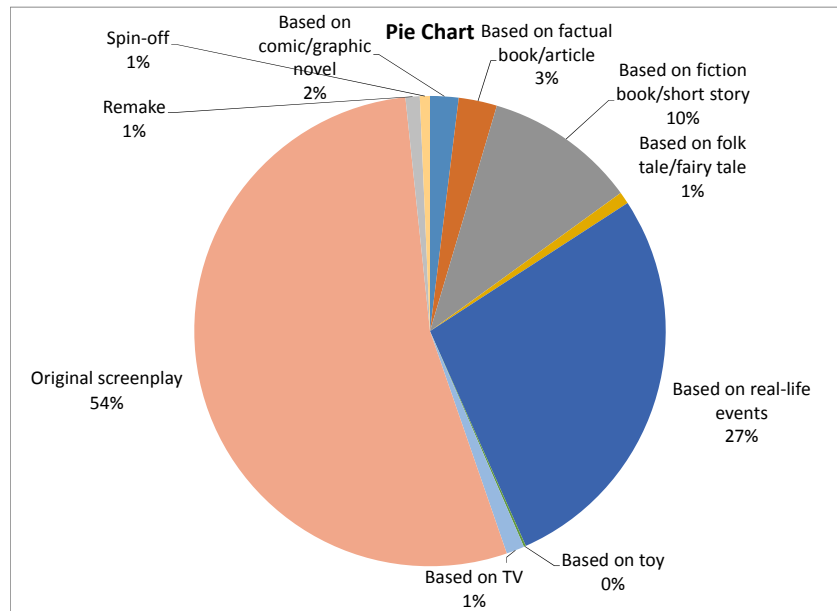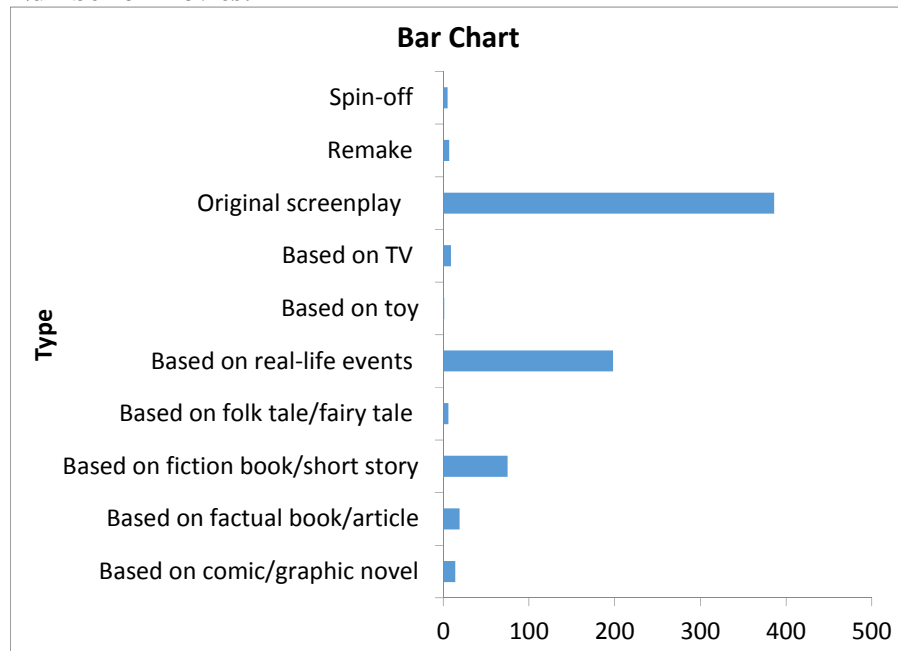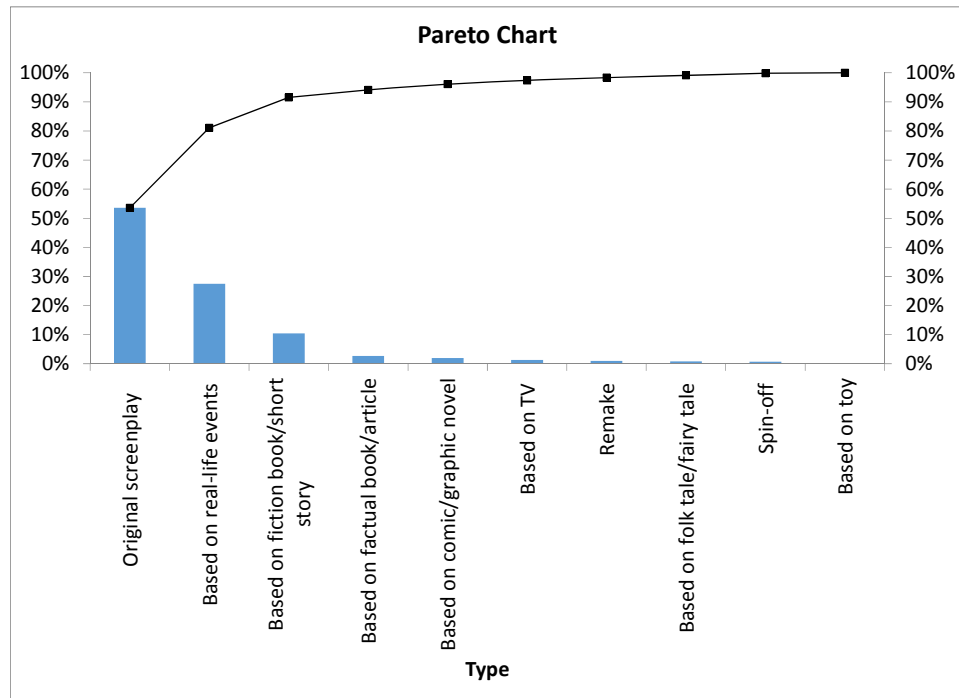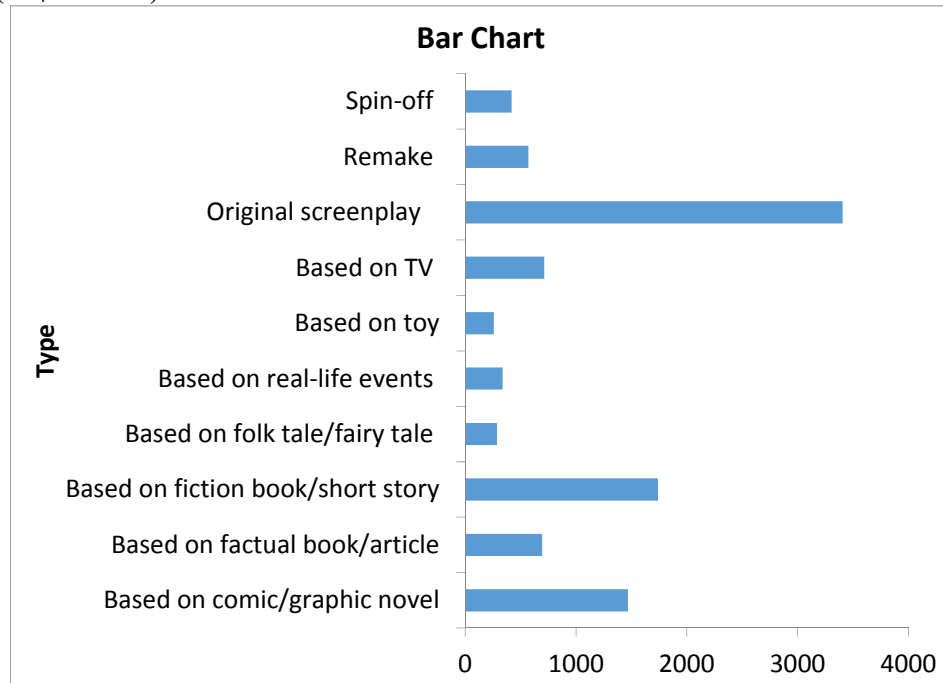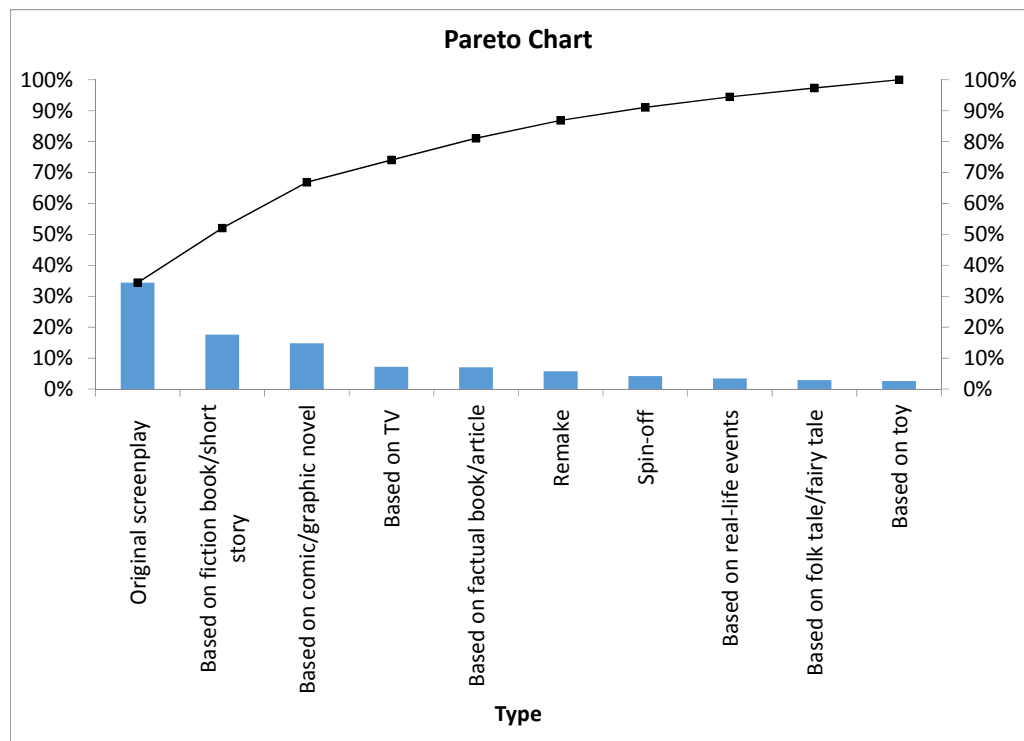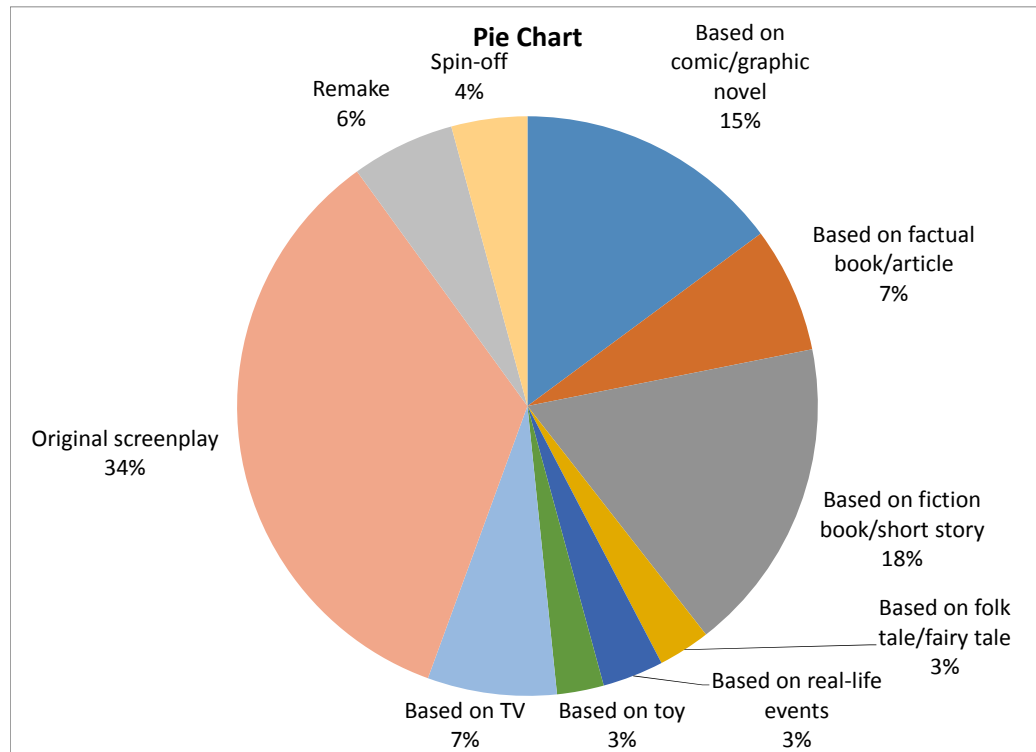
Pareto Chart

2.89   (a)   **Number of Tickets Sold (millions):**
cont.
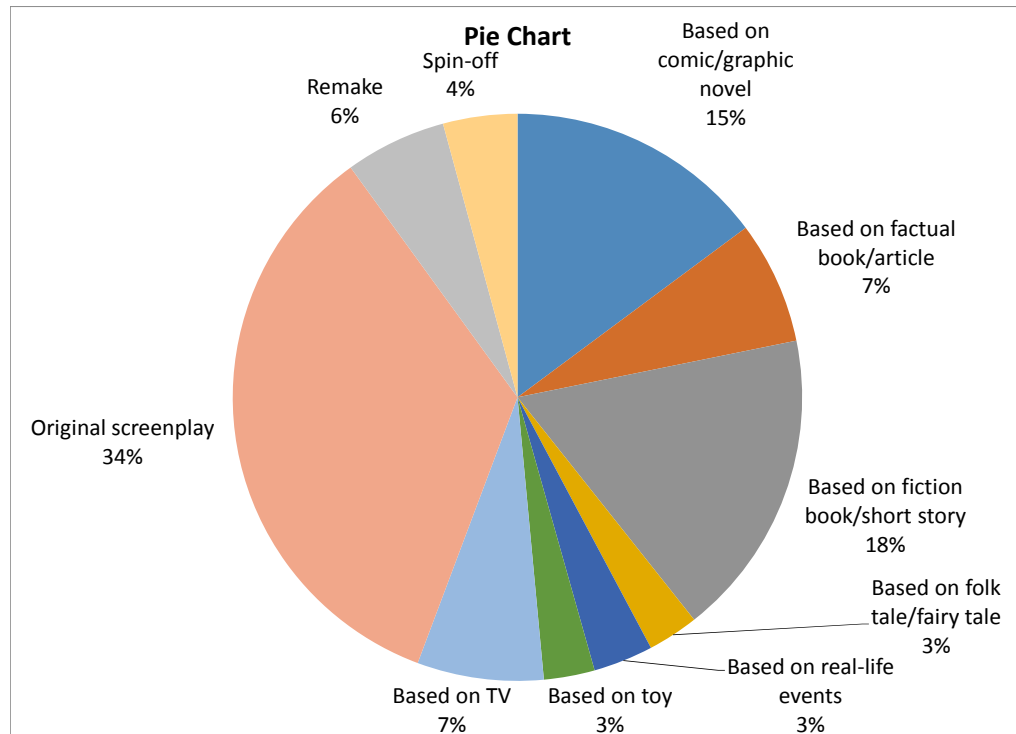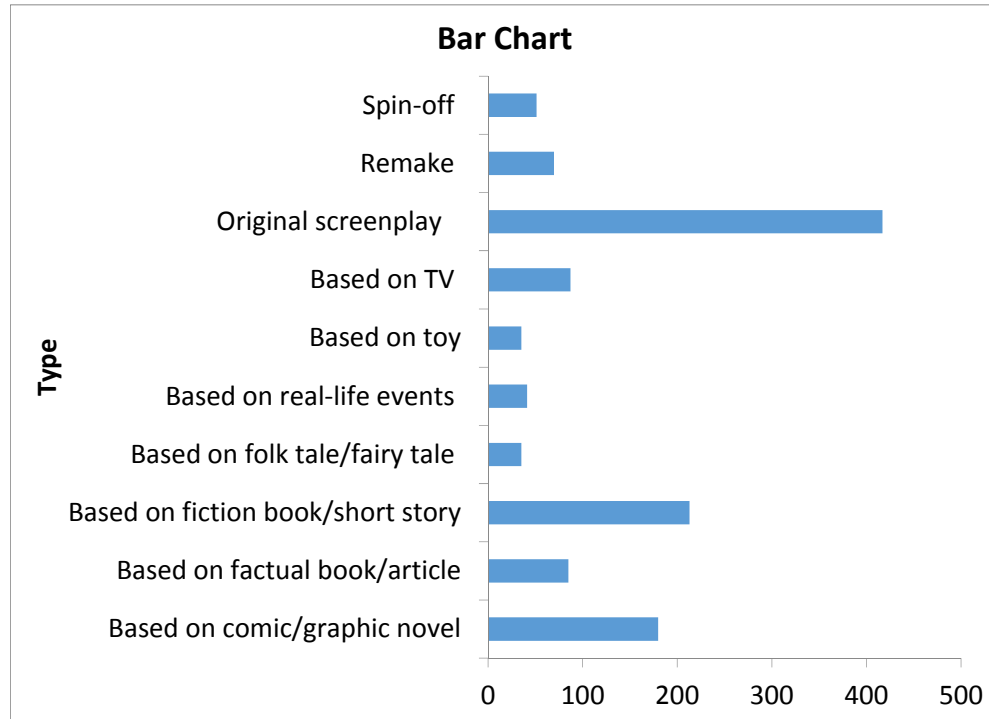
2.89 (a)
cont.
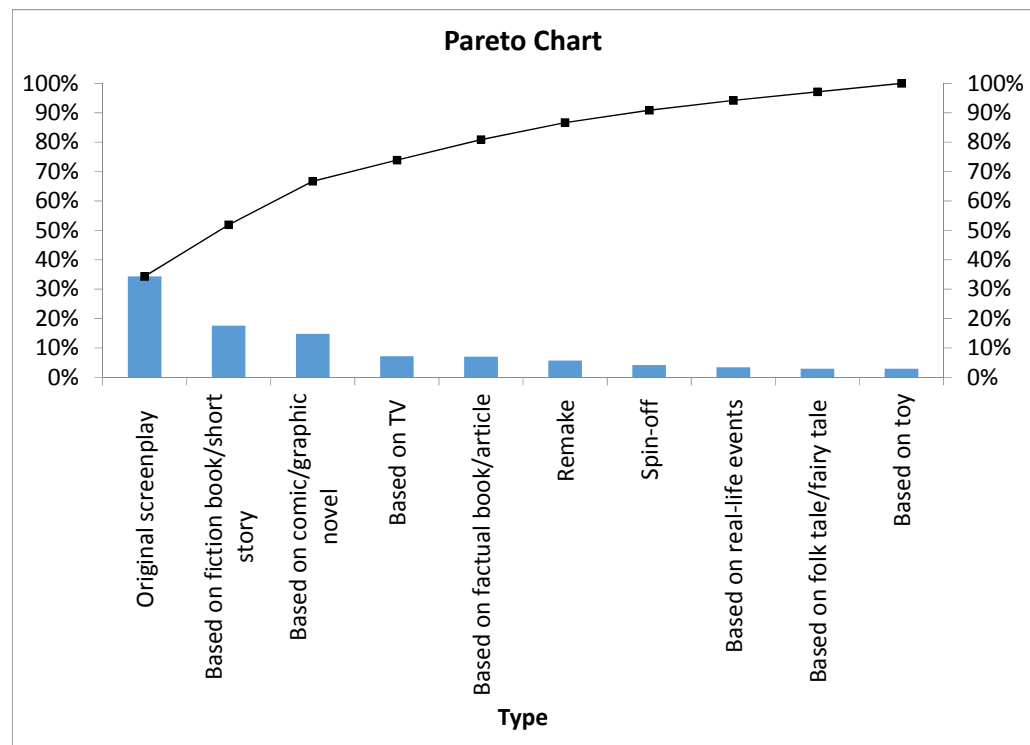
(b)   Based on the Pareto chart for the number of movies, "Original screenplay", "Based on real life events" and "Based on fiction/short story" are the "vital few" and capture about 92% of the market share. According to the Pareto chart for gross (in $millions), "Original screenplay", "Based on fiction book/short story" and "Based on comic/graphic novel" are the "vital few" and capture about 67% of the market share. According to the Pareto chart for number of tickets sold (in millions), "Original screenplay", "Based on fiction book/short story" and "Based on comic/graphic novel" are the "vital few" and capture about 67% of the market share.

2.90   (a) Percentages in decimals as proportions

Bar Chart

Pie Chart

2.90   (a)
cont.

Pareto Chart — Most Often Ways to Find out About New Marketing Agencies
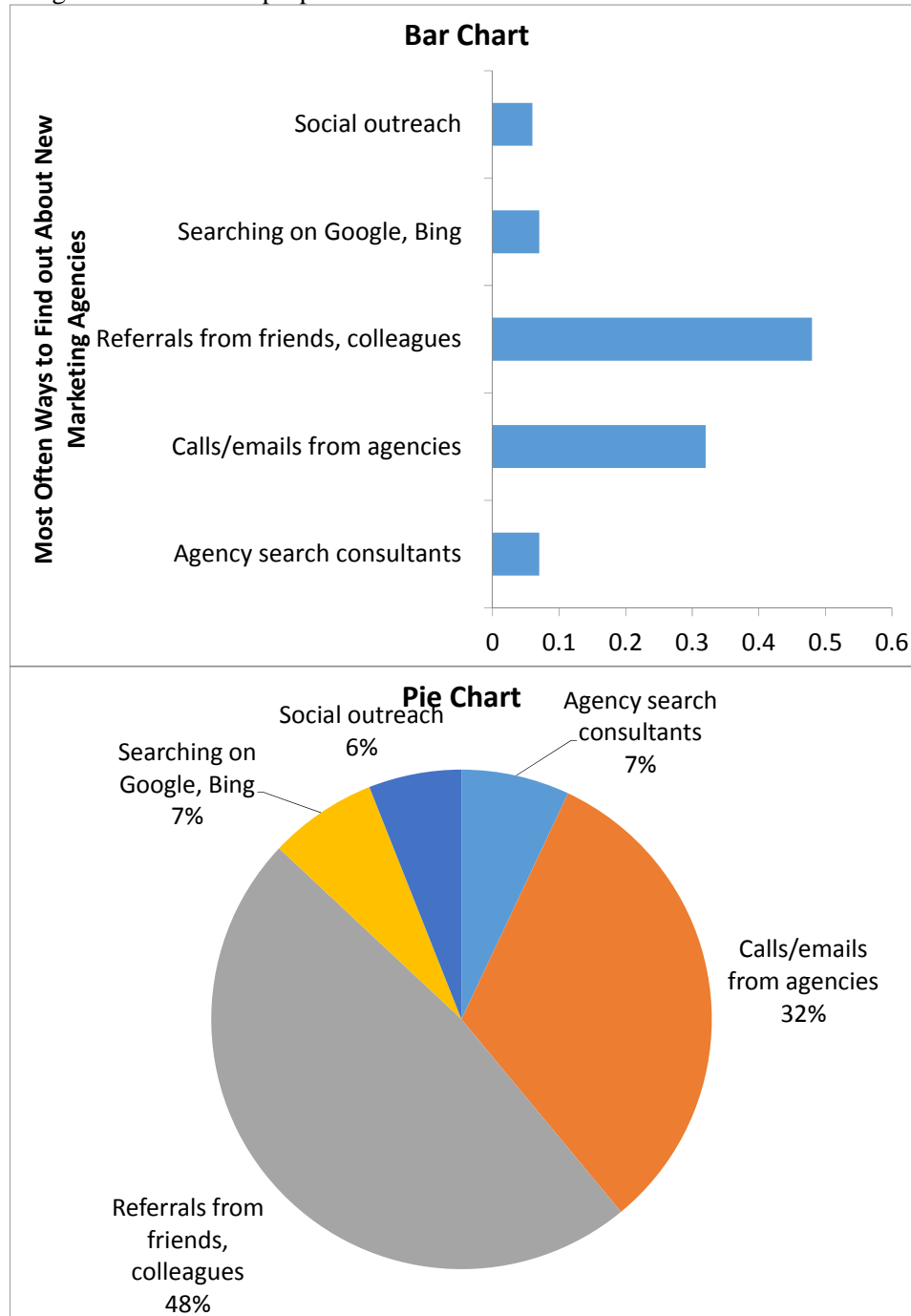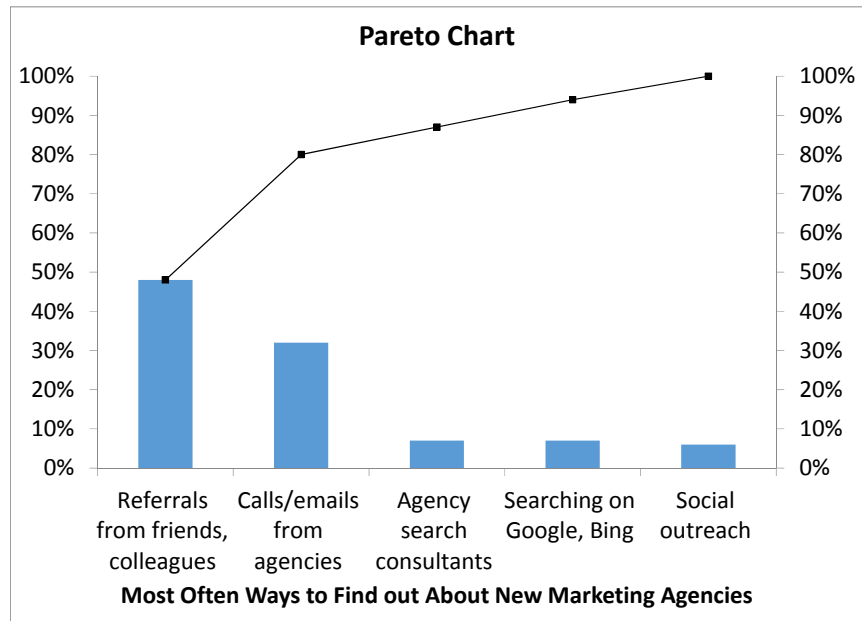
(b)   The pie chart may be best since with only five categories, it enables you to see the portion of the whole in each category.

(c)   Percentages in decimals as proportions

Bar Chart — Importance of Marketing Agency Specializing in Marketer's Industry

2.90   (c)
cont.

Pie Chart

Not at all important 12%

Somewhat important 45%

Very important 43%

Pareto Chart

Importance of Marketing Agency Specializing in Marketer's Industry
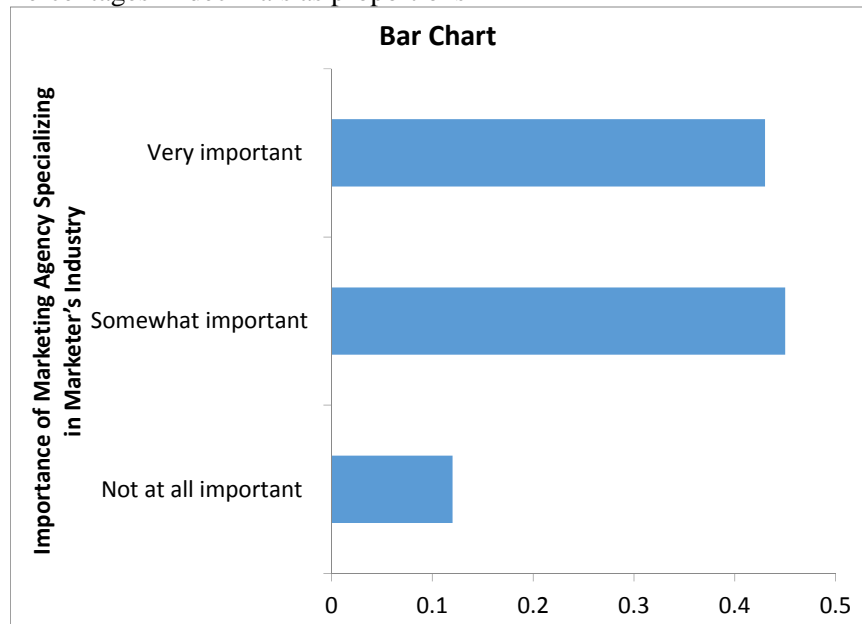
(Somewhat important, Very important, Not at all important)
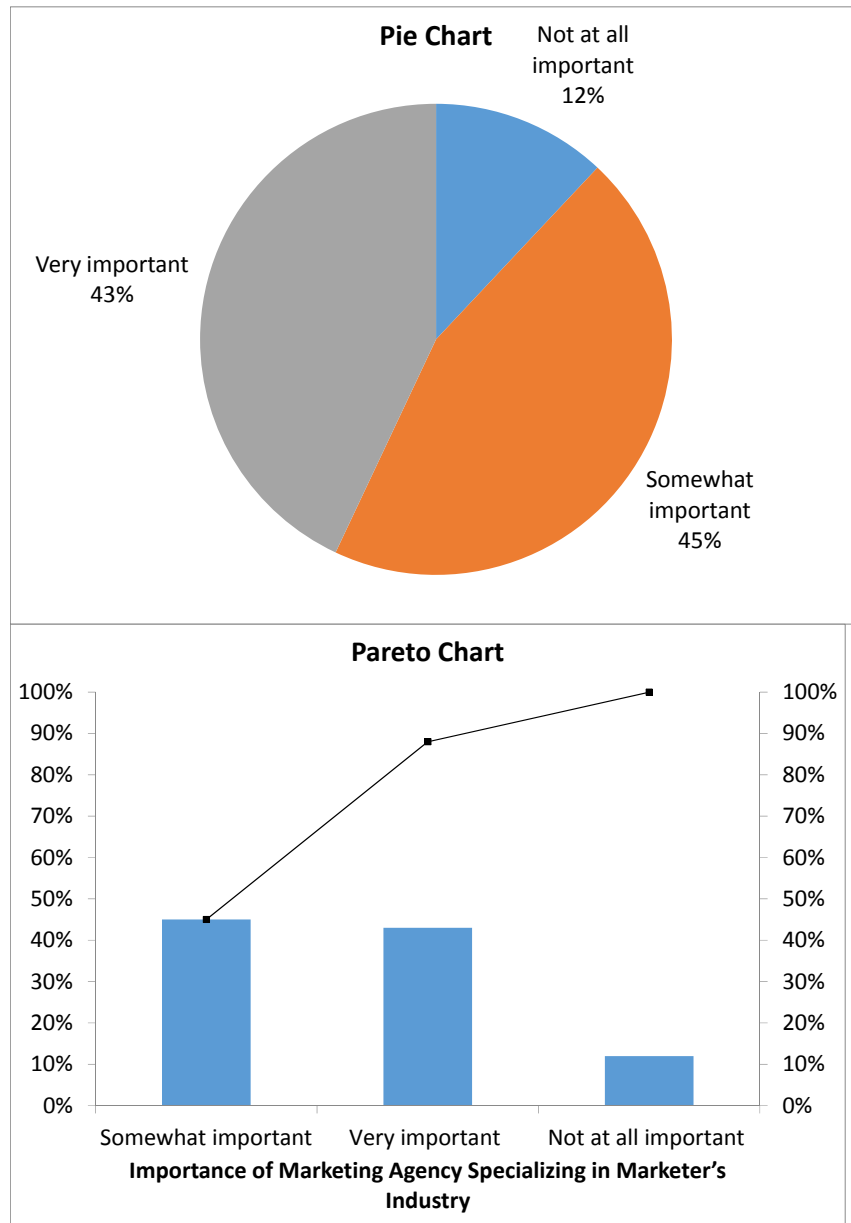
(d)   The pie chart may be best since, with only four categories it enables you to see the portion of the whole in each category.

(e)   Based on the Pareto chart for "Most Often Ways to Find out About New Marketing Agencies", about 80% of the marketers use "referrals from friends/colleagues" and "calls/emails from agencies" to find out about new marketing agencies for hire. Based on the Pareto chart for "Importance of Marketing Agency Specializing in Marketer's Industry", about 88% of the marketers value the marketing agencies that specialize in their industry as "somewhat important" or "very important".

2.91   (a)

Type of Entrée	%	Number S
Beef	29.68%	187
Chicken	16.35%	103
Mixed	4.76%	30
Duck	3.97%	25
Fish	19.37%	122
Pasta	10.00%	63
Shellfish	11.75%	74
Veal	4.13%	26
Total	100.00%	630

(b)
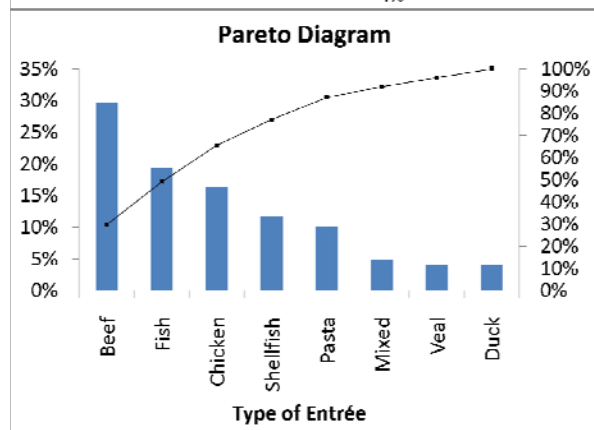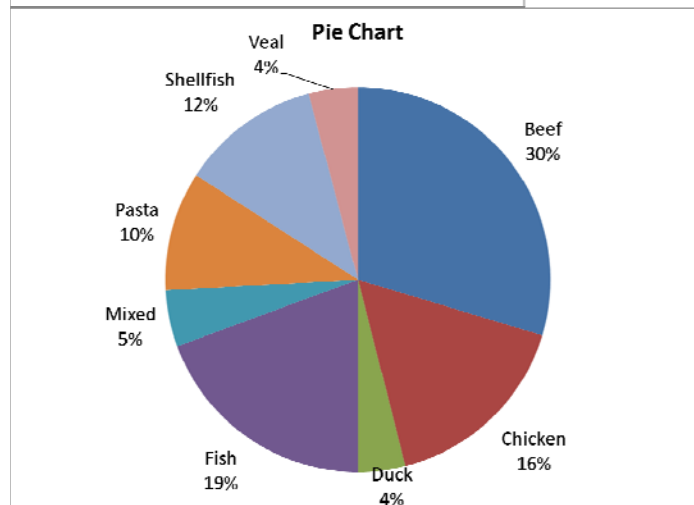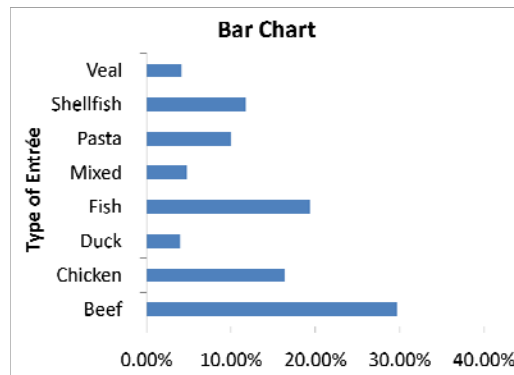
2.91   (c)   The Pareto diagram has the advantage of offering the cumulative percentage view of
cont.   the categories and, hence, enables the viewer to separate the "vital few" from the "trivial
                many".
        (d)   Beef and fish account for nearly 50% of all entrees ordered by weekend patrons
                of a continental restaurant. When chicken is included, nearly two-thirds of the entrees are
                accounted for.

2.92   (a)

Count of Dessert Ordered	Gender		
Desserts Ordered	Male	Female	Grand Total
Yes	34.25%	65.75%	100.00%
No	51.65%	48.35%	100.00%
Grand Total	47.62%	52.38%	100.00%

Count of Dessert Ordered	Gender		
Desserts Ordered	Male	Female	Grand Total
Yes	16.67%	29.09%	23.17%
No	83.33%	70.91%	76.83%
Grand Total	100.00%	100.00%	100.00%

Count of Dessert Ordered	Gender		
Desserts Ordered	Male	Female	Grand Total
Yes	7.94%	15.24%	23.17%
No	39.68%	37.14%	76.83%
Grand Total	47.62%	52.38%	100.00%

Count of Dessert Ordered	Beef Entrée		
Dessert Ordered	Yes	No	Grand Total
Yes	52.11%	47.89%	100.00%
No	25.20%	74.80%	100.00%
Grand Total	31.27%	68.73%	100.00%

Count of Dessert Ordered	Beef Entrée		
Dessert Ordered	Yes	No	Grand Total
Yes	37.56%	15.70%	22.54%
No	62.44%	84.30%	77.46%
Grand Total	100.00%	100.00%	100.00%

Count of Dessert Ordered	Beef Entrée		
Dessert Ordered	Yes	No	Grand Total
Yes	11.75%	10.79%	22.54%
No	19.52%	57.94%	77.46%
Grand Total	31.27%	68.73%	100.00%

2.92   (b)   If the owner is interested in finding out the percentage of joint occurrence of gender
cont.   and ordering of dessert or the percentage of joint occurrence of ordering a beef entrée and
                 a dessert among all patrons, the table of total percentages is most informative. If the
                 owner is interested in the effect of gender on ordering of dessert or the effect of ordering
                 a beef entrée on the ordering of dessert, the table of column percentages will be most
                 informative. Since dessert will usually be ordered after the main entree and the owner
                 has no direct control over the gender of patrons, the table of row percentages is not very
                 useful here.

        (c)   16.67% of the men sampled ordered desserts compared to 29.09% of the women. Women
                 are almost twice as likely to order desserts as men. 37.56% of the patrons ordering a beef
                 entree ordered dessert compared to less than 15.7% of patrons ordering all other entrees.
                 Patrons ordering beef are better than 2.3 times as likely to order dessert as patrons
                 ordering any other entree.

2.93   (a)   **United States Fresh Food Consumed:**
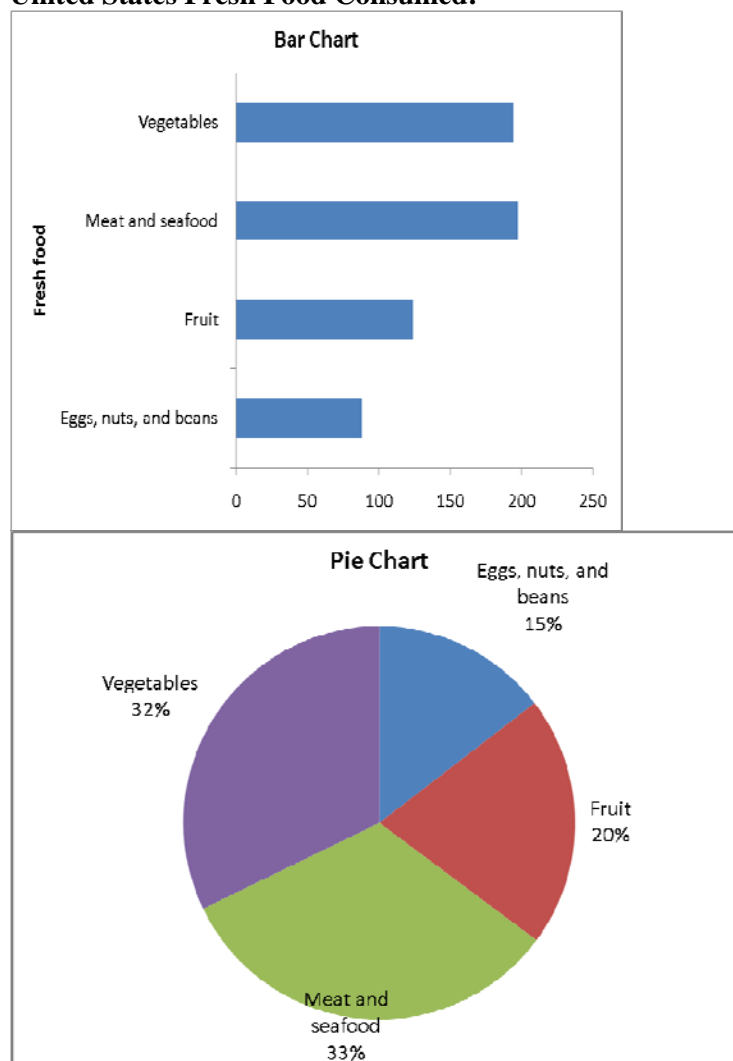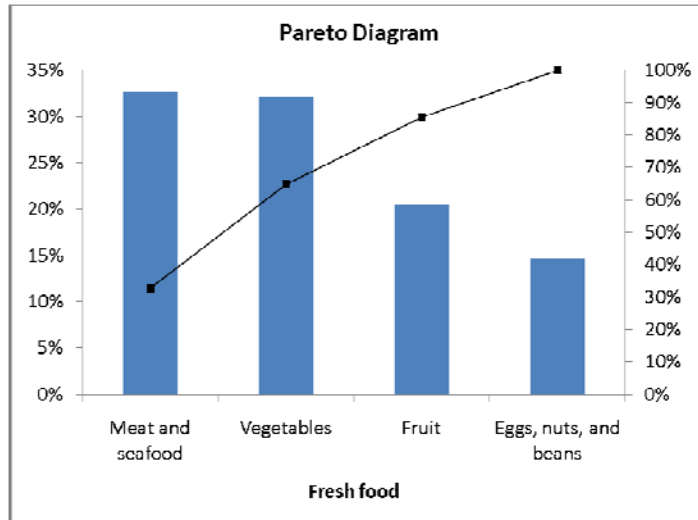
2.93   (a)
cont.

Japan Fresh Food Consumed:
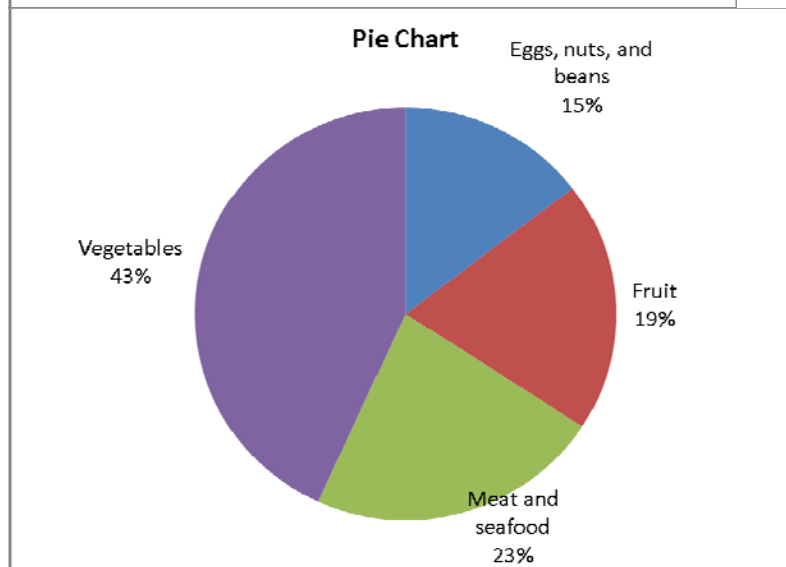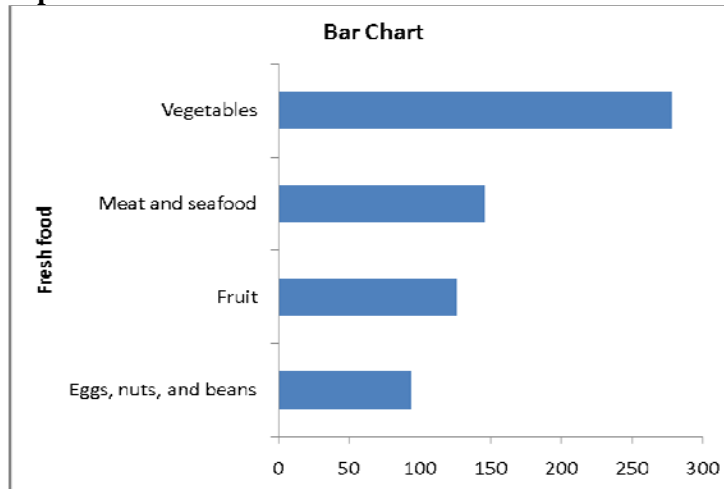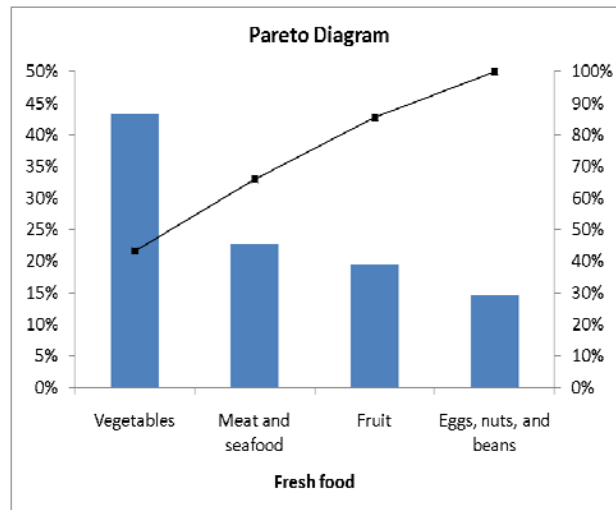
2.93   (a)
cont.

Russia Fresh Food Consumed:

2.93   (a)
cont.

Pareto Diagram
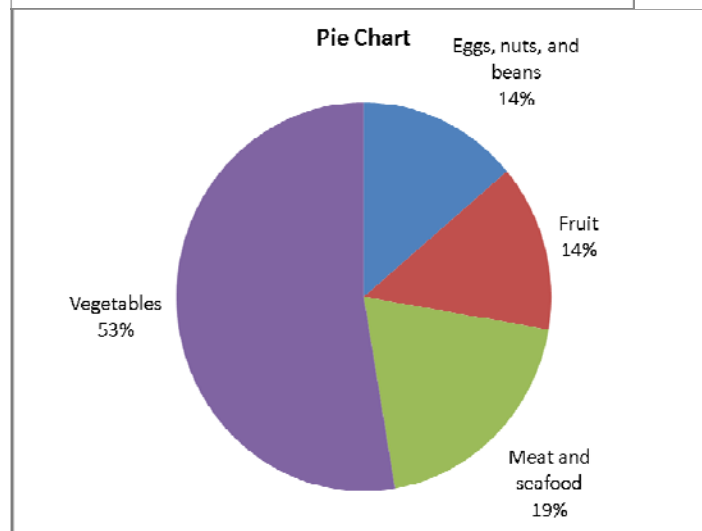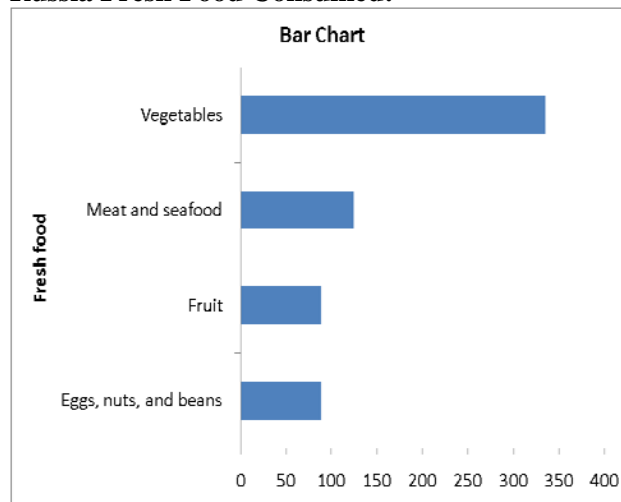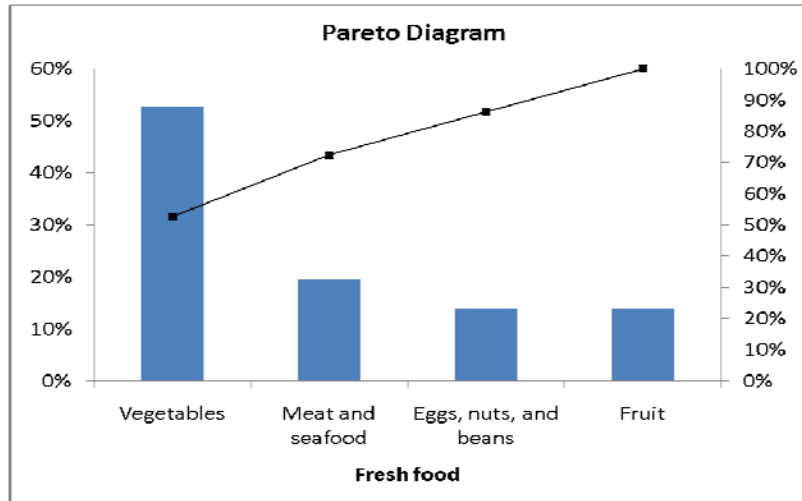
(b)   **United States Packaged Food Consumed:**

Bar Chart

2.93   (b)
cont.

Pie Chart

Pareto Diagram
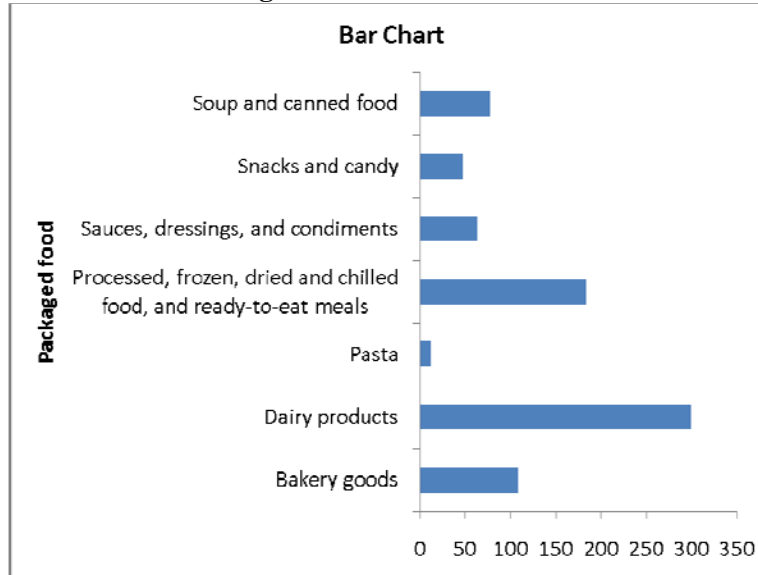
2.93   (b)   **Japan Packaged Food Consumed:**
cont.

2.93   (b)
cont.

Russian Packaged Food Consumed:
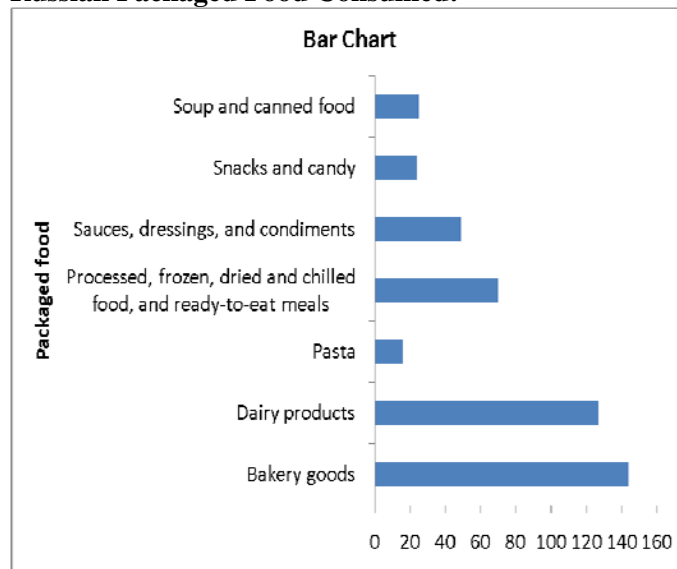
2.93   (b)
cont.
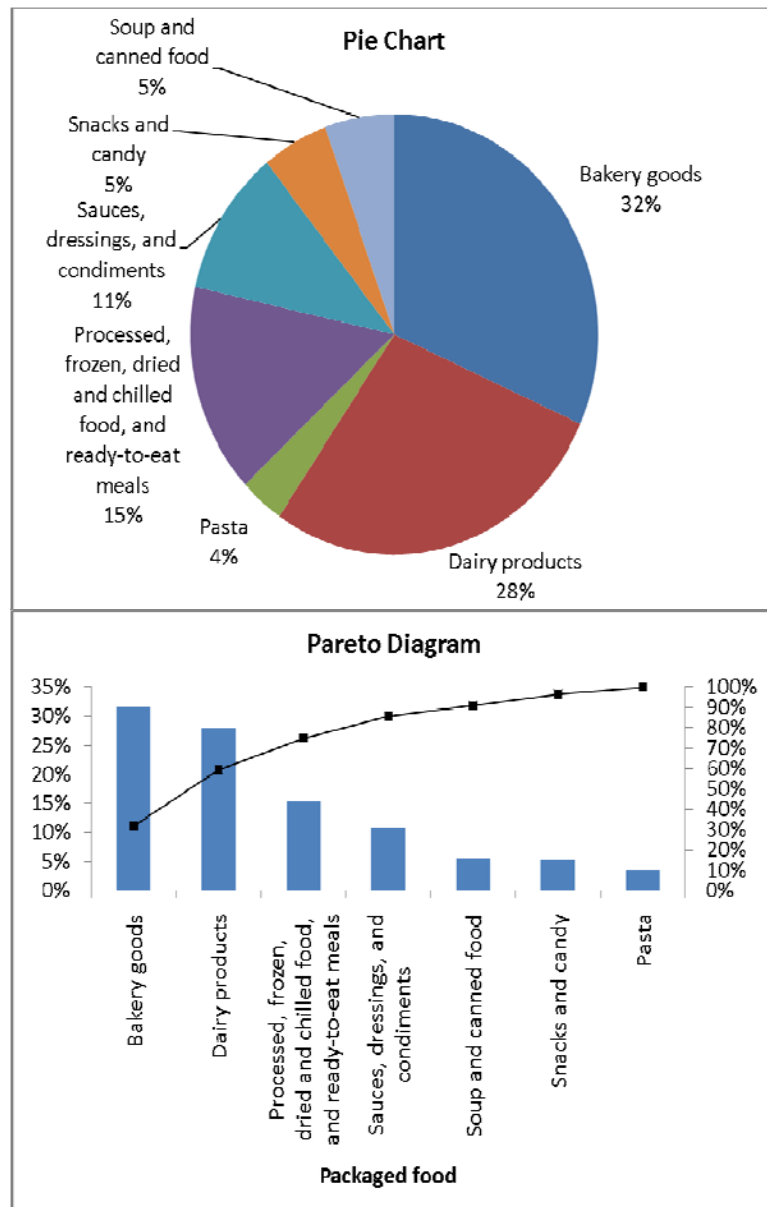
(c)   The fresh food consumption patterns between Japanese and Russians are quite similar
with vegetables taking up the largest share followed by meats and seafood while
Americans consume about the same amount of meats and seafood, and vegetables.
Among the three countries, vegetables, and meats and seafood constitute more than 60%
of the fresh food consumption.
For Americans, dairy products, and processed, frozen, dried and chilled food and ready-
to-eat meals make up slightly more than 60% of the packaged food consumption. For
Japanese, processed, frozen, dried and chilled food, and ready-to-eat meals, and dairy
products constitute more than 60% of their packaged food consumption. For the
Russians, bakery goods and dairy products take up 60% of the share of their package
food consumption.

2.94   (a)
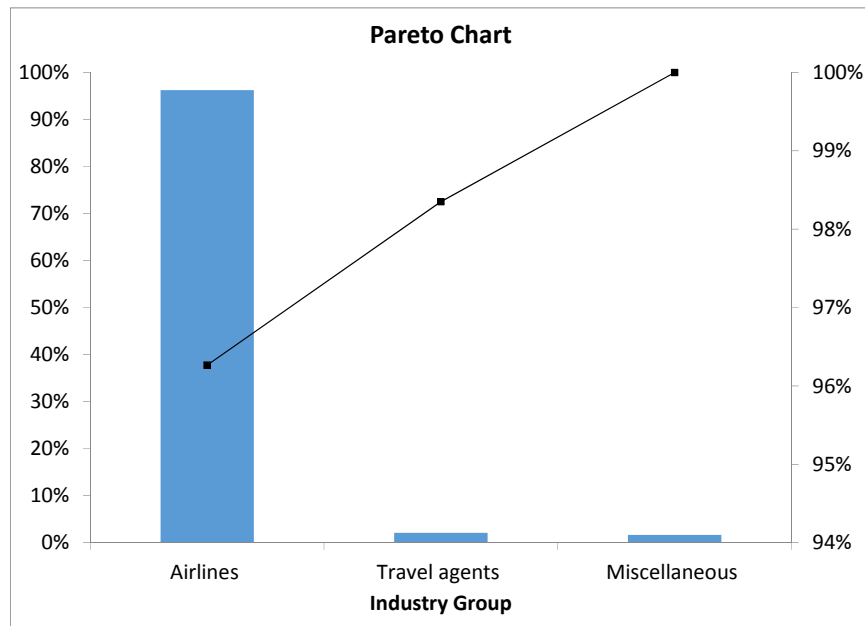
The airline industry accounts for most of the complaints.

(b)
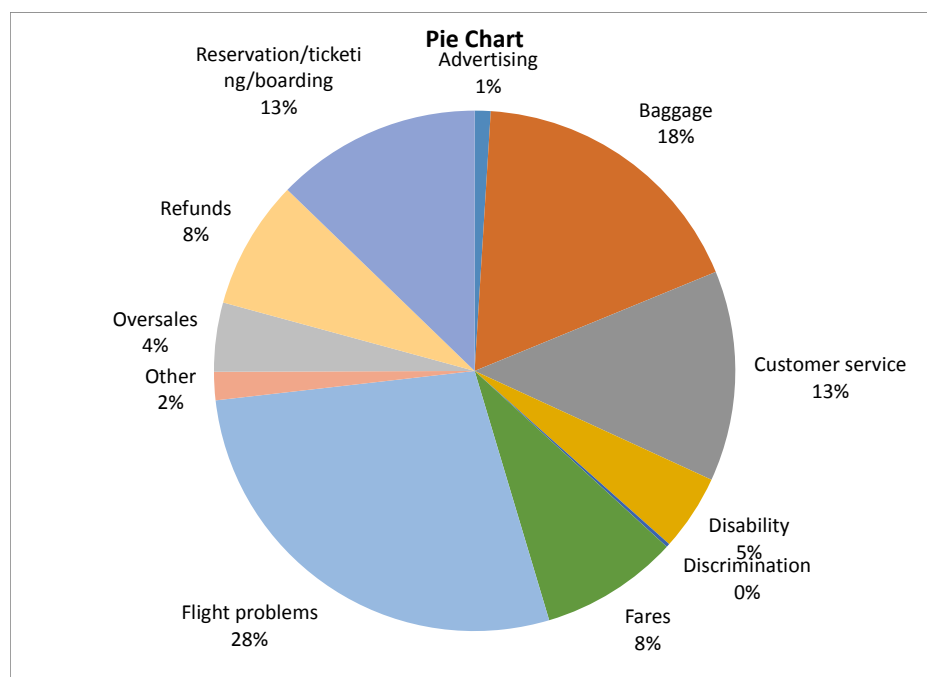
2.94   (b)
cont.
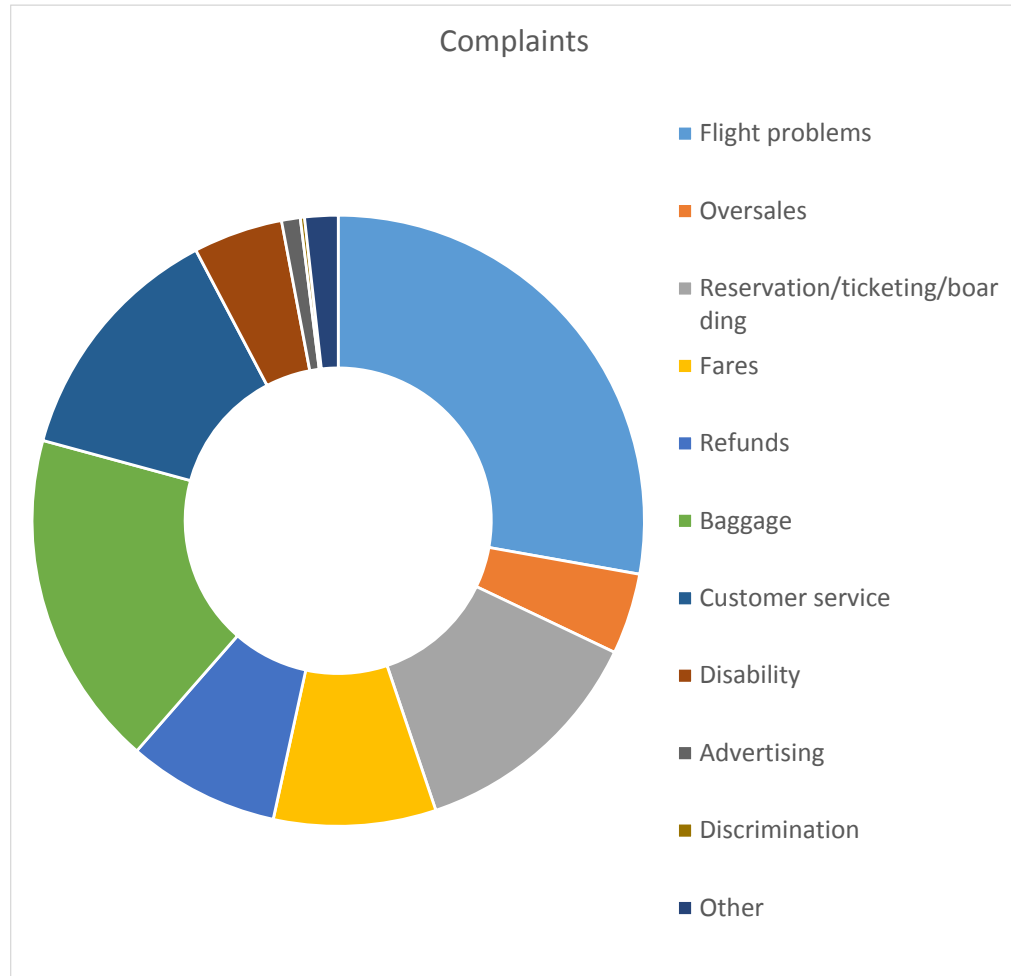
2.95   (a)

Range	Frequency	Percentage
0 but less than 25	17	34%
25 but less than 50	19	38%
50 but less than 75	5	10%
75 but less than 100	2	4%
100 but less than 125	3	6%
125 but less than 150	2	4%
150 but less than 175	2	4%

2.95   (b)
cont.
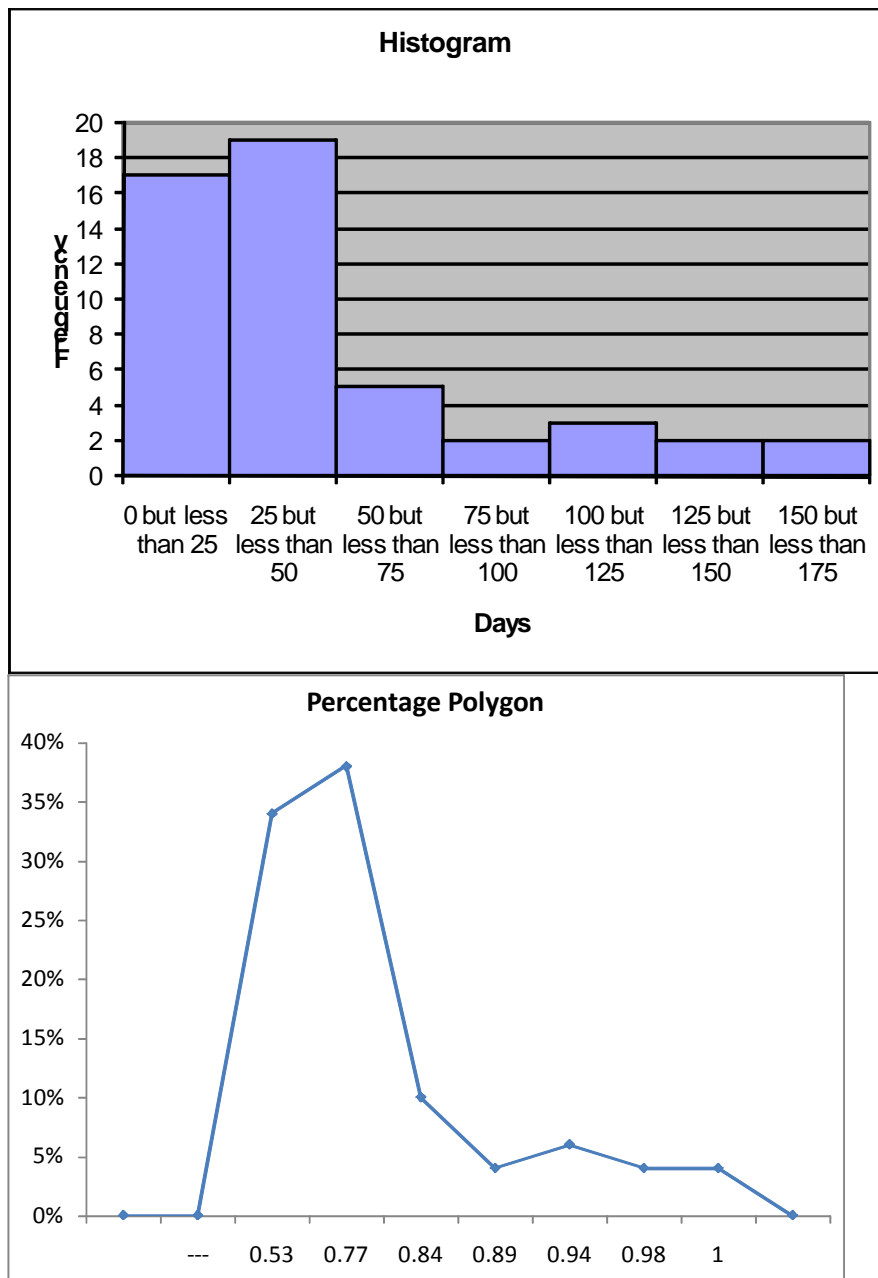
2.95   (c)
cont.

Range	Cumulative %
0 but less than 25	34%
25 but less than 50	72%
50 but less than 75	82%
75 but less than 100	86%
100 but less than 125	92%
125 but less than 150	96%
150 but less than 175	100%

Cumulative Percentage Polygon
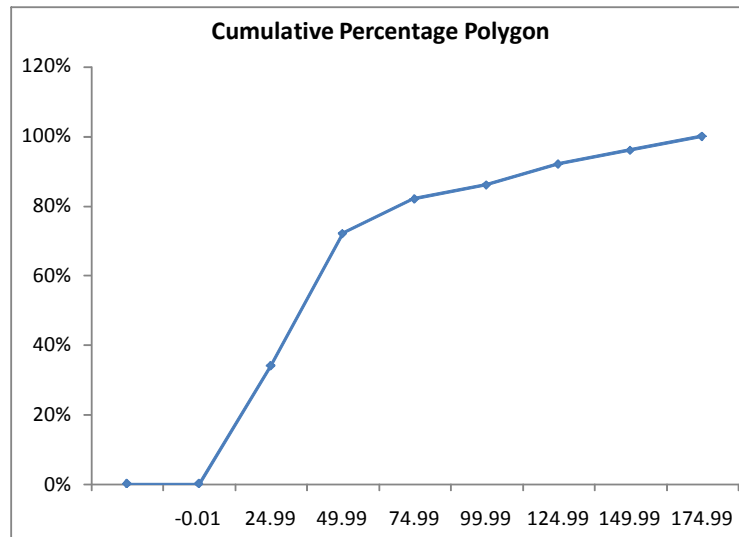
(d)   You should tell the president of the company that over half of the complaints are resolved within a month, but point out that some complaints take as long as three or four months to settle.

2.96   (a)

Percentage Histogram of Alcohol %
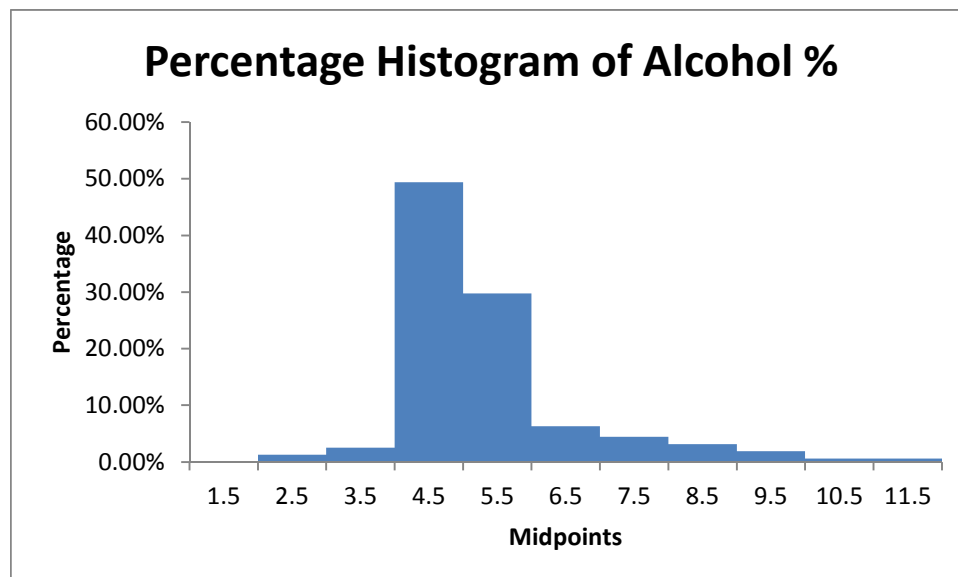
2.96   (a)
cont.
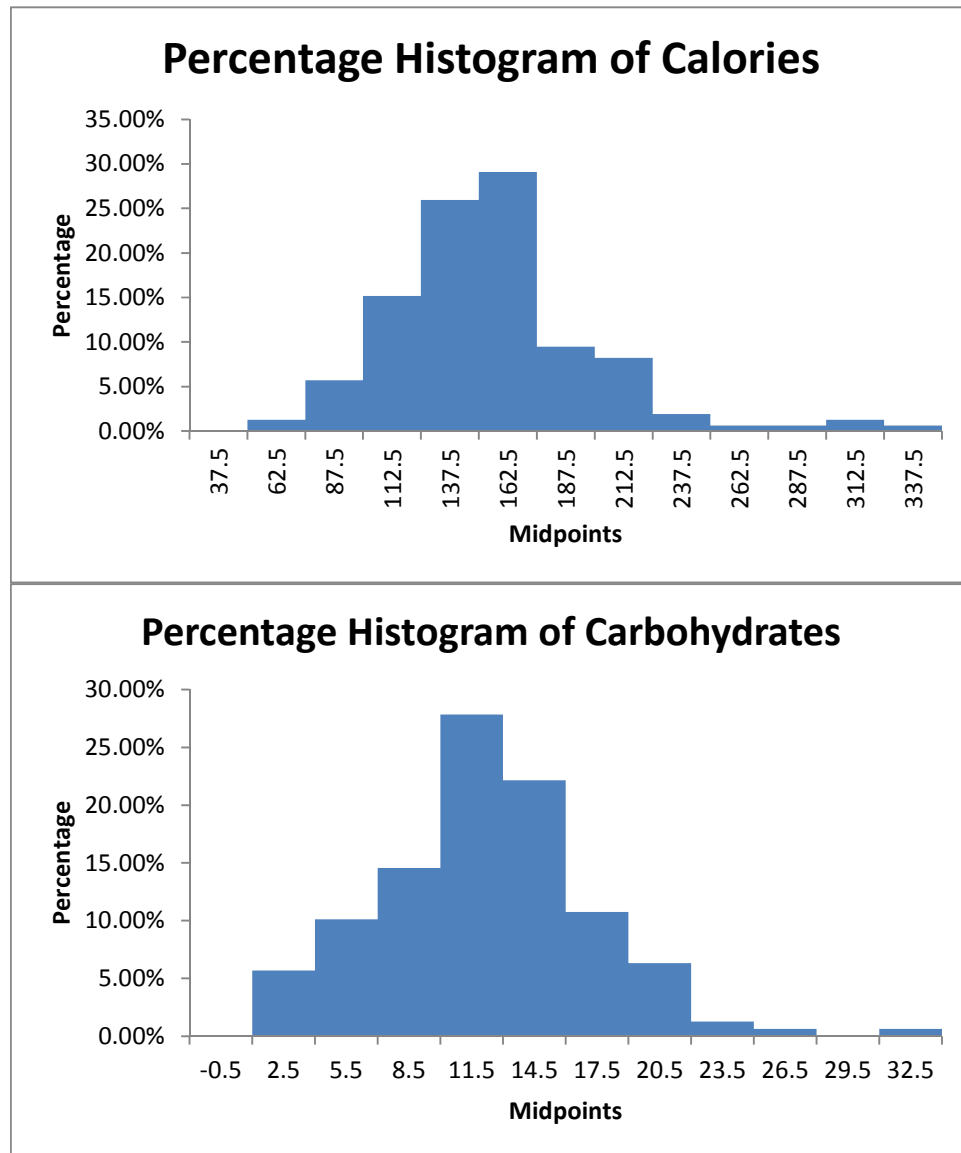
2.96   (b)
cont.
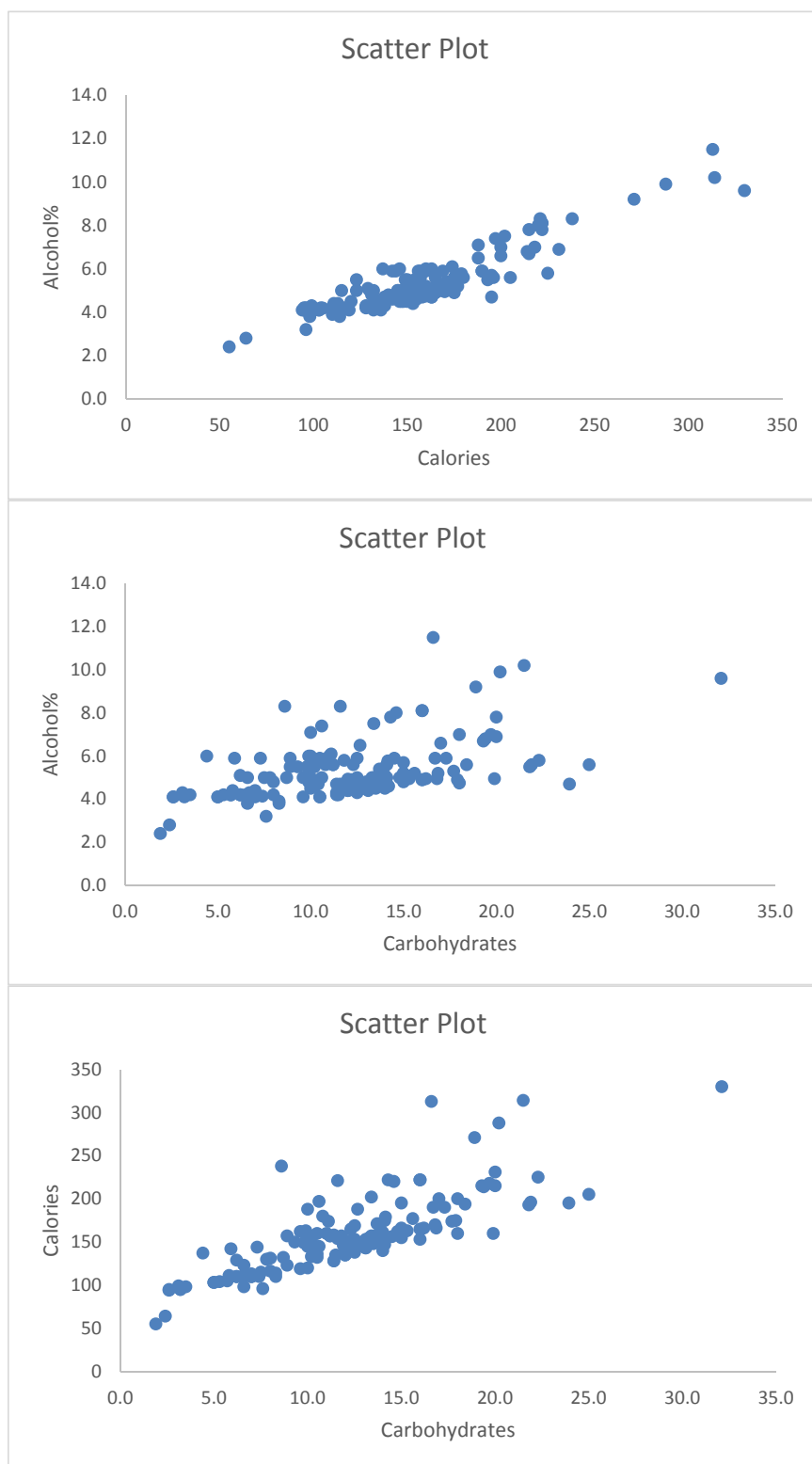
2.96   (c)   The alcohol % is concentrated between 4 and 6, with more between 4 and 5. The calories are concentrated between 125 and 175. The carbohydrates are concentrated between 10 and 16. There are outliers in the percentage of alcohol in both tails. There are a few beers with alcohol content as high as around 11%. There are a few beers with calories content higher than 250 and carbohydrates higher than 31.
There is a strong positive relationship between percentage alcohol and calories, and calories and carbohydrates and a moderately positive relationship between percentage alcohol and carbohydrates.

2.97   (a)   Ordered array:

Cigarette Tax	0.170	0.300	0.360	0.370	0.425	0.440	0.450	0.550	0.570	0.570
	0.600	0.600	0.620	0.640	0.680	0.790	0.800	0.840	0.870	0.995
	1.030	1.150	1.250	1.310	1.339	1.360	1.410	1.530	1.600	1.600
	1.660	1.700	1.700	1.780	1.980	2.000	2.000	2.000	2.000	2.000
	2.520	2.700	2.750	2.900	3.025	3.200	3.400	3.500	3.510	4.350

(b)
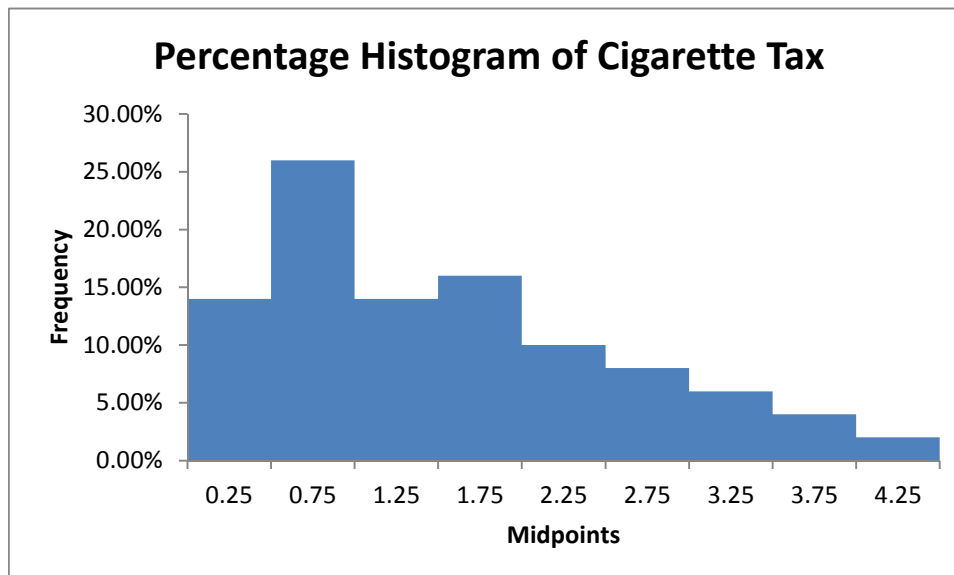
(c)   There is a $4.18 difference in the state cigarette tax between the lowest and highest. The distribution of the cigarette tax is somewhat right-skewed with one state having a cigarette tax higher than $4.00. Majority of the states though have cigarette tax concentrated around $0.75.

2.98   (a)   One-year CD:

			Stem-and-Leaf Disp
			Stem unit 0.1
Statistics			2 3
Sample Size	25		3
Mean	0.8624		4 00
Median	0.9000		5 59
Std. Deviation	0.2893		6 5
Minimum	0.2300		7 015
Maximum	1.3400		8 005
			9 055
			10 00049
			11 59
			12 2
			13 04

5-year CD

			Stem-and-Leaf Display	
			Stem unit 0.1	
Statistics			4 9	
Sample Size	25		5	
Mean	1.6384		6	
Median	1.7300		7	
Std. Deviation	0.4244		8	
Minimum	0.4900		9 3	
Maximum	2.2300		10 5	
			11	
			12 0	
			13 49	
			14 599	
			15 0	
			16 00	
			17 355	
			18 035	
			19 88	
			20 8	
			21 03	
			22 23	

2.98   (b)
cont.
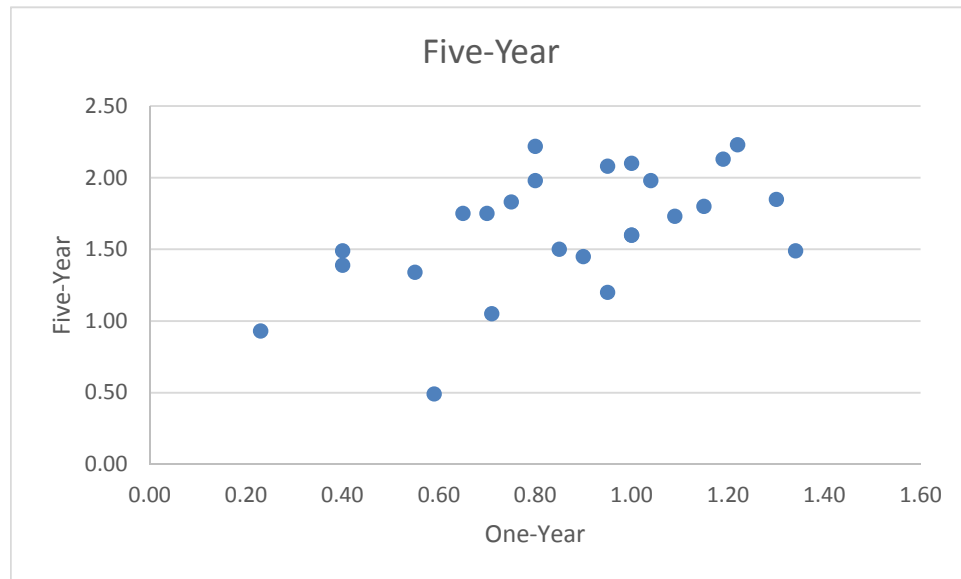
(c)   There appears to be a positive relationship between the yield of the one-year CD and the five-year CD.

2.99   (a),(c)

bin	Frequency	Percentage
0 but less than 5	19	9.50%
5 but less than 10	79	39.50%
10 but less than 15	60	30.00%
15 but less than 20	29	14.50%
20 but less than 25	9	4.50%
25 but less than 30	2	1.00%
30 but less than 35	1	0.50%
35 but less than 40	1	0.50%

(b)

2.99
cont.

(b)

Percentage Polygon

(c)

Cumulative Percentage Polygon
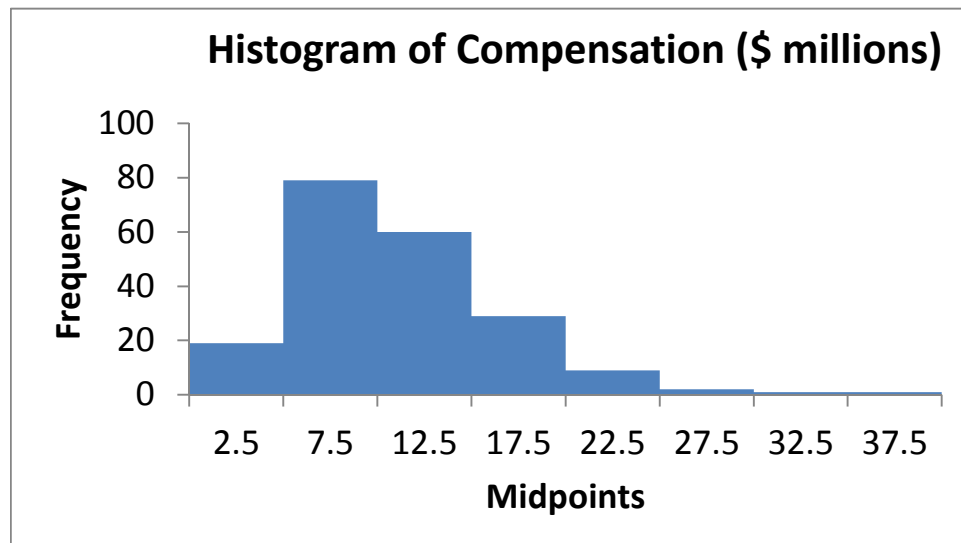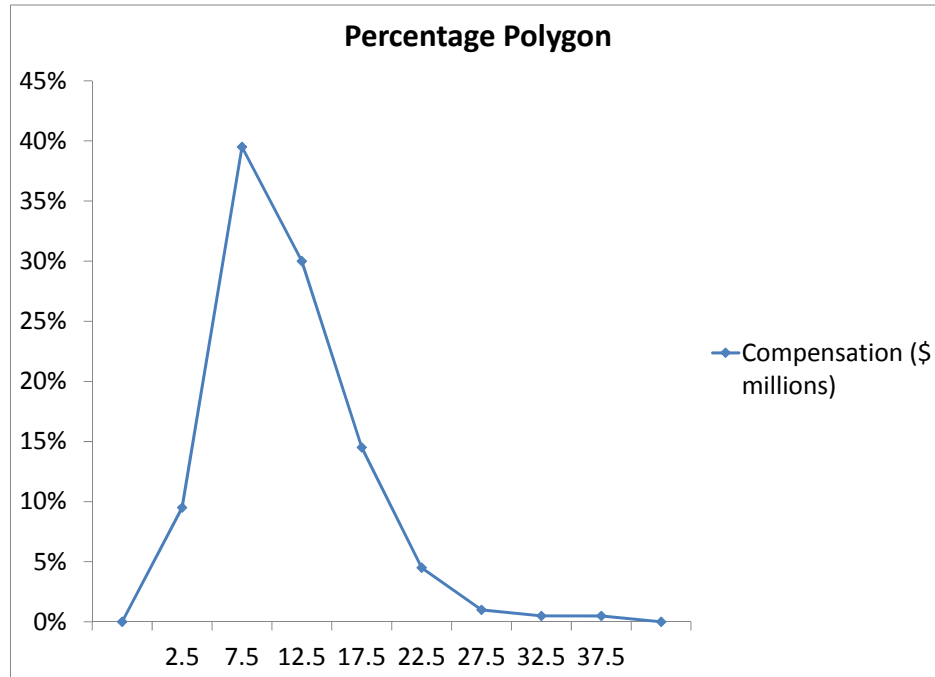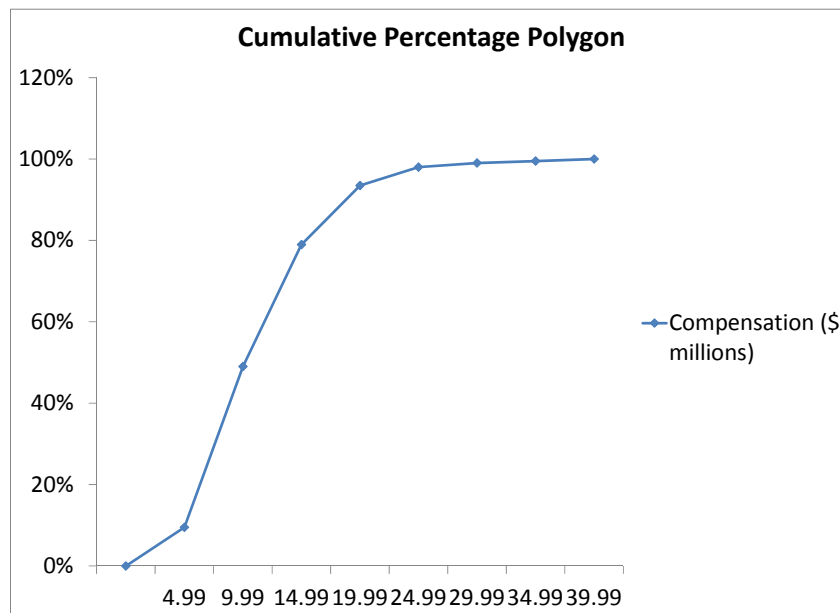
(d)   CEO compensation in 2013 is right skewed. Slightly higher than 80% of the CEOs have compensation lower than $15,000,000

2.99   (e)
cont.
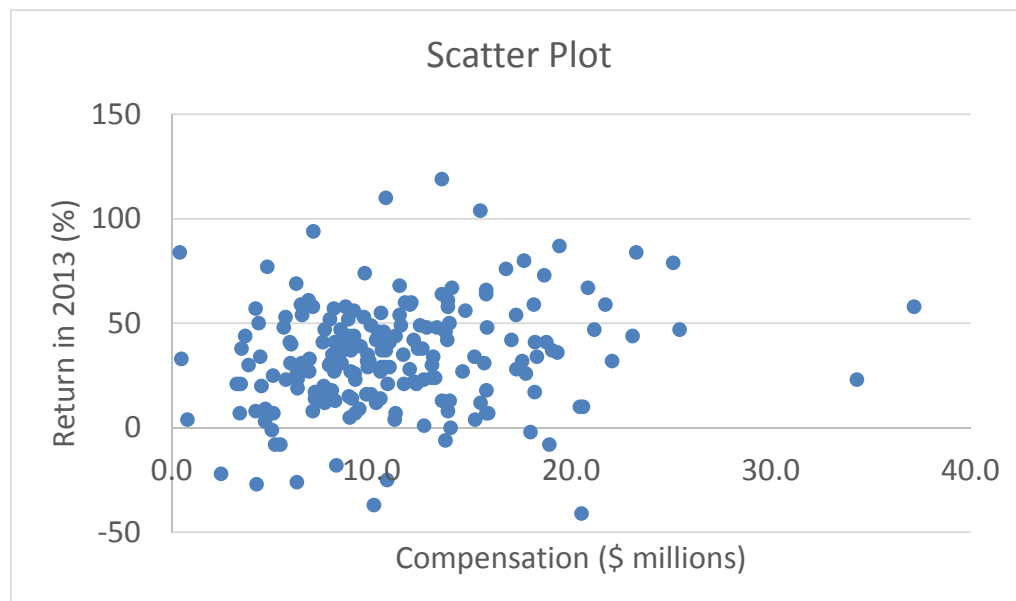
(f)   There is not any obvious relationship between the total compensation and investment
        return in 2013.

2.100   (a)

Frequencies (Boston)

Weight (Boston)	Frequency	Percentage
3015 but less than 3050	2	0.54%
3050 but less than 3085	44	11.96%
3085 but less than 3120	122	33.15%
3120 but less than 3155	131	35.60%
3155 but less than 3190	58	15.76%
3190 but less than 3225	7	1.90%
3225 but less than 3260	3	0.82%
3260 but less than 3295	1	0.27%

(b)

Frequencies (Vermont)

Weight (Vermont)	Frequency	Percentage
3550 but less than 3600	4	1.21%
3600 but less than 3650	31	9.39%
3650 but less than 3700	115	34.85%
3700 but less than 3750	131	39.70%
3750 but less than 3800	36	10.91%
3800 but less than 3850	12	3.64%
3850 but less than 3900	1	0.30%

2.100   (c)
cont.

% Histogram (Boston)

% Histogram (Vermont)
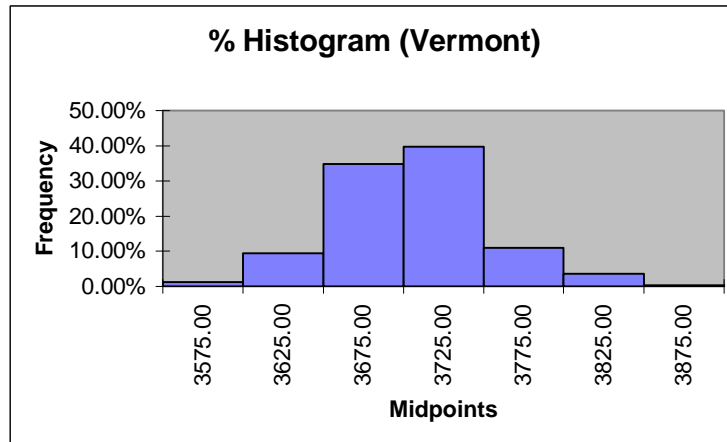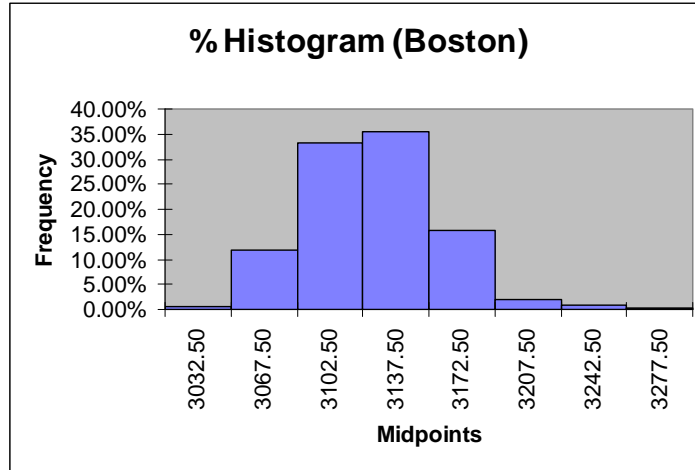
(d)   0.54% of the "Boston" shingles pallets are underweight while 0.27% are overweight.
1.21% of the "Vermont" shingles pallets are underweight while 3.94% are overweight.

2.101   (a),(c) **Two-star:**

	Average price		Frequency	Percentage	Cumulative Pctage.
10	but less than	20	1	2.44%	2.44%
20	but less than	30	1	2.44%	4.88%
30	but less than	40	5	12.20%	17.07%
40	but less than	50	8	19.51%	36.59%
50	but less than	60	4	9.76%	46.34%
60	but less than	70	8	19.51%	65.85%
70	but less than	80	7	17.07%	82.93%
80	but less than	90	4	9.76%	92.68%
90	but less than	100	2	4.88%	97.56%
100	but less than	110	1	2.44%	100.00%

2.101   (a),(c)   **Three-star:**
cont.

	Average price		Frequency	Percentage	Cumulative Pctage.
25	but less than	40	1	2.44%	2.44%
40	but less than	55	5	12.20%	14.63%
55	but less than	70	4	9.76%	24.39%
70	but less than	85	9	21.95%	46.34%
85	but less than	100	11	26.83%	73.17%
100	but less than	115	5	12.20%	85.37%
115	but less than	130	3	7.32%	92.68%
130	but less than	145	1	2.44%	95.12%
145	but less than	160	2	4.88%	100.00%

Four-star:

	Average price		Frequency	Percentage	Cumulative Pctage.
0	but less than	20	1	2.44%	2.44%
20	but less than	40	0	0.00%	2.44%
40	but less than	60	1	2.44%	4.88%
60	but less than	80	7	17.07%	21.95%
80	but less than	100	4	9.76%	31.71%
100	but less than	120	9	21.95%	53.66%
120	but less than	140	7	17.07%	70.73%
140	but less than	160	5	12.20%	82.93%
160	but less than	180	1	2.44%	85.37%
180	but less than	200	5	12.20%	97.56%
200	but less than	220	1	2.44%	100.00%

(b)   **Two-star:**

2.101   (b)
cont.

Percentage Polygon
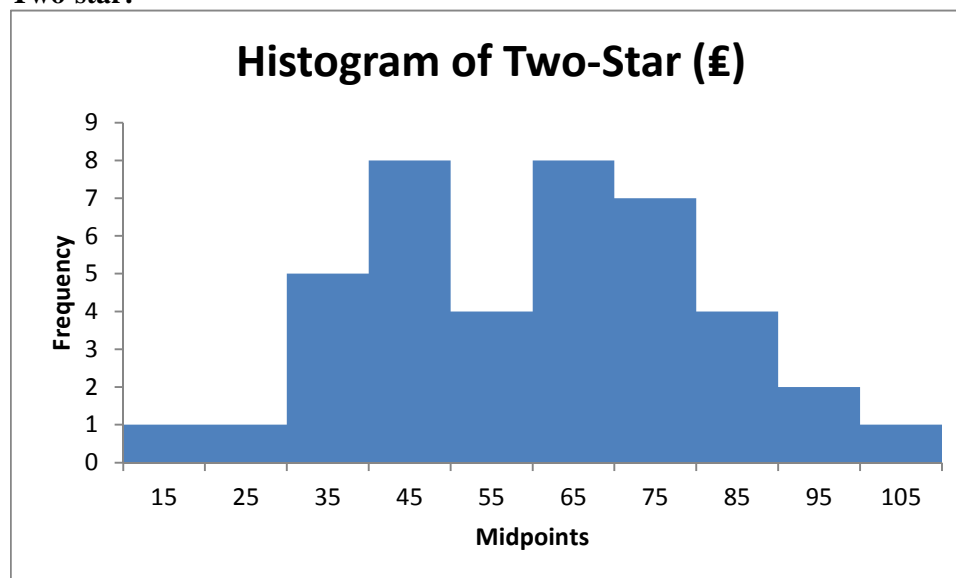
Three-star:
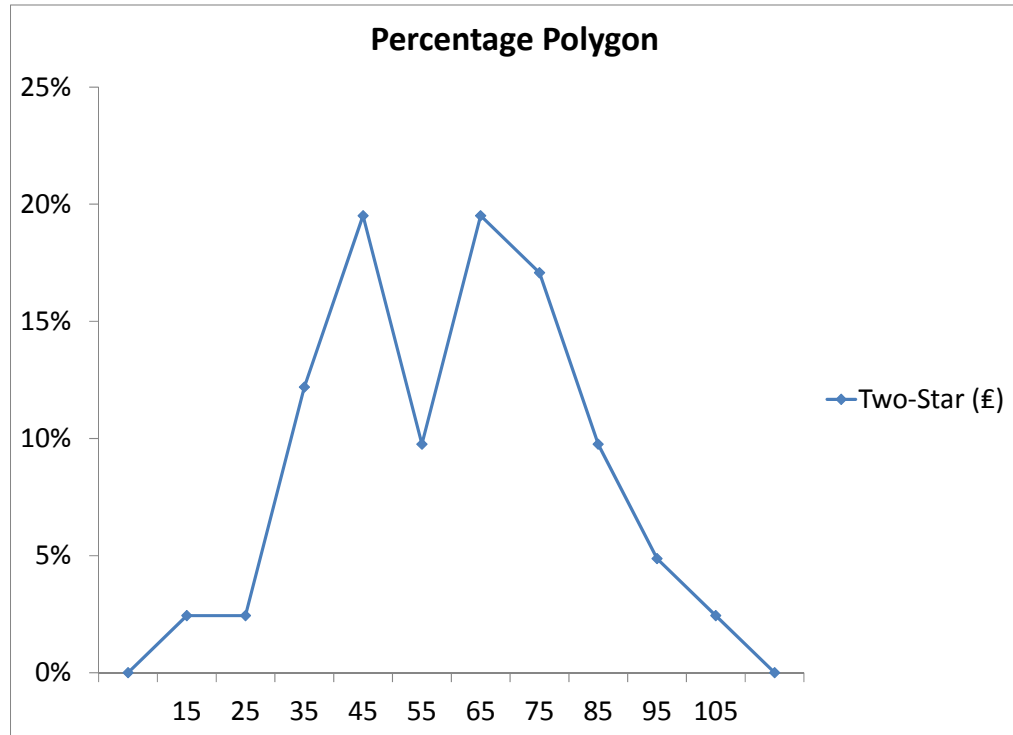
Histogram of Three-Star (£)

2.101   (b)
cont.

Percentage Polygon
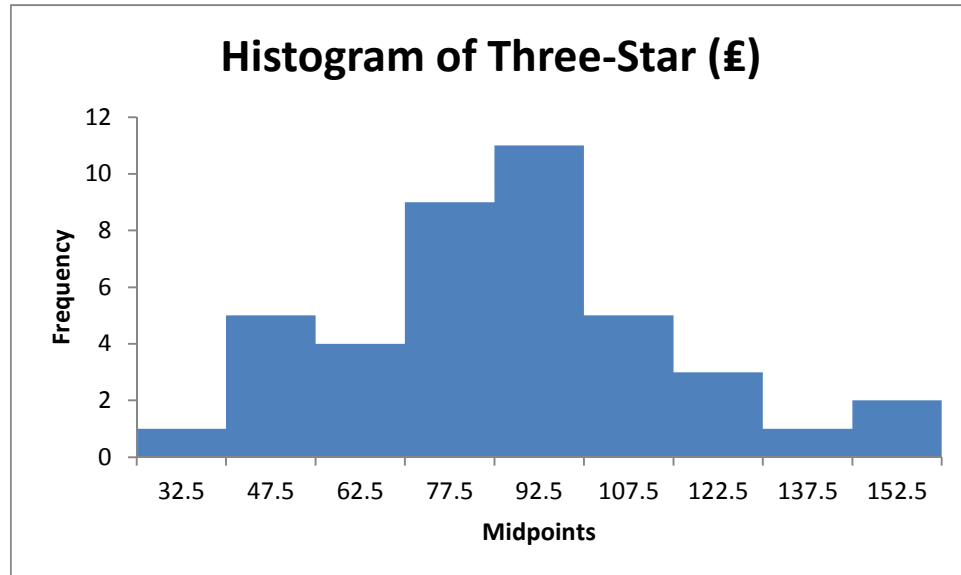
Four-star:
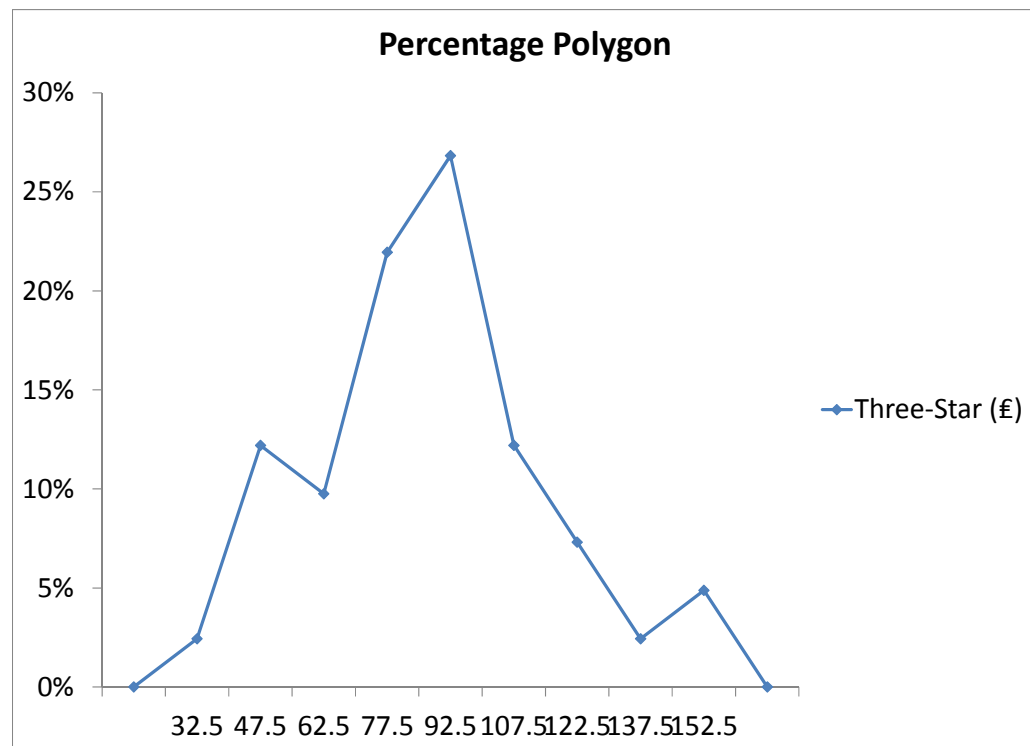
Histogram of Four-Star(£)
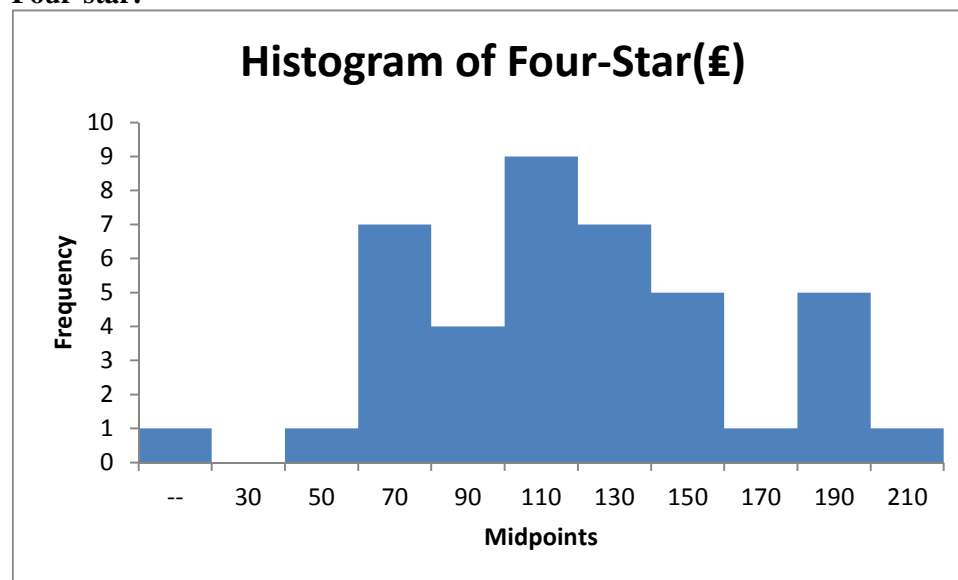
2.101   (b)
cont.
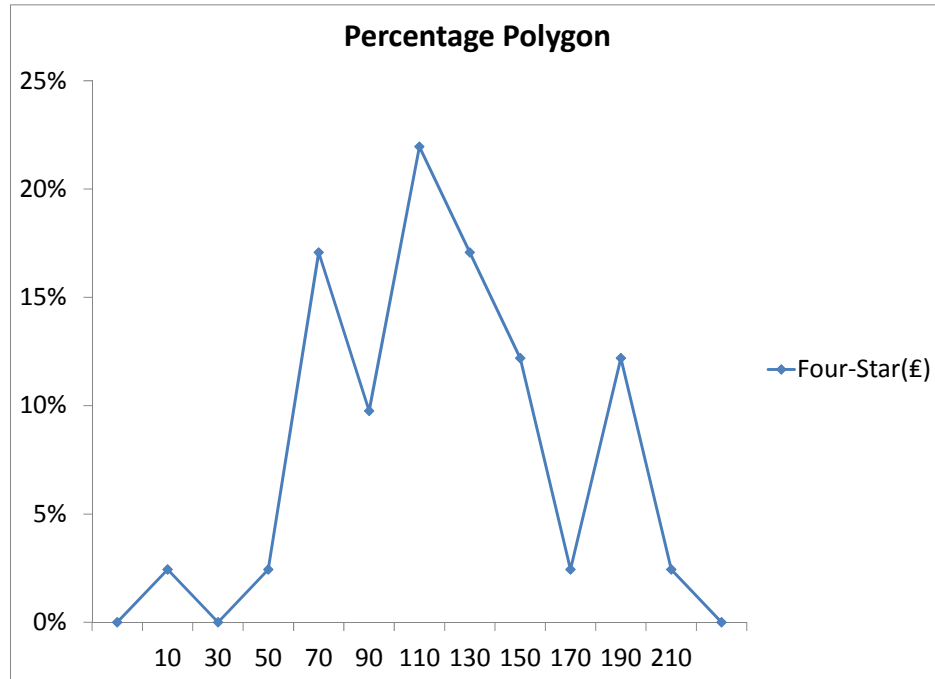
(c)   **Two-star:**
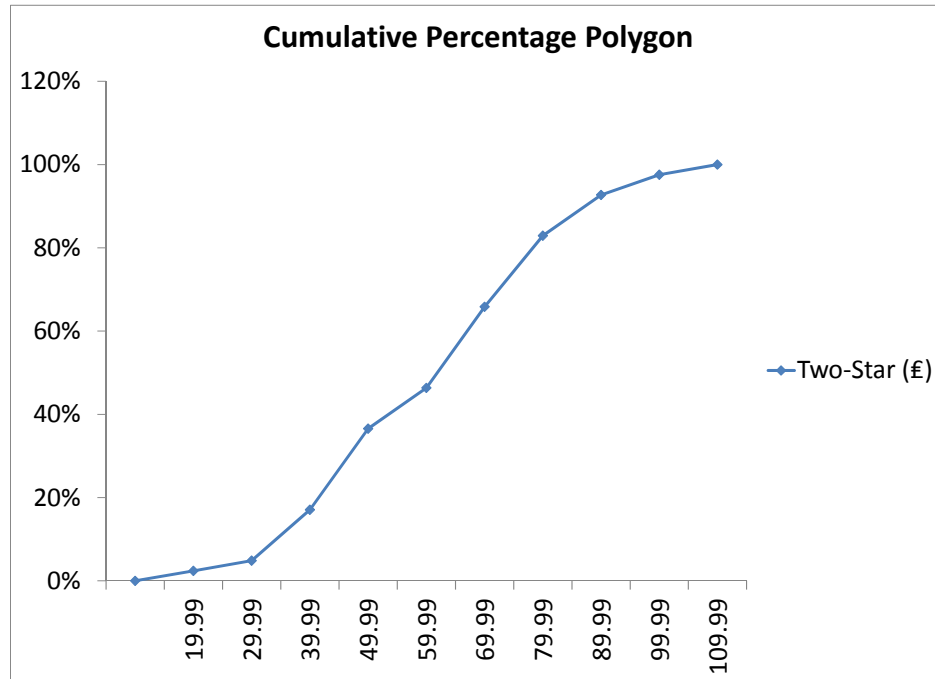
2.101   (c)   **Three-star:**
cont.
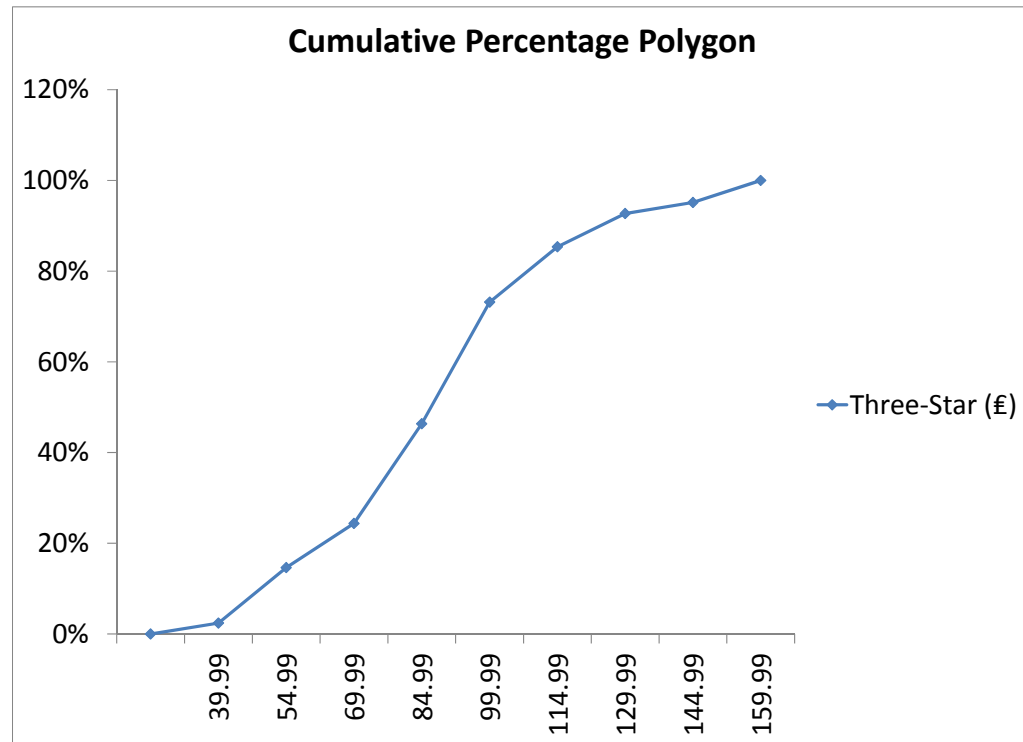
Four-star:
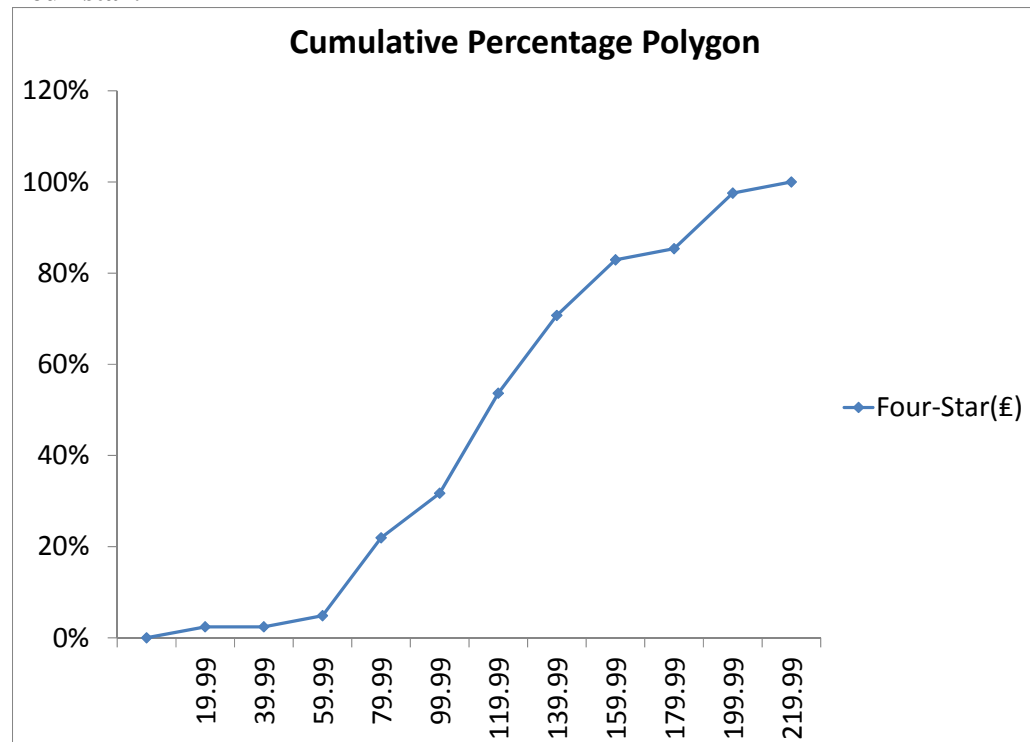
(d)   The price of two-star and four-star hotels are slightly left-skewed while that of three-star isslight right-skewed. The median price of two-star , three-star and four-star hotels is around 65, 92.5, and 110 English pounds, respectively.

2.101   (e)
cont.
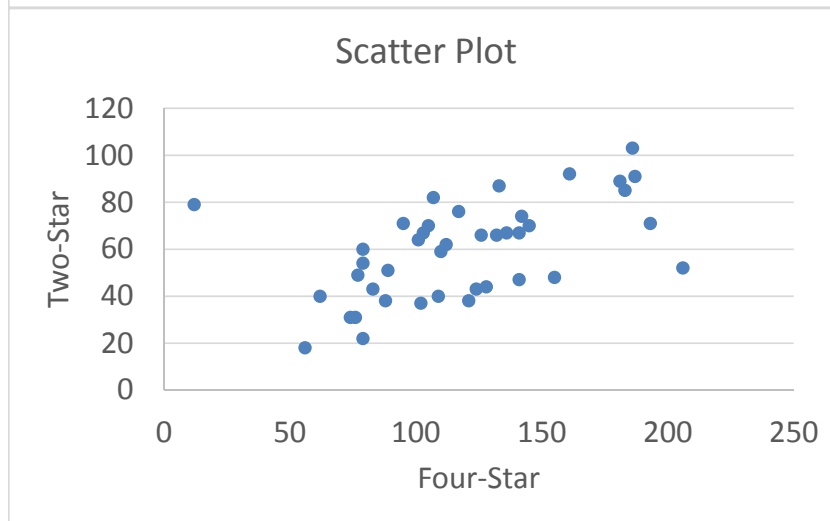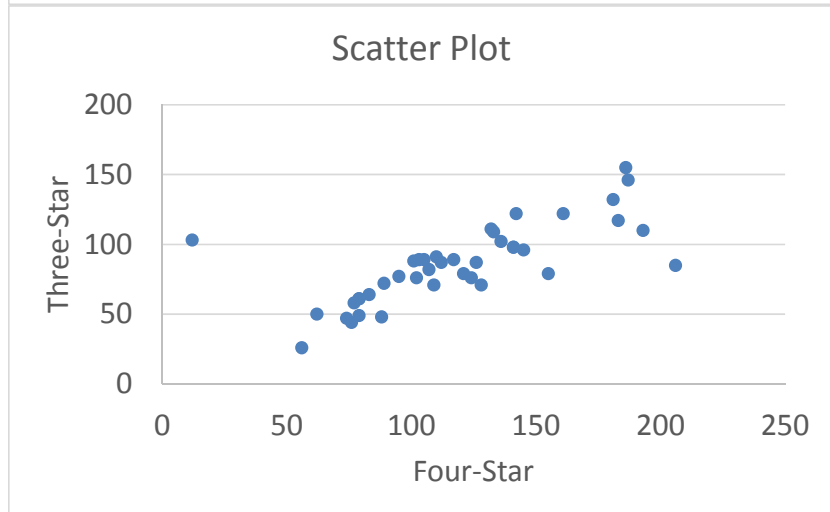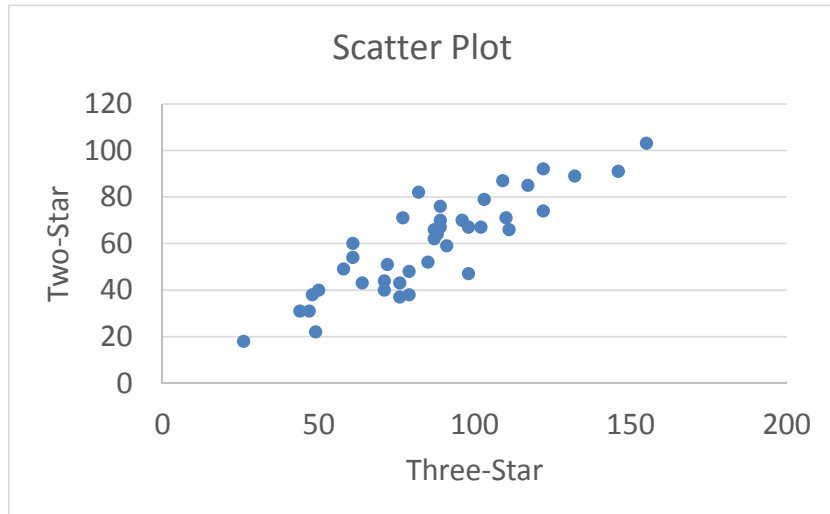
(f)   The relationship of the price between two-star and three-star, three-star and four-star, and
two-star and four-star hotels are all positve.

2.102   (a)

Calories	Frequency	Percentage	Percentage Less Than
50 up to 100	3	12%	12%
100 up to 150	3	12	24
150 up to 200	9	36	60
200 up to 250	6	24	84
250 up to 300	3	12	96
300 up to 350	0	0	96
350 up to 400	1	4	100
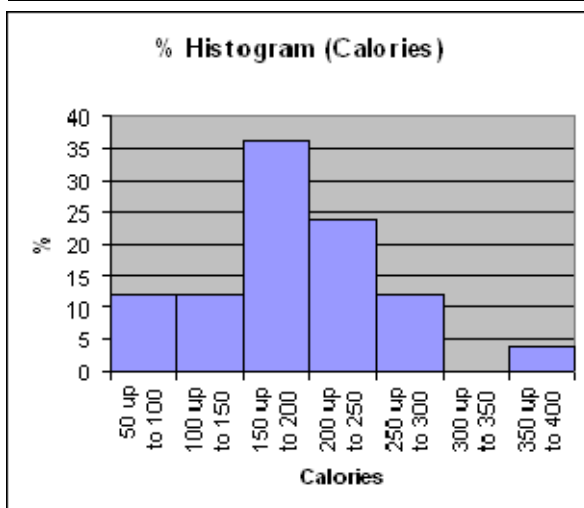
(b)

Cholesterol	Frequency	Percentage	Percentage Less Than
0 up to 50	2	8	8%
50 up to 100	17	68	76
100 up to 150	4	16	92
150 up to 200	1	4	96
200 up to 250	0	0	96
250 up to 300	0	0	96
300 up to 350	0	0	96
350 up to 400	0	0	96
400 up to 450	0	0	96
450 up to 500	1	4	100

2.102 (b)
cont.
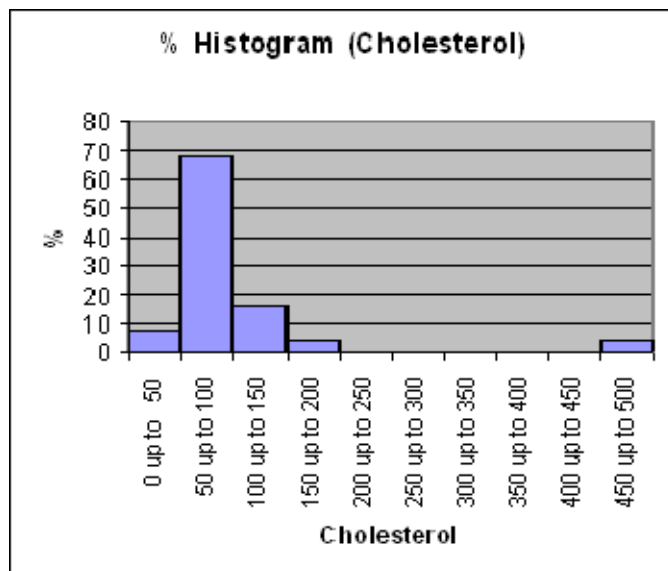
(c)   The sampled fresh red meats, poultry, and fish vary from 98 to 397 calories per serving, with the highest concentration between 150 to 200 calories. One protein source, spareribs, with 397 calories, is more than 100 calories above the next highest caloric food. The protein content of the sampled foods varies from 16 to 33 grams, with 68% of the data values falling between 24 and 32 grams. Spareribs and fried liver are both very different from other foods sampled—the former on calories and the latter on cholesterol content.

2.103 (a)
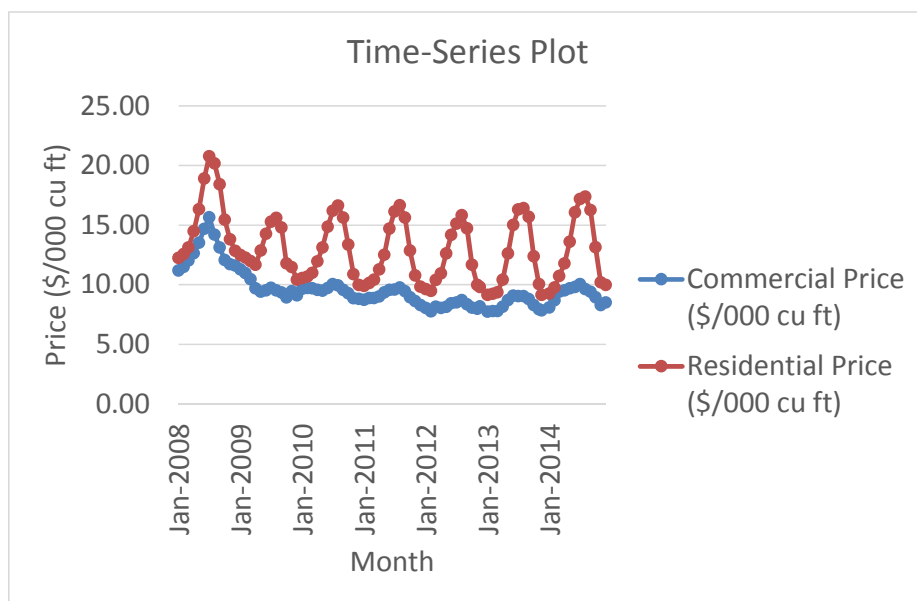
(b)   The commercial average price was highest in the summer of 2008 and had since declined. The residential average price of gasoline in the United States is higher in the summer in general and seems to peak in June.

2.103   (c)
cont.
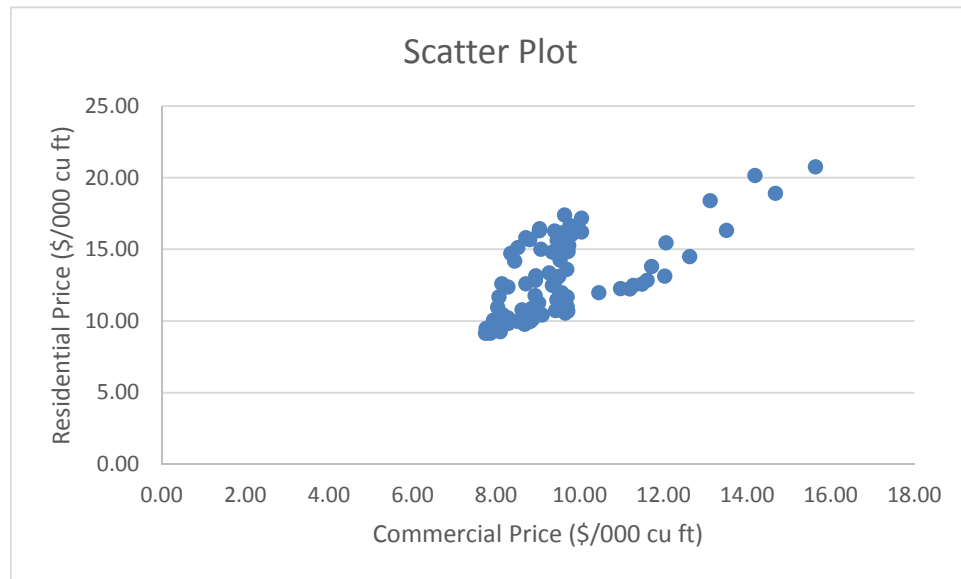
(d)   There appears to be a slight positive relationship between the commercial price and residential price.

2.104   (a)
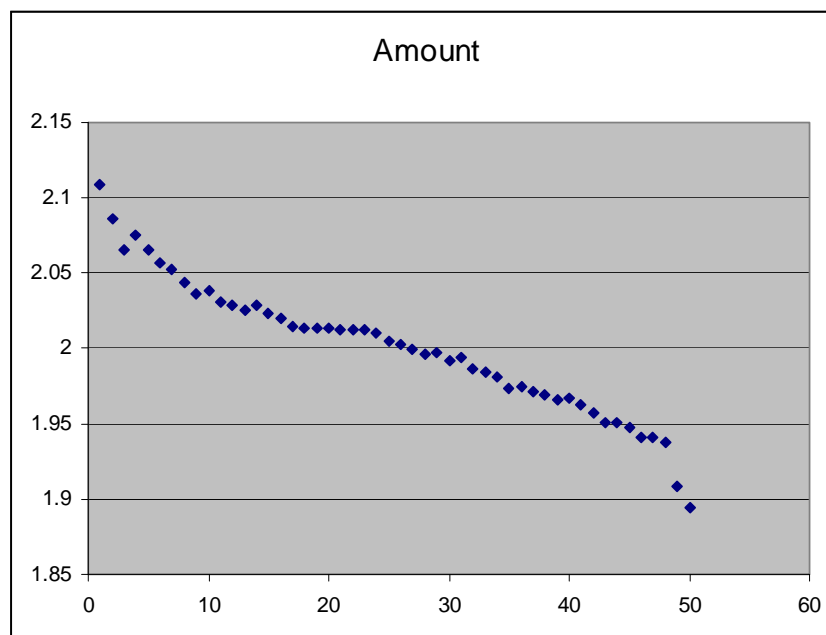
(b)   There is a downward trend in the amount filled.
(c)   The amount filled in the next bottle will most likely be below 1.894 liter.
(d)   The scatter plot of the amount of soft drink filled against time reveals the trend of the data, whereas a histogram only provides information on the distribution of the data.

2.105   (a)
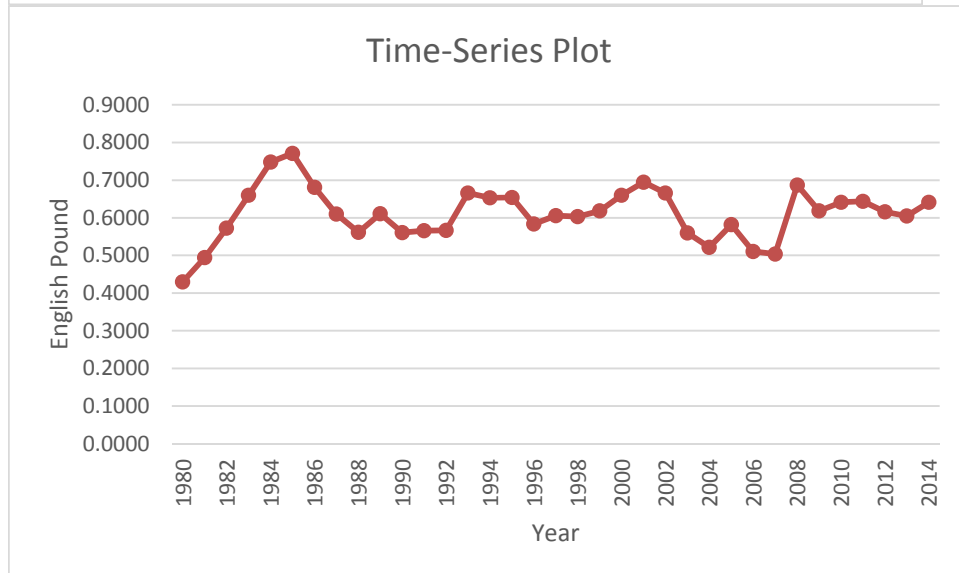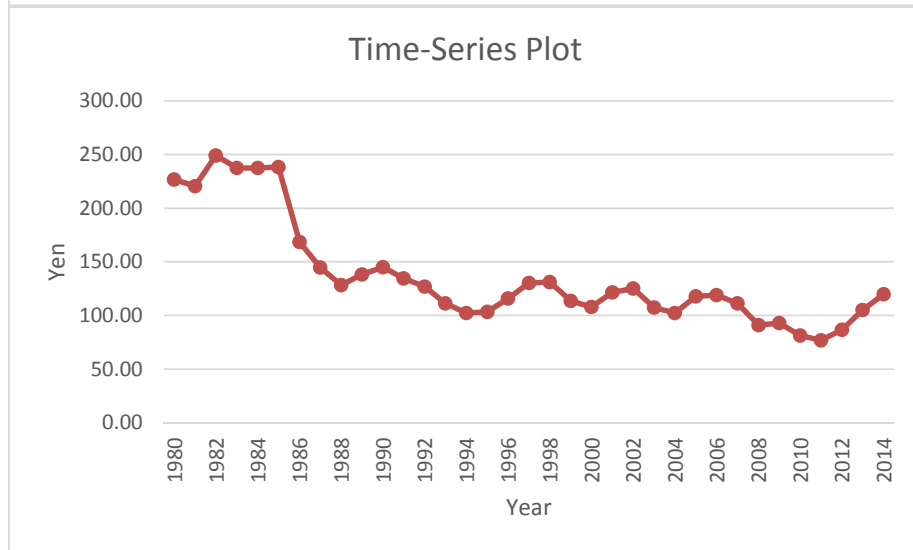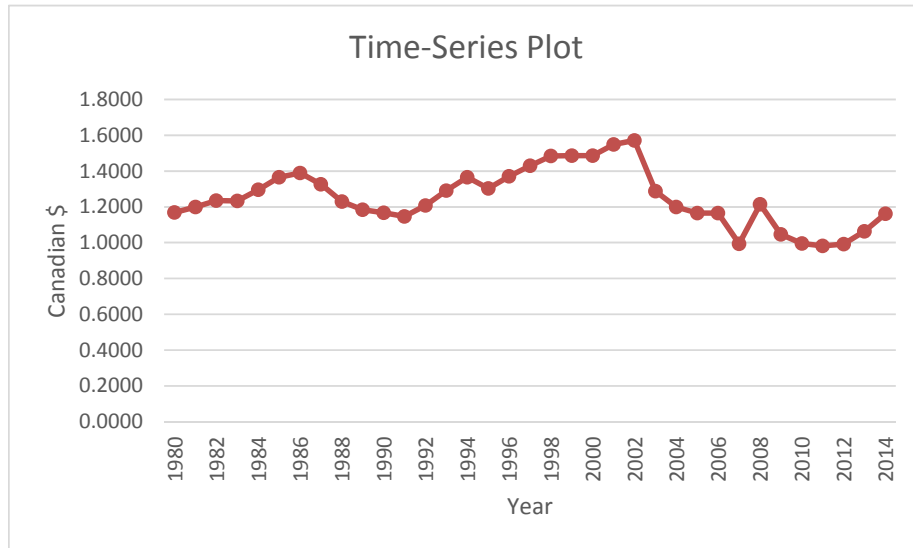
2.105   (b)   The Japanese yen had depreciated against the U.S. dollar since 1982 while the Canadian
cont.   dollar appreciated gradually from 1980 to 1987 and from 1991 to 2002 and then started to
                depreciate since. The English pound to U.S. dollar's exchange rate has been quite stable
                since 1983.
          (c)   The U.S. dollar has appreciated against the Japanese yen since 1980 and appreciated
                against the Canadian dollar since 2002 in general while the exchange rate against the
                English bound has been stable in general.
          (d)
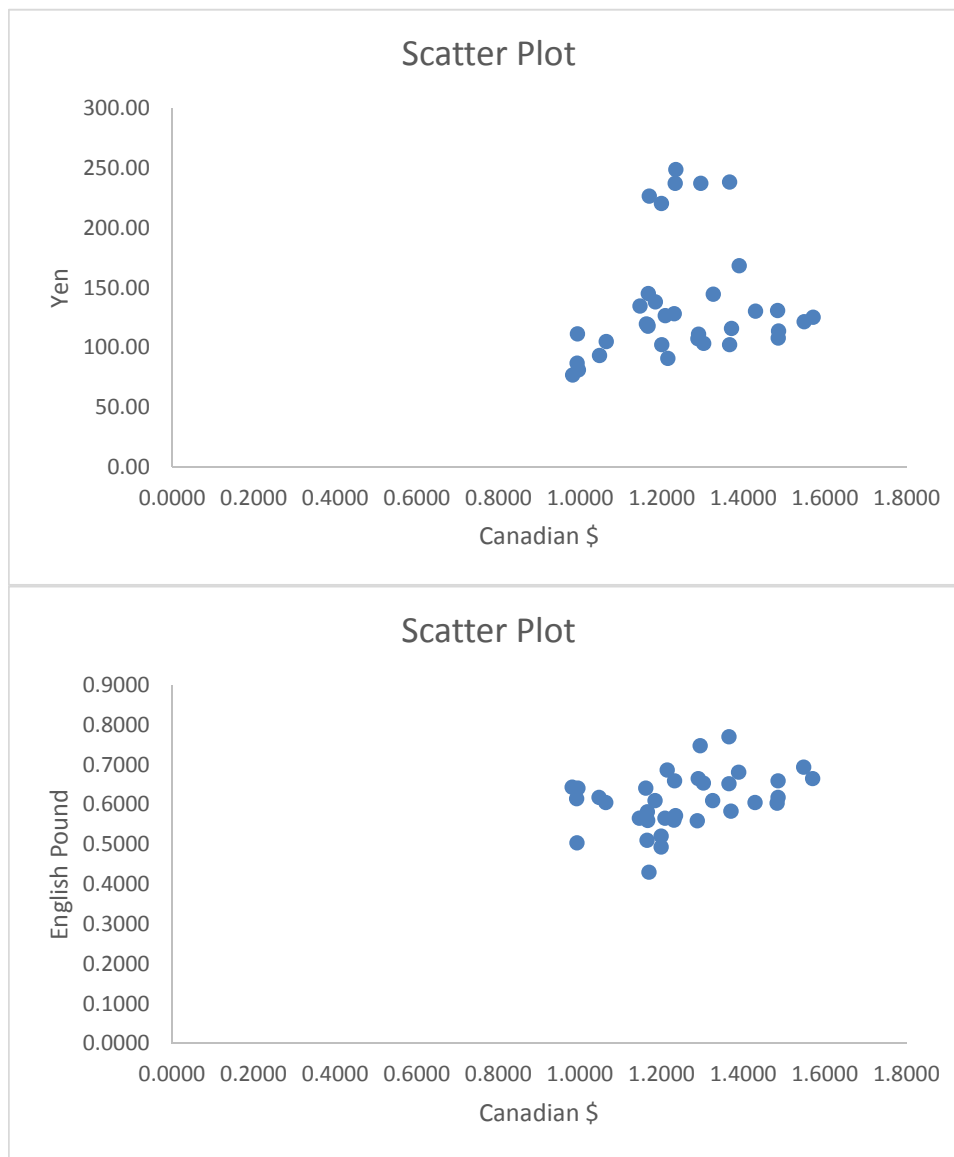
2.105   (d)
cont.
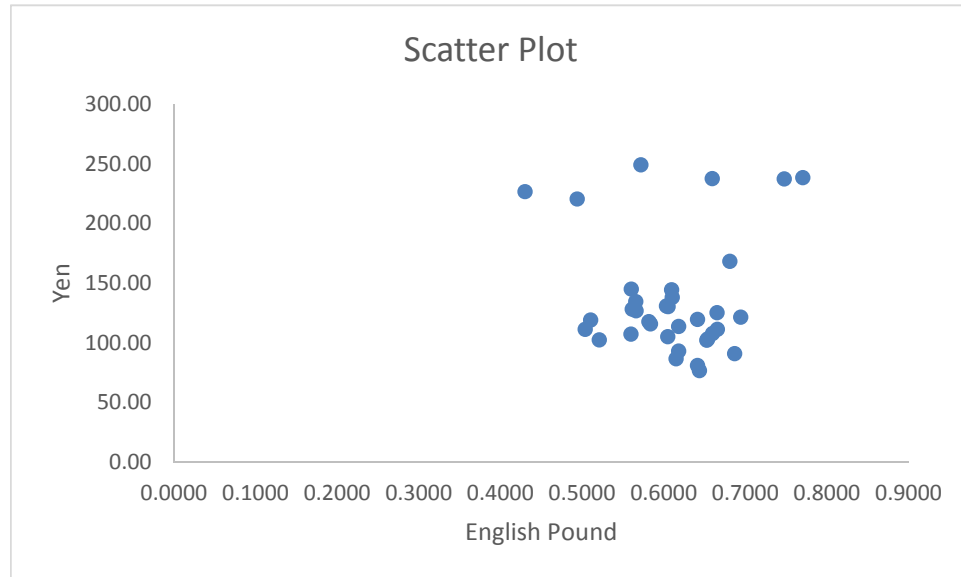
(e)   There is not any obvious relationship between the Canadian dollar and Japanese yen in terms of the U.S. dollar nor any relationship between the Japanese yen and English pound. There is a slightly positive relationship between the Canadian dollar and English pound which reflects the fact that when the Canadian dollar appreciated against the U.S. dollar, so did the English pound.

2.106   (a)

Variations	Percentage of Download
Original Call to Action Button	9.64%
New Call to Action Button	13.64%

(b)
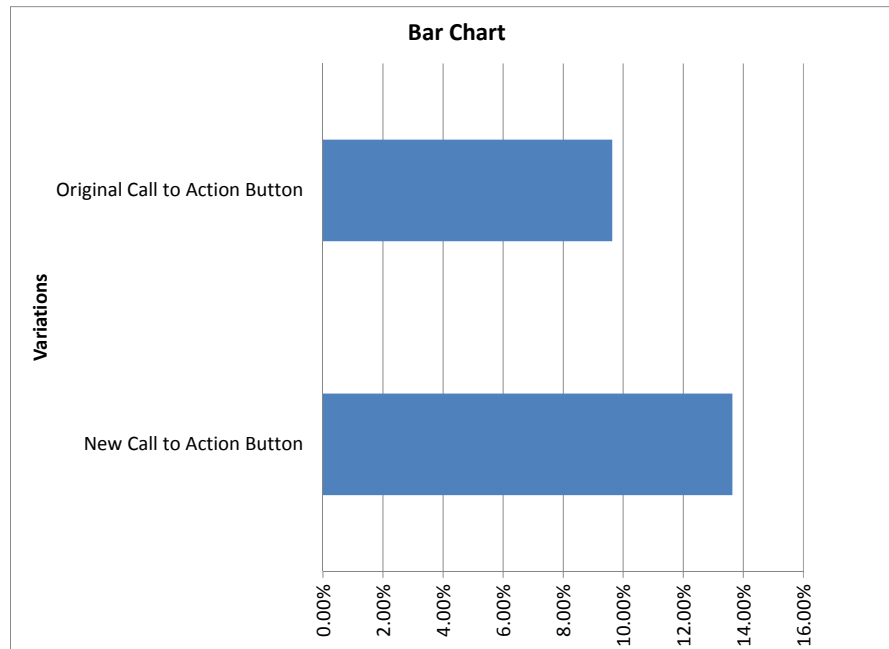
2.106   (c)   The New Call to Action Button has a higher percentage of downloads at 13.64% when
cont.   compared to the Original Call to Action Button with a 9.64% of downloads.

(d)

Variations	Percentage of Downloads
Original web design	8.90%
New web design	9.41%

(e)
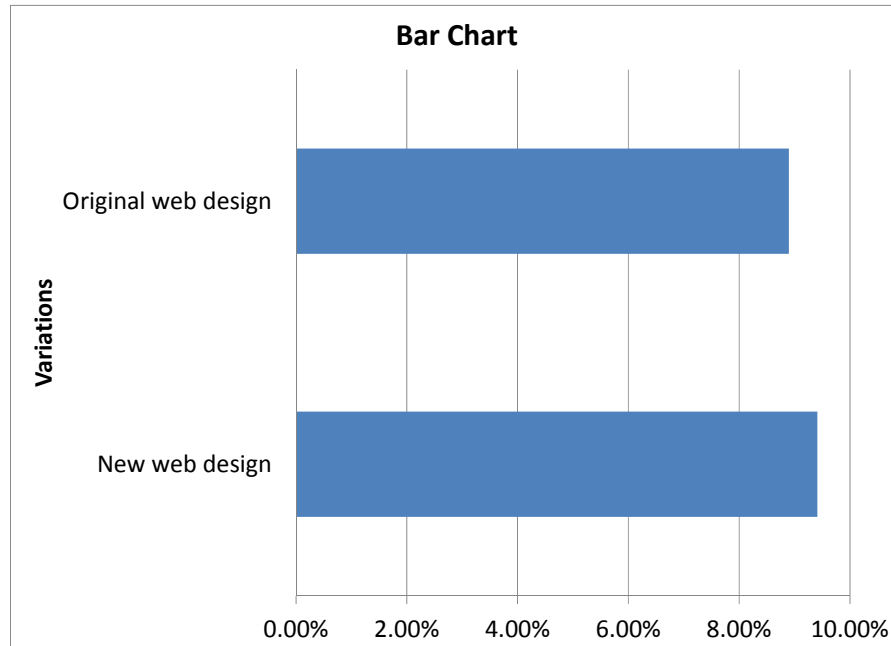
(f)   The New web design has only a slightly higher percentage of downloads at 9.41% when
       compared to the Original web design with an 8.90% of downloads.

(g)   The New web design is only slightly more successful than the Original web design while
       the New Call to Action Button is much more successful than the Original Call to Action
       Button with about 41% higher percentage of downloads.

(h)

Call to Action Button	Web Design	Percentage of Downloads
Old	Old	8.30%
New	Old	13.70%
Old	New	9.50%
New	New	17.00%

(i)   The combination of the New Call to Action Button and the New web design results in
       slightly more than twice as high a percentage of downloads than the combination of the
       Old Call to Action Button and Old web design.

(j)   The New web design is only slightly more successful than the Original web design while
       the New Call to Action Button is much more successful than the Original Call to Action
       Button with about 41% higher percentage of downloads. However, the combination of
       the New Call to Action Button and New web design results in more than twice as high a
       percentage of downloads than the combination of the Old Call to Action Button and Old
       web design.
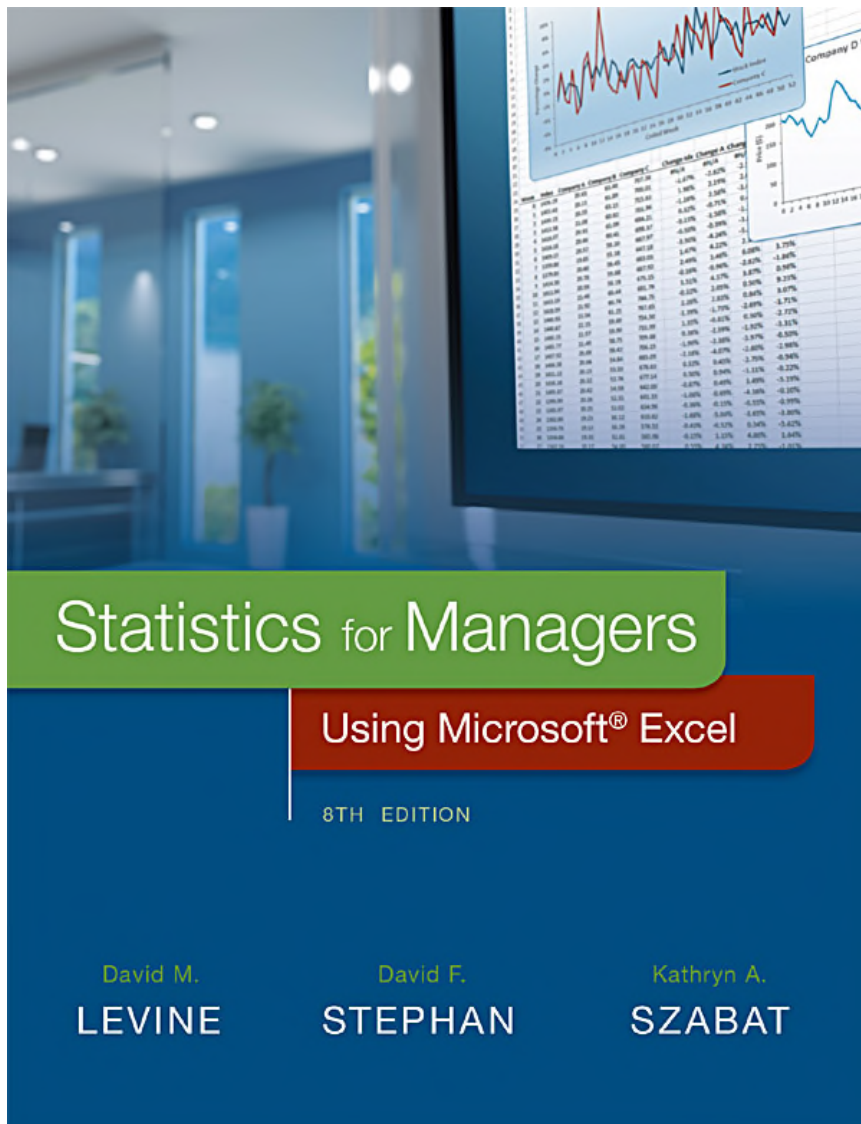
Chapter 2

Organizing and Visualizing Variables

Objectives

In this chapter you learn:

- Organizing categorical variables.

- Organizing numerical variables.

- Visualizing categorical variables.

- Visualizing numerical variables.

- Organizing and visualizing a mix of variables.

- The challenge in organizing and visualizing variables.

Organizing Data Creates Both Tabular And Visual Summaries

DC<u>O</u>VA

- Summaries both guide further exploration and sometimes facilitate decision making.

- Visual summaries enable rapid review of larger amounts of data & show possible significant patterns.

- Often, the **O**rganize and **V**isualize step in **DCOVA** occur concurrently.

Categorical Data Are Organized By Utilizing Tables
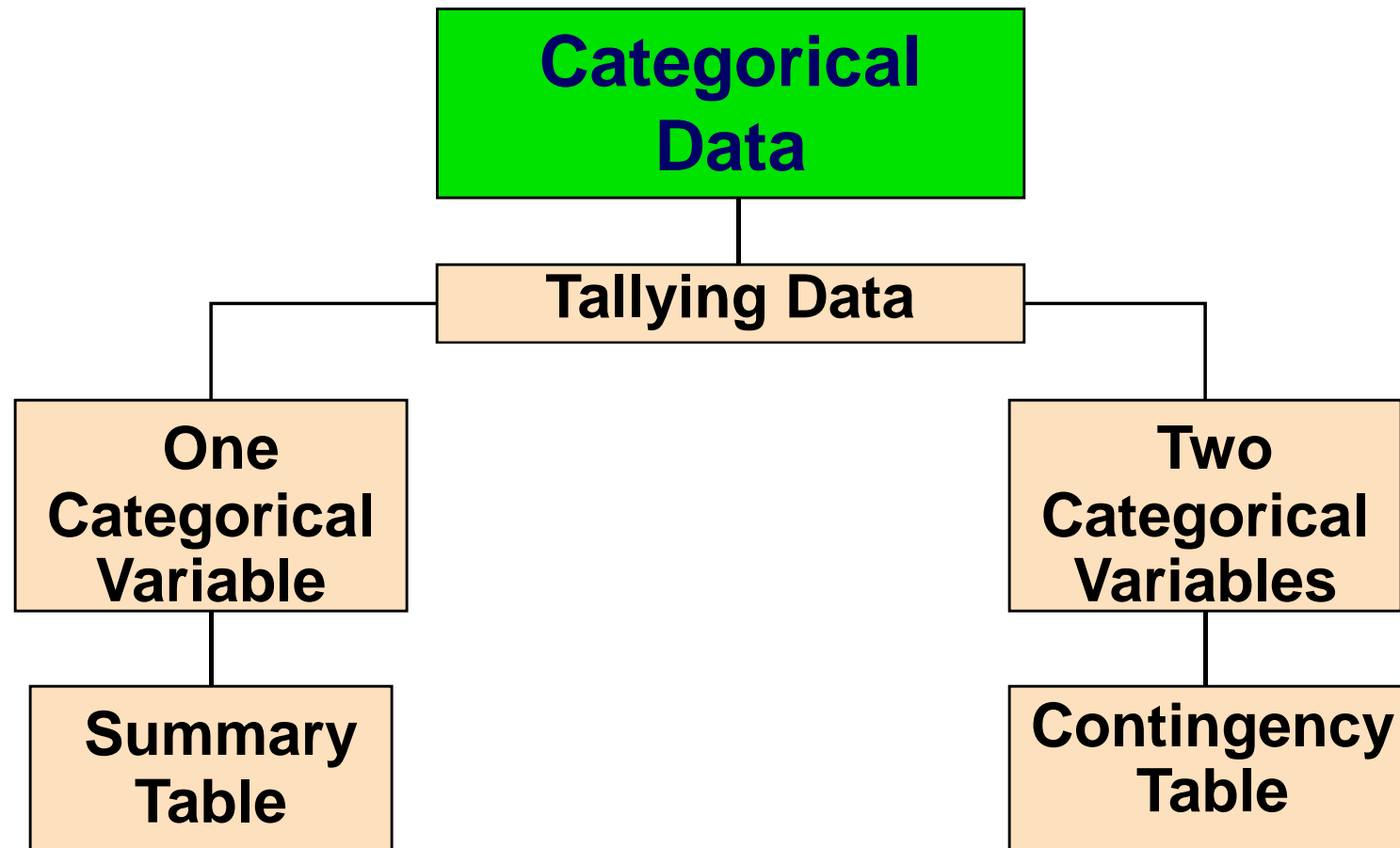
Organizing Categorical Data: Summary Table

- A **summary table** tallies the frequencies or percentages of items in a set of categories so that you can see differences between categories.

Main Reason Young Adults Shop Online

Reason For Shopping Online?	Percent
Better Prices	37%
Avoiding holiday crowds or hassles	29%
Convenience	18%
Better selection	13%
Ships directly	3%

Source: Data extracted and adapted from "Main Reason Young Adults Shop Online?" *USA Today, December 5, 2012, p. 1A.*

A Contingency Table Helps Organize Two or More Categorical Variables

DC<u>O</u>VA

- Used to study patterns that may exist between the responses of two or more categorical variables.

- Cross tabulates or tallies jointly the responses of the categorical variables.

- For two variables the tallies for one variable are located in the rows and the tallies for the second variable are located in the columns.

Contingency Table - Example

- A random sample of 400 invoices is drawn.

- Each invoice is categorized as a small, medium, or large amount.

- Each invoice is also examined to identify if there are any errors.

- This data are then organized in the contingency table to the right.

Contingency Table Showing Frequency of Invoices Categorized By Size and The Presence Of Errors

	No Errors	Errors	Total
Small Amount	170	20	190
Medium Amount	100	40	140
Large Amount	65	5	70
Total	335	65	400

Contingency Table Based On Percentage Of Overall Total

DC<u>O</u>VA

	No Errors	Errors	Total
Small Amount	170	20	190
Medium Amount	100	40	140
Large Amount	65	5	70
Total	335	65	400

42.50% = 170 / 400
25.00% = 100 / 400
16.25% = 65 / 400

	No Errors	Errors	Total
Small Amount	42.50%	5.00%	47.50%
Medium Amount	25.00%	10.00%	35.00%
Large Amount	16.25%	1.25%	17.50%
Total	83.75%	16.25%	100.0%

83.75% of sampled invoices have no errors and 47.50% of sampled invoices are for small amounts.

Contingency Table Based On Percentage of Row Totals

	No Errors	Errors	Total
Small Amount	170	20	190
Medium Amount	100	40	140
Large Amount	65	5	70
Total	335	65	400

89.47% = 170 / 190
71.43% = 100 / 140
92.86% =   65 / 70

	No Errors	Errors	Total
Small Amount	89.47%	10.53%	100.0%
Medium Amount	71.43%	28.57%	100.0%
Large Amount	92.86%	7.14%	100.0%
Total	83.75%	16.25%	100.0%

Medium invoices have a larger chance (28.57%) of having errors than small (10.53%) or large (7.14%) invoices.

Contingency Table Based On Percentage Of Column Totals

	No Errors	Errors	Total
Small Amount	170	20	190
Medium Amount	100	40	140
Large Amount	65	5	70
Total	335	65	400

50.75% = 170 / 335
30.77% =   20 / 65

	No Errors	Errors	Total
Small Amount	50.75%	30.77%	47.50%
Medium Amount	29.85%	61.54%	35.00%
Large Amount	19.40%	7.69%	17.50%
Total	100.0%	100.0%	100.0%

There is a 61.54% chance that invoices with errors are of medium size.
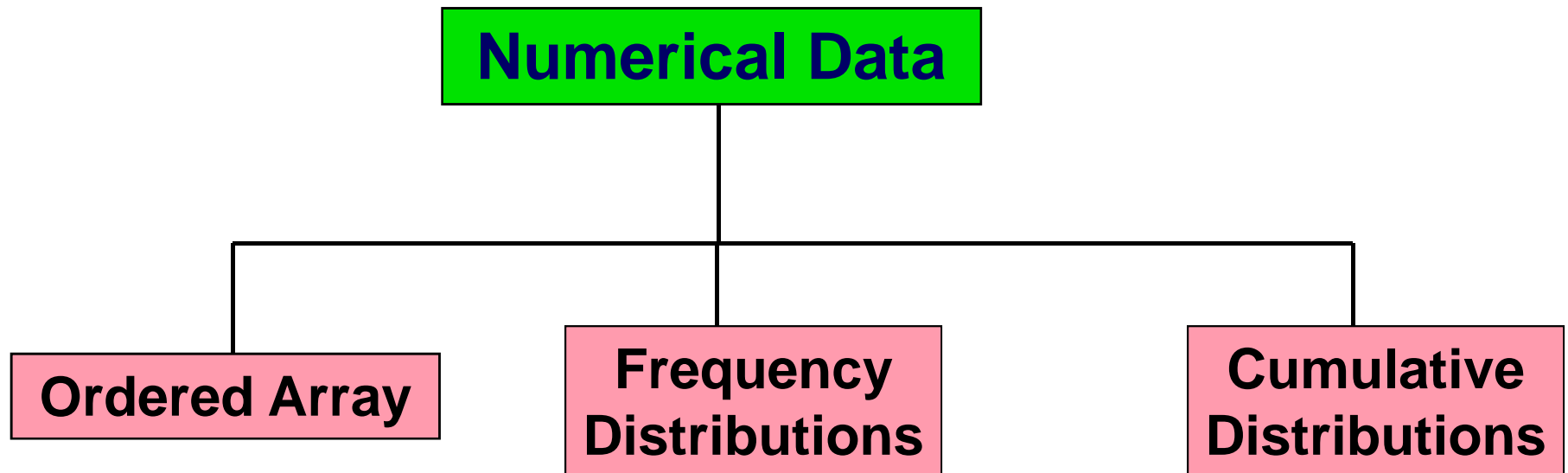
Organizing Numerical Data: Ordered Array

- An **ordered array** is a sequence of data, in rank order, from the smallest value to the largest value.
- Shows range (minimum value to maximum value).
- May help identify outliers (unusual observations).

Age of Surveyed College Students	Day Students					
	16	17	17	18	18	18
	19	19	20	20	21	22
	22	25	27	32	38	42
	Night Students					
	18	18	19	19	20	21
	23	28	32	33	41	45

Organizing Numerical Data: Frequency Distribution

- The **frequency distribution** is a summary table in which the data are arranged into numerically ordered classes.

- You must give attention to selecting the appropriate *number* of **class groupings** for the table, determining a suitable *width* of a class grouping, and establishing the *boundaries* of each class grouping to avoid overlapping.

- The number of classes depends on the number of values in the data. With a larger number of values, typically there are more classes. In general, a frequency distribution should have at least 5 but no more than 15 classes.

- To determine the **width of a class interval,** you divide the **range** (Highest value–Lowest value) of the data by the number of class groupings desired.

Organizing Numerical Data: Frequency Distribution Example

Example: A manufacturer of insulation randomly selects 20 winter days and records the daily high temperature.

24, 35, 17, 21, 24, 37, 26, 46, 58, 30, 32, 13, 12, 38, 41, 43, 44, 27, 53, 27

Organizing Numerical Data: Frequency Distribution Example

DC**O**VA

- Sort raw data in ascending order:
  12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58.
- Find range: **58 - 12 = 46.**
- Select number of classes: **5 (usually between 5 and 15).**
- Compute class interval (width): **10 (46/5 then round up).**
- Determine class boundaries (limits):
  - **Class 1: 10 but less than 20.**
  - **Class 2: 20 but less than 30.**
  - **Class 3: 30 but less than 40.**
  - **Class 4: 40 but less than 50.**
  - **Class 5: 50 but less than 60.**
- Compute class midpoints: **15, 25, 35, 45, 55.**
- Count observations & assign to classes.

Organizing Numerical Data: Frequency Distribution Example

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class	Midpoints	Frequency
10 but less than 20	15	3
20 but less than 30	25	6
30 but less than 40	35	5
40 but less than 50	45	4
50 but less than 60	55	2
Total		20

Organizing Numerical Data: Relative & Percent Frequency Distribution Example

Class	Frequency	Relative Frequency	Percentage
10 but less than 20	3	.15	15%
20 but less than 30	6	.30	30%
30 but less than 40	5	.25	25%
40 but less than 50	4	.20	20%
50 but less than 60	2	.10	10%
Total	20	1.00	100%

Relative Frequency = Frequency / Total,   e.g. 0.10 = 2 / 20

Organizing Numerical Data: Cumulative Frequency Distribution Example

DC<u>O</u>VA

Class	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
10 but less than 20	3	15%	3	15%
20 but less than 30	6	30%	9	**45%**
30 but less than 40	5	25%	14	70%
40 but less than 50	4	20%	18	90%
50 but less than 60	2	10%	20	100%
Total	**20**	**100%**	**20**	**100%**

Cumulative Percentage = Cumulative Frequency / Total * 100   e.g. 45% = 100*9/20

Why Use a Frequency Distribution?

DC<u>O</u>VA

- It condenses the raw data into a more useful form.

- It allows for a quick visual interpretation of the data.

- It enables the determination of the major characteristics of the data set including where the data are concentrated / clustered.

Frequency Distributions: Some Tips

- Different class boundaries may provide different pictures for the same data (especially for smaller data sets).

- Shifts in data concentration may show up when different class boundaries are chosen.

- As the size of the data set increases, the impact of alterations in the selection of class boundaries is greatly reduced.

- When comparing two or more groups with different sample sizes, you must use either a relative frequency or a percentage distribution.

Chapter 2 - 20

Going From Classes To Excel Bins

- Microsoft Excel creates distribution tables using bins (named by their upper limit) rather than classes.

Class	Excel Bin Name
	9.99
10 but less than 20	19.99
20 but less than 30	29.99
30 but less than 40	39.99
40 but less than 50	49.99
50 but less than 60	59.99

An extra bin added in Excel slightly less than the smallest observation

Visualizing Categorical Data Through Graphical Displays

Categorical Data

Visualizing Data

Summary Table For One Variable

Contingency Table For Two Variables

Bar Chart

Pie or Doughnut Chart

Pareto Chart

Side By Side Bar Chart

Doughnut Chart
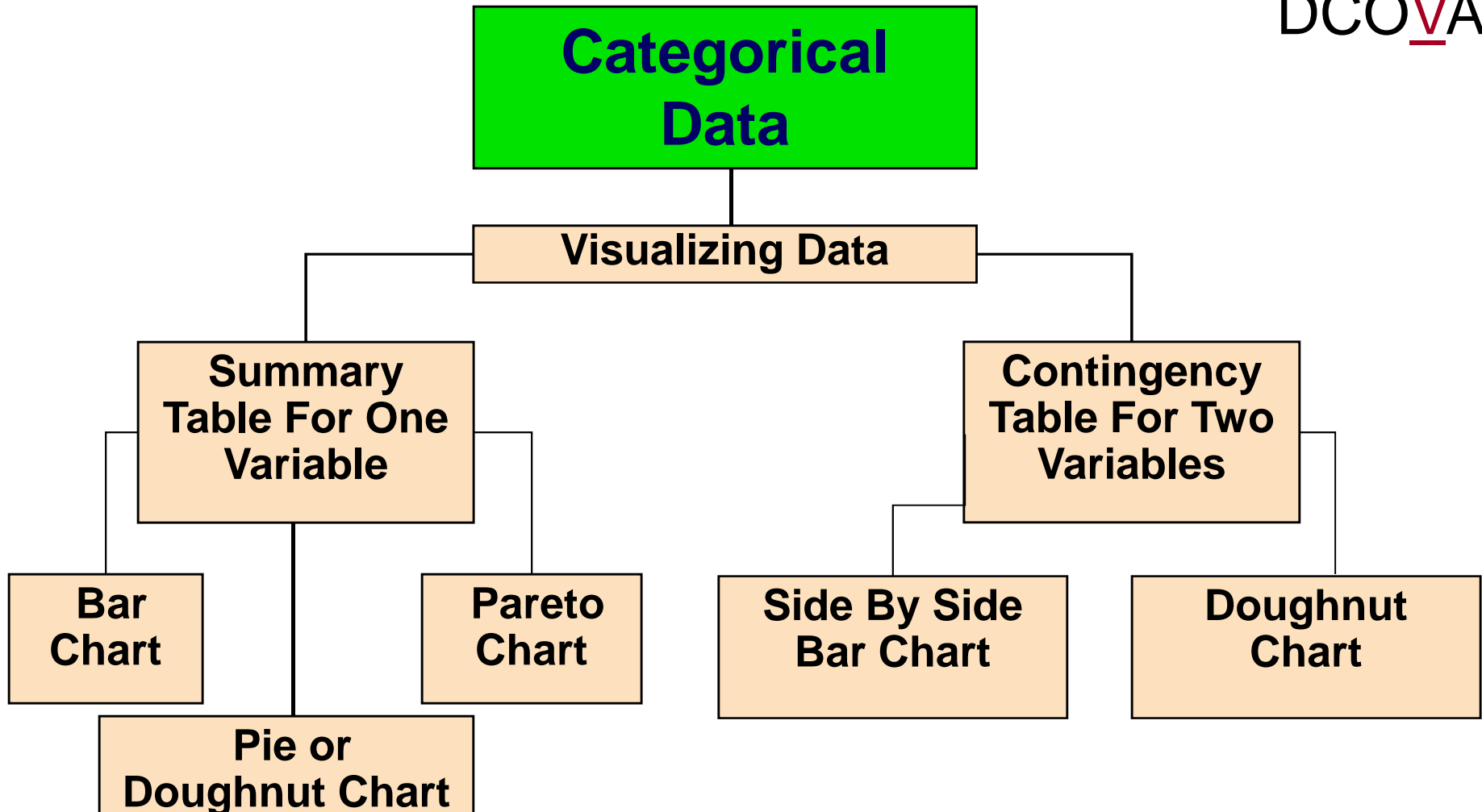
Visualizing Categorical Data:
The Bar Chart

- The **bar chart** visualizes a categorical variable as a series of bars. The length of each bar represents either the frequency or percentage of values for each category. Each bar is separated by a space called a gap.

Reason For Shopping Online?	Percent
Better Prices	37%
Avoiding holiday crowds or hassles	29%
Convenience	18%
Better selection	13%
Ships directly	3%

Bar Chart of Reasons to Shop Online

Visualizing Categorical Data: The Pie Chart

▪ The **pie chart** is a circle broken up into slices that represent categories. The size of each slice of the pie varies according to the percentage in each category.

Reason For Shopping Online?	Percent
Better Prices	37%
Avoiding holiday crowds or hassles	29%
Convenience	18%
Better selection	13%
Ships directly	3%

Pie Chart of Reasons to Shop Online

Visualizing Categorical Data: The Doughnut Chart

▪ The **doughnut chart** is the outer part of a circle broken up into pieces that represent categories. The size of each piece of the doughnut varies according to the percentage in each category.

Reason For Shopping Online?	Percent
Better Prices	37%
Avoiding holiday crowds or hassles	29%
Convenience	18%
Better selection	13%
Ships directly	3%

Doughnut Chart of Reasons to Shop Online

Ships directly 3%
Convenience 18%
Avoiding holiday crowds or hassles 29%
Better selection 13%
Better prices 37%
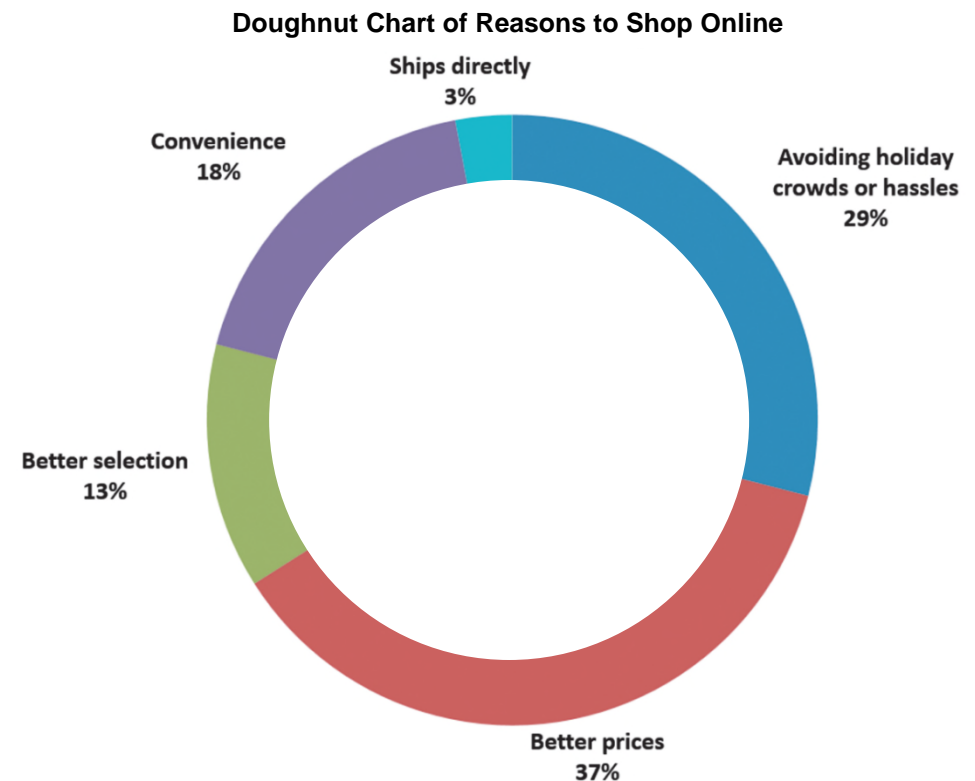
Visualizing Categorical Data: The Pareto Chart

DCO<u>V</u>A

- Used to portray categorical data (nominal scale).

- A vertical bar chart, where categories are shown in descending order of frequency.

- A cumulative polygon is shown in the same graph.

- Used to separate the "vital few" from the "trivial many."

PEARSON   Chapter 2 - 26

Visualizing Categorical Data: The Pareto Chart (con't)

Ordered Summary Table For Causes Of Incomplete ATM Transactions

Cause	Frequency	Percent	Cumulative Percent
Warped card jammed	365	50.41%	50.41%
Card unreadable	234	32.32%	82.73%
ATM malfunctions	32	4.42%	87.15%
ATM out of cash	28	3.87%	91.02%
Invalid amount requested	23	3.18%	94.20%
Wrong keystroke	23	3.18%	97.38%
Lack of funds in account	19	2.62%	100.00%
Total	**724**	**100.00%**	

Source: Data extracted from A. Bhalla, "Don't Misuse the Pareto Principle," *Six Sigma Forum Magazine, May 2009, pp. 15–18.*

Visualizing Categorical Data:
The Pareto Chart (con't)
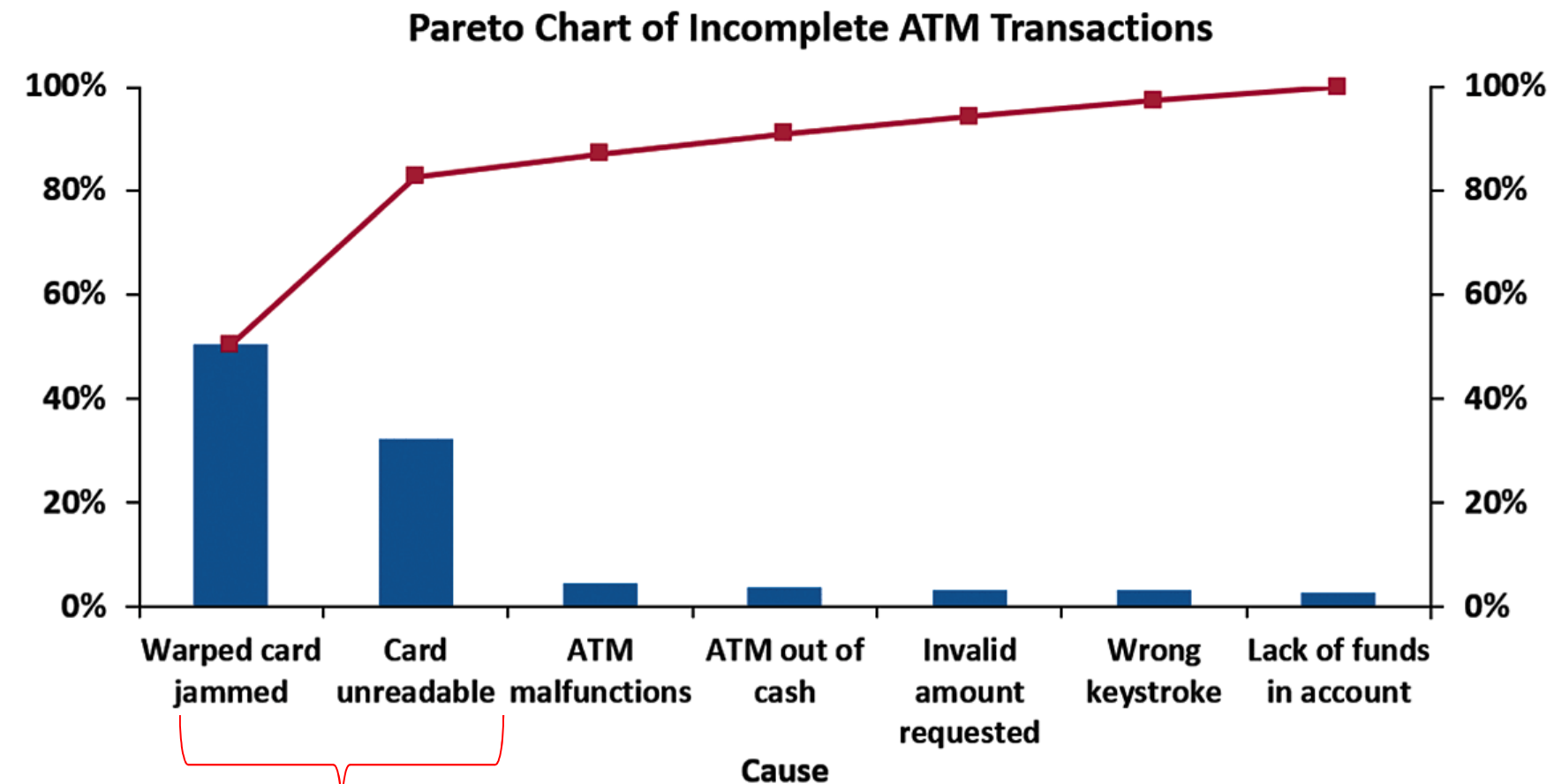
The "Vital Few"

Visualizing Categorical Data: Side By Side Bar Charts

▪ The **side by side bar chart** represents the data from a contingency table.

	No Errors	Errors	Total
Small Amount	50.75%	30.77%	47.50%
Medium Amount	29.85%	61.54%	35.00%
Large Amount	19.40%	7.69%	17.50%
Total	100.0%	100.0%	100.0%

Invoice Size Split Out By Errors & No Errors
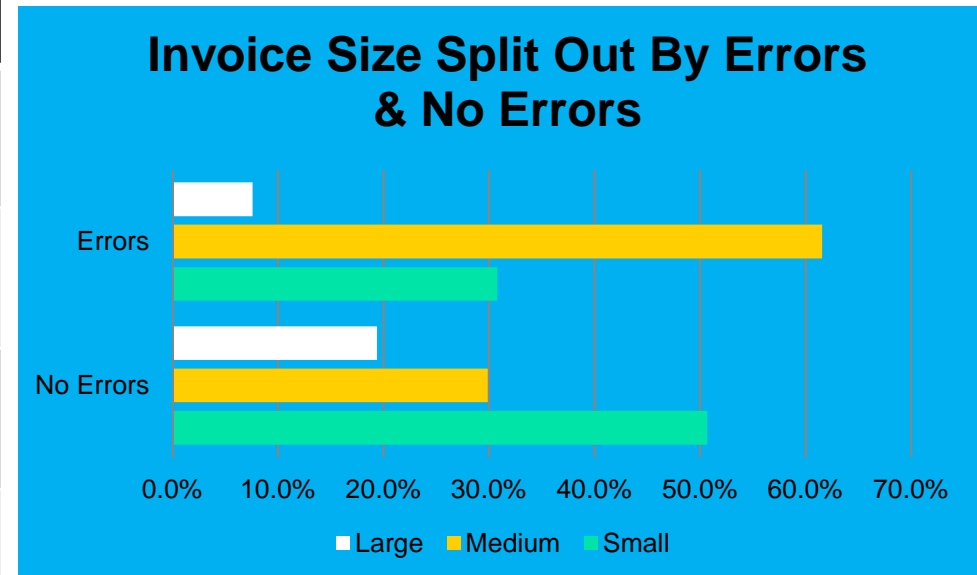
Invoices with errors are much more likely to be of medium size (61.5% vs 30.8% & 7.7%).

Visualizing Categorical Data: Doughnut Charts

▪ A **Doughnut Chart** can be used to represent the data from a contingency table.

	No Errors	Errors	Total
Small Amount	50.75%	30.77%	47.50%
Medium Amount	29.85%	61.54%	35.00%
Large Amount	19.40%	7.69%	17.50%
Total	100.0%	100.0%	100.0%

Invoice Size & Errors
Inner Ring With Errors, Outer Ring No Errors

19.4%
7.7%
30.8%
30.8%
61.5%
29.9%

☐ Small ☐ Medium ☐ Large
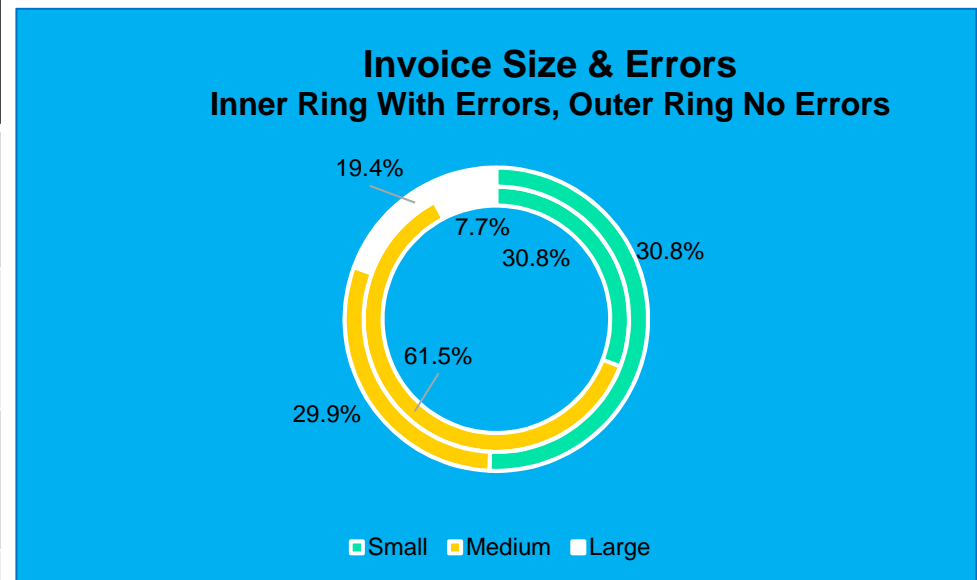
Invoices with errors are much more likely to be of medium size (61.5% vs 30.8% & 7.7%).

Visualizing Numerical Data By Using Graphical Displays
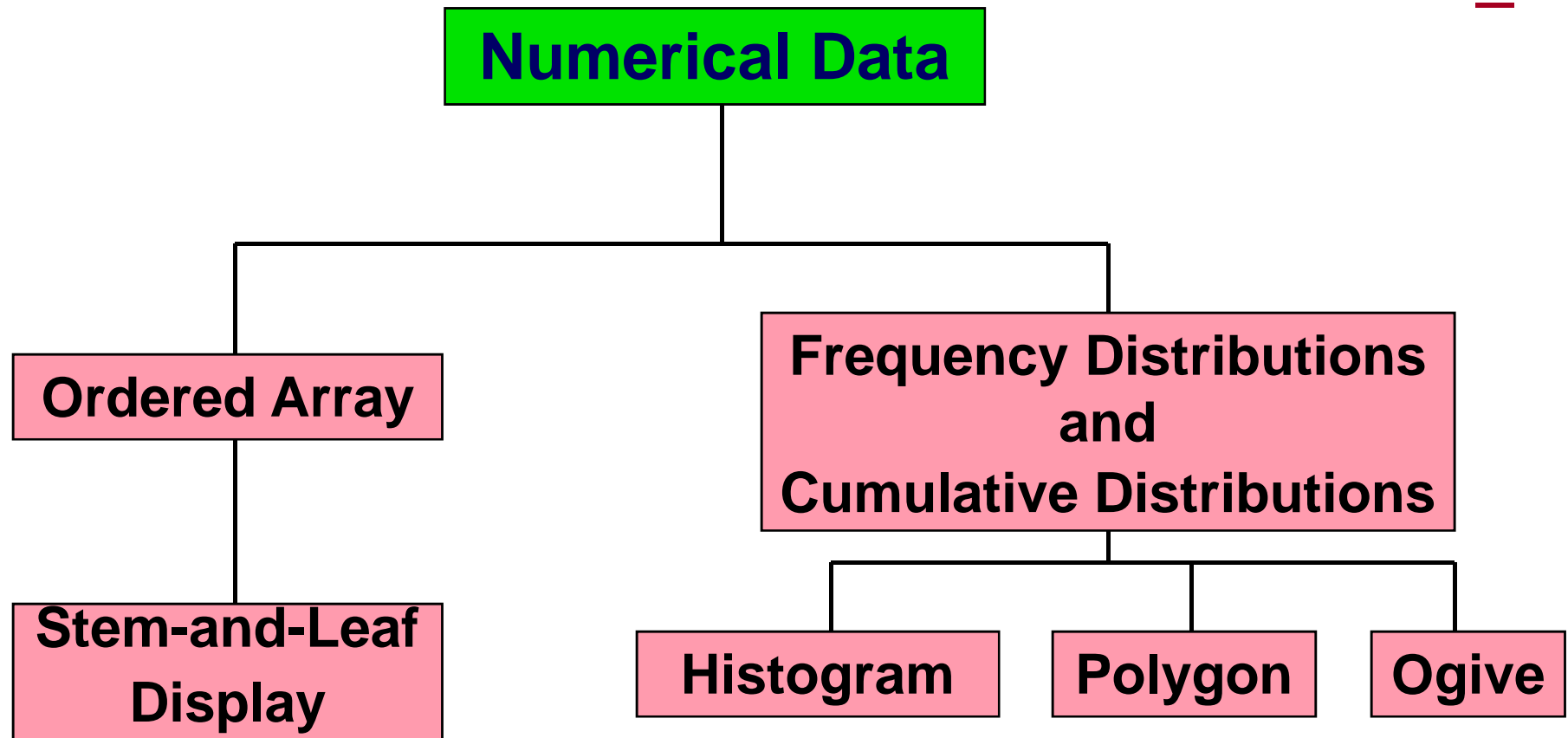
DCOVA

Stem-and-Leaf Display

- A simple way to see how the data are distributed and where concentrations of data exist.

METHOD: Separate the sorted data series into leading digits (the **stems**) and the trailing digits (the **leaves**).

Organizing Numerical Data: Stem and Leaf Display

- A **stem-and-leaf display** organizes data into groups (called stems) so that the values within each group (the leaves) branch out to the right on each row.

Age of College Students

Age of Surveyed College Students	Day Students					
	16	17	17	18	18	18
	19	19	20	20	21	22
	22	25	27	32	38	42
	Night Students					
	18	18	19	19	20	21
	23	28	32	33	41	45

Day Students

Stem	Leaf
1	67788899
2	0012257
3	28
4	2

Night Students

Stem	Leaf
1	8899
2	0138
3	23
4	15

Visualizing Numerical Data:
The Histogram

- A vertical bar chart of the data in a frequency distribution is called a **histogram.**

- In a histogram there are no gaps between adjacent bars.

- The **class boundaries** (or **class midpoints**) are shown on the horizontal axis.

- The vertical axis is either **frequency, relative frequency,** or **percentage**.

- The height of the bars represent the frequency, relative frequency, or percentage.

Visualizing Numerical Data: The Histogram

Class	Frequency	Relative Frequency	Percentage
10 but less than 20	3	.15	15
20 but less than 30	6	.30	30
30 but less than 40	5	.25	25
40 but less than 50	4	.20	20
50 but less than 60	2	.10	10
Total	20	1.00	100

(In a percentage histogram the vertical axis would be defined to show the percentage of observations per class).

Histogram: Temperature
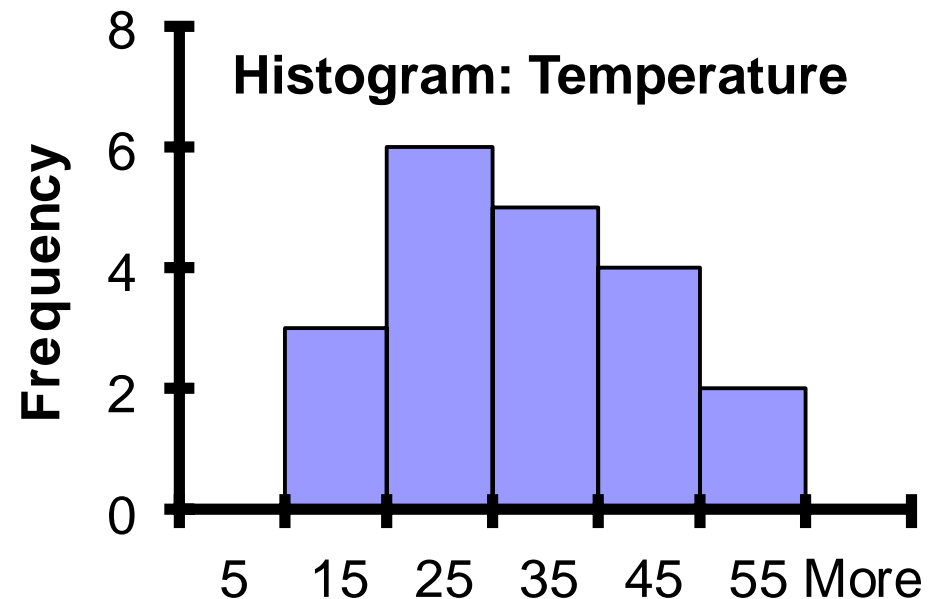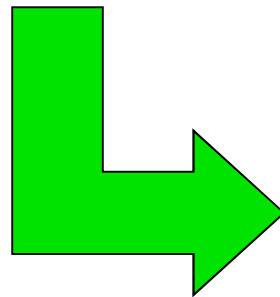
Visualizing Numerical Data: The Polygon

- A **percentage polygon** is formed by having the midpoint of each class represent the data in that class and then connecting the sequence of midpoints at their respective class percentages.

- The **cumulative percentage polygon,** or **ogive,** displays the variable of interest along the X axis, and the cumulative percentages along the Y axis.

- Useful when there are two or more groups to compare.

Visualizing Numerical Data:
The Frequency Polygon

Useful When Comparing Two or More Groups
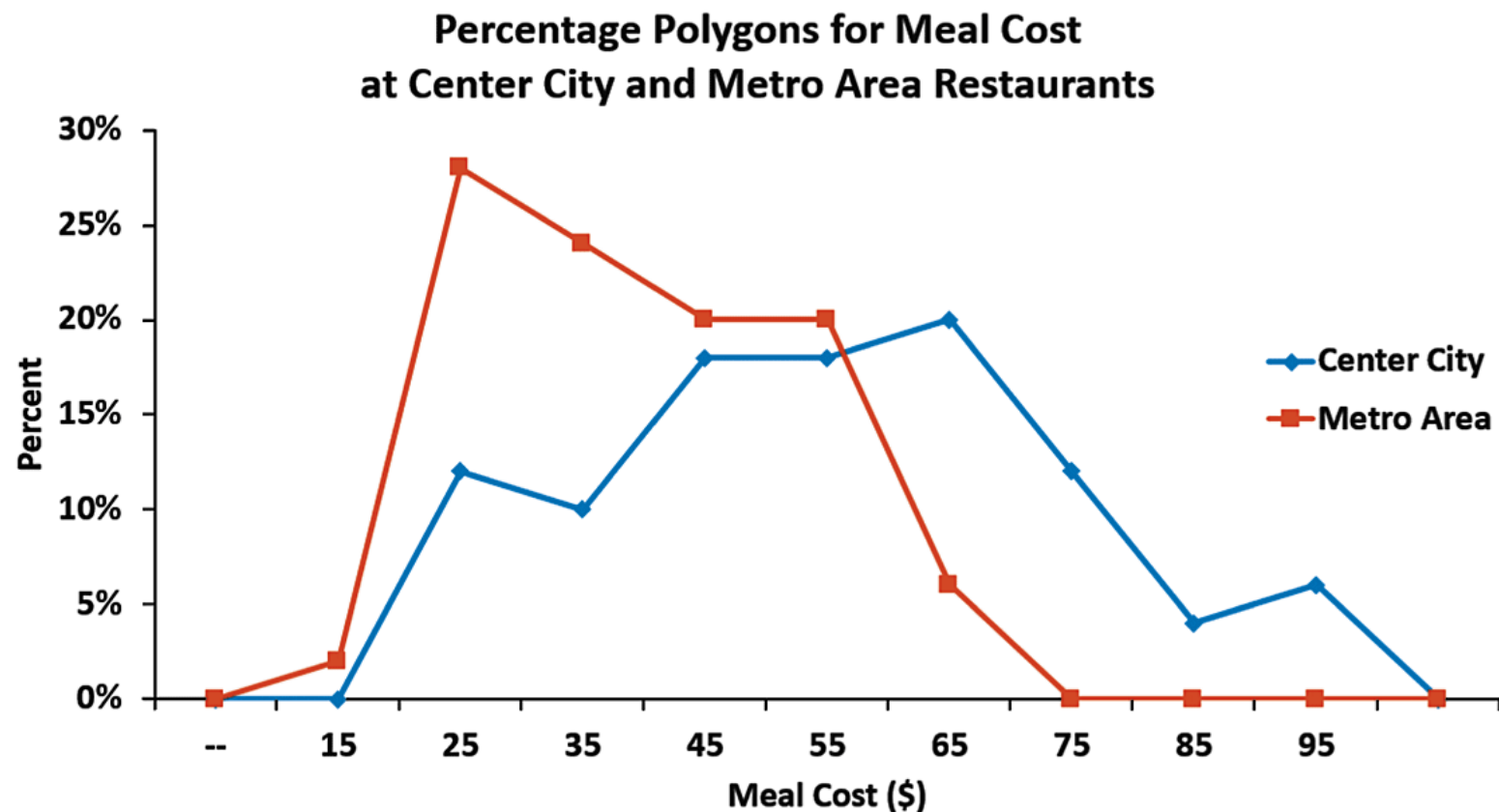
Visualizing Numerical Data: The Percentage Polygon

DCOVA

Percentage Polygons for One-Year Return Percentage for the Growth and Value Funds
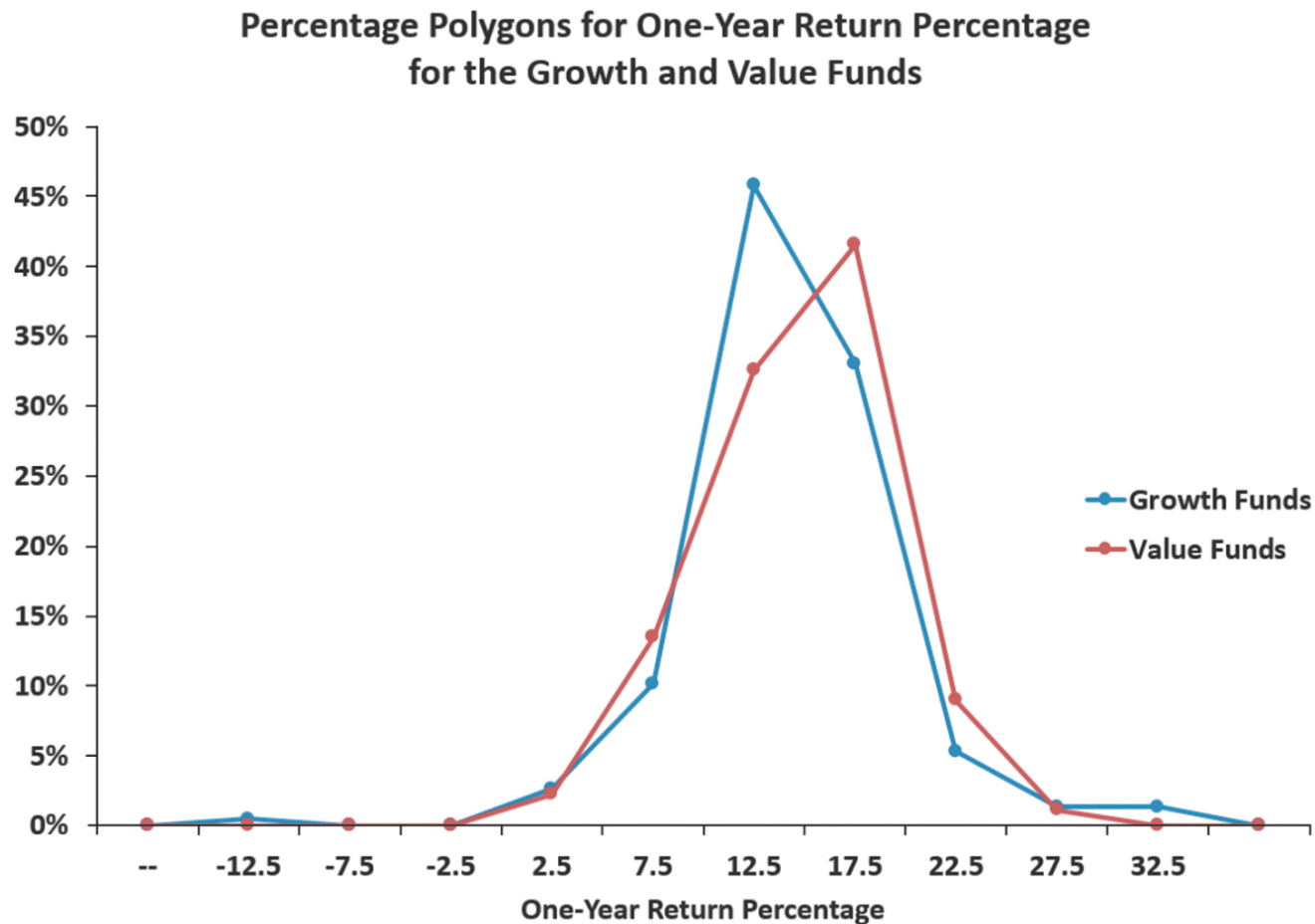
Visualizing Two Numerical Variables By Using Graphical Displays
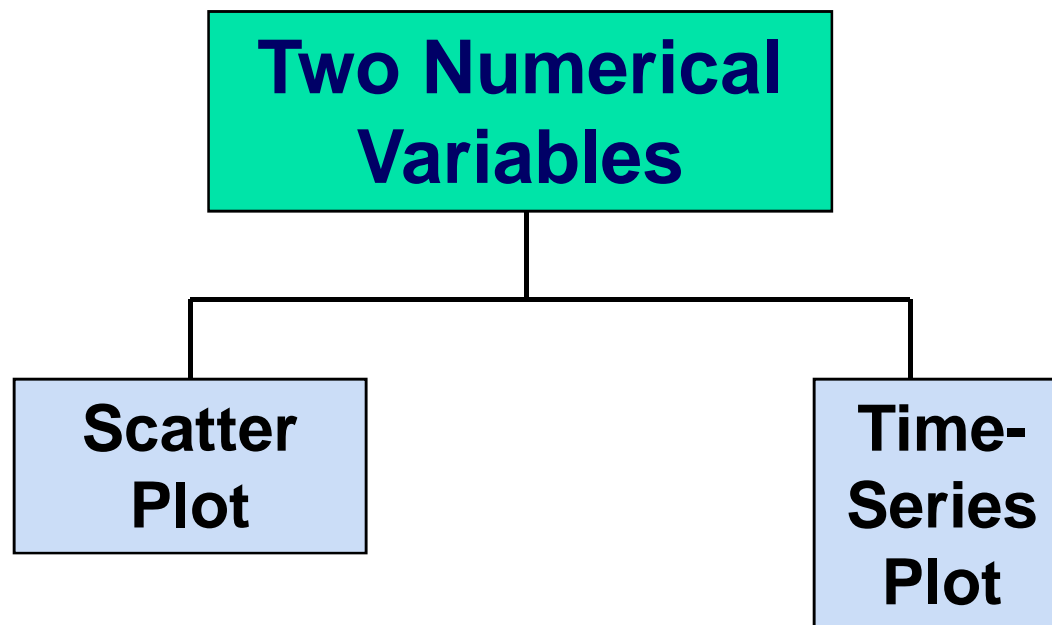
DCO<u>V</u>A

Visualizing Two Numerical Variables: The Scatter Plot

DCO<u>V</u>A

- **Scatter plots** are used for numerical data consisting of paired observations taken from two numerical variables.

- One variable is measured on the vertical axis and the other variable is measured on the horizontal axis.

- Scatter plots are used to examine possible relationships between two numerical variables.

Scatter Plot Example

Volume per day	Cost per day
23	125
26	140
29	146
33	160
38	167
42	170
50	188
55	195
60	200

Cost per Day vs. Production Volume
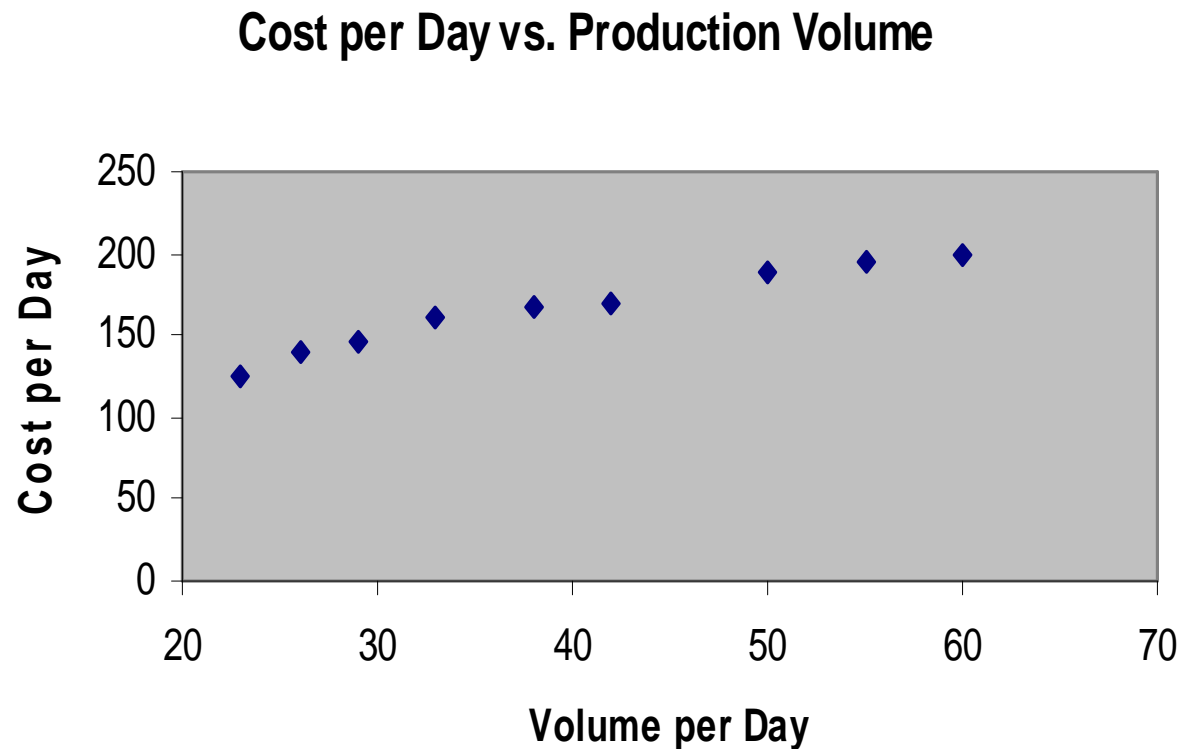
Visualizing Two Numerical Variables: The Time Series Plot

DCO<u>V</u>A

- A Time-Series Plot is used to study patterns in the values of a numeric variable over time.

- The Time-Series Plot:
  - Numeric variable is measured on the vertical axis and the time period is measured on the horizontal axis.

Time Series Plot Example

DCO<u>V</u>A

Year	Number of Franchises
2007	43
2008	54
2009	60
2010	73
2011	82
2012	95
2013	107
2014	99
2015	95

Number of Franchises, 2007 to 2015
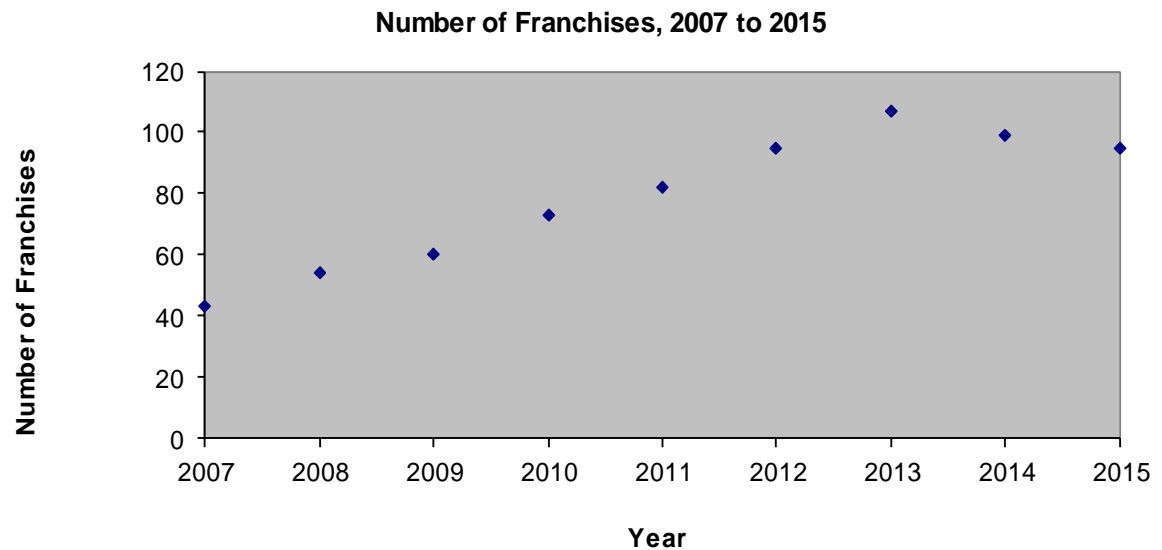
Organizing Many Categorical Variables: The Multidimensional Contingency Table

- A multidimensional contingency table is constructed by tallying the responses of three or more categorical variables.

- In Excel you create a Pivot Table to yield an interactive display of this type.

Using Excel Pivot Tables To Organize & Visualize Many Variables

DC<u>OV</u>A

A Pivot Table:

- Summarizes variables as a multidimensional summary table.

- Allows interactive changing of the level of summarization and formatting of the variables.

- Allows you to interactively "slice" your data to summarize subsets of data that meet specified criteria.

- Can be used to discover possible patterns and relationships in multidimensional data that simpler tables and charts would fail to make apparent.

A Multidimensional Contingency Table Tallies Responses Of Three or More Categorical Variables

DC<u>O</u>VA

Two Dimensional Table Showing The Mean 10 Year Return % Broken Out By Type Of Fund & Risk Level.

Mean 10YrReturn%	Risk ▾			
Type ▾	Low	Average	High	Grand Total
Growth	7.59	8.07	7.63	7.95
Value	6.48	6.93	7.02	6.86
Grand Total	6.91	7.70	7.49	7.58

Three Dimensional Table Showing The Mean 10 Year Return % Broken Out By Type Of Fund, Market Cap, &Risk Level.

Mean 10YrReturn%	Risk ▾			
Type ▾	Low	Average	High	Grand Total
⊟ Growth	7.59	8.07	7.63	7.95
Large	7.35	7.83	10.14	7.89
Mid-Cap	8.61	8.49	7.41	8.32
Small		7.91	7.27	7.50
⊟ Value	6.48	6.93	7.02	6.86
Large	6.52	6.46	4.10	6.45
Mid-Cap	6.05	7.84	7.28	7.68
Small		7.24	7.21	7.22
Grand Total	6.91	7.70	7.49	7.58

Data Discovery Methods Can Yield Initial Insights Into Data
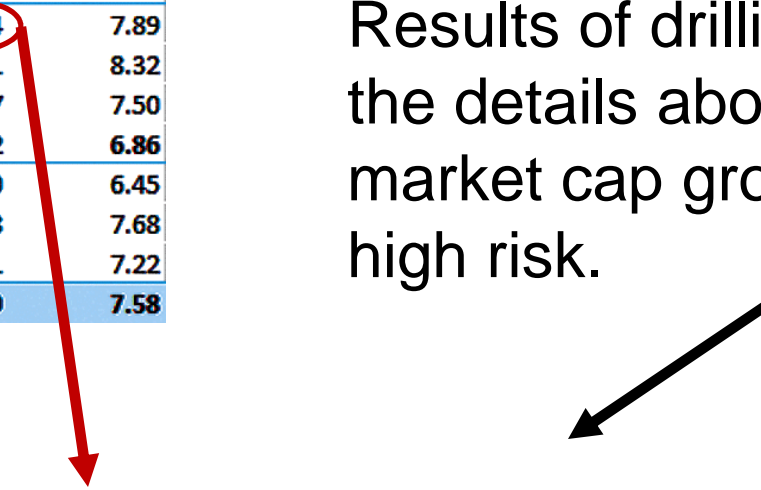
- Data discovery methods enable the performance of preliminary analyses by manipulating interactive summarizations.

- Are used to:

  - Take a closer look at historical or status data.

  - Review data for unusual values.

  - Uncover new patterns in data.

- Drill-down is perhaps the simplest form of data discovery.

Drill-Down Reveals The Data Underlying A Higher-Level Summary

DC<u>O</u>VA

Mean 10YrReturn% Risk ▾				
Type ▾	**Low**	**Average**	**High**	**Grand Total**
⊟ **Growth**	7.59	8.07	7.63	7.95
Large	7.35	7.83	**10.14**	7.89
Mid-Cap	8.61	8.49	7.41	8.32
Small		7.91	7.27	7.50
⊟ **Value**	6.48	6.93	7.02	6.86
Large	6.52	6.46	4.10	6.45
Mid-Cap	6.05	7.84	7.28	7.68
Small		7.24	7.21	7.22
Grand Total	6.91	7.70	7.49	7.58

Results of drilling down to the details about large market cap growth funds with high risk.

	F	G	H	I	J	K	L	M	N
1	Turnover Ratio(%) ▾	SD ▾	Sharpe Ratio ▾	1YrReturn% ▾	3YrReturn% ▾	5YrReturn% ▾	10YrReturn% ▾	Expense Ratio ▾	Star Rating ▾
2	49	16.94	1.01	0.53	16.18	15.16	9.50	1.3	Three
3	146	13.05	1.23	8.73	16.39	12.80	8.11	0.98	Two
4	39	14.51	1.17	2.90	17.23	17.02	11.13	1.27	Three
5	21	13.17	1.19	5.65	15.95	15.44	9.71	1.22	Two
6	31	13.62	1.35	6.42	19.02	18.05	10.26	0.96	Four
7	106	14.07	1.42	14.70	20.80	19.36	12.14	0.77	Five

Some Business Analytics Methods Start With Many Variables

DC<u>O</u>VA

- These methods allow you to filter data by exploring specific combinations of categorical values or numerical ranges.

- In Excel *slicers* are used to do this filtering.

- By clicking buttons in the slicer panels you ask questions about your data.

Chapter 2 - 49

Example Of Slicers For The Retirement Funds Workbook

DC<u>OV</u>A

With the four slicers below, you can ask questions such as:

1. What are the attributes of the fund(s) with the lowest expense ratio?
2. Which fund(s) in the sample have the highest expense ratio?
3. What are the expense ratios associated with mid-cap funds that have a star rating of five?

Answering Questions One & Three

DC<u>O</u>VA

The answer to question 1 is a growth fund with a large market cap and a five-star rating.

For question 3, you observe that 0.88 and 1.44 are the expense ratios associated with mid-cap funds that have a star rating of five.

More Complex Displays Are Also Possible

A seven-slicer panel in which the five highest turnover ratio percentages have been selected to reveal the identity and the attributes of the funds associated with those percentages.
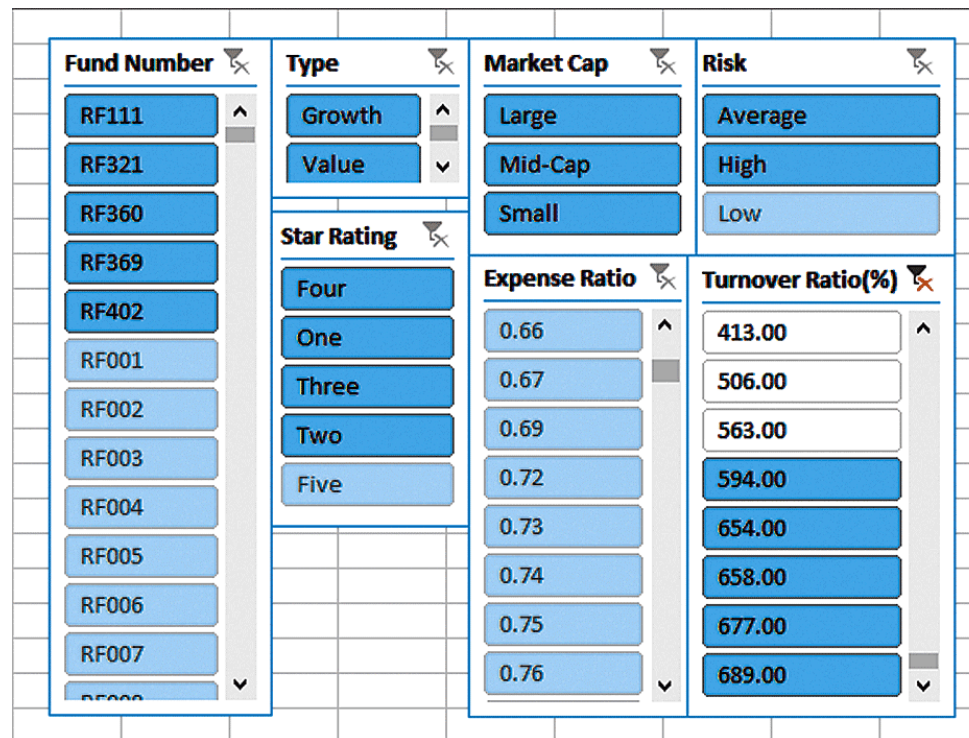
A PivotChart Filters Data By Exploring Specific Combinations Of Categorical Variables Visually

Mean 10YrReturn%	Risk ▾			
Type ▾	Low	Average	High	Grand Total
⊟ Growth	7.59	8.07	7.63	7.95
Large	7.35	7.83	10.14	7.89
Mid-Cap	8.61	8.49	7.41	8.32
Small		7.91	7.27	7.50
⊟ Value	6.48	6.93	7.02	6.86
Large	6.52	6.46	4.10	6.45
Mid-Cap	6.05	7.84	7.28	7.68
Small		7.24	7.21	7.22
Grand Total	6.91	7.70	7.49	7.58

Mean Ten-Year Return Percentage by Fund Type, Market Cap, and Risk

Mean Ten-Year Return Percentage by Fund Type, Market Cap, and Risk

Sparklines Are Compact Time-Series Visualizations Of Numerical Variables

DC<u>O</u>VA

Movie revenues by week per month
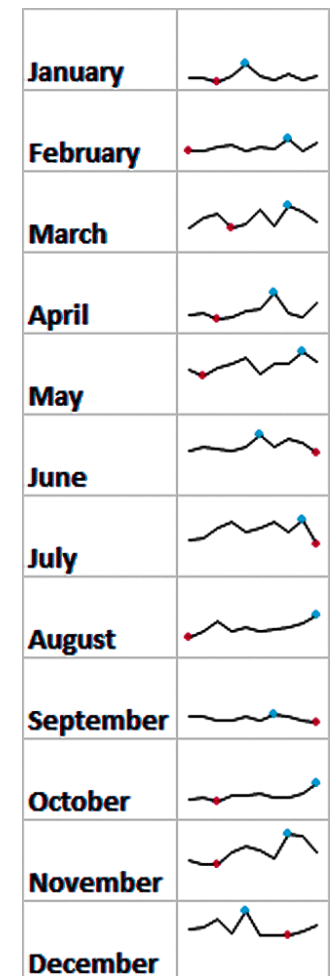
The Challenges in Organizing and Visualizing Variables

DC<u>O</u>VA

- When organizing and visualizing data need to be mindful of:
    - The limits of other's ability to perceive and comprehend.
    - Presentation issues that can undercut the usefulness of methods from this chapter.
- It is easy to create summaries that:
    - Obscure the data or
    - Create false impressions.

An Example Of Obscuring Data, Information Overload

DCOVA

Side-by-Side Bar Chart of Retirement Funds Sample
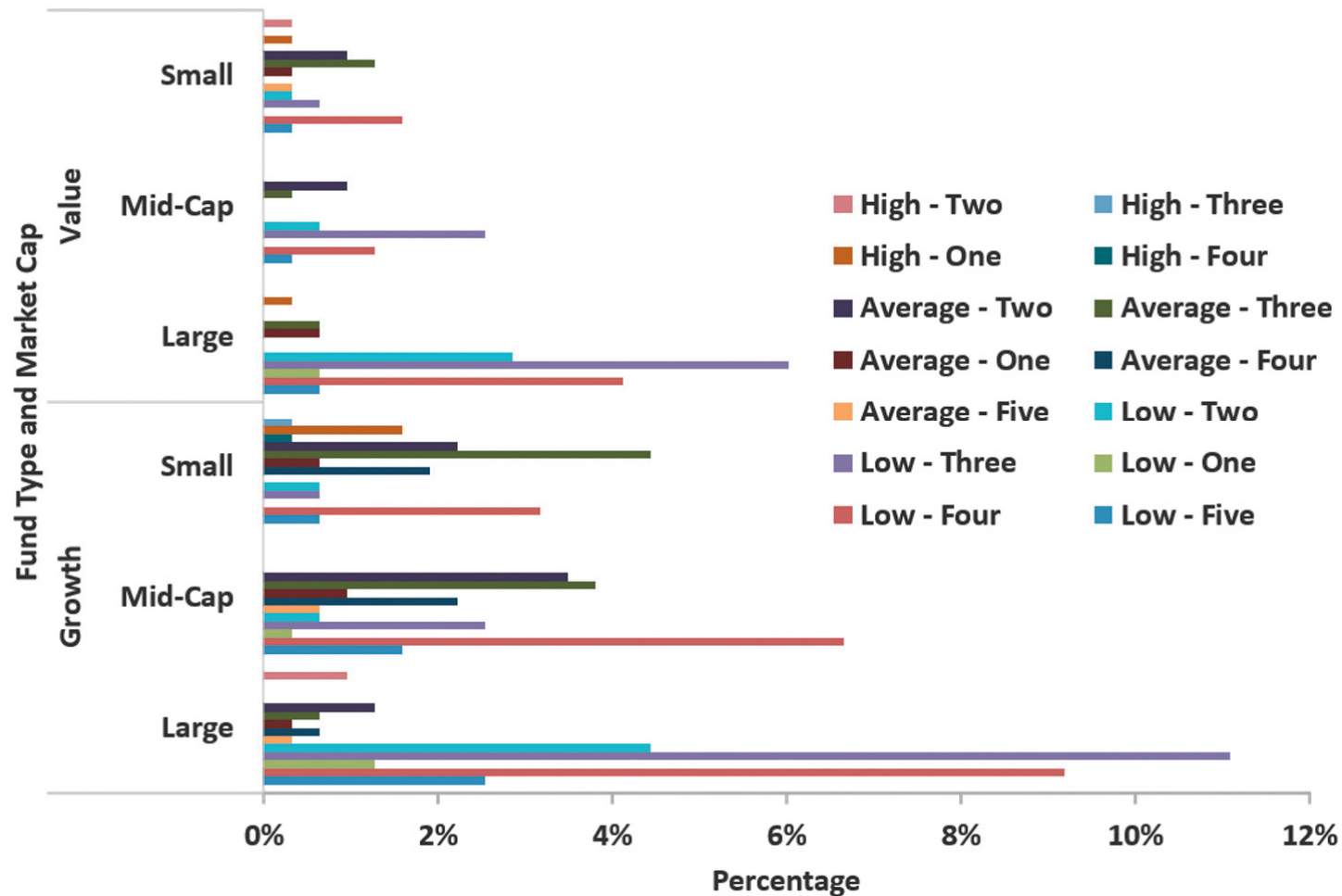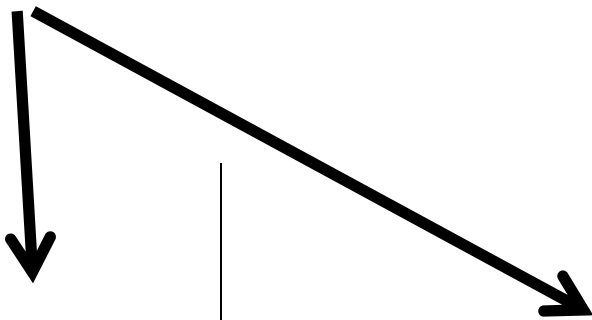
False Impressions Can Be Created In Many Ways

DC<u>O</u>VA

- Selective summarization:
    - Presenting only part of the data collected.

- Improperly constructed charts:
    - Potential pie chart issues.
    - Improperly scaled axes.
    - A Y axis that does not begin at the origin or is a broken axis missing intermediate values.

- Chartjunk.

An Example of Selective Summarization, These Two Summarizations Tell Totally Different Stories

DC<u>OV</u>A

Company	Change from Prior Year	Company	Year 1	Year 2	Year 3
A	+7.2%	A	-22.6%	-33.2%	+7.2%
B	+24.4%	B	-4.5%	-41.9%	+24.4%
C	+24.9%	C	-18.5%	-31.5%	+24.9%
D	+24.8%	D	-29.4%	-48.1%	+24.8%
E	+12.5%	E	-1.9%	-25.3%	+12.5%
F	+35.1%	F	-1.6%	-37.8%	+35.1%
G	+29.7%	G	+7.4%	-13.6%	+29.7%

How Obvious Is It That Both Pie Charts Summarize The Same Data?

DC<u>O</u>VA
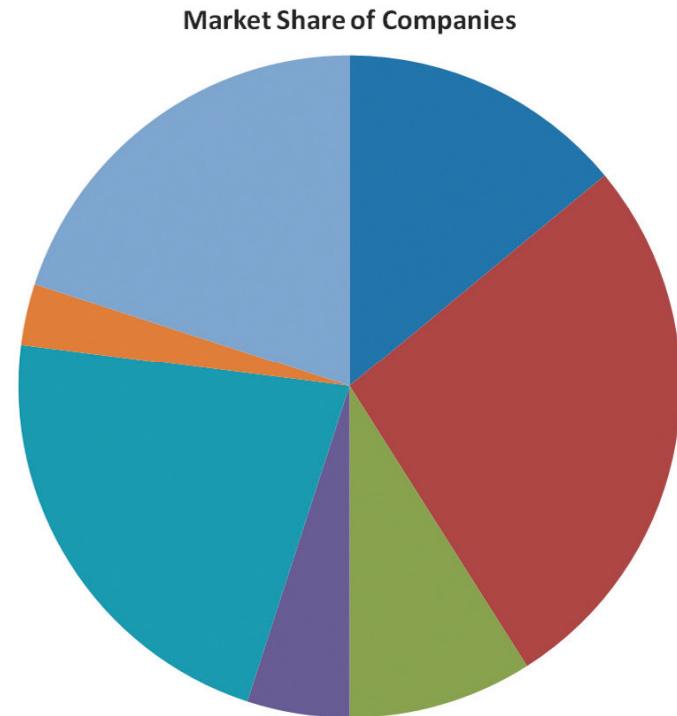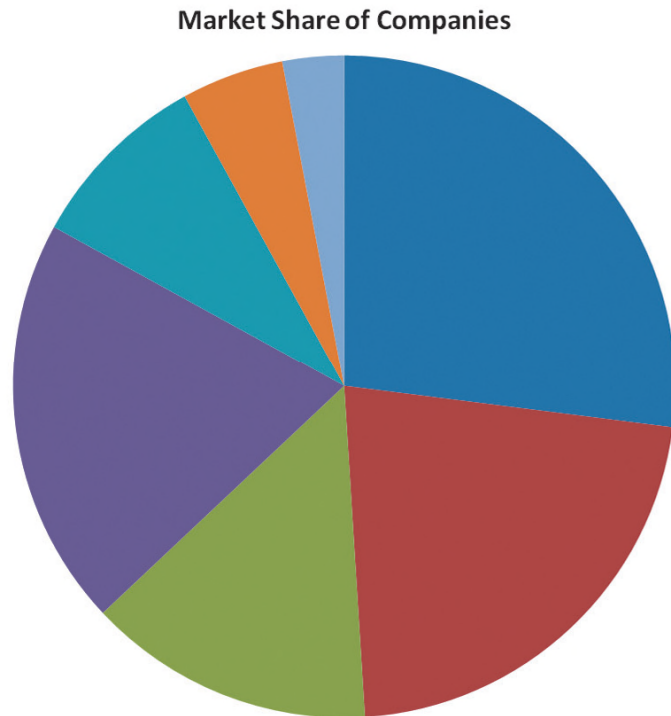
Why is it hard to tell? What would you do to improve?

Graphical Errors:
No Relative Basis

DC<u>O</u>VA

Bad Presentation

A's received by students.

✓ Good Presentation

A's received by students.

Freq.

300

200

100

0

FR   SO   JR   SR

%

30%

20%

10%

0%

FR   SO   JR   SR
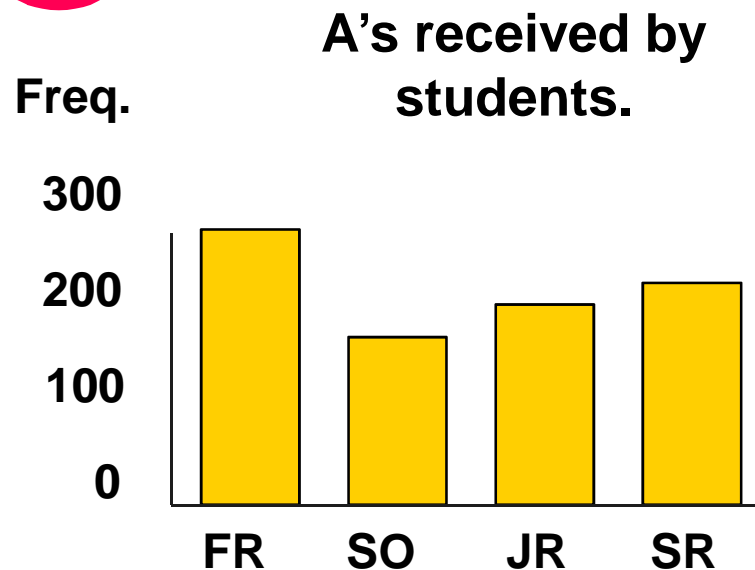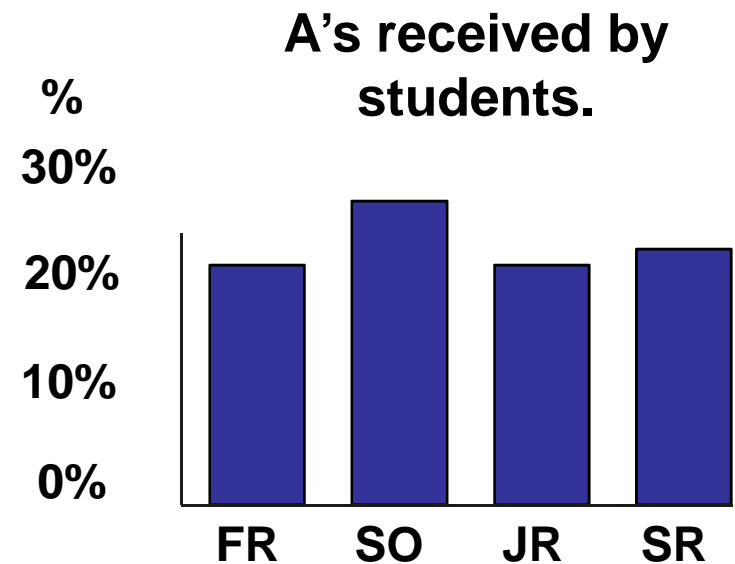
FR = Freshmen, SO = Sophomore, JR = Junior, SR = Senior

Graphical Errors: Compressing the Vertical Axis

🚫 **Bad Presentation**   ✓ **Good Presentation**

Quarterly Sales

Quarterly Sales
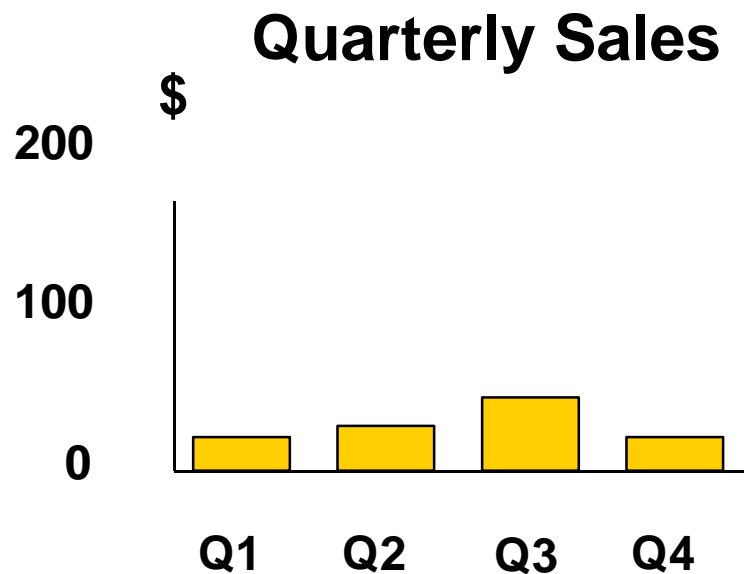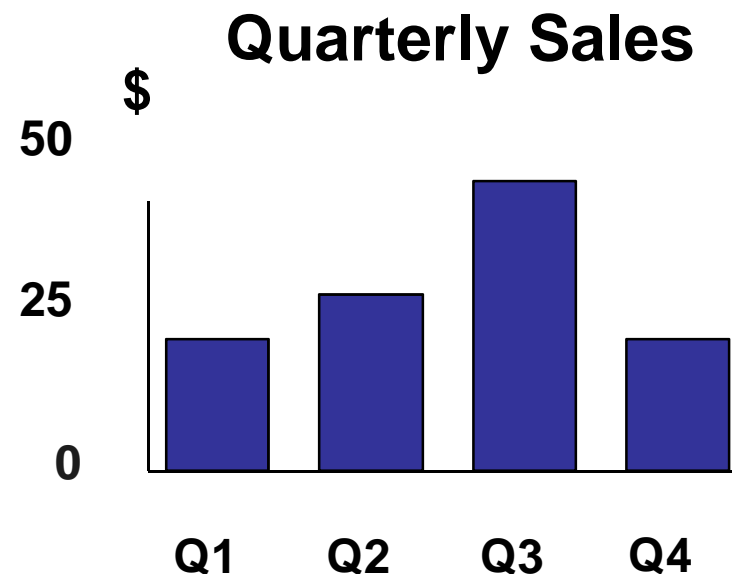
Graphical Errors: No Zero Point on the Vertical Axis

🚫 **Bad Presentation**

Monthly Sales
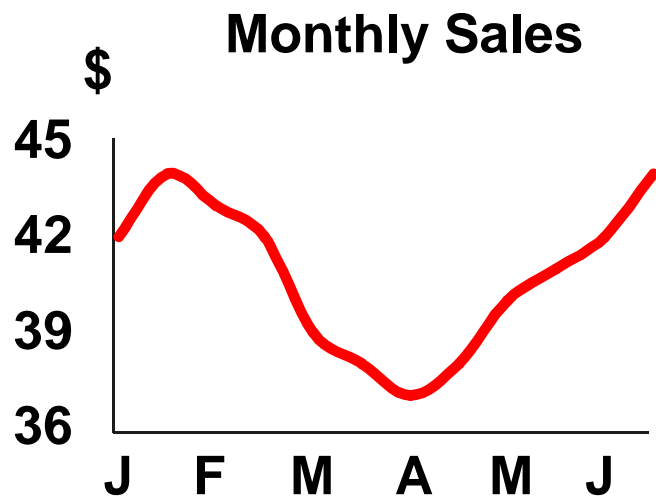
✓ **Good Presentations**

Monthly Sales
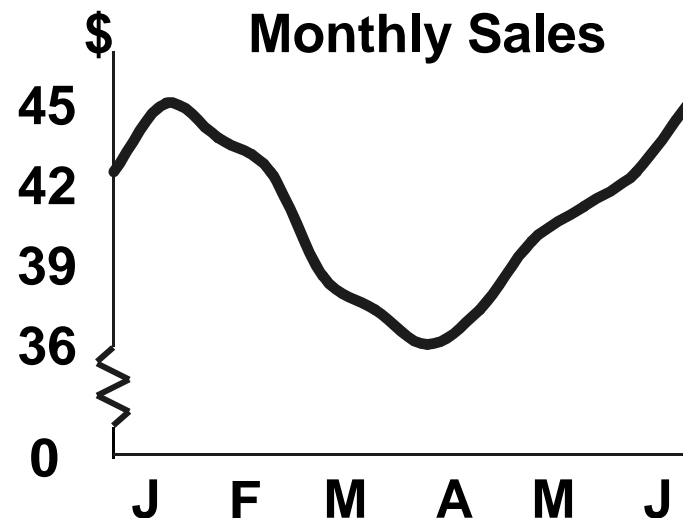
Graphing the first six months of sales

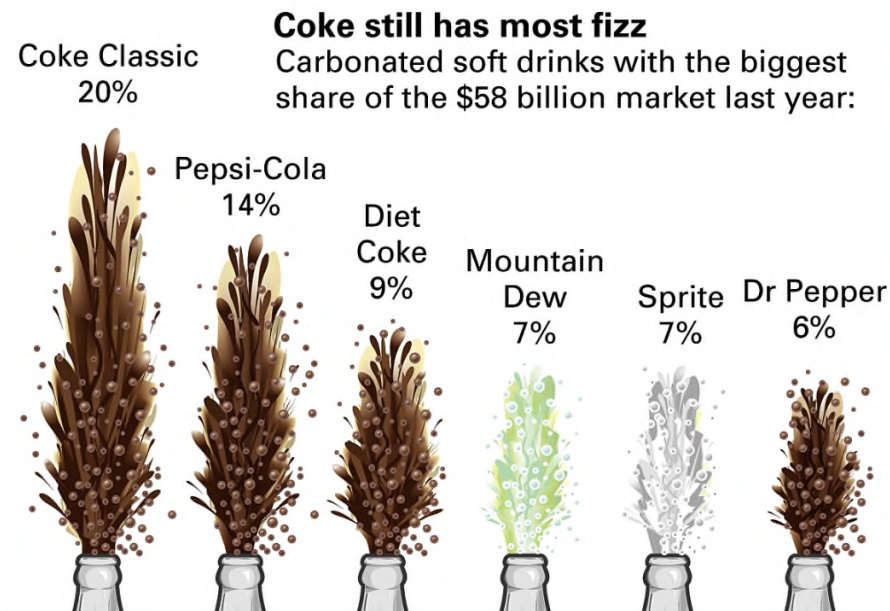# Graphical Errors: Chart Junk, Can You Identify The Junk?

DC<u>OV</u>A

🚫 **Bad Presentation**   ✔ **Good Presentation**

Graphical Errors: Chart Junk, Can You Identify The Junk?

🚫 **Bad Presentation**   ✔ **Good Presentation**

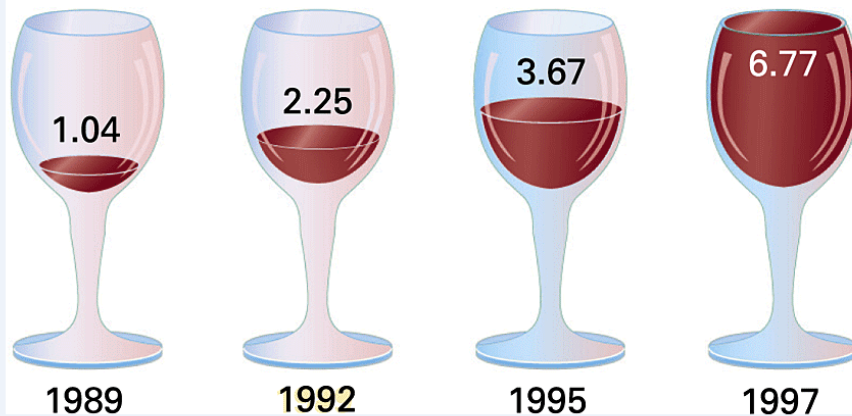# Graphical Errors: Chart Junk, Can You Identify The Junk?

...they're growing more...
Amount of land planted with grapes for the wine industry

1997–1998
243,644
acres

1979–1980
172,075
acres

1949–1950
135,326
acres

1969–1970
150,300
acres

1989–1990
146,204
acres

1959–1960
130,201
acres

Graphical Errors: Chart Junk, Can You Identify The Junk?

🚫 **Bad Presentation**   ✔ **Good Presentation**

Minimum Wage

1960: $1.00

1970: $1.60

1980: $3.10

1990: $3.80

Minimum Wage

$

4

2

0

1960 1970 1980 1990
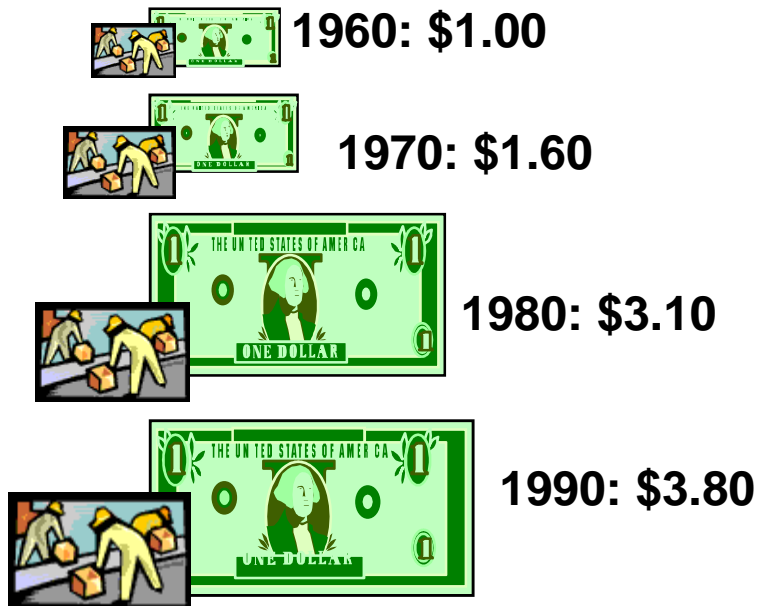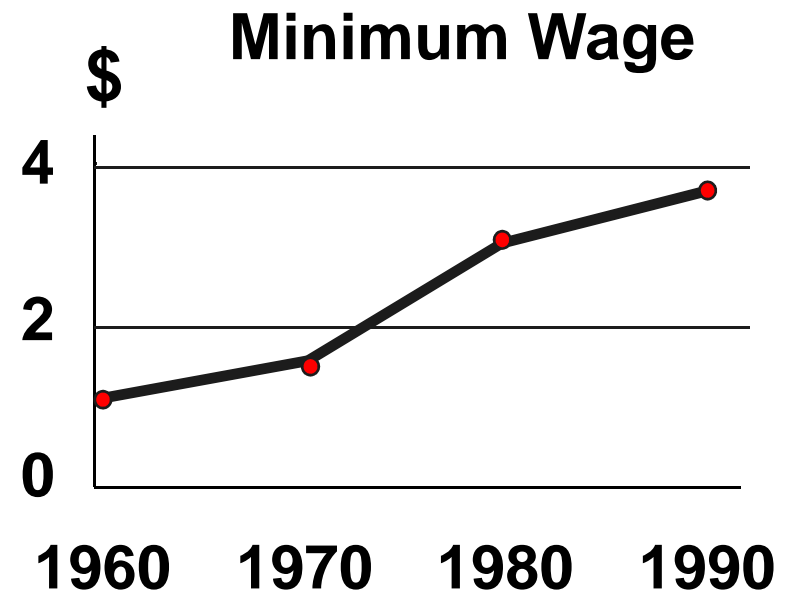
In Excel It Is Easy To Inadvertently Create Distortions

- Excel often will create a graph where the vertical axis does not start at 0.

- Excel offers the opportunity to turn simple charts into 3-D charts and in the process can create distorted image.

- Unusual charts offered as choices by Excel will most often create distorted images.

Best Practices for Constructing Visualizations

DC<u>OV</u>A

- Use the simplest possible visualization.

- Include a title & label all axes.

- Include a scale for each axis if the chart contains axes.

- Begin the scale for a vertical axis at zero & use a constant scale.

- Avoid 3D or "exploded" effects & the use of chartjunk.

- Use consistent colorings in charts meant to be compared.

- Avoid using uncommon chart types including radar, surface, bubble, cone, and pyramid charts.

Chapter Summary

In this chapter we covered:

- Organizing categorical variables.

- Organizing numerical variables.

- Visualizing categorical variables.

- Visualizing numerical variables.

- Organizing and visualizing a mix of variables.

- The challenge in organizing and visualizing variables.