Chapter 2—Organizing and Visualizing Data

2.1 The answer depends on the chosen data set.

2.2 The answer depends on the specific story.

2.3 The supermarket chain should use primary data collected through an observation study of the shopping behavior of their customers.

2.4 Government records.

2.5 (a)

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

(b) Category "B" is the majority.

2.6 (a) Tally

	Yes = Y	No = N	Without opinion = W	Total
Male	7	13	5	25
Female	8	11	6	25
Total	15	24	11	50

(b) Percentage based on total of 50 responses

	Yes = Y	No = N	Without opinion = W	Total
Male	14%	26%	10%	50%
Female	16%	22%	12%	50%
Total	30%	48%	22%	100%

Row Percentages

	Yes = Y	No = N	Without opinion = W	Total
Male	28%	52%	20%	100%
Female	32%	44%	24%	100%
Total	30%	48%	22%	100%

Column Percentages

	Yes = Y	No = N	Without opinion = W	Total
Male	46.67%	54.17%	45.45%	50%
Female	53.33%	45.83%	54.55%	50%
Total	100%	100%	100%	100%

2.7 (a)

Category	Frequency	Percentage
Flammables/Irritants	8,350	59.26%
Knives and blades	4,134	29.34%
Prohibited tools	753	5.34%
Sharp objects	497	3.53%
Other	357	2.53%
Total	14,091	100.00%

(b) Flammables, irritants, knives and blades made up almost 90% of the banned items.

2.8 (a)

Source of Electricity	Net Electricity Generation in millions of megawatt hours	Percentage
Coal	1,994.40	48.52%
Hydroelectric	248.1	6.04%
Natural gas	876.9	21.33%
Nuclear	806.2	19.61%
Other	184.7	4.49%
Total	4,110.3	100.00%

(b) Three sources of electricity dominate the U.S. electricity generation with coal being the major source at 48.52% followed by natural gas at 21.33% and nuclear at 19.61%.

2.9 (a)

Category	Cost per Household	Percentage
Civil servant retirement	15,851	2.90%
Federal debt	54,537	9.97%
Medicare	284,288	52.00%
Military retirement	29,694	5.43%
Social Security	160,216	29.30%
Other	2,172	0.40%
Total	546,758	100.00%

(b) Medicare at 52% and Social Security at 29.3% together made up more than 80% of the debt.

2.10 (a) Percentages based on the total of rows and columns

Undergraduate Major

Graduate School	Business	Engineering	Total
Yes	28.33%	18.33%	46.67%
No	30.33%	23.00%	53.33%
Total	58.76%	41.33%	100.00%

Percentages based on the total of the rows

Graduate School	Business	Engineering	Total
Yes	60.71%	39.29%	100.00%
No	56.88%	43.13%	100.00%
Total	58.67%	41.33%	100.00%

Percentages based on the total of the columns

Graduate School	Business	Engineering	Total
Yes	48.30%	44.35%	46.67%
No	51.70%	55.65%	53.33%
Total	100.00%	100.00%	100.00%

*Students should be aware that percentages have been rounded to the nearest figure.

(b) The business students have the highest desire of attending graduate school as compared to the engineering students.

60.71% of the business students compared to 39.29% of the engineering students said "Yes" to attending graduate school while 56.88% of the business students and 43.13% of the engineering students said "No" to attending graduate school.

2.11 (a)

Table of total percentages

	Shift		
	Day	Evening	
Nonconforming	1.6%	2.4%	4%
Conforming	65.4%	30.6%	96%
Total	67%	33%	100%

Table of row percentages

	Shift		
	Day	Evening	
Nonconforming	40%	60%	100%
Conforming	68%	32%	100%
Total	67%	33%	100%

Table of column percentages

	Shift		
	Day	Evening	
Nonconforming	2%	7%	4%
Conforming	98%	93%	96%
Total	100%	100%	100%

(b) The row percentages allow us to block the effect of disproportionate group size and show us that the pattern for day and evening tests among the nonconforming group is very different from the pattern for day and evening tests among the conforming group. Where 40% of the nonconforming group was tested during the day, 68% of the conforming group was tested during the day.

(c) The director of the lab may be able to cut the number of nonconforming tests by reducing the number of tests run in the evening, when there is a higher percent of tests run improperly.

2.12 Table of row percentages

	Need => 3 Clicks		
Year	Yes	No	
2009	39%	61%	100%
2008	7%	93%	100%

According to the row percentages table, 32% more online retailers were requiring three or more clicks in 2009 than in 2008.

2.13 Ordered array: 101 103 104 108 124 125 138

2.14 Ordered array: 73 78 78 78 85 88 91

2.15 (a) 4% (b) 32% (c) 36% (d) 100%

2.16 (a) The class boundaries of the 9 classes can be "10 to less than 20", "20 to less than 30", "30 to less than 40", "40 to less than 50", "50 to less than 60", "60 to less than 70", "70 to less than 80", "80 to less than 90", and "90 to less than 100".

(b) The class-interval width is $= \dfrac{97.8 - 11.6}{9} = 9.58 \cong 10$.

(c) The nine class midpoints are: 15, 25, 35, 45, 55, 65, 75, 85, and 95.

2.17 (a) Ordered array: Cost($)114, 135, 141, 145, 146, 151, 158, 161, 162, 164, 165, 166, 170, 170, 172, 180, 185, 187, 205, 210, 215, 216, 220, 222, 223, 224, 259, 305, 326, 411

(b) PHStat output:

Bin Cell	*Frequency*	*Percentage*
110 but less than 150	5	16.67%

150 but less than 190	13	43.33%
190 but less than 230	8	26.67%
230 but less than 270	1	3.33%
270 but less than 310	1	3.33%
310 but less than 350	1	3.33%
350 but less than 390	0	0.00%
390 but less than 430	1	3.33%

(c) The costs of attending a baseball game is concentrating around $170 for thirteen of the teams have costs in between $150 and $190.

2.18 (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.

2.19 (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)

Bulb Life (hrs)	*Frequency Manufacturer A*		*Bulb Life (hrs)*	*Frequency Manufacturer B*
650 -- 749	3		750 -- 849	2
750 -- 849	5		850 -- 949	8
850 -- 949	20		950 -- 1049	16
950 -- 1049	9		1050 -- 1149	9
1050 -- 1149	3		1150 -- 1249	5

(a), (b)

Bulb Life (hrs)	*A*		*B*	
	Percentage	*Cumulative %*	*Percentage*	*Cumulative %*
650 – 749	7.50%	7.50%	.00%	0.00%
750 – 849	12.50%	20.00%	5.00%	5.00%
850 – 949	50.00%	70.00%	20.00%	25.00%
950 – 1049	22.50%	92.50%	40.00%	65.00%
1050 – 1149	7.50%	100.00%	22.50%	87.50%
1150 – 1249	0.00%	100.00%	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 1,050 hours, contrasted with 70% of Manufacturer A's bulbs, which lasted less than 950 hours. None of Manufacturer A's bulbs lasted more than 1,149 hours, but 12.5% of Manufacturer B's bulbs lasted between 1,150 and 1,249 hours. At the same time, 7.5% of Manufacturer A's bulbs lasted less than 750 hours, whereas all of Manufacturer B's bulbs lasted at least 750 hours

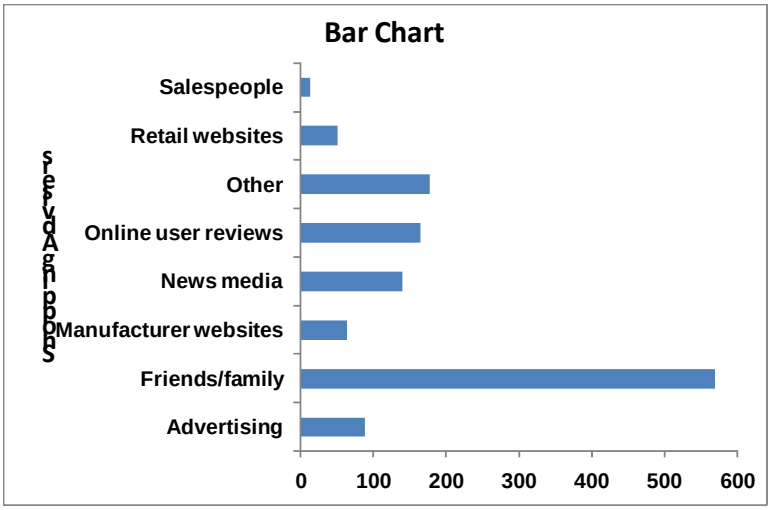
2.23 (a)

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

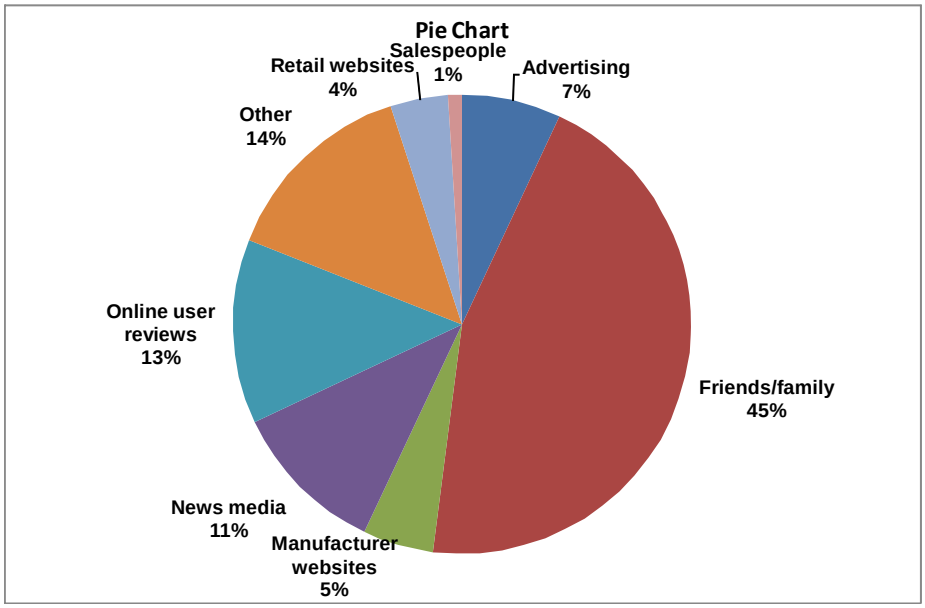
Amount of Soft Drink Less Than	Frequency Less Than	Percentage Less Than
1.899	1	2%
1.949	6	12
1.999	24	48
2.049	43	86
2.099	49	98
2.149	50	100

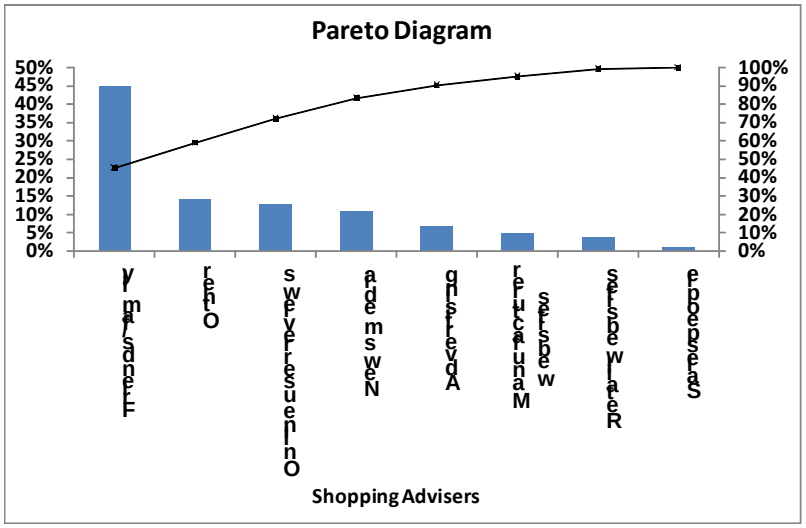
(b) The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.24 (a) Note: %s converted to counts. $n = 1264$



2.24 cont. (a)





(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 friends/family account for the largest percentage of 45%. When other, news media, and online user reviews are added to friends/family, this accounts for 83%.

2.25 (a)

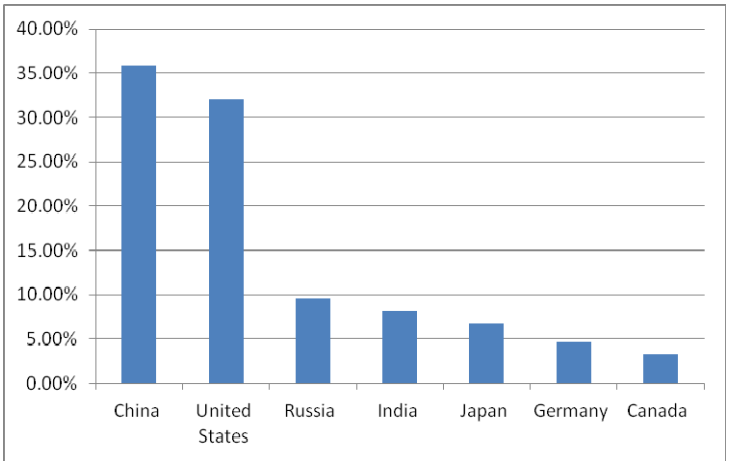


Figure 1 Carbon Dioxide Emission by Country

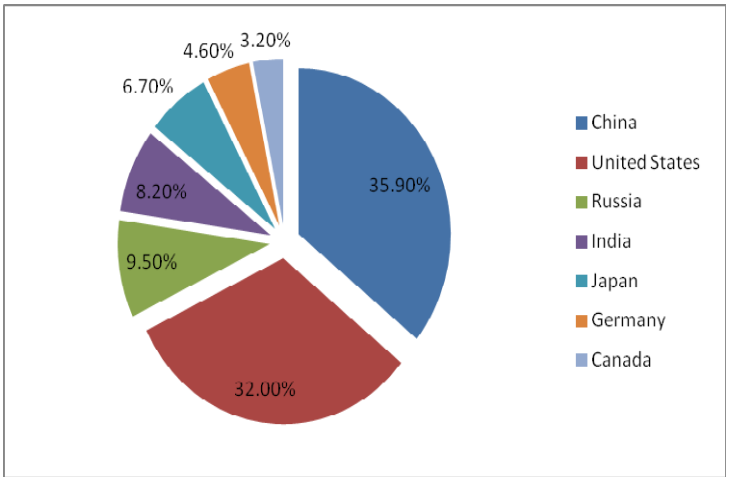


Figure 2: Percent Emission by Country

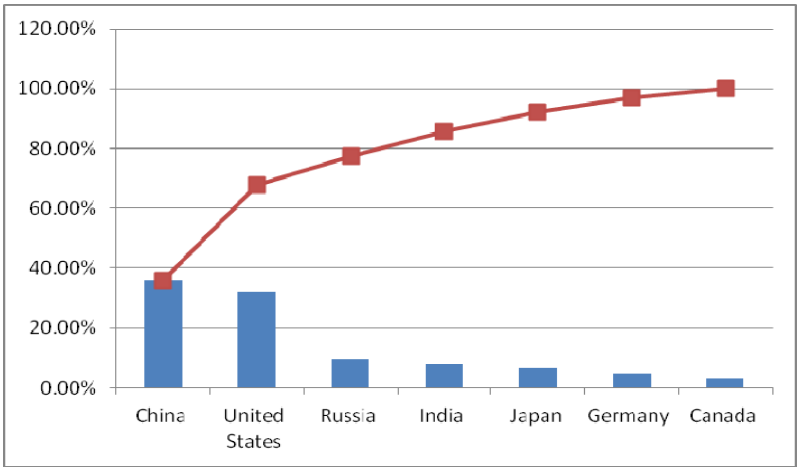
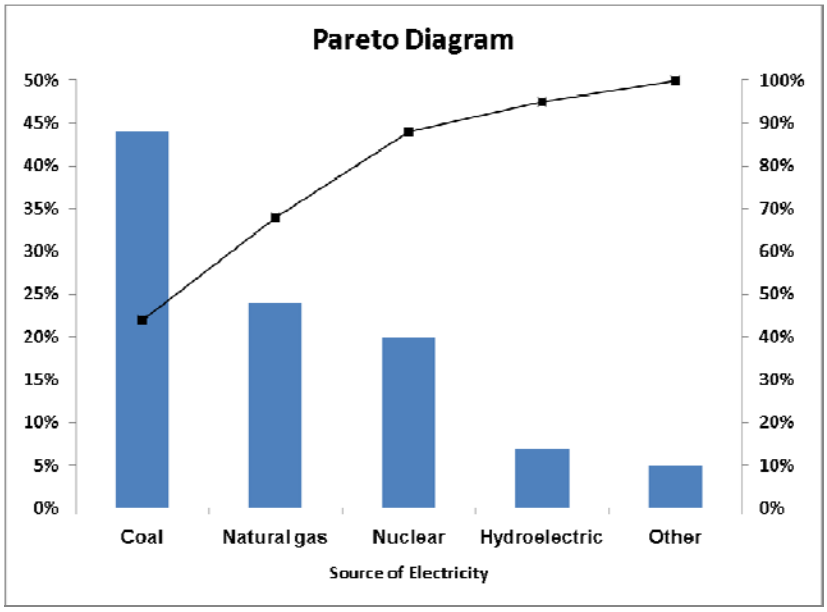


Figure 3: Percent Emission by Country

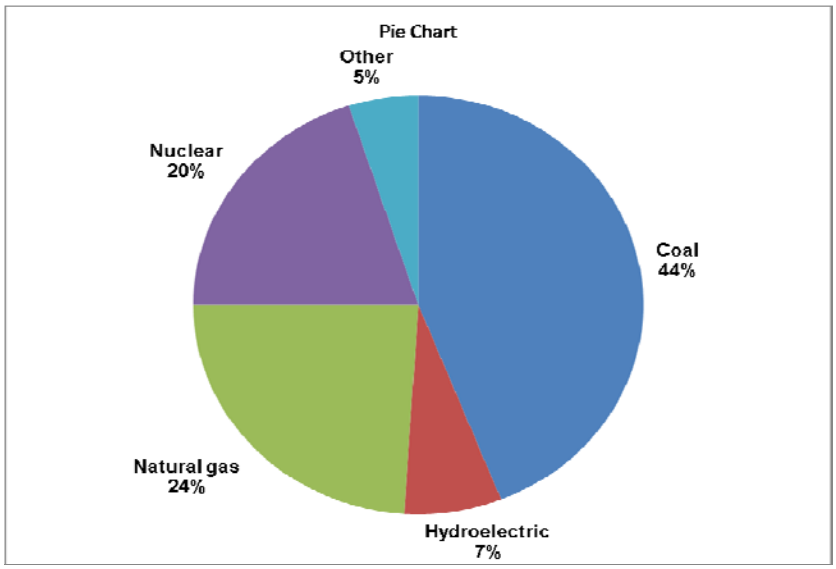
(b) The bar graph shows the distribution for all countries considered in this study so this is the preferred graphical method.

2.26 (a)



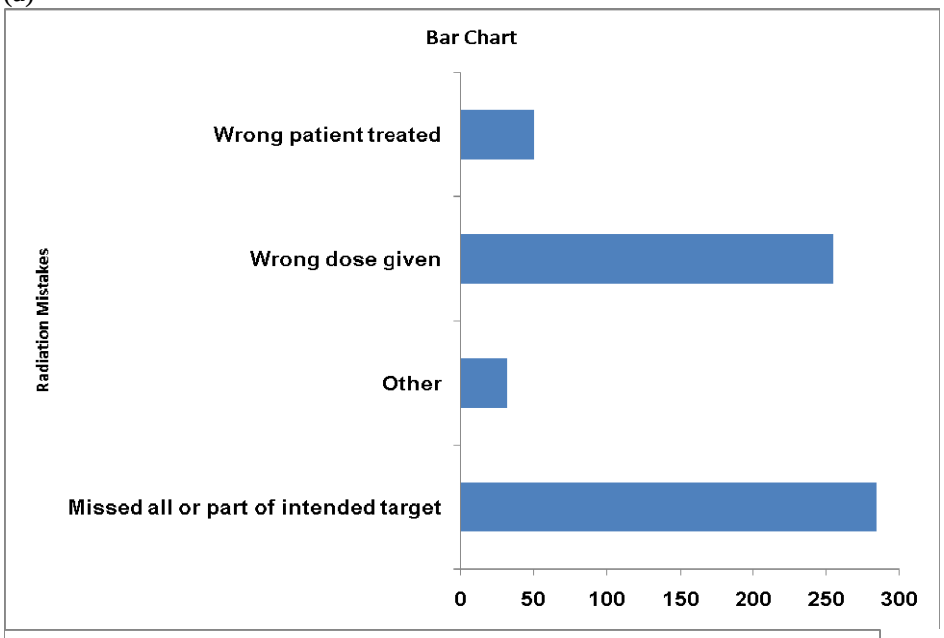
(b) According to the Pareto chart, slightly less than 90% of the power is derived from coal, nuclear, or natural gas.

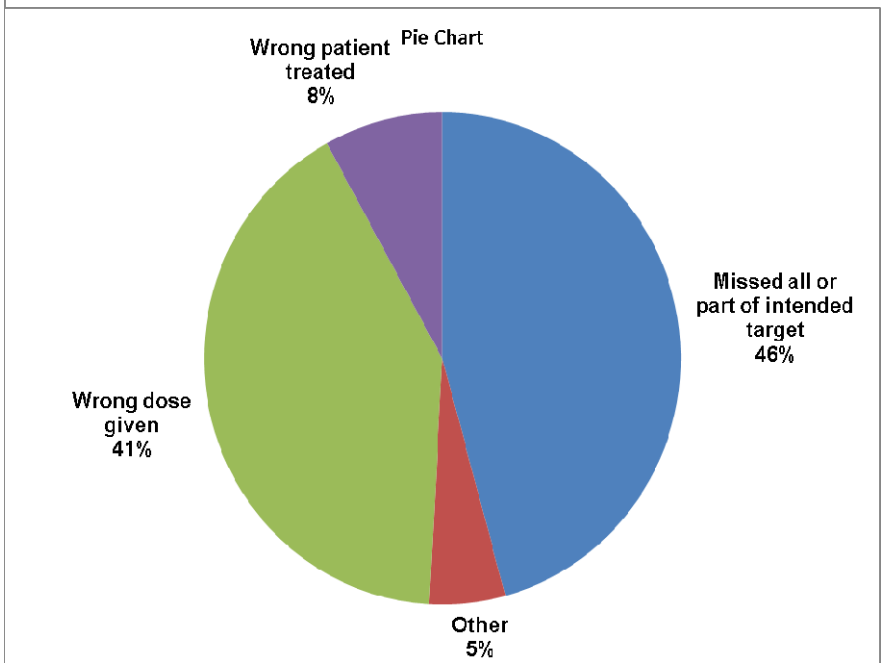
(c)



(d) You will prefer using the Pareto chart over the pie chart because the Pareto chart not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

2.27 (a)

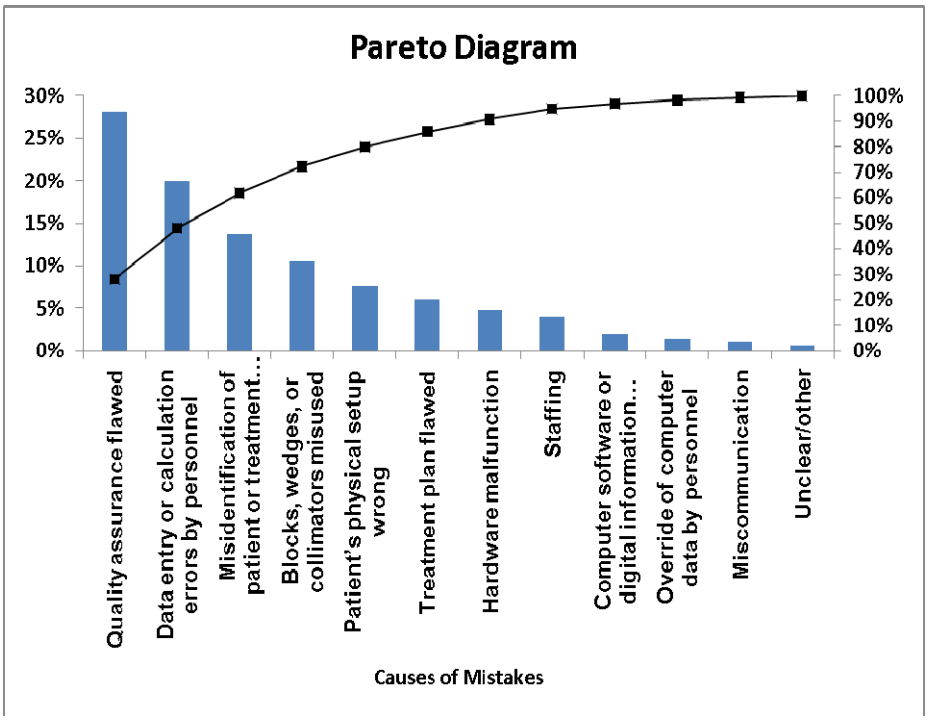




(b) The bar chart is more suitable if the purpose is to compare the categories. The pie chart is more suitable if the main objective is to investigate the portion of the whole that is in a particular category. *

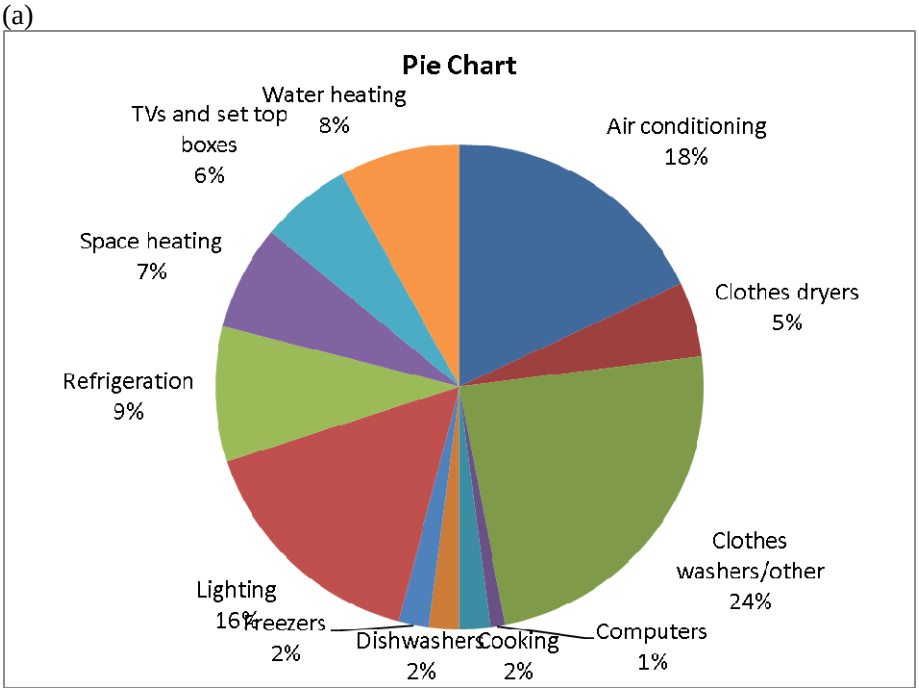
* Note: This is one of the many possible solutions for the question.

2.27 cont. (c)

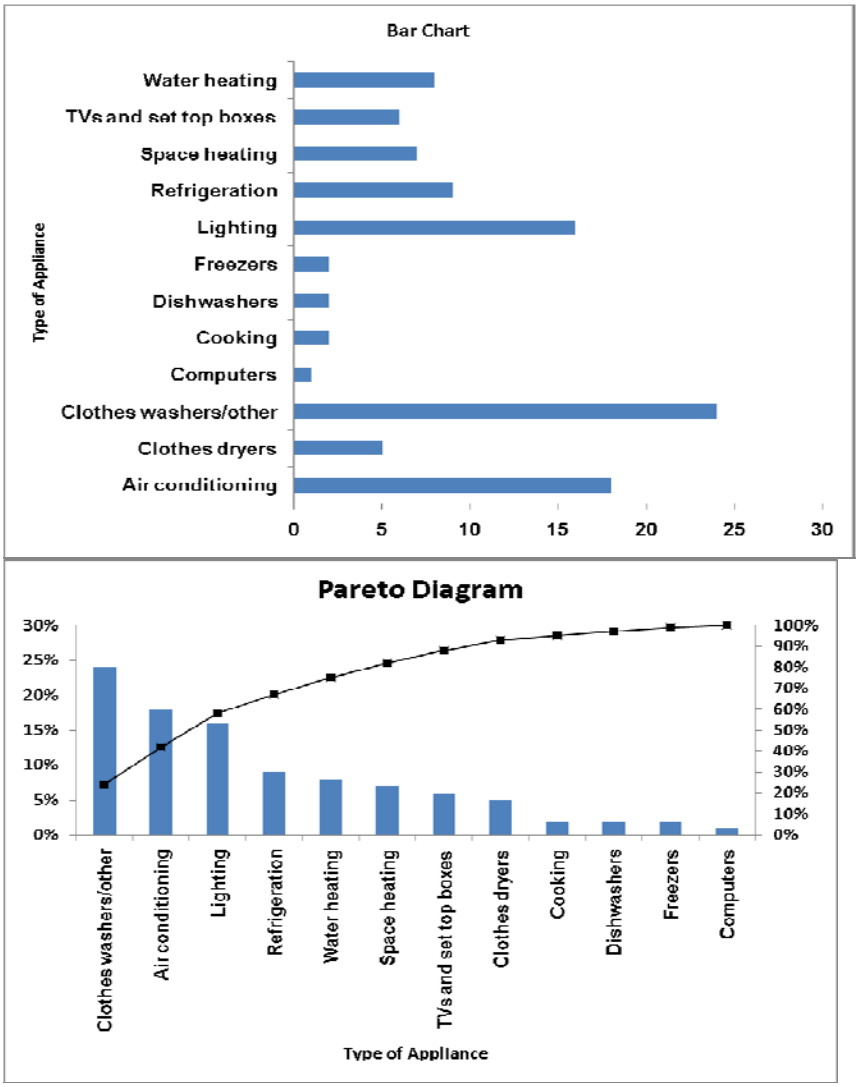


(d) The "vital few" reasons for the causes of mistakes are "Quality assurance flawed", "Data entry or calculation errors by personnel", and "Misidentification of patient or treatment location" which account for more than 60% of the mistakes. The remaining causes are the "trivial many" which make up less than 40% of the mistakes.

2.28 (a)



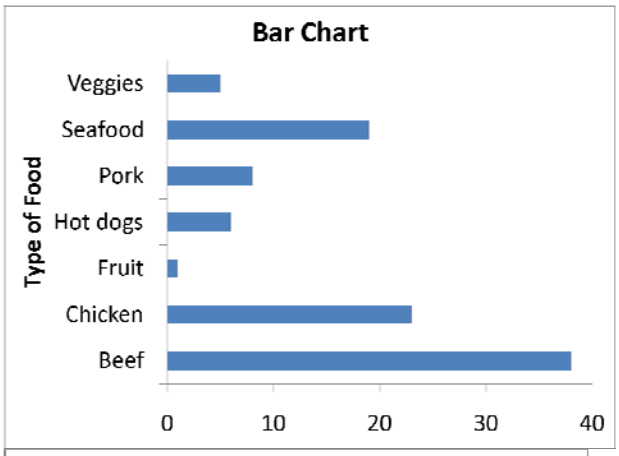
2.28 cont. (a)

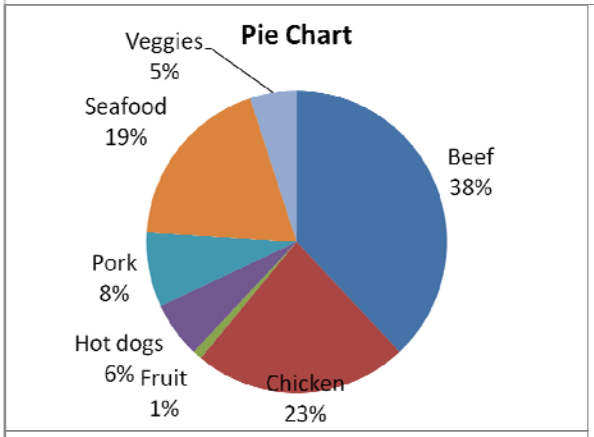


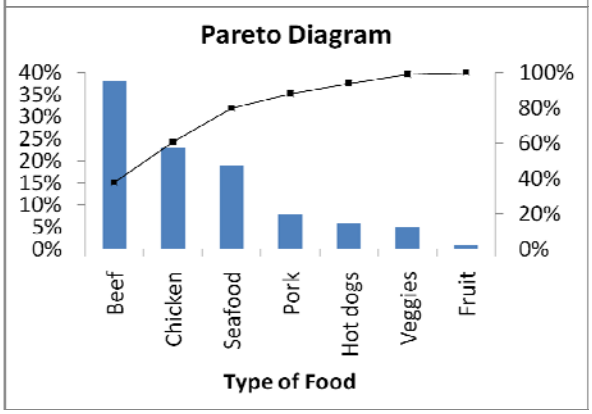
(b) The Pareto diagram is better than the pie chart and bar chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

(c) Almost 60% of the residential electricity consumption in the United States is on "Clothes washers/other", "Air conditioning", and "Lighting".

2.29 (a)



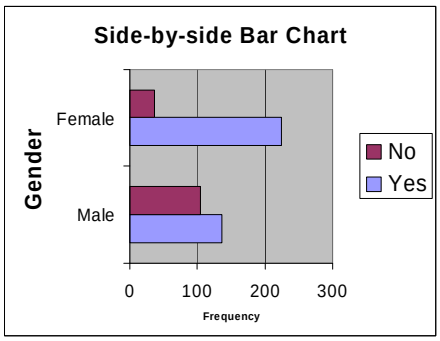




(b) The Pareto diagram is better than the pie chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

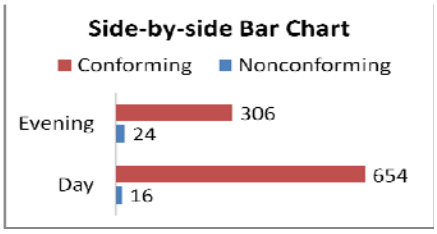
(c) From the Pareto chart, beef, chicken and seafood make up 80% of what folks want sizzling on the grill during barbecue season.

2.30 (a)



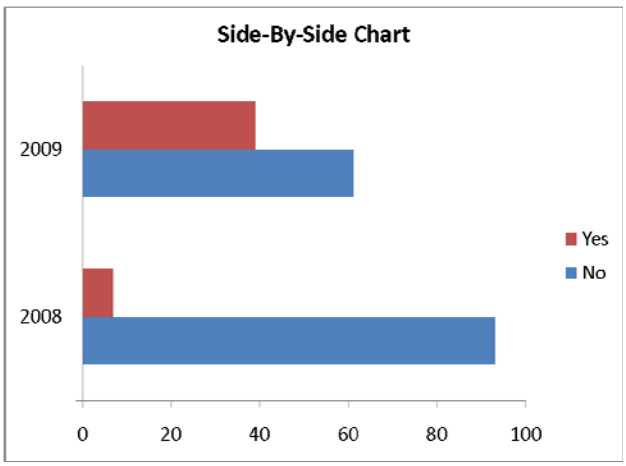
(b) A higher percentage of females enjoy shopping for clothing.

2.31 (a)



(b) The director of the lab may be able to cut the number of nonconforming tests by reducing the number of tests run in the evening, when there is a higher percent of tests run improperly.

2.32 (a)



(b) 32% more online retailers were requiring three or more clicks in 2009 than in 2008.

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: 1, 1, 1.1, 1.2, 1.3, 1.4, 1.7, 1.8, 2, 2.1, 2.1, 2.1, 2.2, 2.2, 2.3, 2.3, 2.9, 3.1, 3.2, 4.1

(b) The stem and leaf display provides information about the maximum value, minimum value, location of where the data is concentrated, and the original data.

(d) The data is concentrated between the values of 1 to 2 minutes waiting time for the sample of customers analyzed.

.

2.36 (a) **Stem-and-Leaf Display**

Stem unit: 10

Statistics	
Sample Size	30
Mean	196.9333
Median	176
Std. Deviation	62.26857
Minimum	114
Maximum	411

11	4
12	
13	5
14	1 5 6
15	1 8
16	1 2 4 5 6
17	0 0 2
18	0 5 7
19	
20	5
21	0 5 6
22	0 2 3 4
23	
24	
25	9
26	
27	
28	
29	
30	5
31	
32	6
33	
34	
35	
36	
37	
38	
39	
40	
41	1

(b) The results are concentrated between $160 and $225.

2.37 (a) Ordered array: Cost($) 0.55, 0.57, 0.57, 0.68, 0.72, 0.77, 0.86, 0.90, 0.92, 0.94, 1.14, 1.41, 1.42, 1.51

(b)

Stem-and-Leaf Display

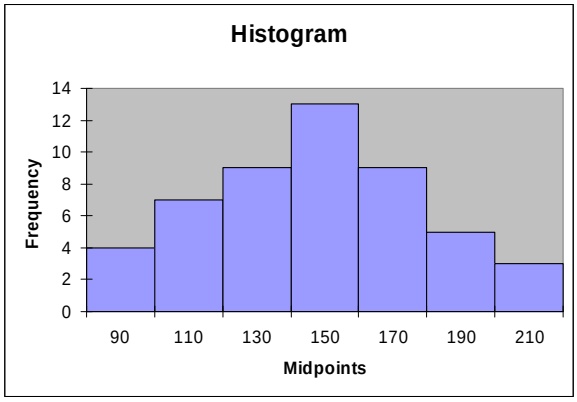
Stem unit: 0.1

5	5 7 7
6	8
7	2 7
8	6
9	0 2 4
10	
11	4
12	
13	
14	1 2
15	1

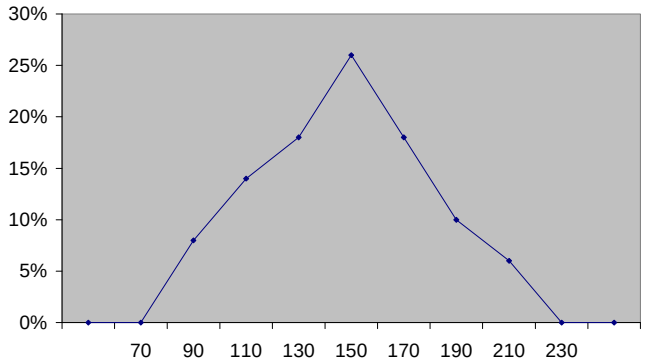
(c) The stem-and-leaf display conveys more information than the ordered array. We can more readily determine the arrangement of the data from the stem-and-leaf display than we can from the ordered array. We can also obtain a sense of the distribution of the data from the stem-and-leaf display.

(d) The cost does not appear to be concentrated around any value.

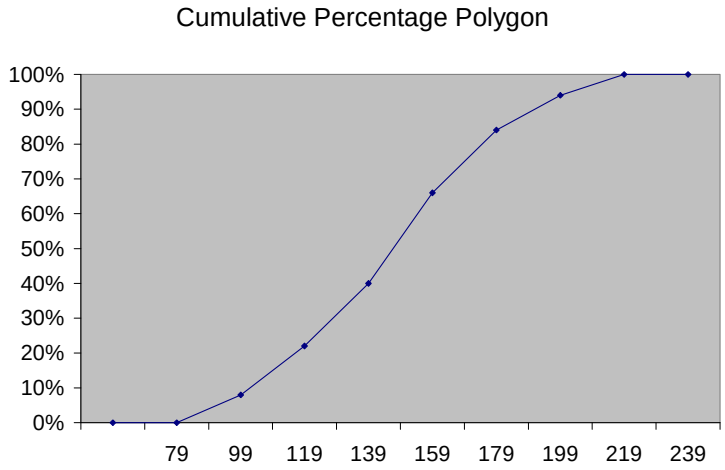
2.38 (a)



Percentage Polygon



2.38 (b) cont.

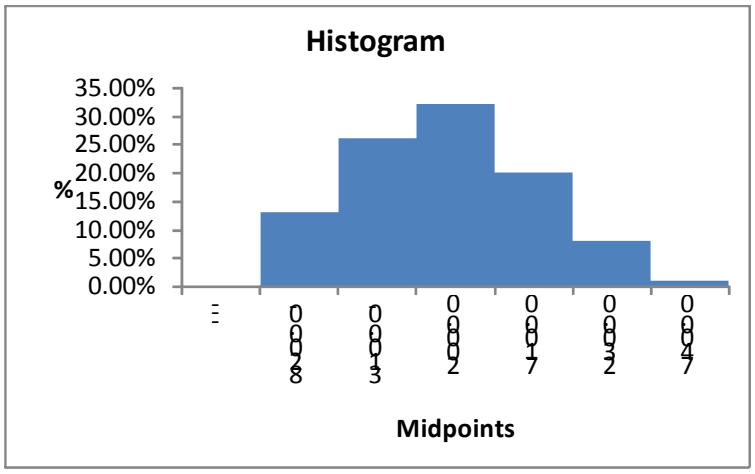


(c) The majority of utility charges are clustered between $120 and $180.

2.39 The costs of attending a baseball game is concentrating around $160 for nine of the teams. Six teams have costs centered around $220. There are a few outliers in the right tail with one team having a cost higher than $410.

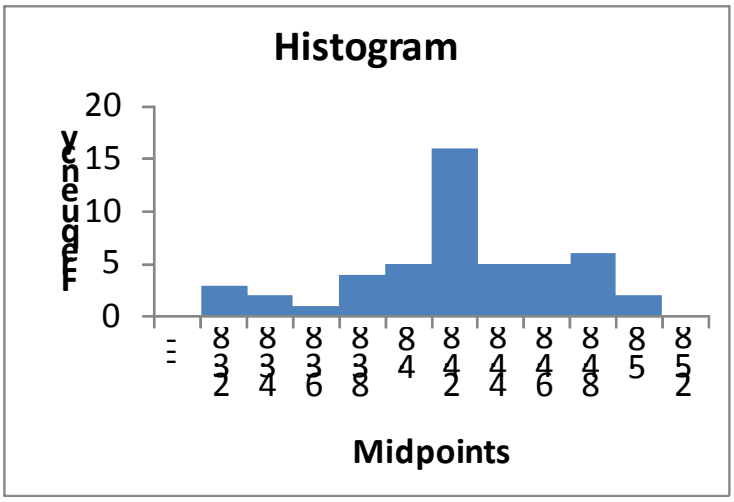
2.40 The property taxes per capita appear to be right-skewed with approximately 90% falling between $399 and $1,700, and the remaining 10% fall between $1,700 and $2,100. The center is at about $1,000.

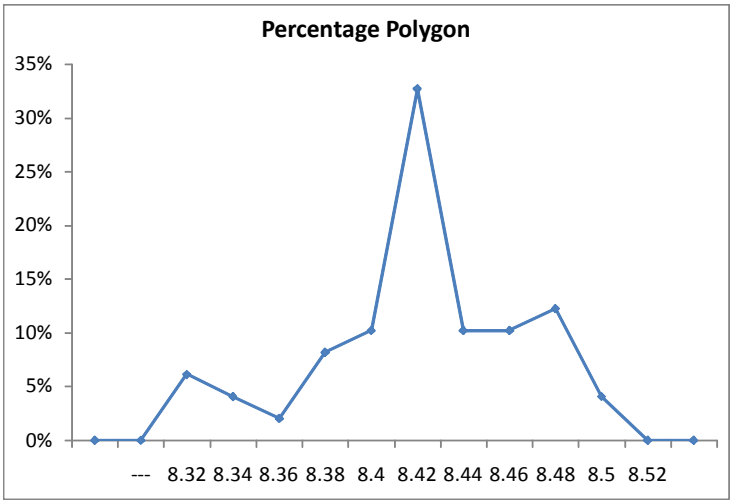
2.41 (a)



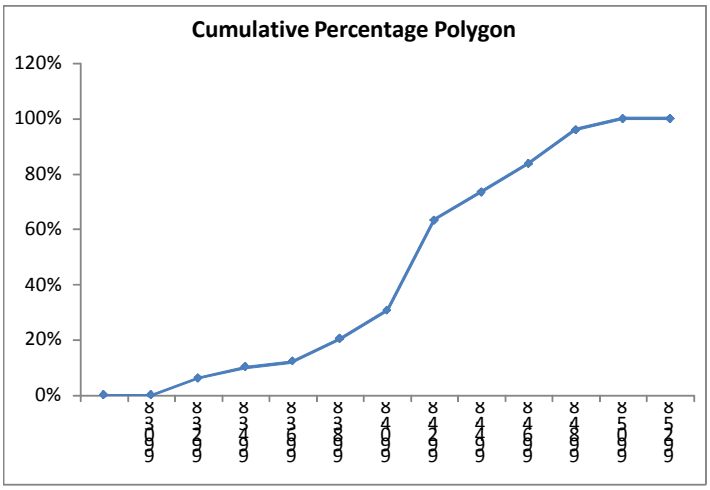
(b) Yes, the steel mill is doing a good job at meeting the requirement as there is only one steel part out of a sample of 100 that is as much as 0.005 inches longer than the specified requirement.

2.42 (a)



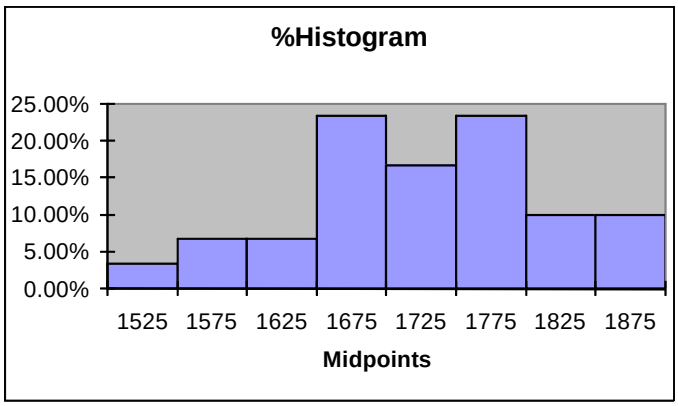


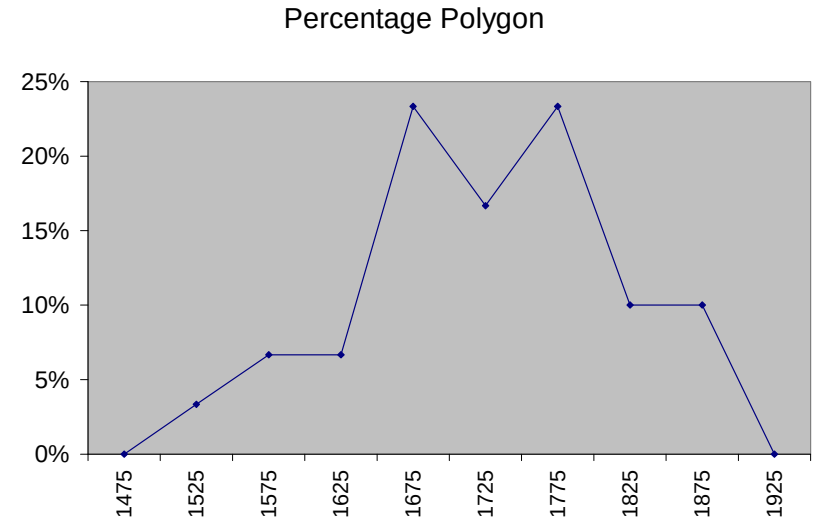
(b)



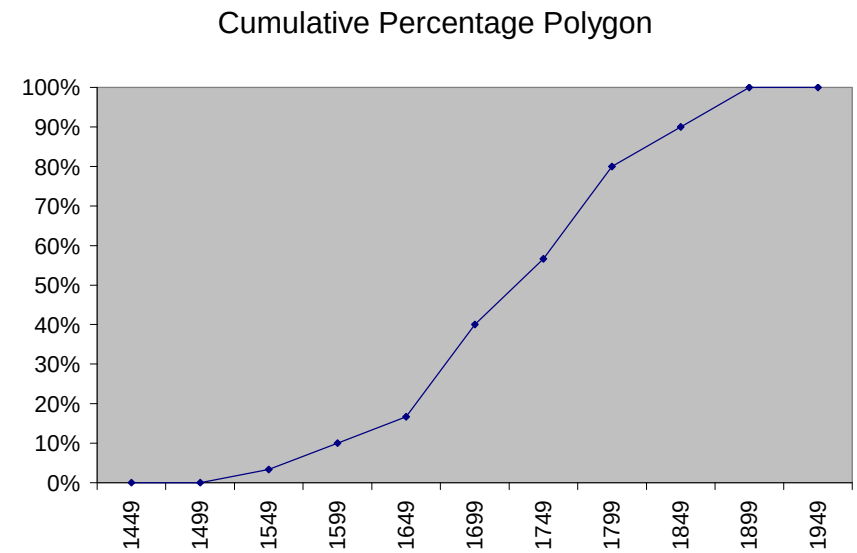
(c) All the troughs will meet the company's requirements of between 8.31 and 8.61 inches wide.

2.43 (a)



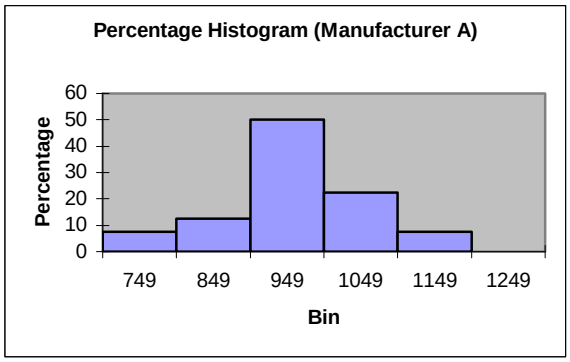


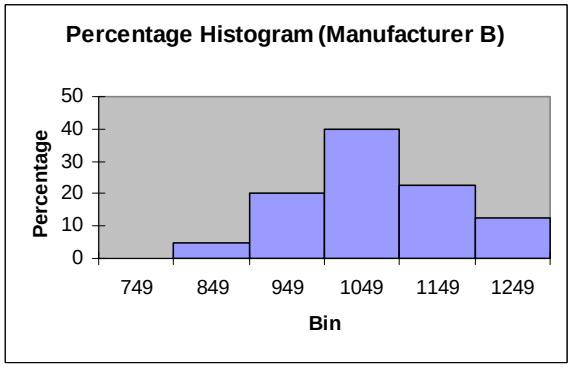
(b)

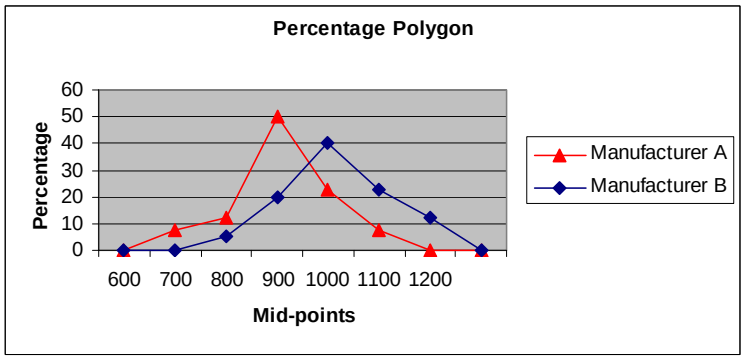


(c) The strength of all the insulators meets the company's requirement of at least 1500 lbs.

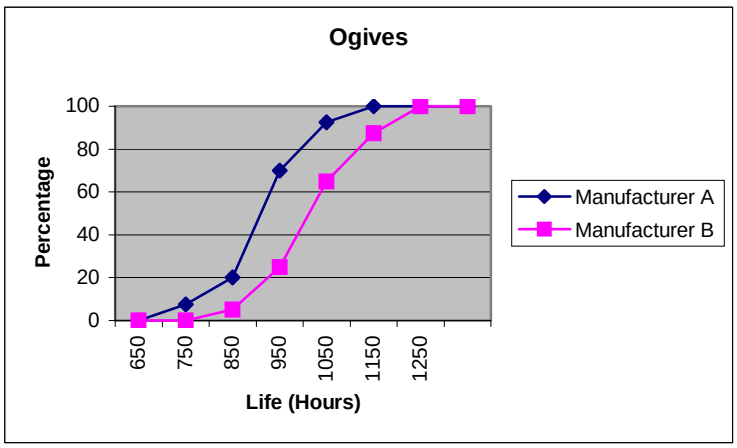
2.44 (a)





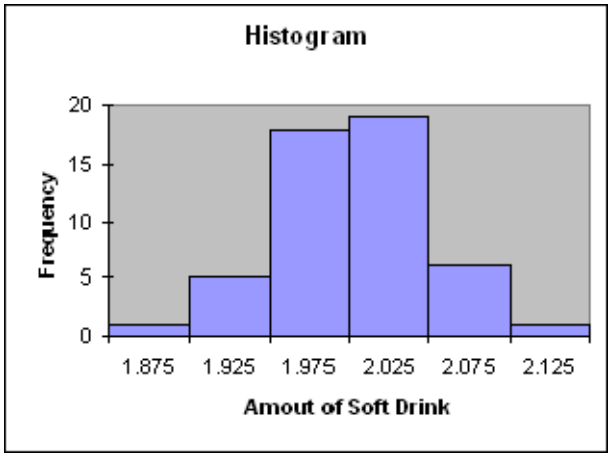


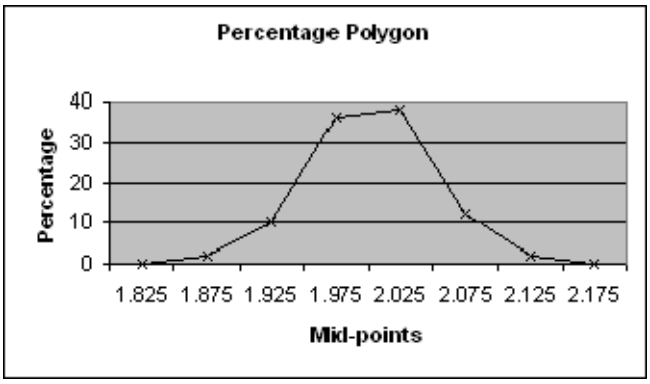
(b)



2.44 cont. (c) Manufacturer B produces bulbs with longer lives than Manufacturer A. The cumulative percentage for Manufacturer B shows 65% of their bulbs lasted 1049 hours or less contrasted with 70% of Manufacturer A's bulbs which lasted 949 hours or less. None of Manufacturer A's bulbs lasted more than 1149 hours, but 12.5% of Manufacturer B's bulbs lasted between 1150 and 1249 hours. At the same time, 7.5% of Manufacturer A's bulbs lasted less than 750 hours, while all of Manufacturer B's bulbs lasted at least 750 hours.

2.45 (a)

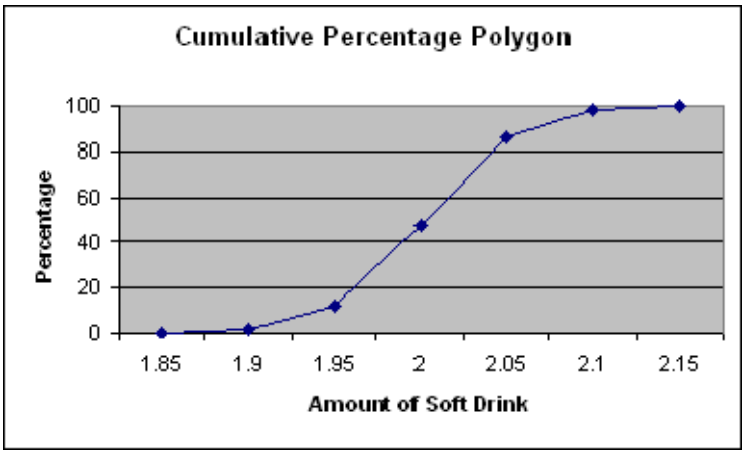




(b)

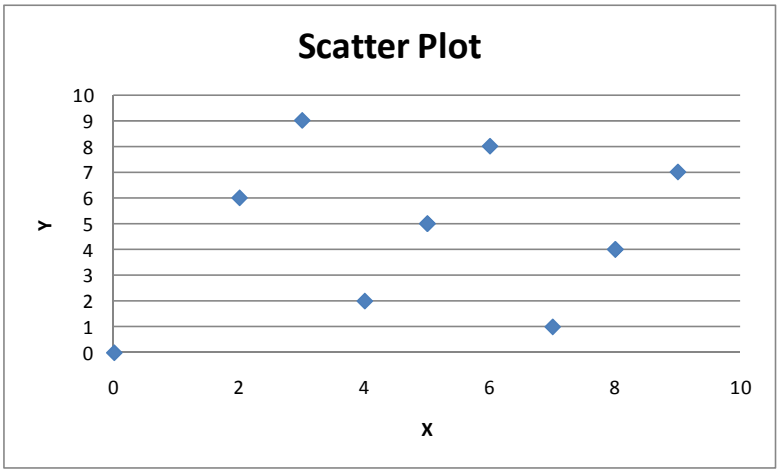
Amount of Soft Drink	Frequency Less Than	Percentage Less Than
1.899	1	2%
1.949	6	12
1.999	24	48
2.049	43	86
2.099	49	98
2.149	50	100

2.45 cont. (b)



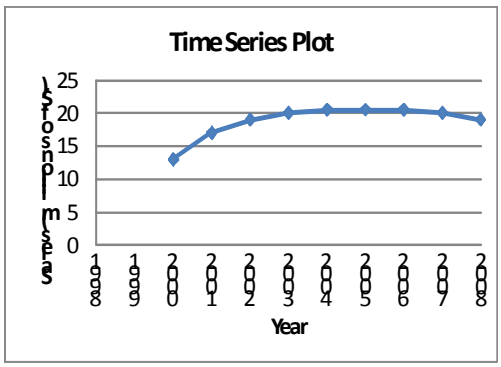
(c) The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.46 (a)



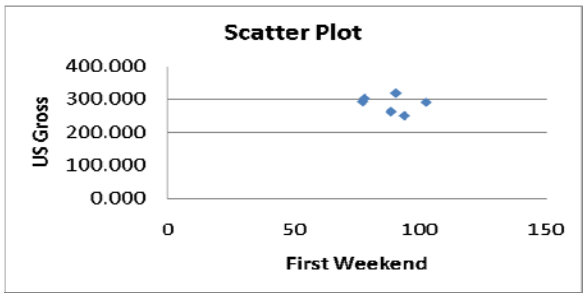
(b) There is no relationship between *X* and *Y*.

2.47 (a)

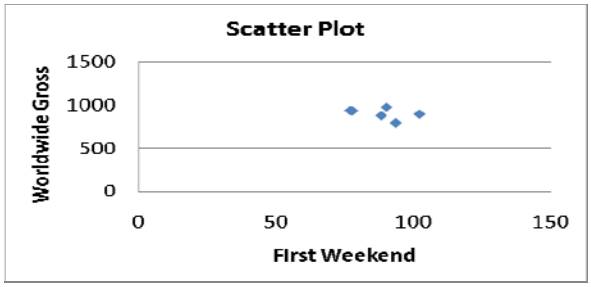


(b) Annual sales appear to be increasing in the earlier years before 2002 but start to decline after 2006.

2.48 (a)

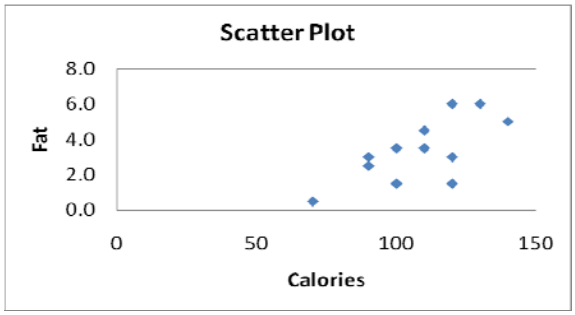


(b)



(c) There appears to be a rather weak negative relationship between first weekend gross and U. S. gross and between first weekend gross and worldwide gross. However, due to the small sample size, the relationships should not be taken as conclusive.

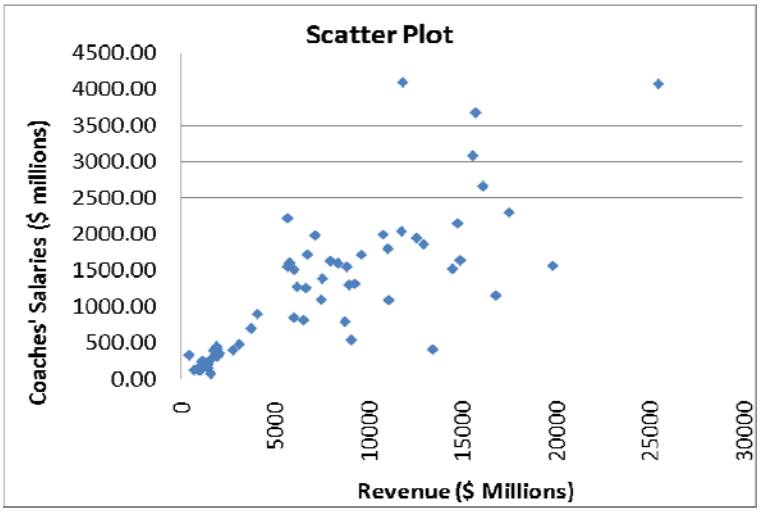
2.49 (a)



(b) There appears to be a positive relationship between the calories and total fat in veggie burgers.

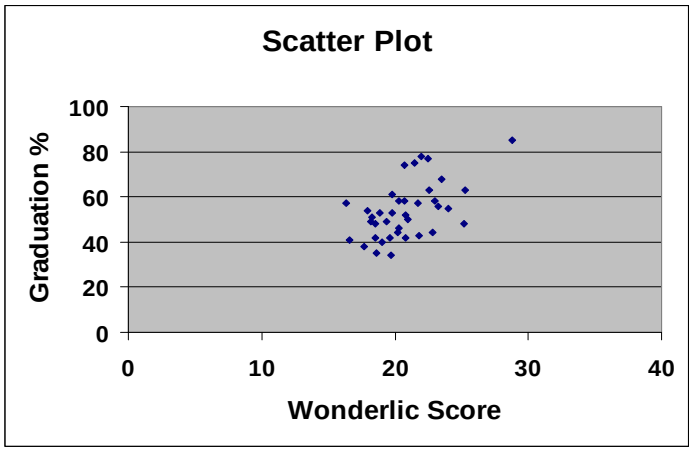
2.50 (a) Yes, schools with higher revenues will also have higher coaches' salaries.

(b)



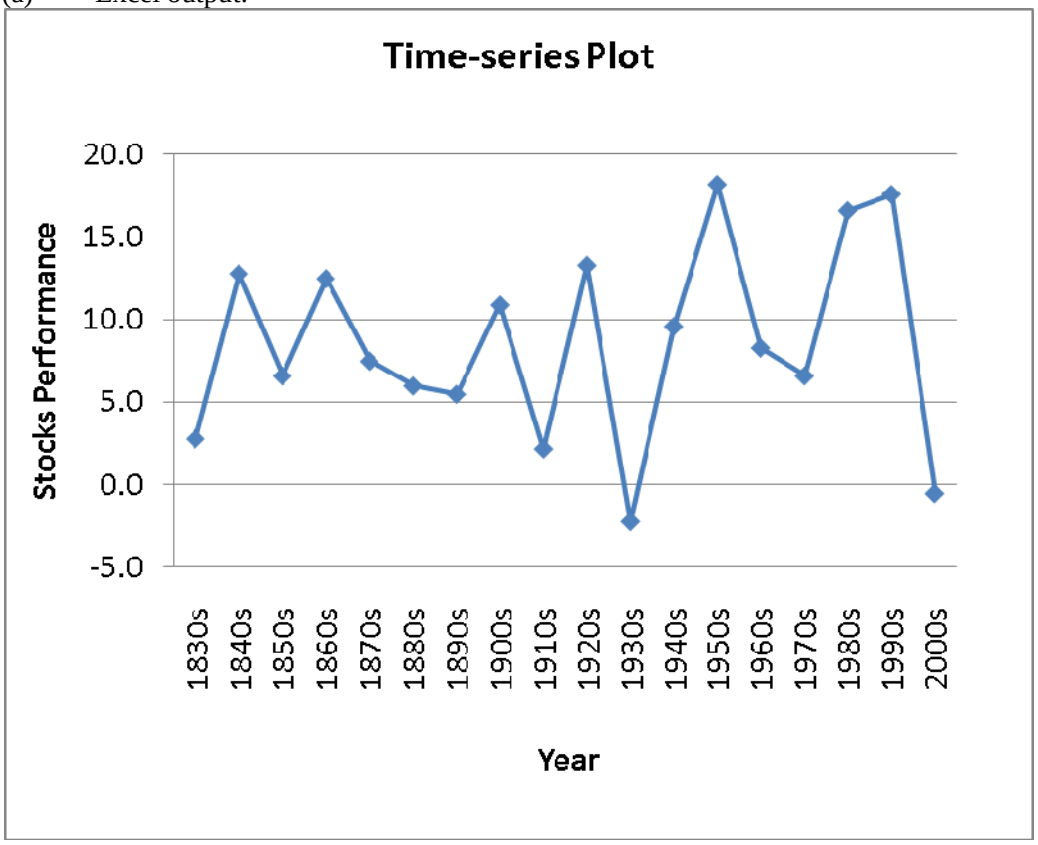
(c) There appears to be a positive relationship between coaches' salary and revenue. Yes, this is borne out by the data.

2.51 (a)



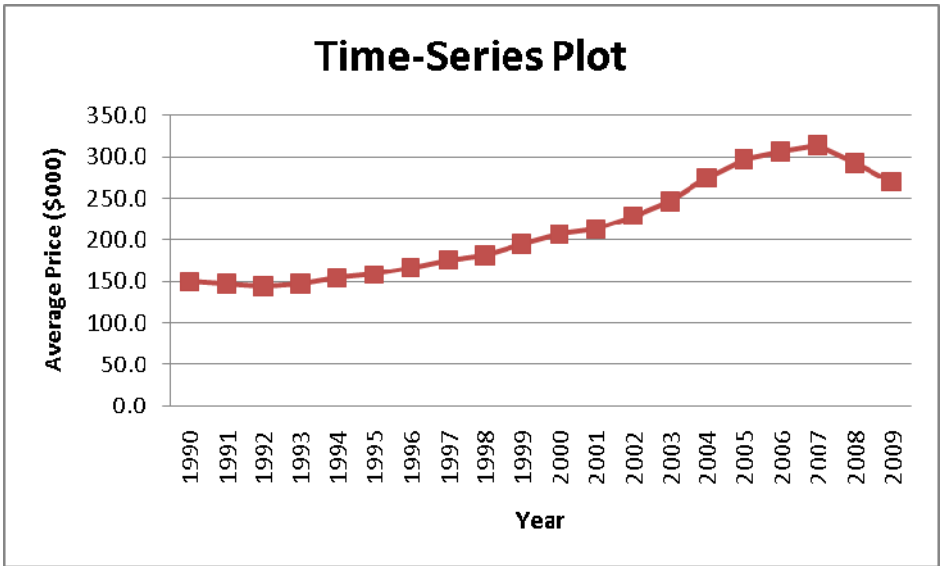
(b) There is a positive relationship between Wonderlic score and graduation rate.

2.52 (a) Excel output:



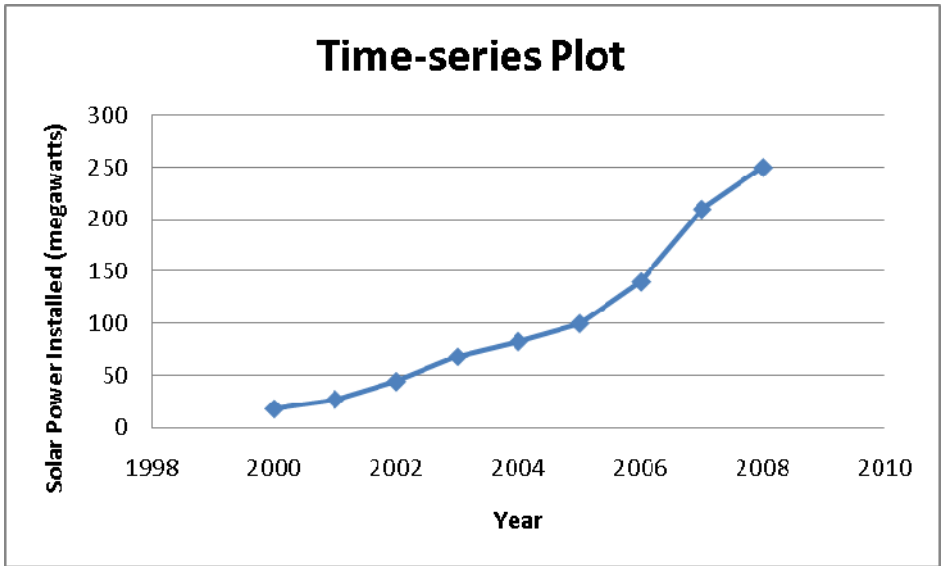
(b) There is no obvious pattern in the data.

2.53 (a)



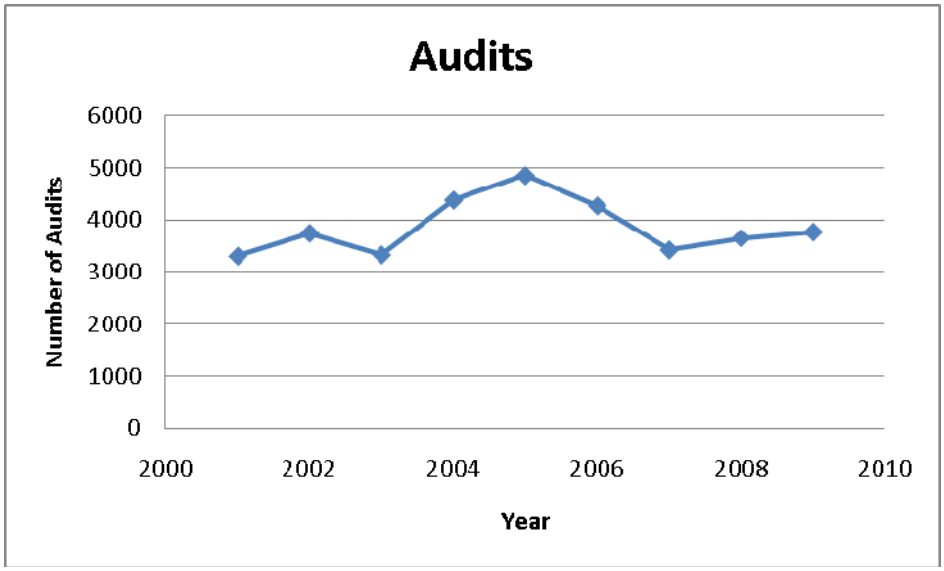
(b) There is an upward trend on the average price till 2007 and the average price started a downward trend from then on.

2.54 (a)



(b) There was a steady increase in the amount of solar power installed in the United States between 2000 and 2008

2.55 (a)



(b) There is no obvious pattern in the data.

2.56 (a)

Count of Risk		Fees		
Risk	Type	No	Yes	Grand Total
Above Average	Intermediate Government	9.44%	8.89%	18.33%
	Short Term Corporate	11.11%	2.78%	13.89%
Above Average Total		20.56%	11.67%	32.22%
Average	Intermediate Government	10.56%	5.56%	16.11%
	Short Term Corporate	12.78%	4.44%	17.22%
Average Total		23.33%	10.00%	33.33%
Below Average	Intermediate Government	10.56%	5.00%	15.56%
	Short Term Corporate	16.67%	2.22%	18.89%
Below Average Total		27.22%	7.22%	34.44%
Grand Total		71.11%	28.89%	100.00%

(b) Although the ratio of fee-yes to fee-no bond funds for intermediate government category seems to be about 2-to-3 (19% to 31%), the ratio for above average risk intermediate government bond funds is closer to 1-to-1 (8.9% to 9.4%). While the group "intermediate government funds that do not charge a fee" has nearly equal numbers of above average risk, average risk, and below risk funds, the group "short term corporate bond funds that do not charge a fee" contains about fifty percent more below average risk funds than above average ones. The pattern of risk percentages differs between the fee-yes and fee-no funds in each bond fund category.

(c) The results for type, fee, and risk, in the two years are similar.

2.57 (a)

Count of Fee		Objective		
Category	Fees	Growth	Value	Grand Total
Large Cap	No	137	114	251
	Yes	94	105	199
Large Cap Total		231	219	450
Mid Cap	No	58	39	97
	Yes	53	24	77
Mid Cap Total		111	63	174
Small Cap	No	71	81	152
	Yes	51	41	92
Small Cap Total		122	122	244
Grand Total		464	404	868

Count of Fee		Objective		
Category	Fees	Growth	Value	Grand Total
Large Cap	No	15.78%	13.13%	28.92%
	Yes	10.83%	12.10%	22.93%
Large Cap Total		26.61%	25.23%	51.84%
Mid Cap	No	6.68%	4.49%	11.18%
	Yes	6.11%	2.76%	8.87%
Mid Cap Total		12.79%	7.26%	20.05%
Small Cap	No	8.18%	9.33%	17.51%
	Yes	5.88%	4.72%	10.60%
Small Cap Total		14.06%	14.06%	28.11%
Grand Total		53.46%	46.54%	100.00%

(b) The large cap constitutes the largest percentage among all combinations of objective and fees.

2.58 (a)

Count of Risk		Fees		
Category	Risk	No	Yes	Grand Total
Large Cap	Average	95	79	174
	High	76	51	127
	Low	80	69	149
Large Cap Total		251	199	450
Mid Cap	Average	33	22	55
	High	41	45	86
	Low	23	10	33
Mid Cap Total		97	77	174
Small Cap	Average	52	30	82
	High	84	58	142
	Low	16	4	20
Small Cap Total		152	92	244
Grand Total		500	368	868

Count of Risk		Fees		
Category	Risk	No	Yes	Grand Total
Large Cap	Average	10.94%	9.10%	20.05%
	High	8.76%	5.88%	14.63%
	Low	9.22%	7.95%	17.17%
Large Cap Total		28.92%	22.93%	51.84%
Mid Cap	Average	3.80%	2.53%	6.34%
	High	4.72%	5.18%	9.91%
	Low	2.65%	1.15%	3.80%
Mid Cap Total		11.18%	8.87%	20.05%
Small Cap	Average	5.99%	3.46%	9.45%
	High	9.68%	6.68%	16.36%
	Low	1.84%	0.46%	2.30%
Small Cap Total		17.51%	10.60%	28.11%
Grand Total		57.60%	42.40%	100.00%

(b) Large cap funds without fees are fairly evenly spread in risk while large cap funds with fees are more likely to have average or low risk. Mid cap and small cap funds regardless of fees are more likely to have average or high risk.

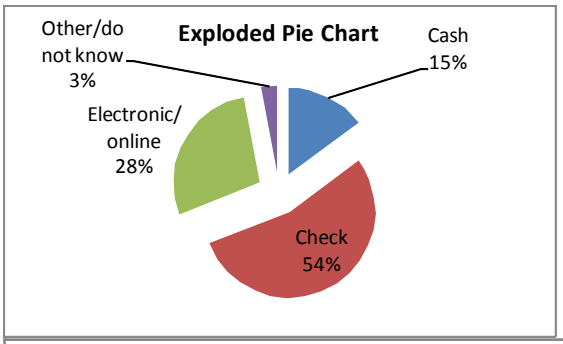
2.59 (a)

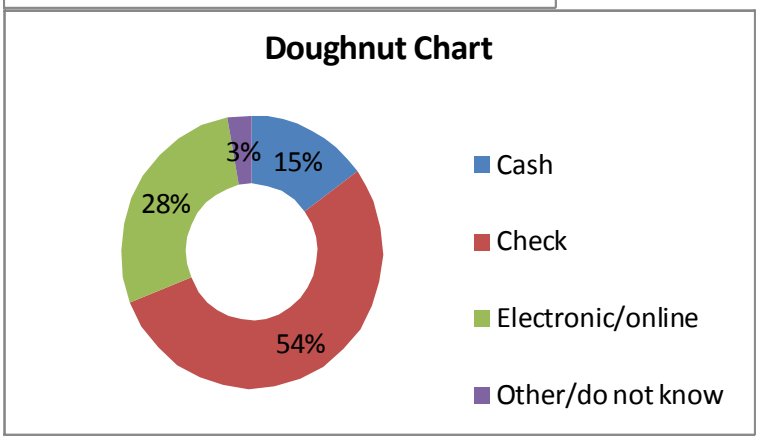
Count of Risk		Objective	Risk							
		Growth			Growth Total	Value			Value Total	Grand Total
Category	Fees	Average	High	Low		Average	High	Low		
Large Cap	No	59	68	10	137	36	8	70	114	251
	Yes	38	48	8	94	41	3	61	105	199
Large Cap Total		97	116	18	231	77	11	131	219	450
Mid Cap	No	22	36		58	11	5	23	39	97
	Yes	10	40	3	53	12	5	7	24	77
Mid Cap Total		32	76	3	111	23	10	30	63	174
Small Cap	No	9	61	1	71	43	23	15	81	152
	Yes	2	49		51	28	9	4	41	92
Small Cap Total		11	110	1	122	71	32	19	122	244
Grand Total		140	302	22	464	171	53	180	404	868

Count of Risk		Objective	Risk							
		Growth			Growth Total	Value			Value Total	Grand Total
Category	Fees	Average	High	Low		Average	High	Low		
Large Cap	No	6.80%	7.83%	1.15%	15.78%	4.15%	0.92%	8.06%	13.13%	28.92%
	Yes	4.38%	5.53%	0.92%	10.83%	4.72%	0.35%	7.03%	12.10%	22.93%
Large Cap Total		11.18%	13.36%	2.07%	26.61%	8.87%	1.27%	15.09%	25.23%	51.84%
Mid Cap	No	2.53%	4.15%	0.00%	6.68%	1.27%	0.58%	2.65%	4.49%	11.18%
	Yes	1.15%	4.61%	0.35%	6.11%	1.38%	0.58%	0.81%	2.76%	8.87%
Mid Cap Total		3.69%	8.76%	0.35%	12.79%	2.65%	1.15%	3.46%	7.26%	20.05%
Small Cap	No	1.04%	7.03%	0.12%	8.18%	4.95%	2.65%	1.73%	9.33%	17.51%
	Yes	0.23%	5.65%	0.00%	5.88%	3.23%	1.04%	0.46%	4.72%	10.60%
Small Cap Total		1.27%	12.67%	0.12%	14.06%	8.18%	3.69%	2.19%	14.06%	28.11%
Grand Total		16.13%	34.79%	2.53%	53.46%	19.70%	6.11%	20.74%	46.54%	100.00%

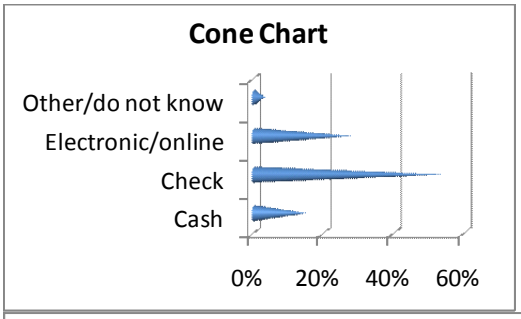
(b) The large cap constitute the largest percentage among the various combinations of fees, risk factor, and objective except the high risk, growth and fee; average risk, value and no fee; high risk, value and fee combinations that are dominated by the small cap.

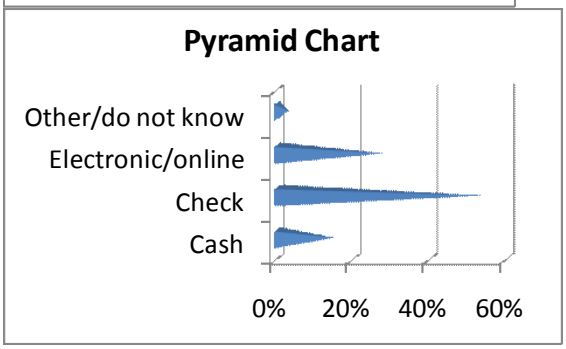
2.67 (a)





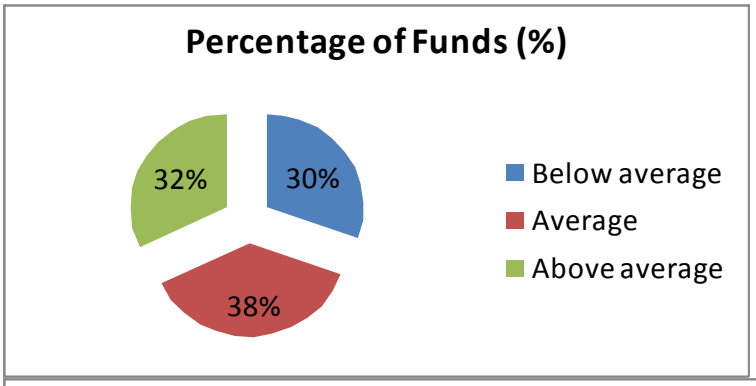
2.67 (a) cont.

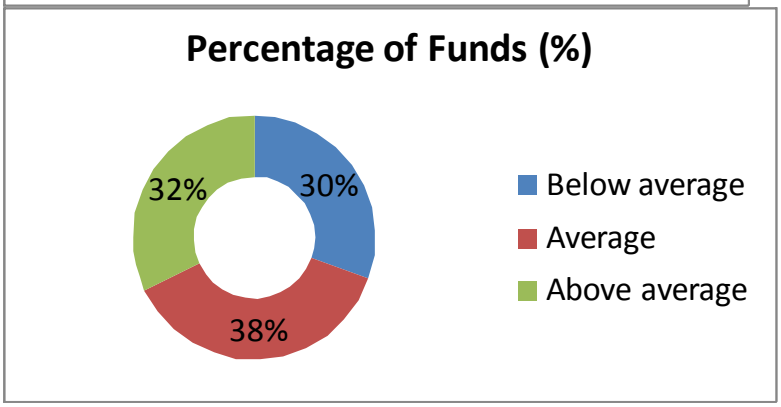




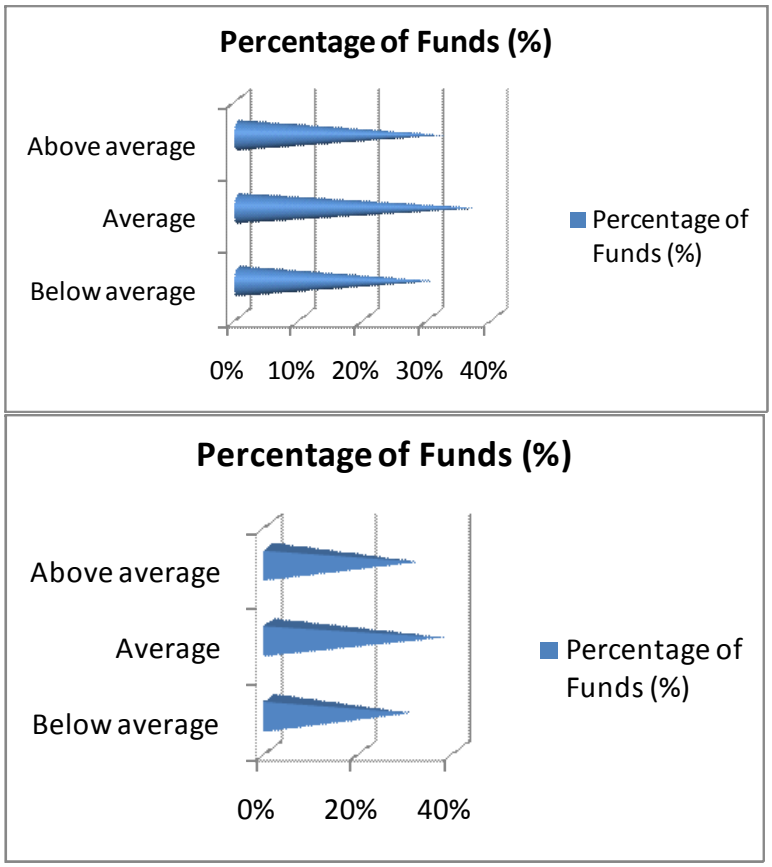
(b) The bar chart and the pie chart should be preferred over the exploded pie chart, doughnut chart, the cone chart and the pyramid chart since the former set is simpler and easier to interpret.

2.68 (a)





2.68 cont. (a)



(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.

2.69 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.

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

2.71 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.

2.72 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.

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

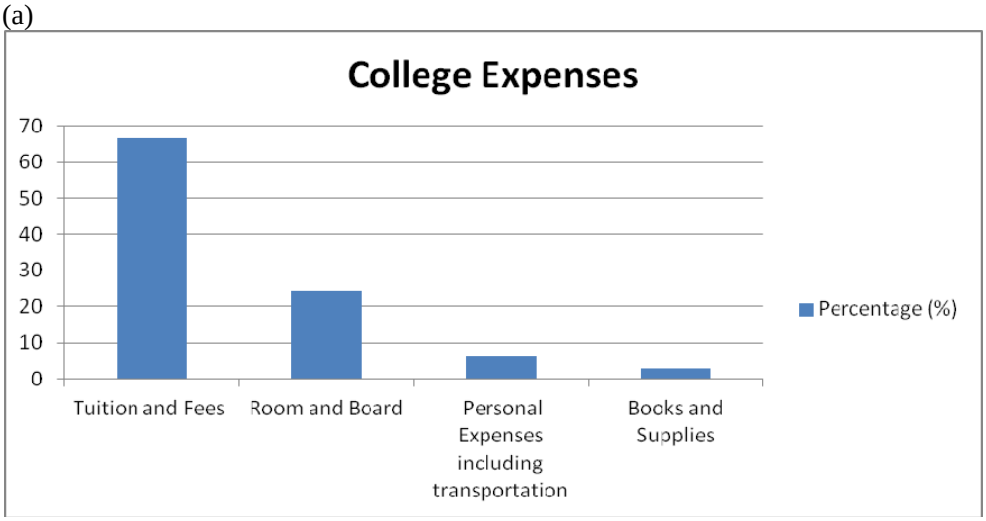
2.74 Because the categories are arranged according to frequency or importance, it allows the user to focus attention on the categories that have the greatest frequency or importance.

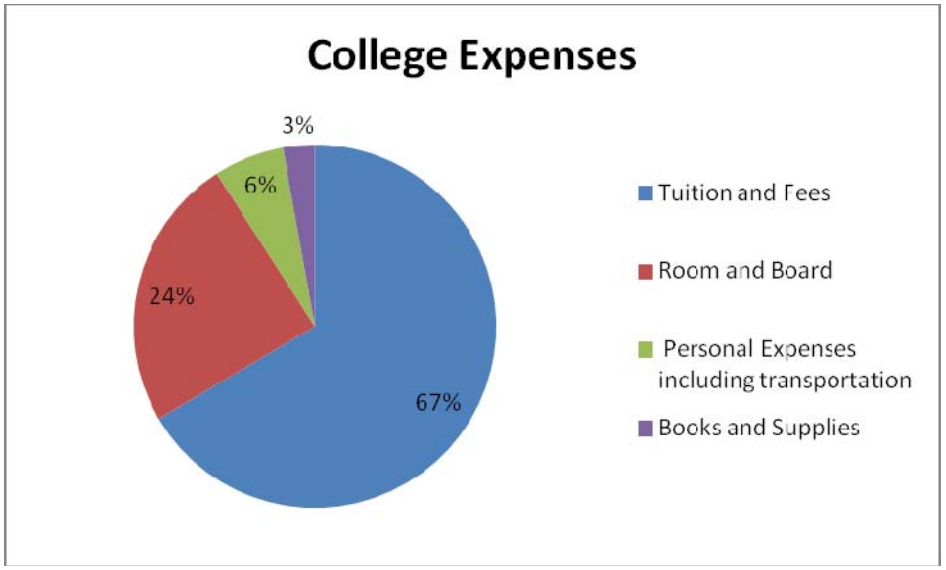
2.75 Percentage breakdowns according to the total percentage, the row percentage, and/or the column percentage allow the interpretation of data in a two-way contingency table from several different perspectives.

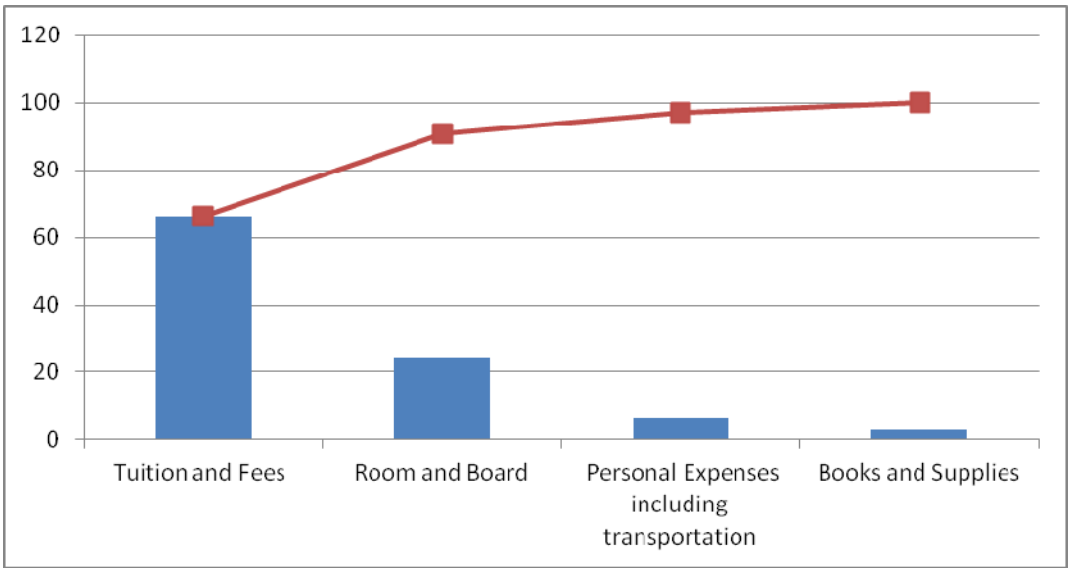
2.76 A contingency table contains information on two categorical variables whereas a multidimensional table can display information on more than two categorical variables.

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

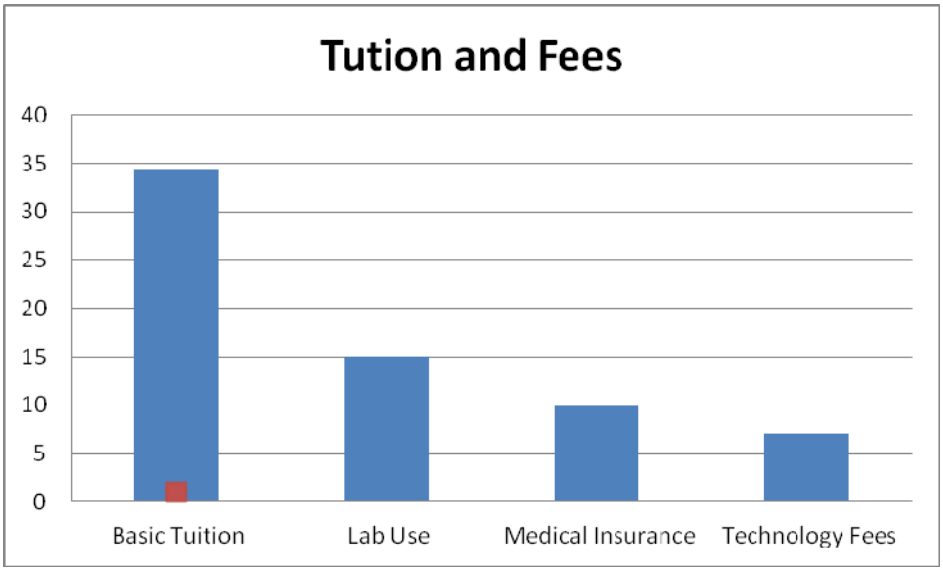
2.78 (a)

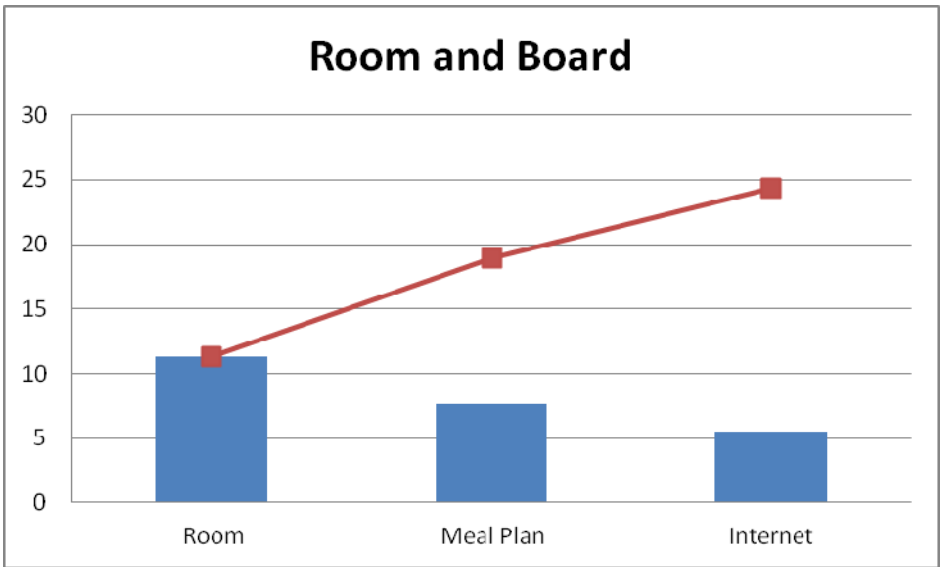






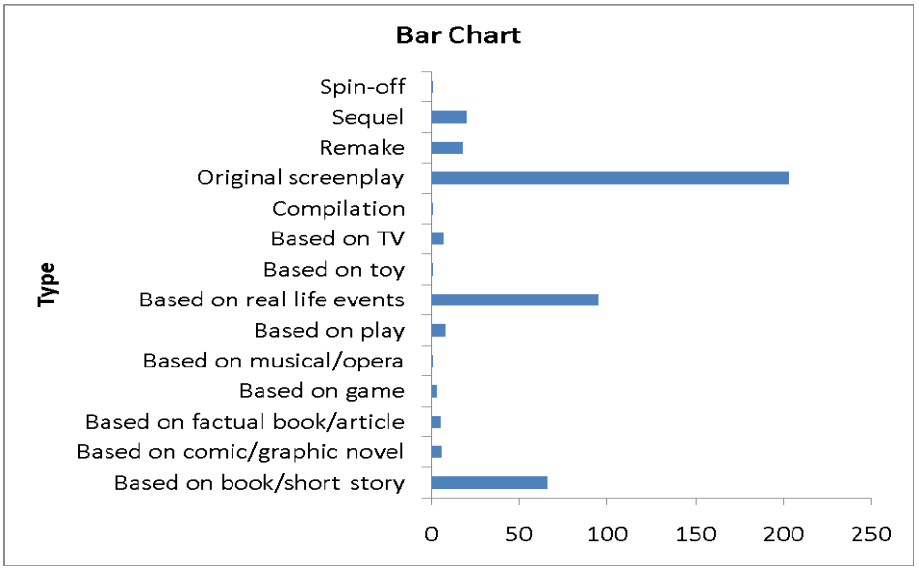
(b)

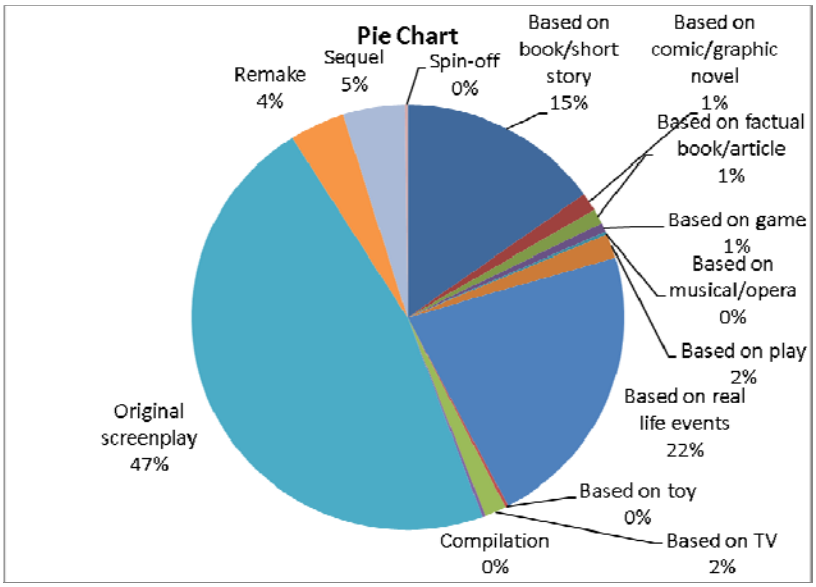




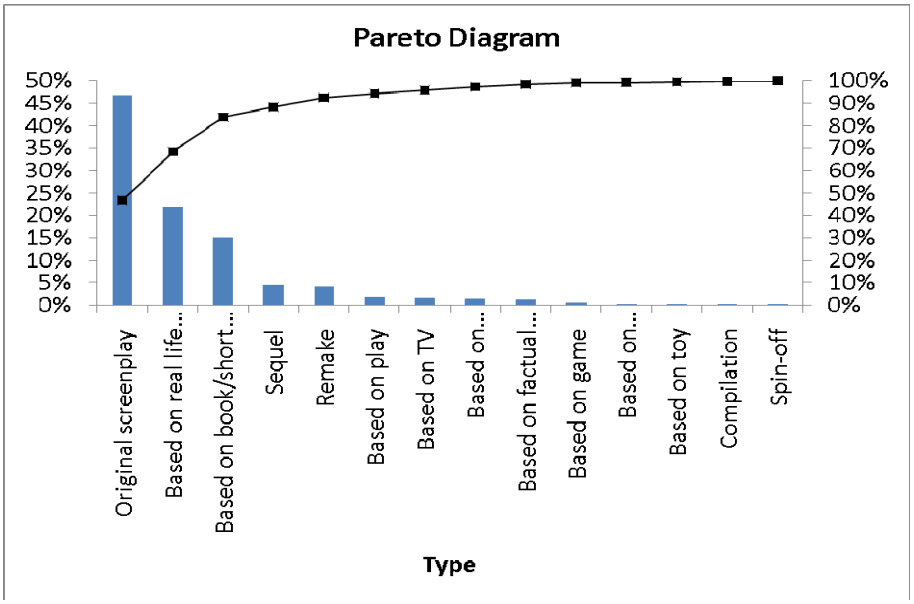
(c) The highest cost is tuitions and fees followed by room and board then followed by personal expenses and books and supplies. The tuitions and fees are divided with the highest cost for tuition followed by lab use and medical insurance. The last item in this category is for technology fees. The category of room and board is divided with the highest cost for room followed by the meal plan and internet. The results are as expected with the rising cost of education.

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

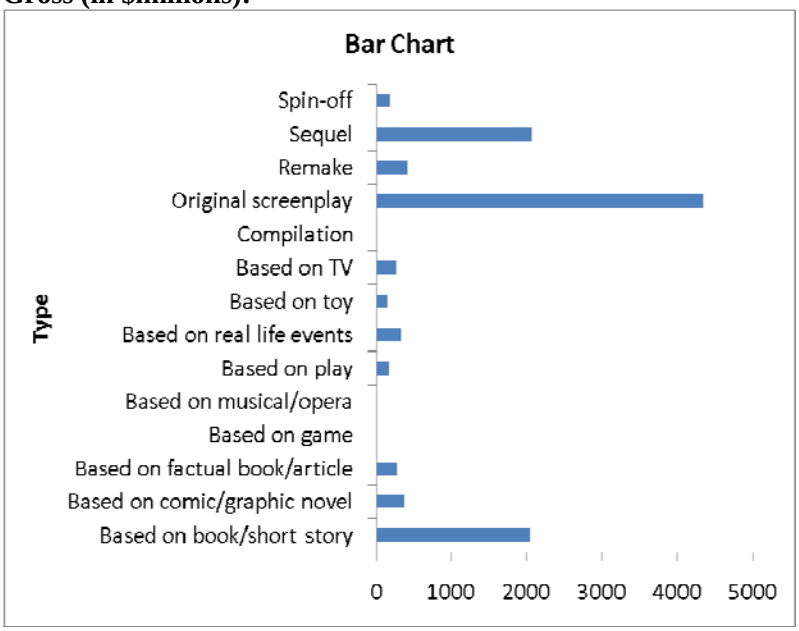




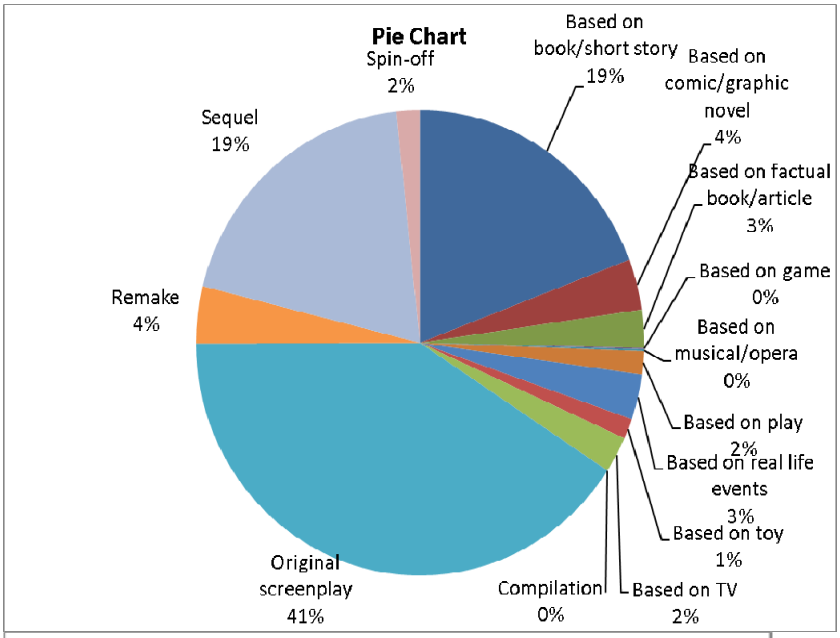
2.79 (a)
cont.

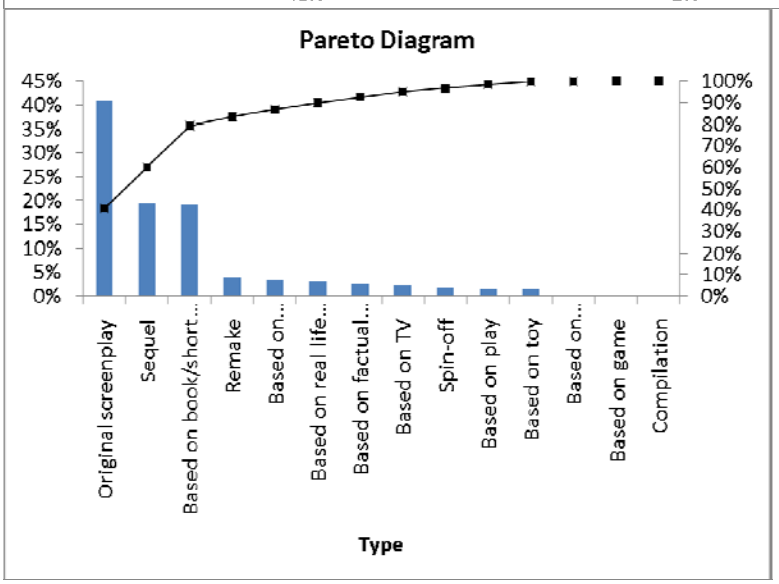


Gross (in $millions):

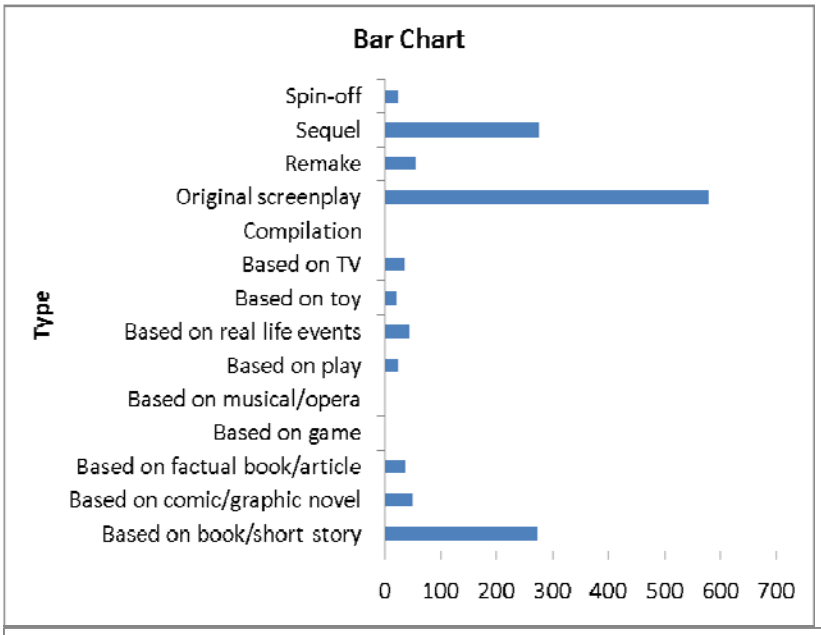


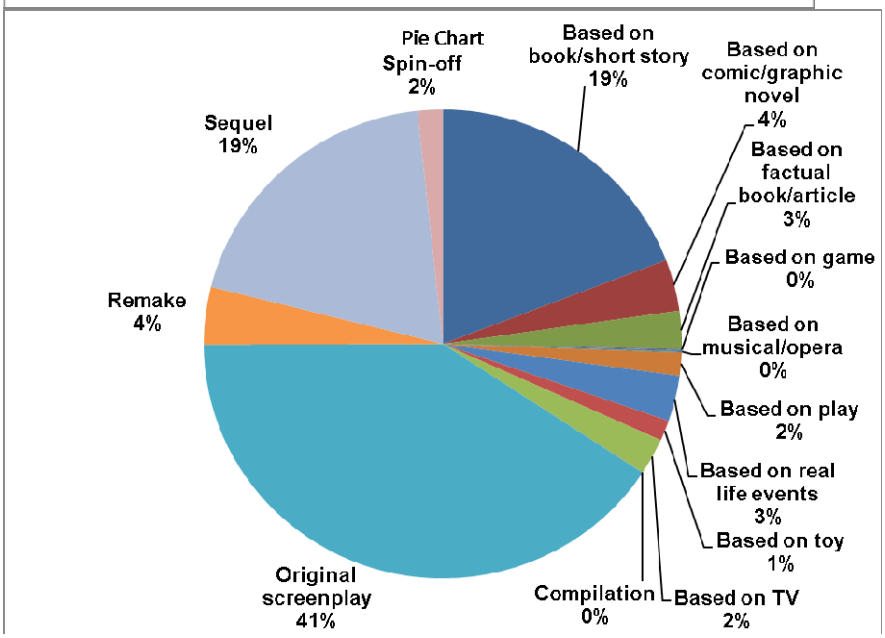
2.79 (a) cont.



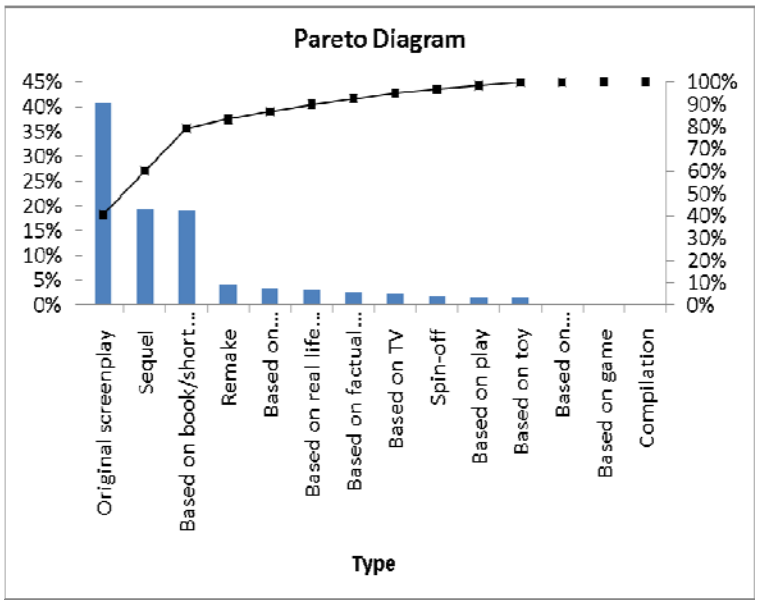


2.79 cont. (a) **Number of Tickets Sold (millions):**



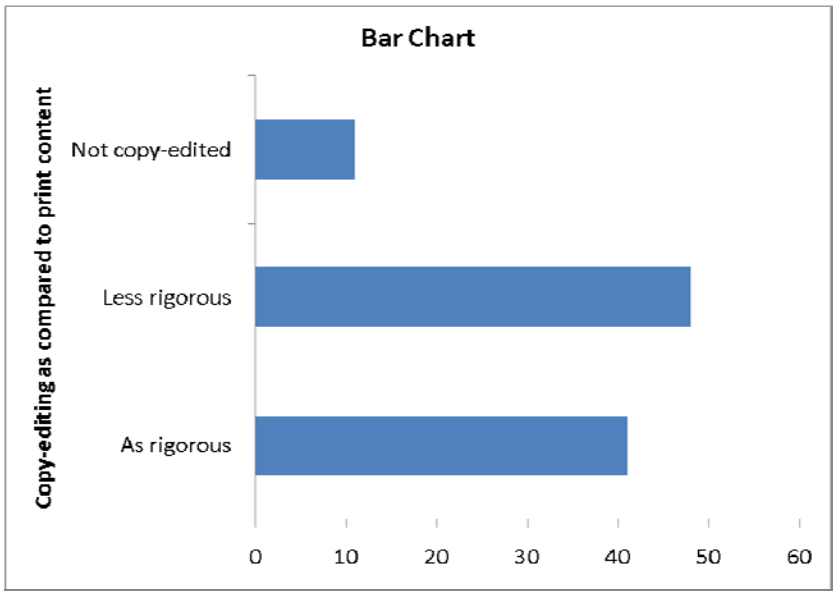


2.79 cont. (a)

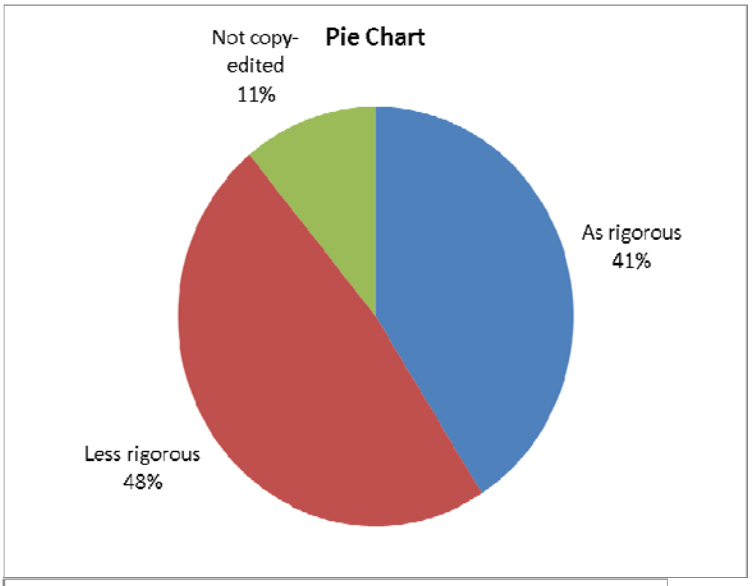


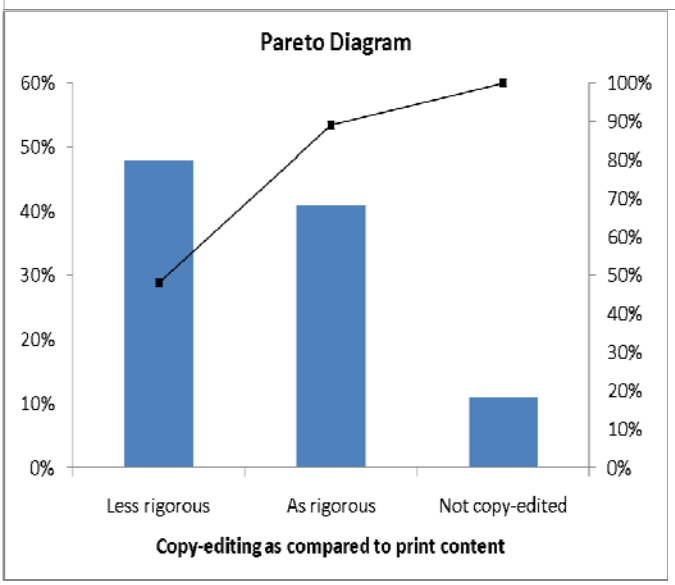
(b) Based on the Pareto chart for the number of movies, "Original screenplay", "Based on real life events" and "Based on book/short story" are the "vital few" and capture more than 80% of the market share. According to the Pareto chart for gross (in $millions) and number of ticket sold in millions, "Original screenplay", "Sequel" and "Based on book/short story" are the "vital few" and capture about 80% of the market share.

2.80 (a)



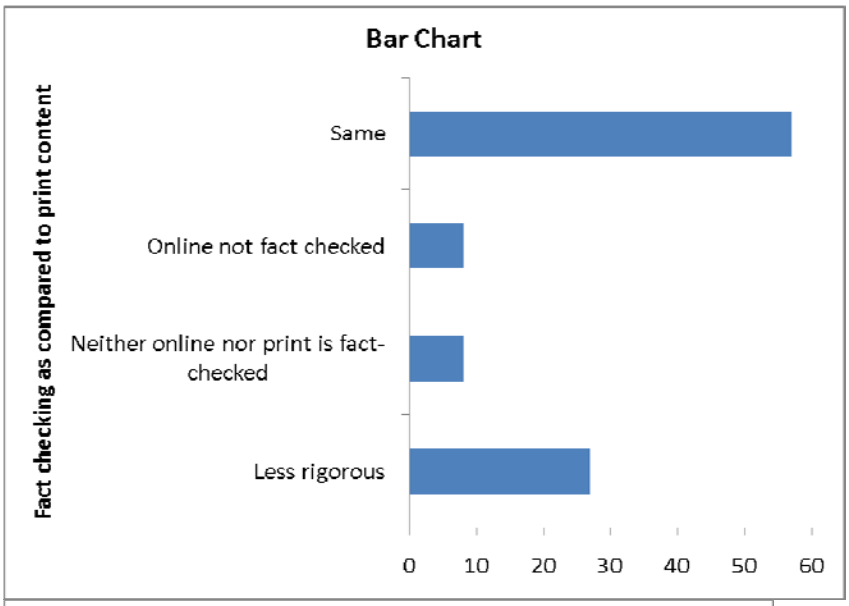
2.80 cont. (a)

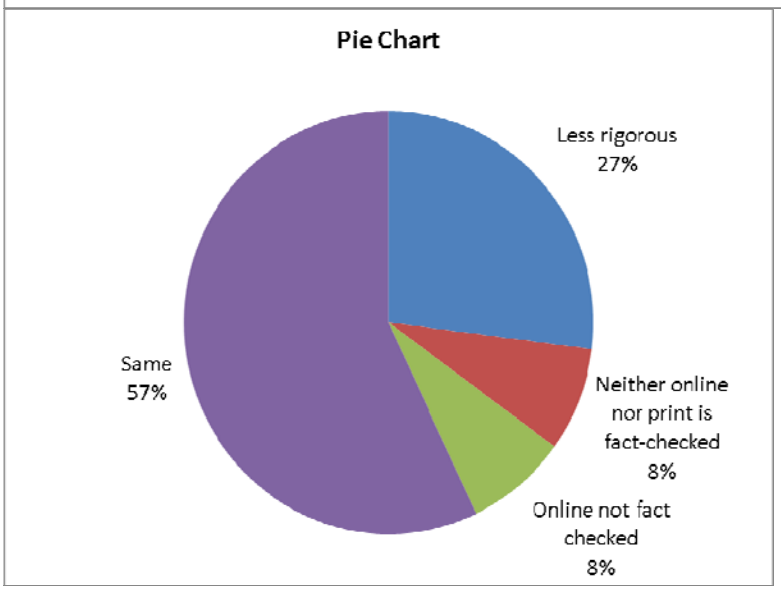




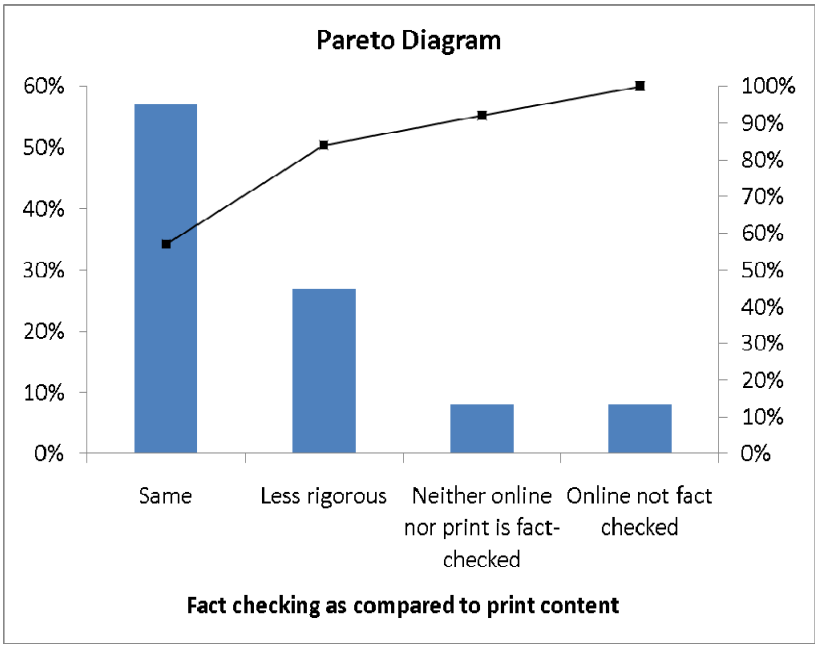
(b) Since there are only three categories, all the three graphical methods are capable of portraying these data well. The Pareto diagram, however, is better than the pie chart and bar chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

2.80 (c) cont.





2.80 (c) cont.



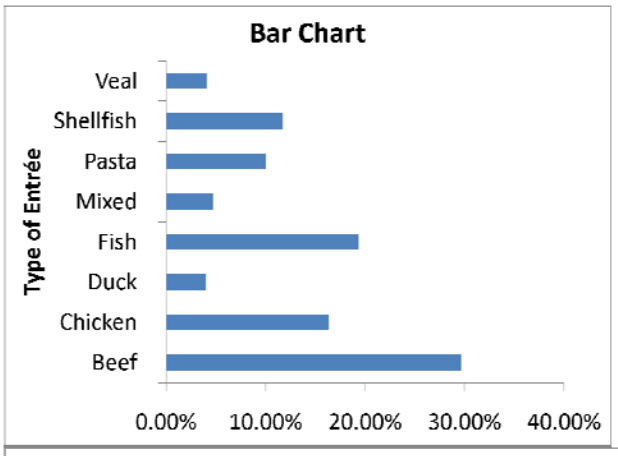
(d) Since there are only four categories, all the three graphical methods are capable of portraying these data well. The Pareto diagram, however, is better than the pie chart and bar chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

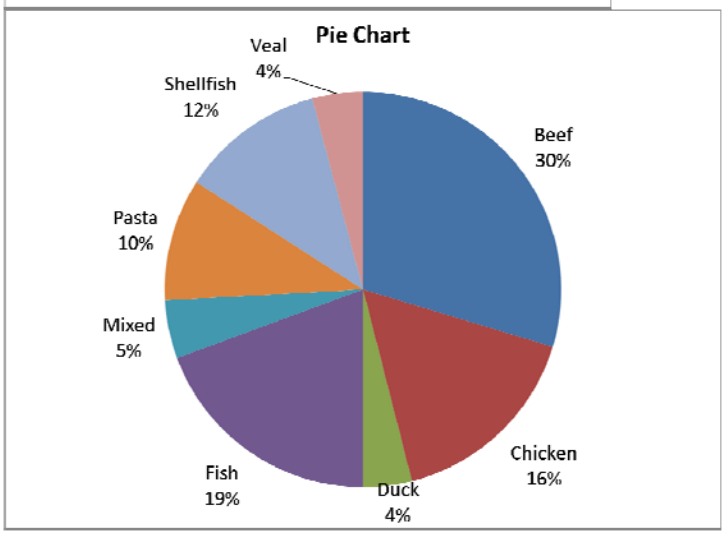
(e) Based on the Pareto chart for copy-editing, about 50% of the contents in online consumer magazines receive less rigorous copy-editing. Based on the Pareto chart for fact-checking, more than 50% of the contents in online consumer magazines receive the same amount of fact-checking.

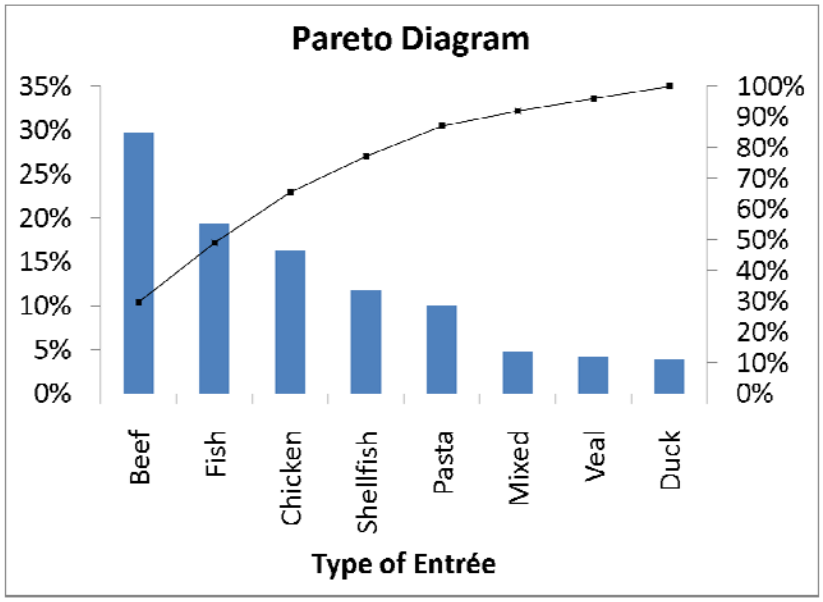
2.81 (a)

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

2.81 (b) cont.







2.81 cont. (c) The Pareto diagram has the advantage of offering the cumulative percentage view of 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.

2.82 (a)

Gender

Dessert Ordered	Male	Female	Total
Yes	71%	29%	100%
No	48%	52%	100%
Total	53%	47%	100%

Beef Entrée

Dessert Ordered	Yes	No	Total
Yes	52%	48%	100%
No	25%	75%	100%
Total	31%	69%	100%

Gender

Dessert Ordered	Male	Female	Total
Yes	30%	14%	23%
No	70%	86%	77%
Total	100%	100%	100%

Beef Entrée

Dessert Ordered	Yes	No	Total
Yes	38%	16%	23%
No	62%	84%	77%
Total	100%	100%	100%

Gender

Dessert Ordered	Male	Female	Total
Yes	16%	7%	23%
No	37%	40%	77%
Total	53%	47%	100%

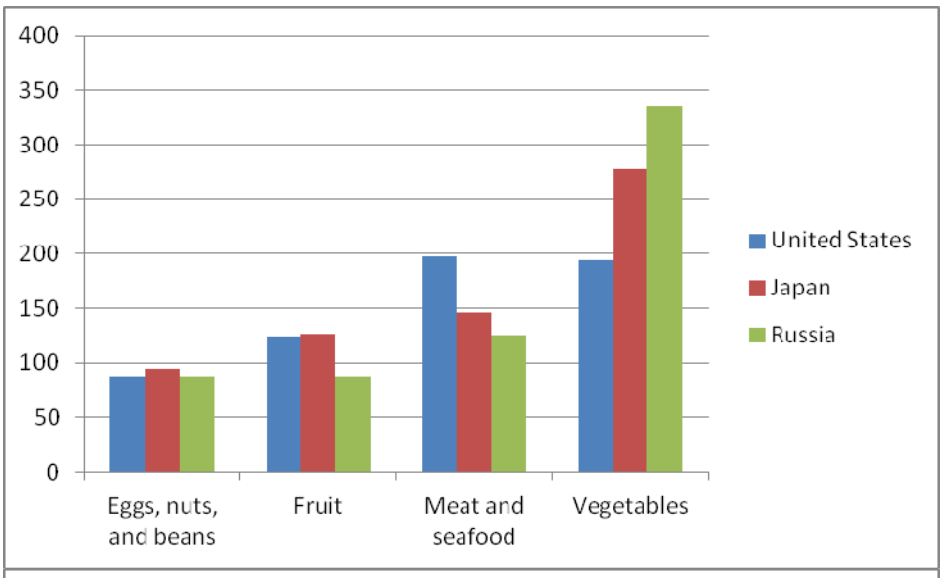
Beef Entrée

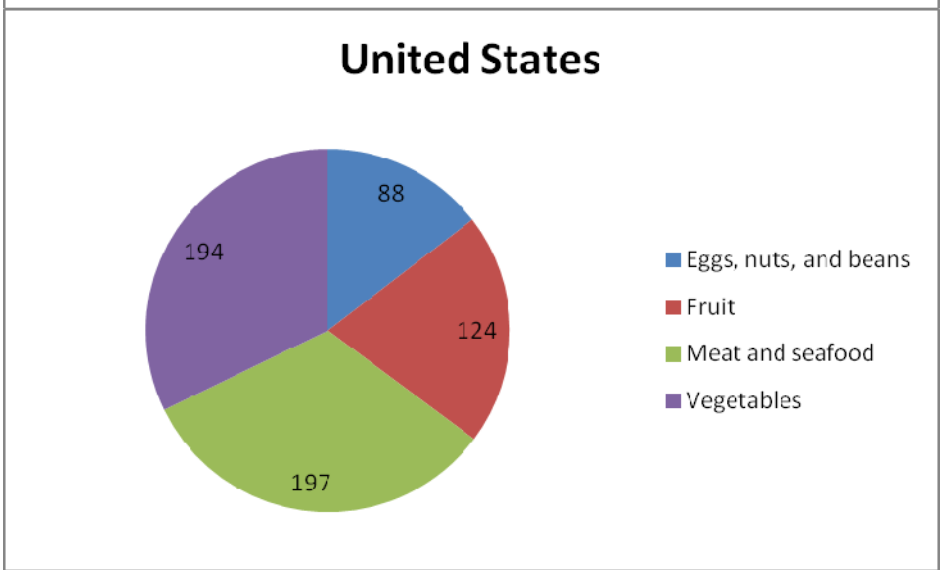
Dessert Ordered	Yes	No	Total
Yes	12%	11%	23%
No	19%	58%	77%
Total	31%	69%	100%

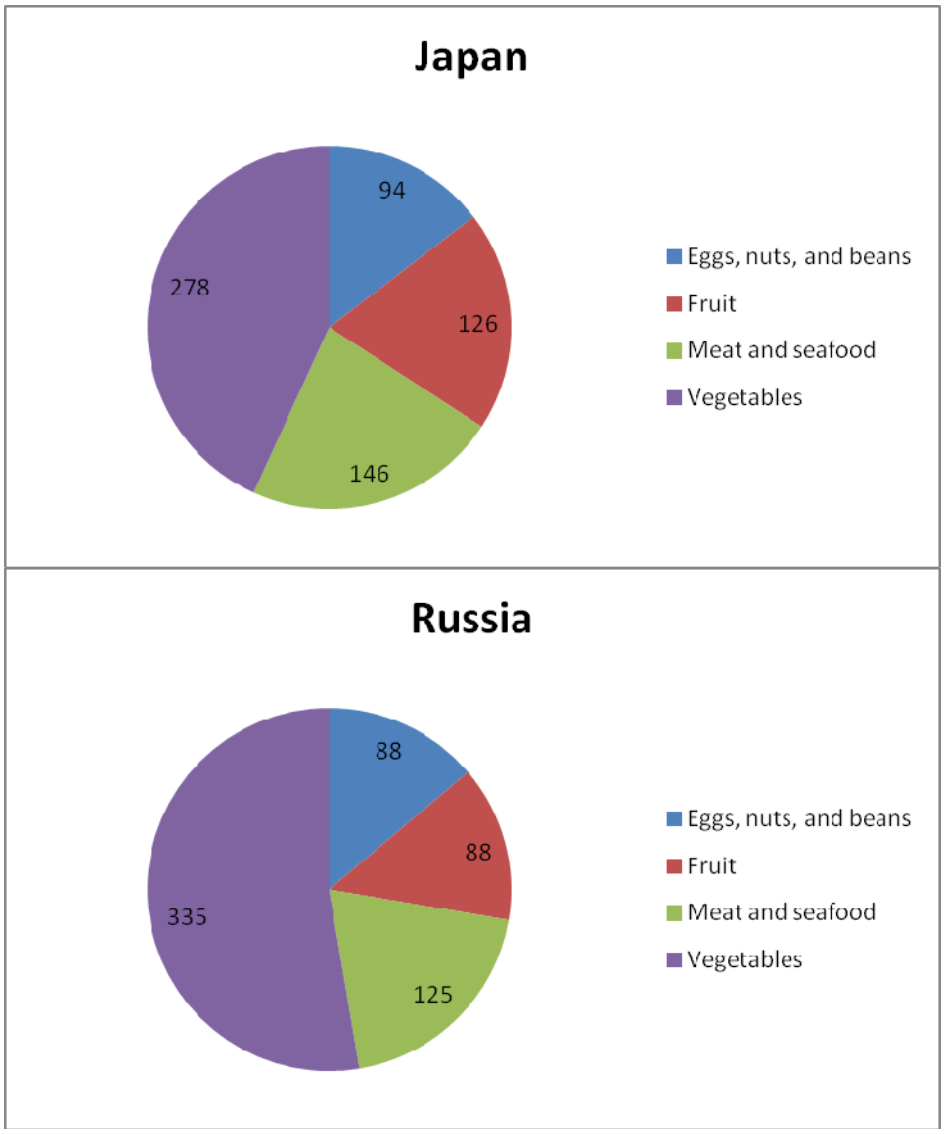
(b) If the owner is interested in finding out the percentage of joint occurrence of gender and ordering of dessert or the percentage of joint occurrence of ordering a beef entrée and a dessert among all patrons, the table of total percentages is most informative. If the owner is interested in the effect 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) 30% of the men sampled ordered desserts compared to 14% of the women. Men are more than twice as likely to order desserts as women. Almost 38% of the patrons ordering a beef entree ordered dessert compared to less than 16% 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.

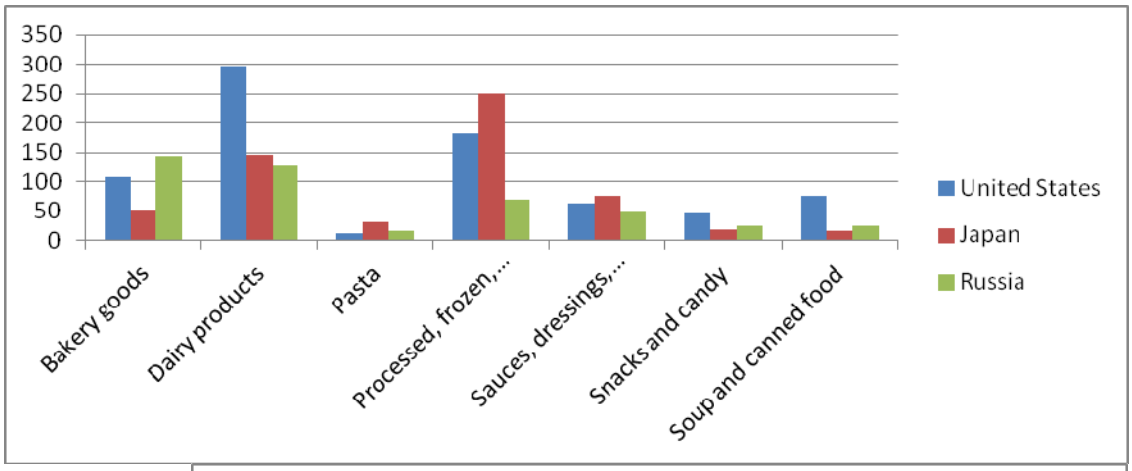
2.83 (a)

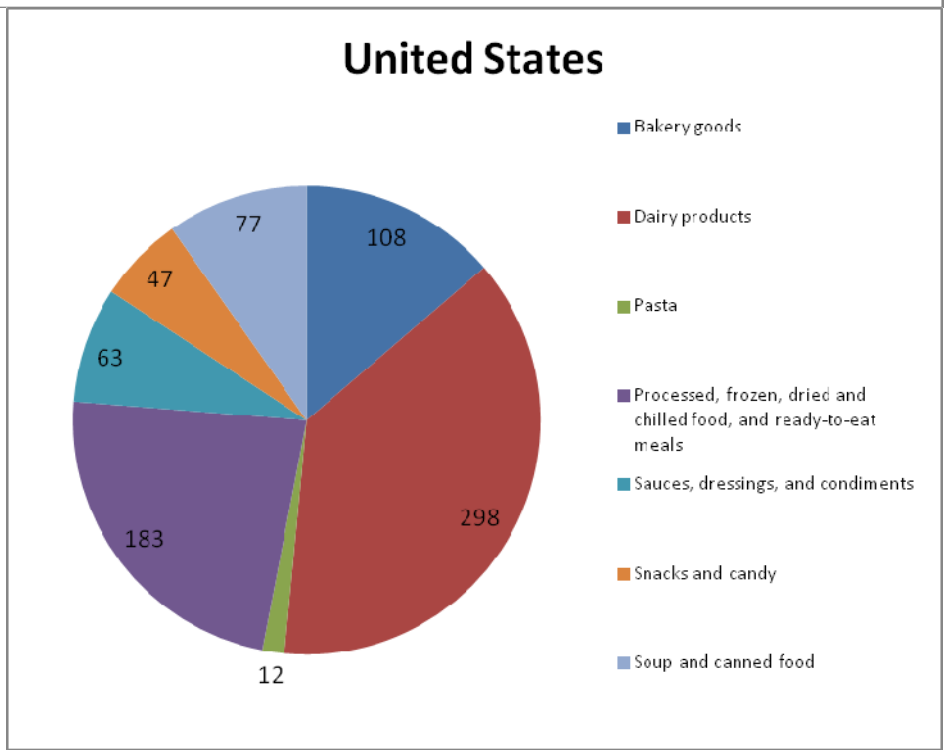


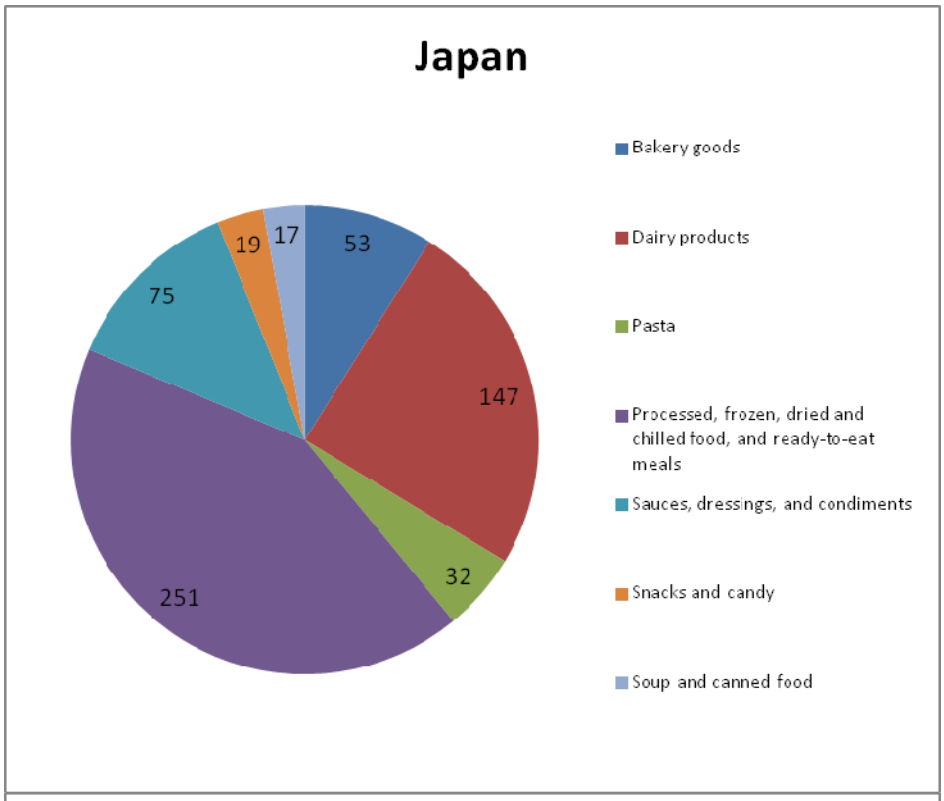


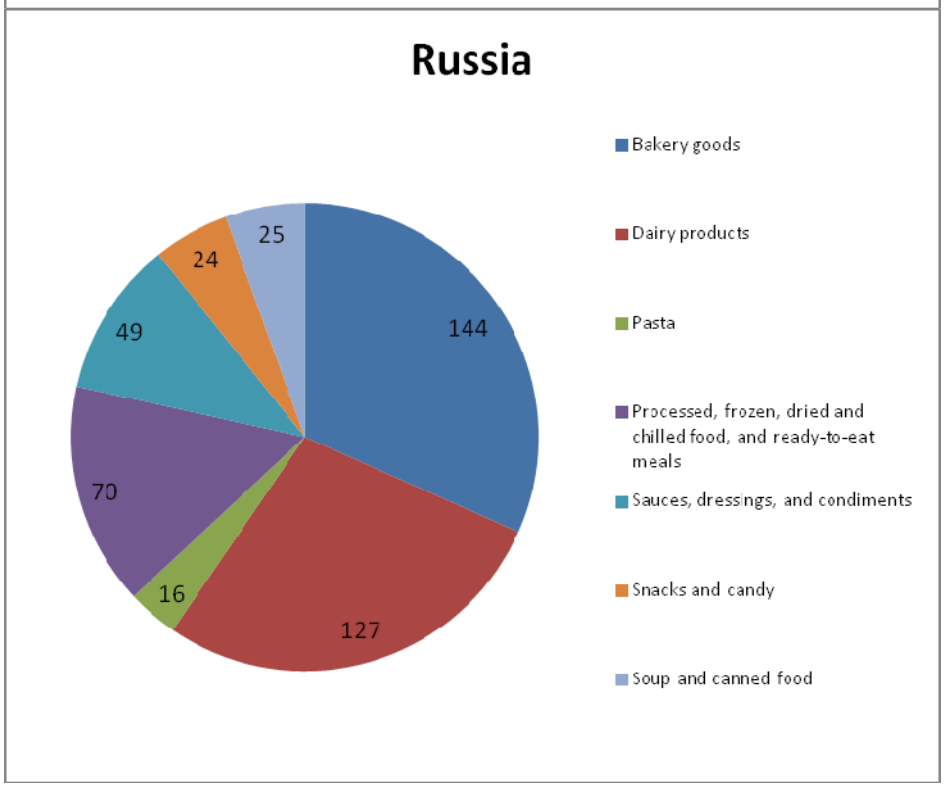


2.83 (b)









2.83 (c) Fresh food consumption per capita

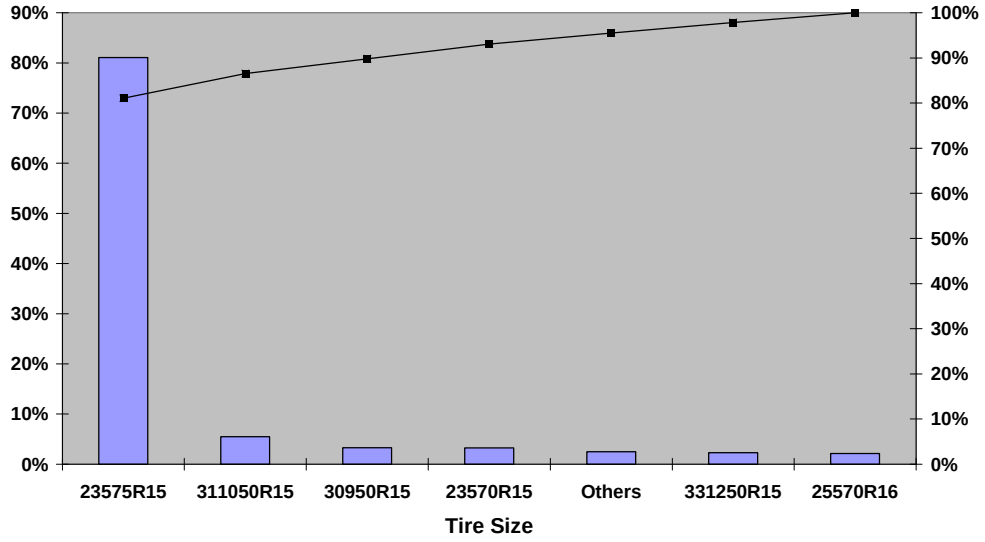
Ranking

Fresh Food	Highest Consumption	2nd Highest	3rd Highest
Vegetables	Russia	Japan	USA
Meat	USA	Japan	Russia
Fruit	USA =Japan	Russia	
Eggs, Nuts, and Beans	Almost the same for all		

Ranking

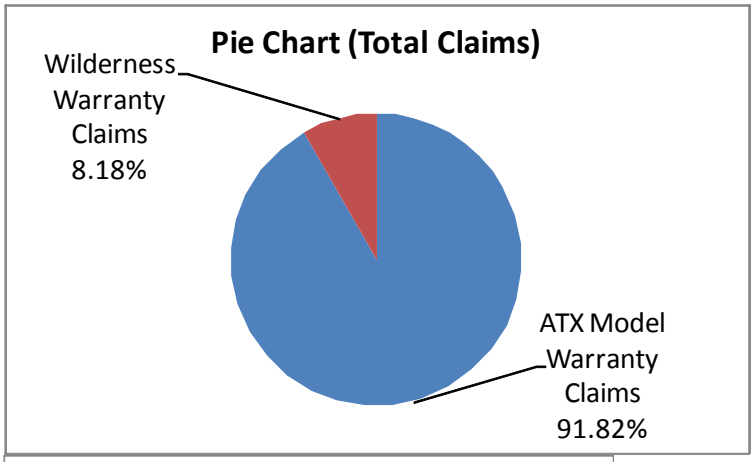
Packaged Food	Highest Consumption	2nd Highest	3rd Highest
Bakery	Russia	USA	Japan
Dairy	USA	Japan	Russia
Pasta	USA =Japan=Russia		
Processed Food	Japan	USA	Russia
Soup and Canned Food	USA	Japan = Russia	
Snacks and Candy	USA	Japan = Russia	

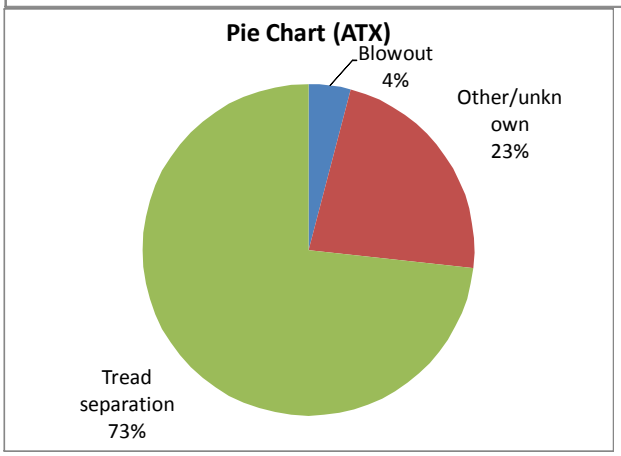
2.84 (a)



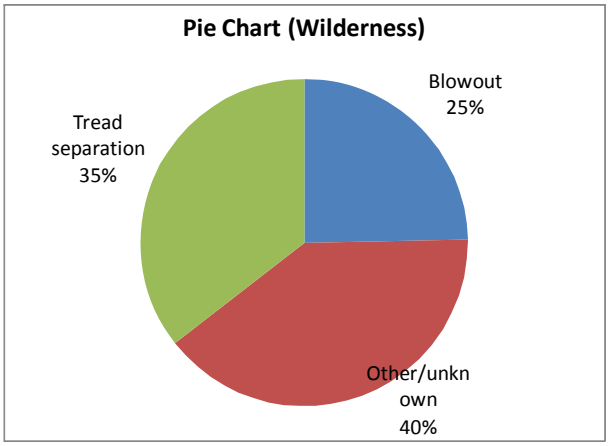
23575R15 accounts for over 80% of the warranty claims.

(b)



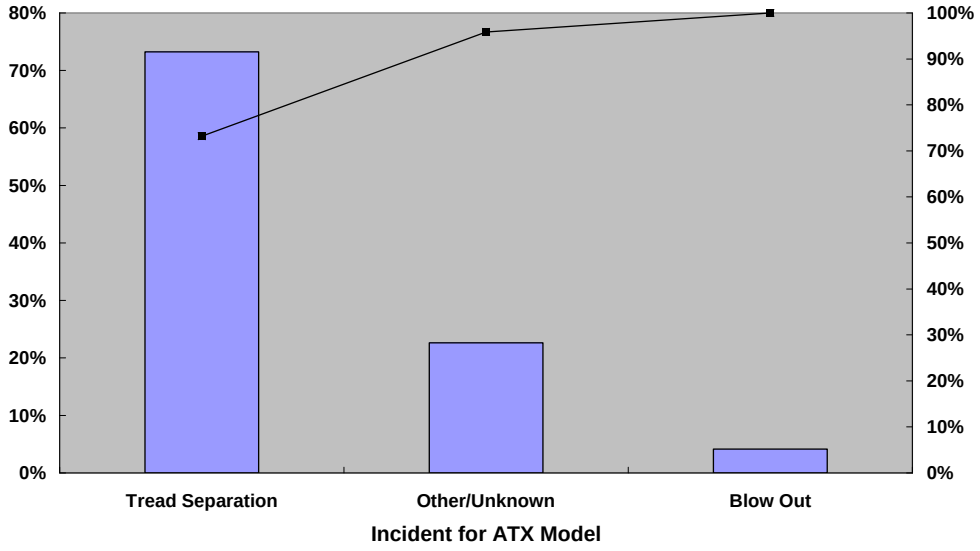


2.84 cont. (b)



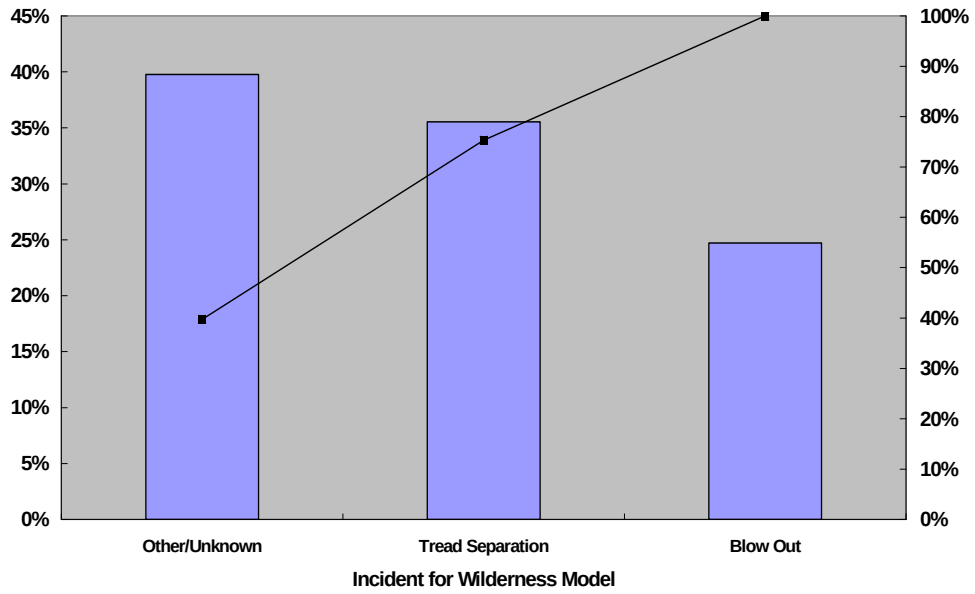
91.82% of the warranty claims are from the ATX model.

(c)



Tread separation accounts for 73.23% of the warranty claims among the ATX model..

2.84 (d)
cont.

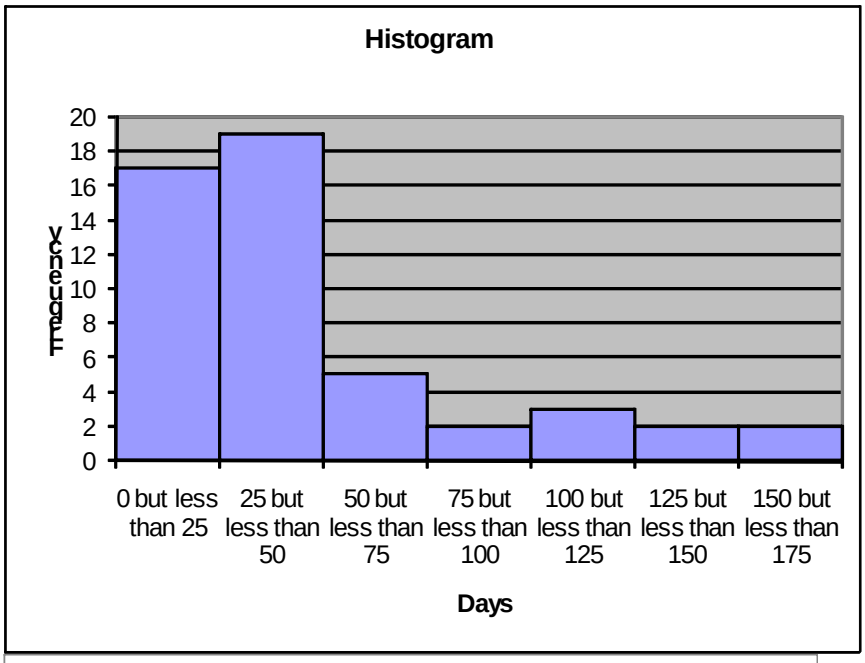


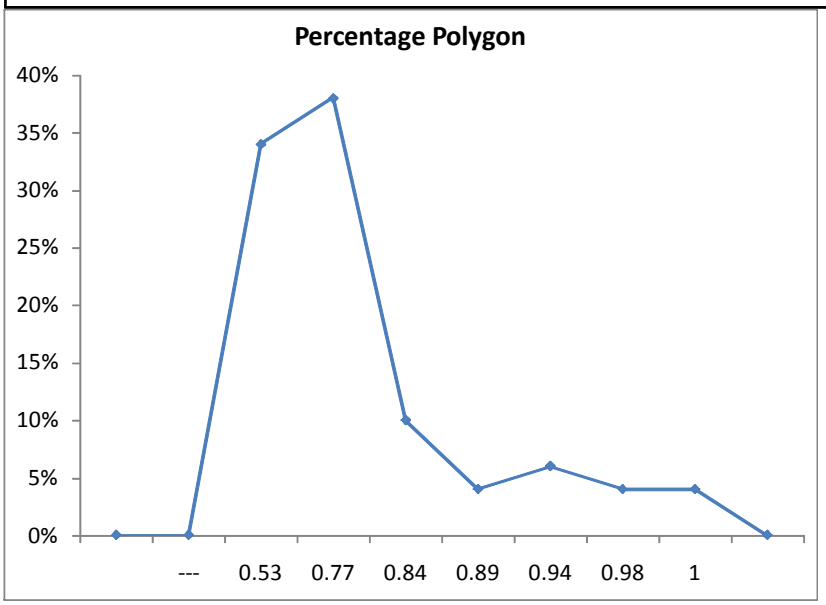
The number of claims is fairly evenly distributed among the three incidents; other/unknown incidents account for almost 40% of the claims, tread separation accounts for about 35% of the claims, and blowout accounts for about 25% of the claims.

2.85 (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%

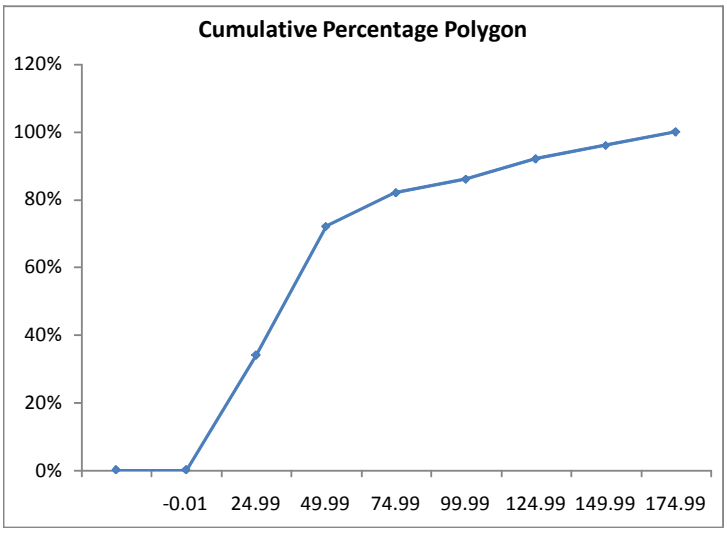
2.85 cont. (b)





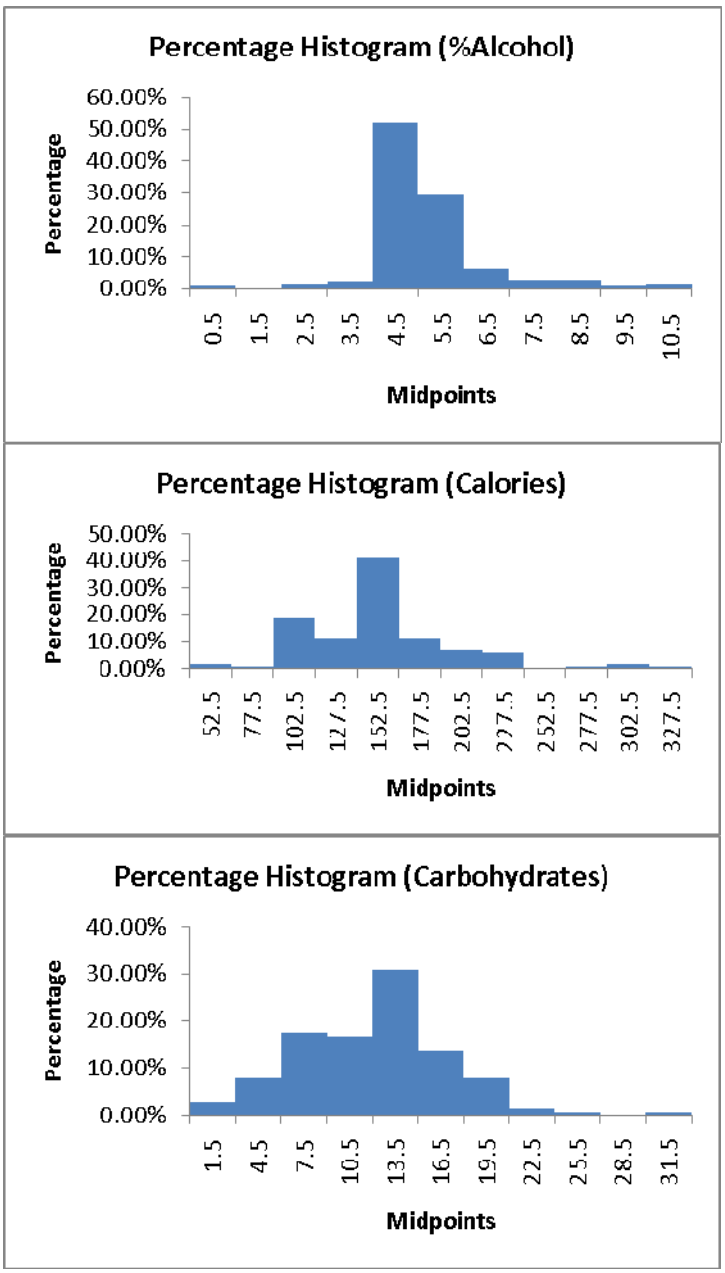
2.85 cont. (c)

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%

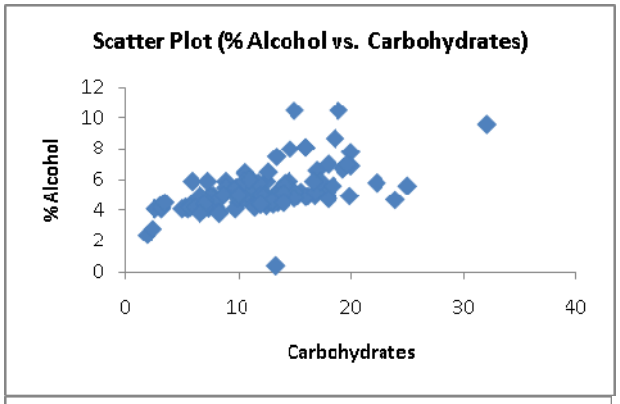


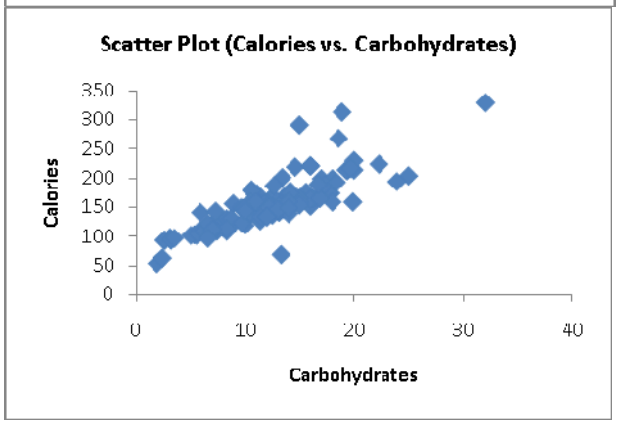
(d) You should tell the president of the company that over half of the complaints are resolved within a month, but point out that some complaints take as long as three or four months to settle.

2.86 (a)



2.86 cont. (b)





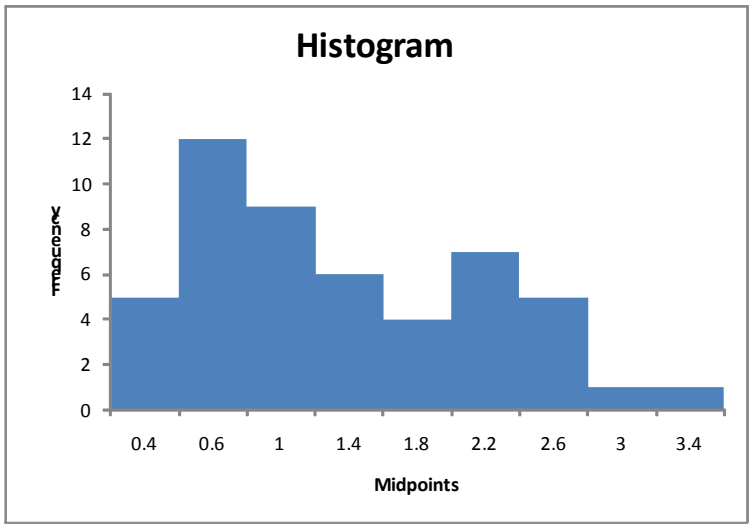
(c) The alcohol % is concentrated between 4 and 6, with more between 4 and 5. The calories are concentrated between 140 and 160. The carbohydrates are concentrated between 12 and 15. There are outliers in the percentage of alcohol in both tails. The outlier in the lower tail is due to the non-alcoholic beer O'Doul's with only a 0.4% alcohol content. There are a few beers with alcohol content as high as around 10.5%. There are a few beers with calories content as high as around 302.5 and carbohydrates as high as 31.5.

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.

2.87 (a) Ordered array:

0.070	0.170	0.300	0.360	0.370	0.425	0.440	0.450	0.550	0.570
0.600	0.600	0.620	0.640	0.680	0.695	0.790	0.800	0.840	0.870
0.910	0.980	0.995	1.030	1.150	1.180	1.230	1.250	1.339	1.360
1.410	1.530	1.600	1.600	1.700	1.780	2.000	2.000	2.000	2.000
2.000	2.025	2.240	2.510	2.520	2.600	2.700	2.750	3.000	3.460

(b)



(c) There is a 3.39% difference in the state cigarette tax between the lowest and highest. The distribution of the cigarette tax is somewhat right-skewed with a few states having a cigarette tax as high as around 2.8% to 3.6%. Majority of the states though have cigarette tax concentrated around 0.8%.

2.88 (a) Money market:

Statistics	
Sample Size	25
Mean	0.9948
Median	1
Std. Deviation	0.322892
Minimum	0.2
Maximum	1.4

Stem-and-Leaf Display

Stem unit: 0.1

2	0
3	
4	0
5	0
6	5 5
7	5
8	5
9	0 5 5 5
10	0 0
11	0 1
12	0 1 5 6
	9
13	0 0 5 5
14	0

2.88 cont. (a) 5-year CD

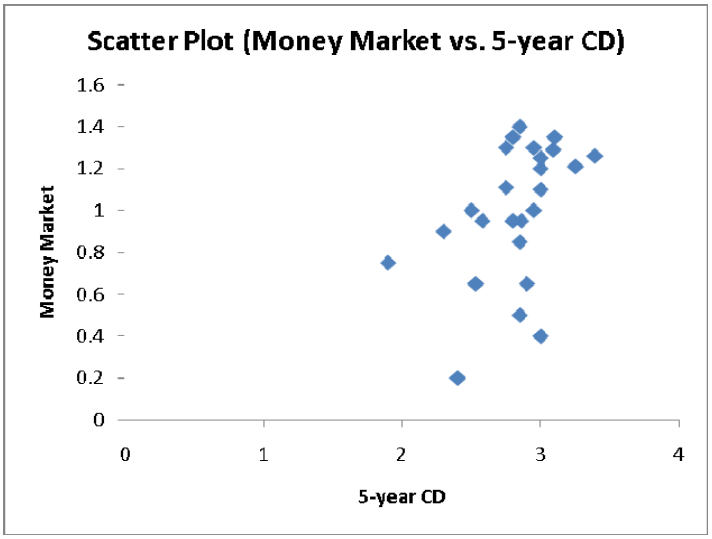
Statistics	
Sample Size	25
Mean	2.814
Median	2.85
Std. Deviation	0.315053
Minimum	1.9
Maximum	3.39

Stem-and-Leaf Display

Stem unit: 0.1

19	0
20	
21	
22	
23	0
24	0
25	0 3 8
26	
27	5 5
28	0 0 5 5 5 6
29	0 5 5
30	0 0 0 0 9
31	0
32	5
33	9

(b)

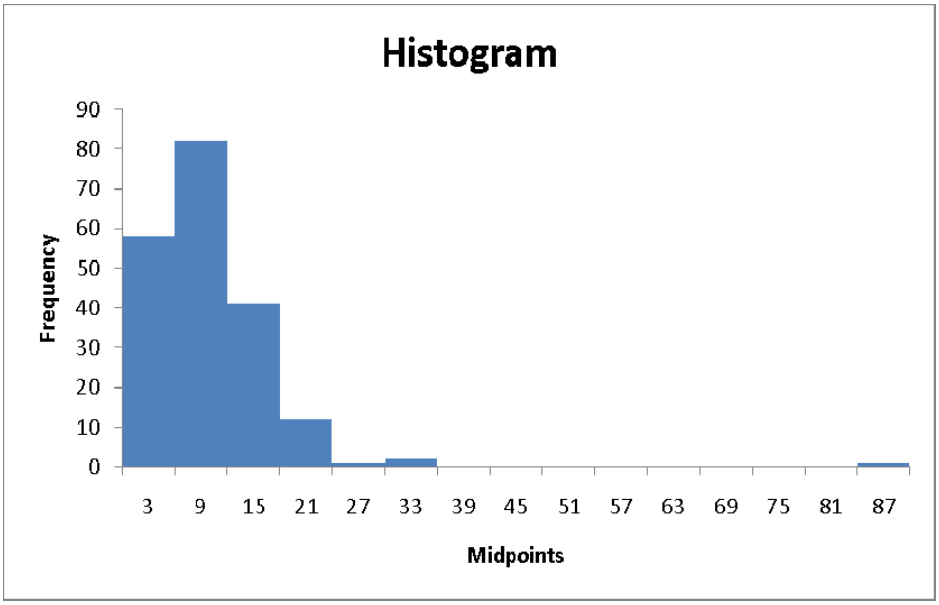


(c) The money market yield is concentrated between 0.95 and 1.35. The five-year CD is concentrated between 2.8 and 3.1. In general, the five-year CD has the higher yield. There appears to be a positive relationship between the yield of the money market and the five-year CD.

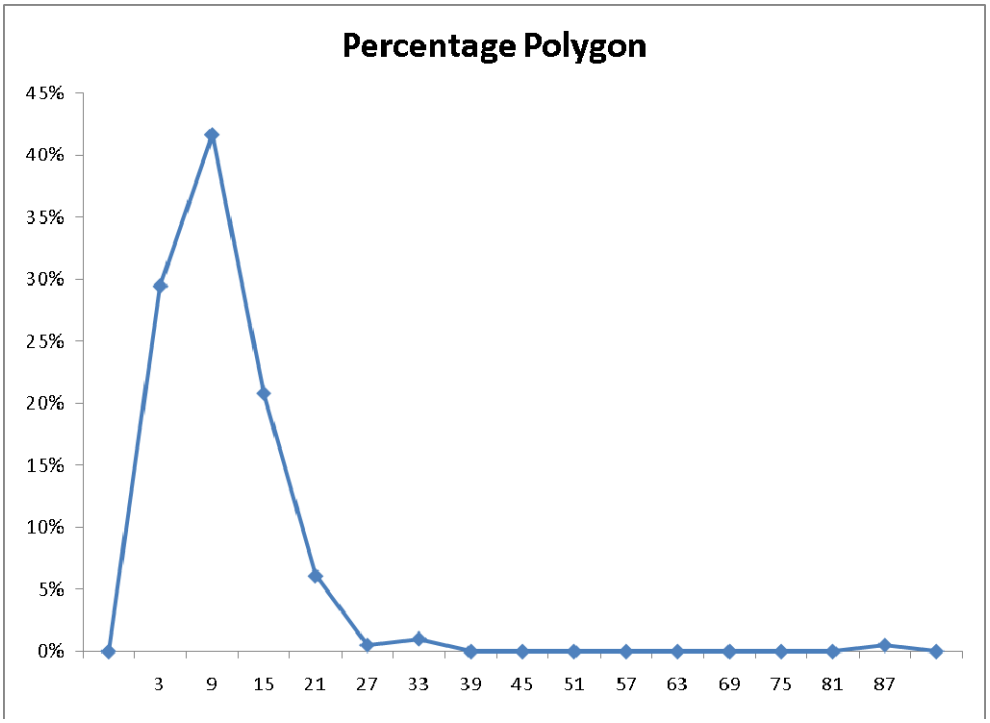
2.89 (a),(c)

bin	*Frequency*	*Percentage*	*Cumulative %*	*Midpts*
0 but less than 6	58	29.44%	29.44%	*3*
6 but less than 12	82	41.62%	71.07%	9
12 but less than 18	41	20.81%	91.88%	15
18 but less than 24	12	6.09%	97.97%	21
24 but less than 30	1	0.51%	98.48%	27
30 but less than 36	2	1.02%	99.49%	33
36 but less than 42	0	0.00%	99.49%	39
42 but less than 48	0	0.00%	99.49%	45
48 but less than 54	0	0.00%	99.49%	51
54 but less than 60	0	0.00%	99.49%	57
60 but less than 66	0	0.00%	99.49%	63
66 but less than 72	0	0.00%	99.49%	69
72 but less than 78	0	0.00%	99.49%	75
78 but less than 84	0	0.00%	99.49%	81
84 but less than 90	1	0.51%	100.00%	87

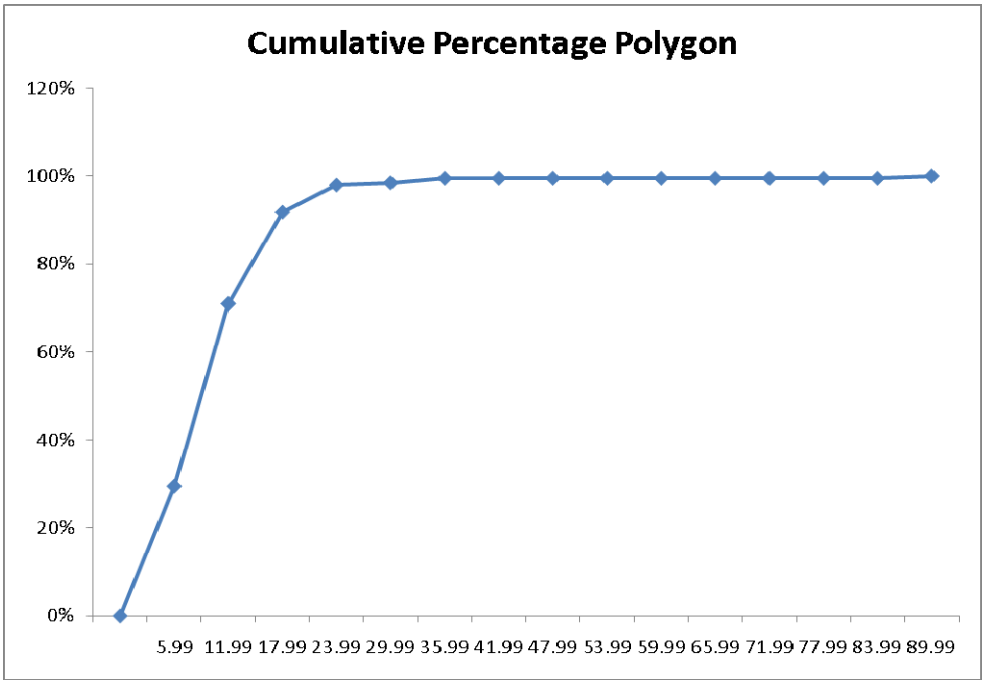
(b)



2.89 cont. (b)

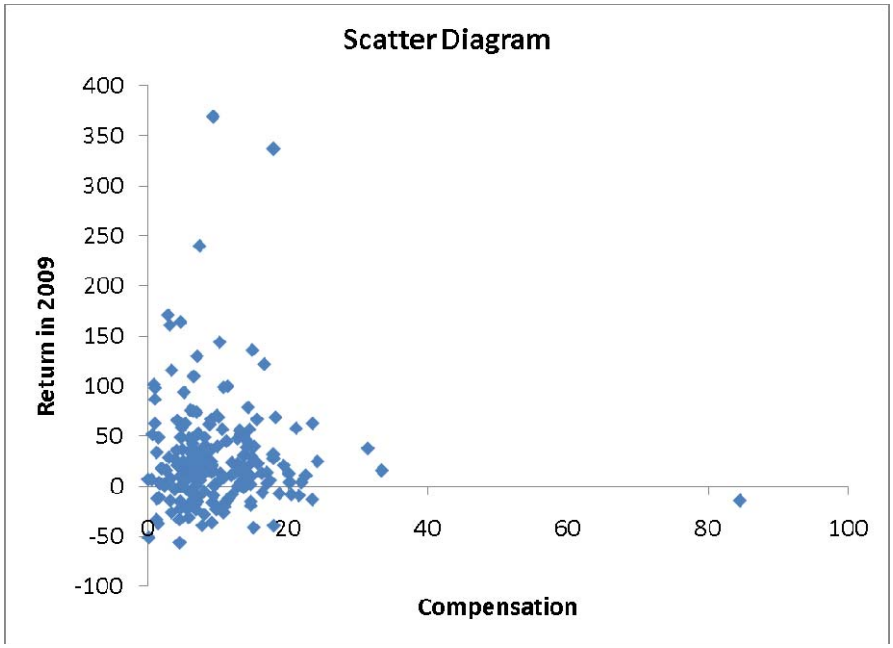


(c)



(d) CEO compensation in 2009 is extremely right skewed. More than 90% of the CEOs have compensation lower than $18,000,000. On the other end, 0.51% of the CEOs have compensation higher than $84,000,000.

2.89 cont. (e)



(f) There is not any obvious relationship between the total compensation and investment return in 2009.

2.90 (a)

Frequencies (Boston)

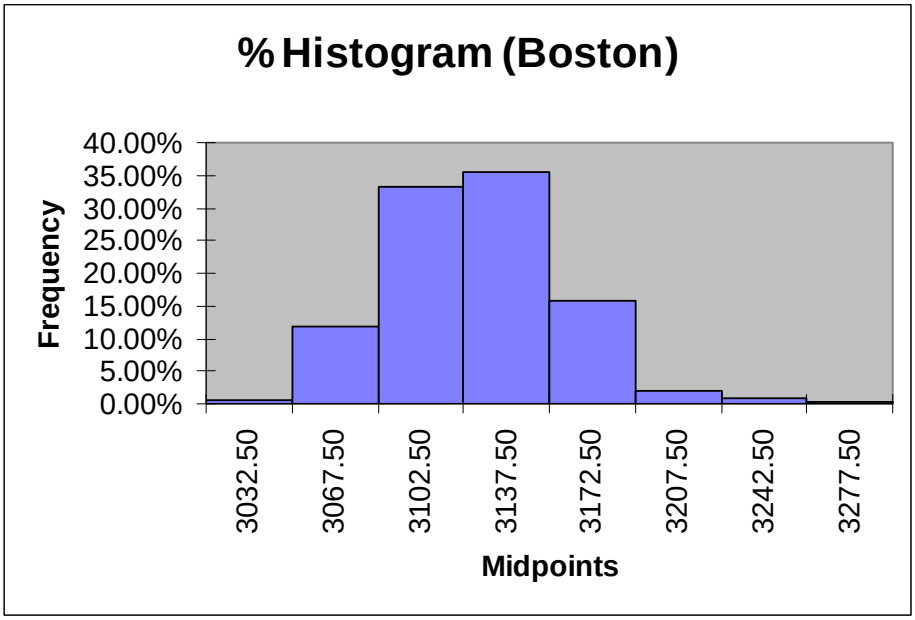
Weight (Boston)	*Frequency*	*Percentage*
3015 but less than 3050	2	0.54%
3050 but less than 3085	44	11.96%
3085 but less than 3120	122	33.15%
3120 but less than 3155	131	35.60%
3155 but less than 3190	58	15.76%
3190 but less than 3225	7	1.90%
3225 but less than 3260	3	0.82%
3260 but less than 3295	1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

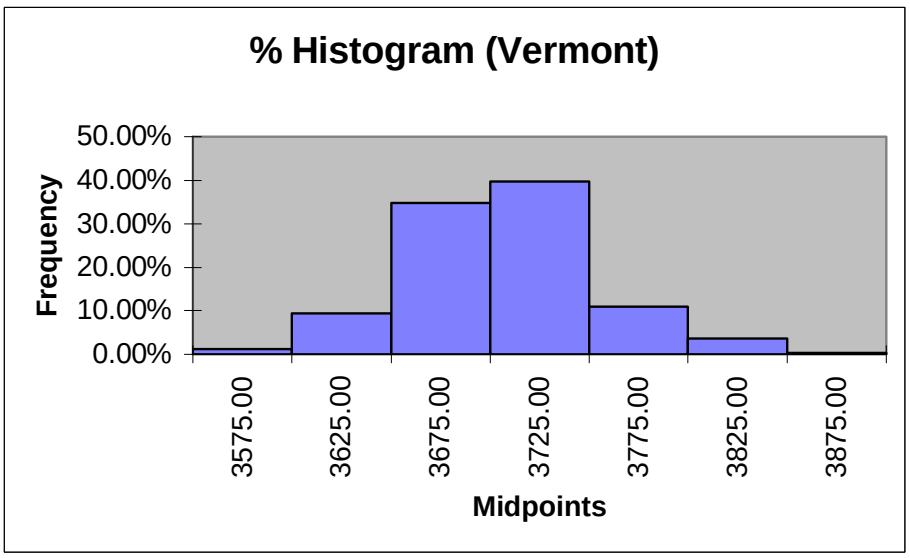
(b)

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%

2.90 cont. (c)





(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.

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

bin	*Frequency*	*Percentage*	*Cumulative %*	*Midpts*
20 but less than 30	3	6.38%	6.38%	*25*
30 but less than 40	6	12.77%	19.15%	35
40 but less than 50	3	6.38%	25.53%	45
50 but less than 60	8	17.02%	42.55%	55
60 but less than 70	15	31.91%	74.47%	65
70 but less than 80	5	10.64%	85.11%	75
80 but less than 90	4	8.51%	93.62%	85
90 but less than 100	1	2.13%	95.74%	95
100 but less than 110	0	0.00%	95.74%	105
110 but less than 120	1	2.13%	97.87%	115
120 but less than 130	0	0.00%	97.87%	125
130 but less than 140	1	2.13%	100.00%	135

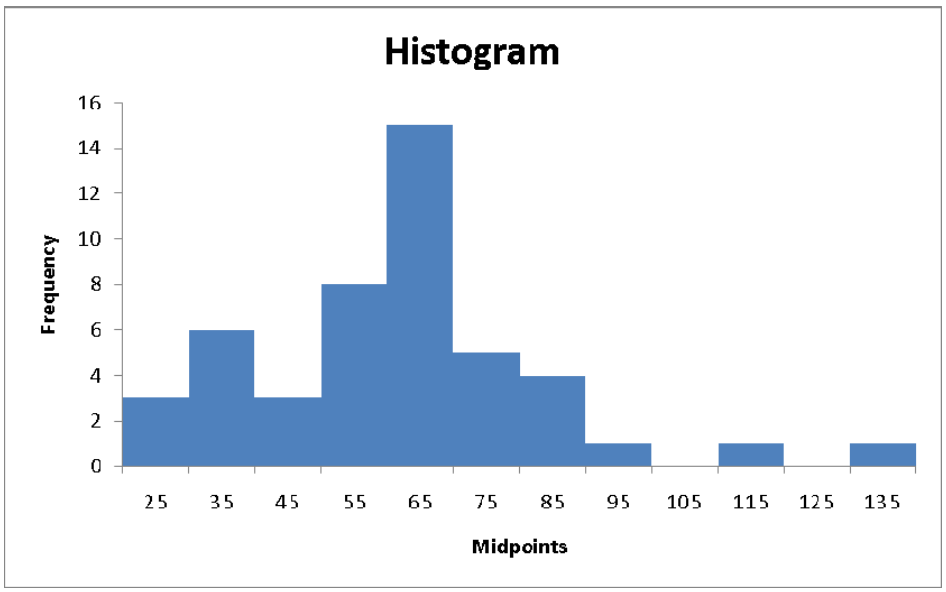
Three-star:

bin	*Frequency*	*Percentage*	*Cumulative %*	*Midpts*
30 but less than 40	2	4.26%	4.26%	*35*
40 but less than 50	4	8.51%	12.77%	45
50 but less than 60	4	8.51%	21.28%	55
60 but less than 70	7	14.89%	36.17%	65
70 but less than 80	5	10.64%	46.81%	75
80 but less than 90	11	23.40%	70.21%	85
90 but less than 100	5	10.64%	80.85%	95
100 but less than 110	5	10.64%	91.49%	105
110 but less than 120	2	4.26%	95.74%	115
120 but less than 130	1	2.13%	97.87%	125
130 but less than 140	0	0.00%	97.87%	135
140 but less than 150	1	2.13%	100.00%	145

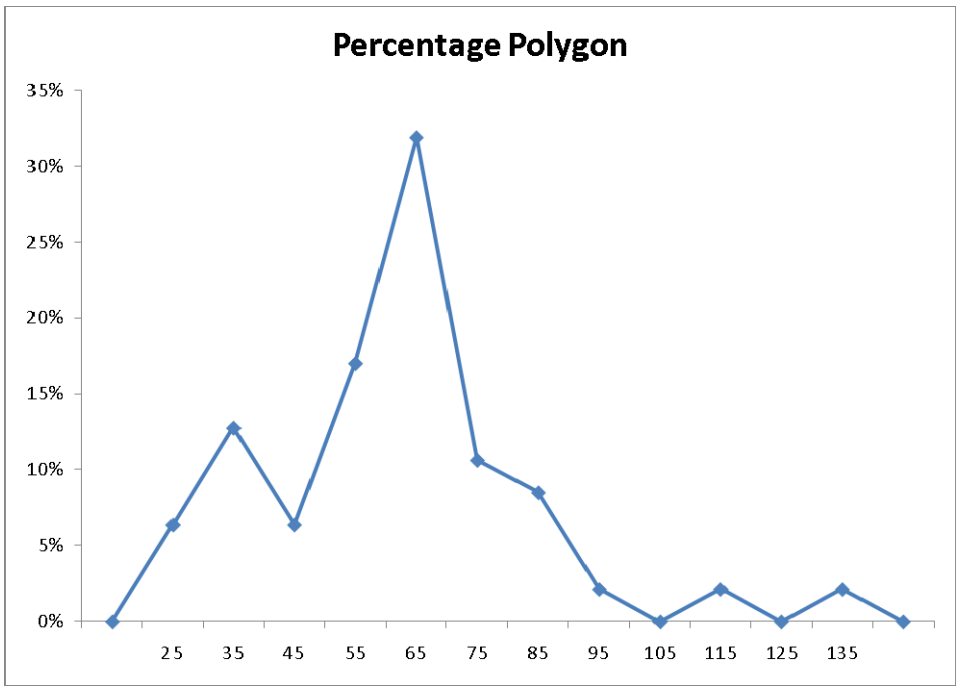
2.91 cont. (a),(c) **Four-star:**

bin	*Frequency*	*Percentage*	*Cumulative %*	*Midpts*
60 but less than 70	3	6.38%	6.38%	*65*
70 but less than 80	4	8.51%	14.89%	75
80 but less than 90	5	10.64%	25.53%	85
90 but less than 100	9	19.15%	44.68%	95
100 but less than 110	6	12.77%	57.45%	105
110 but less than 120	7	14.89%	72.34%	115
120 but less than 130	1	2.13%	74.47%	125
130 but less than 140	1	2.13%	76.60%	135
140 but less than 150	5	10.64%	87.23%	145
150 but less than 160	2	4.26%	91.49%	155
160 but less than 170	2	4.26%	95.74%	165
170 but less than 180	1	2.13%	97.87%	175
180 but less than 190	0	0.00%	97.87%	185
190 but less than 200	0	0.00%	97.87%	195
200 but less than 210	1	2.13%	100.00%	205

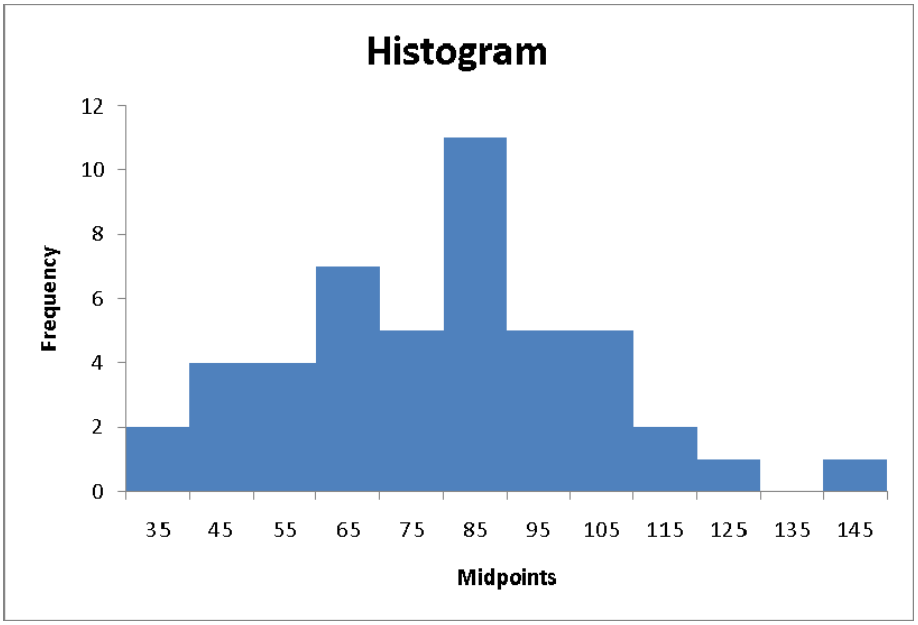
(b) **Two-star:**



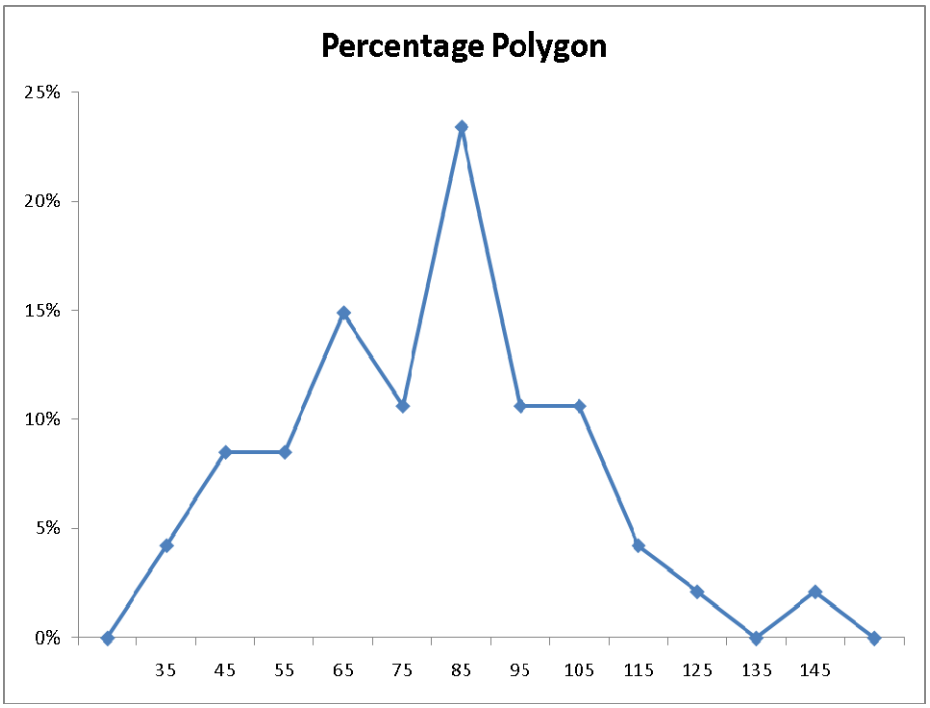
2.91 (b) cont.



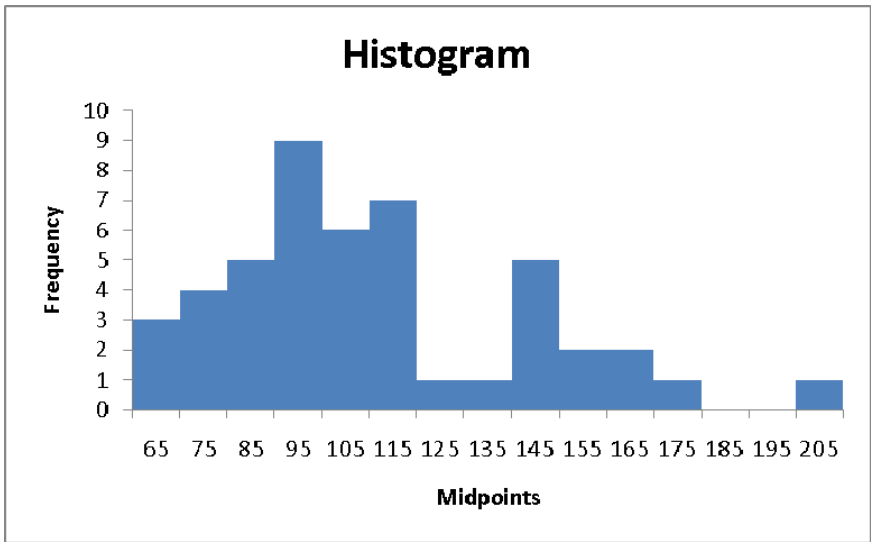
Three-star:



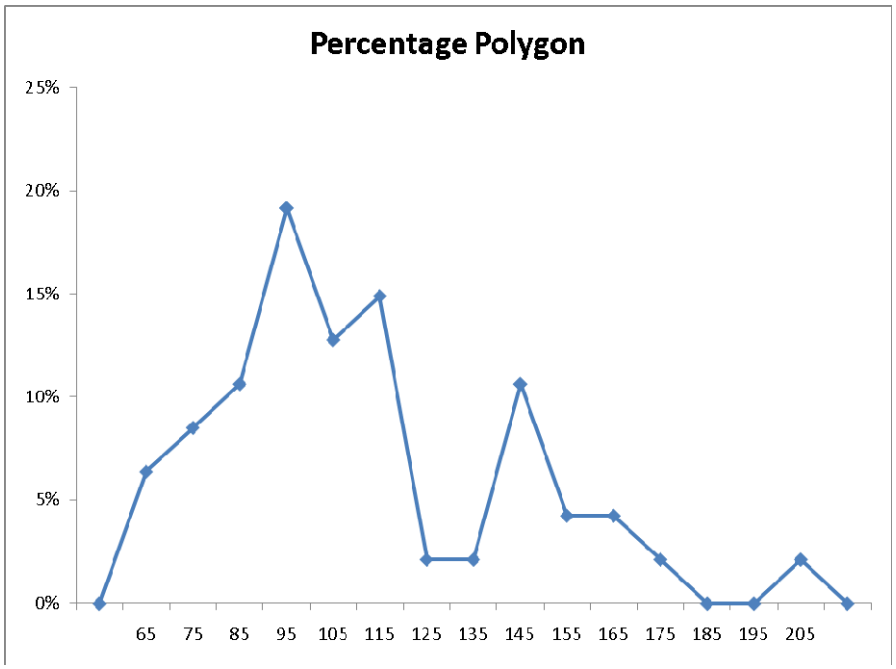
2.91 (b)
cont.



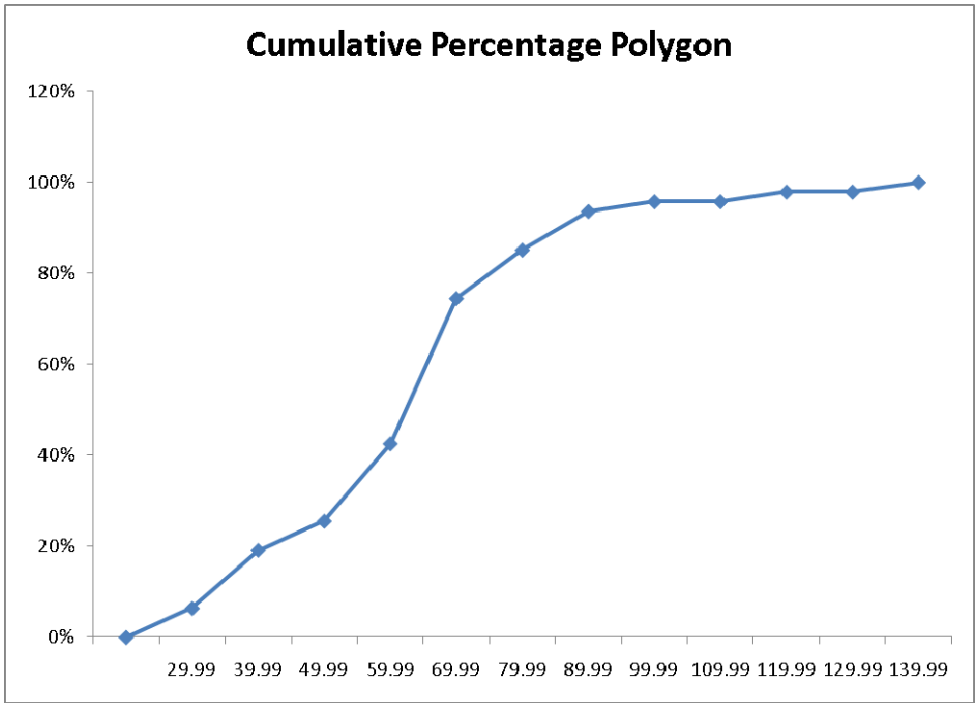
Four-star:



2.91 cont. (b)

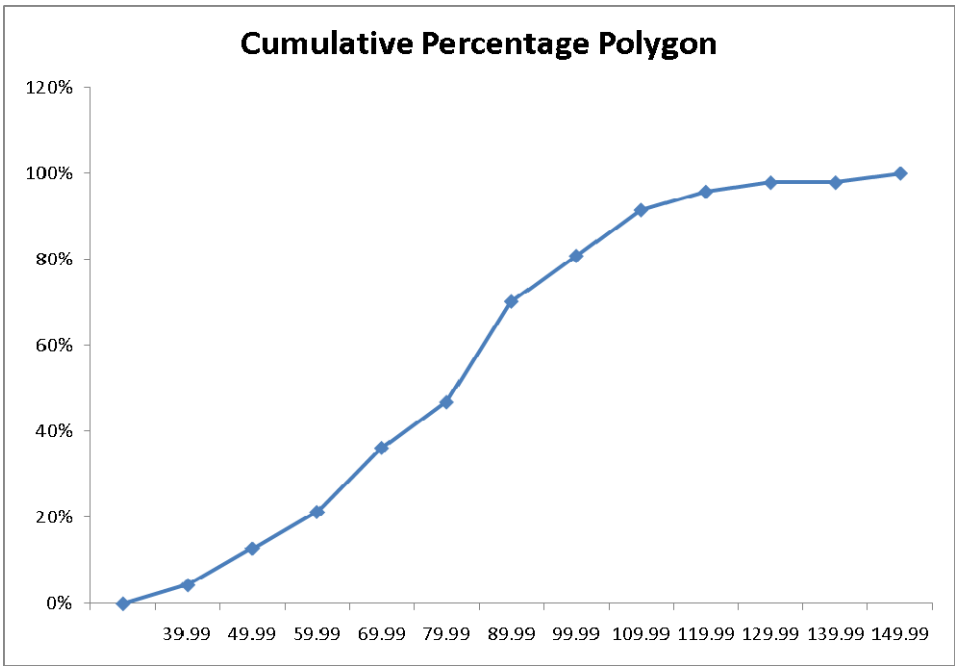


(c) **Two-star:**

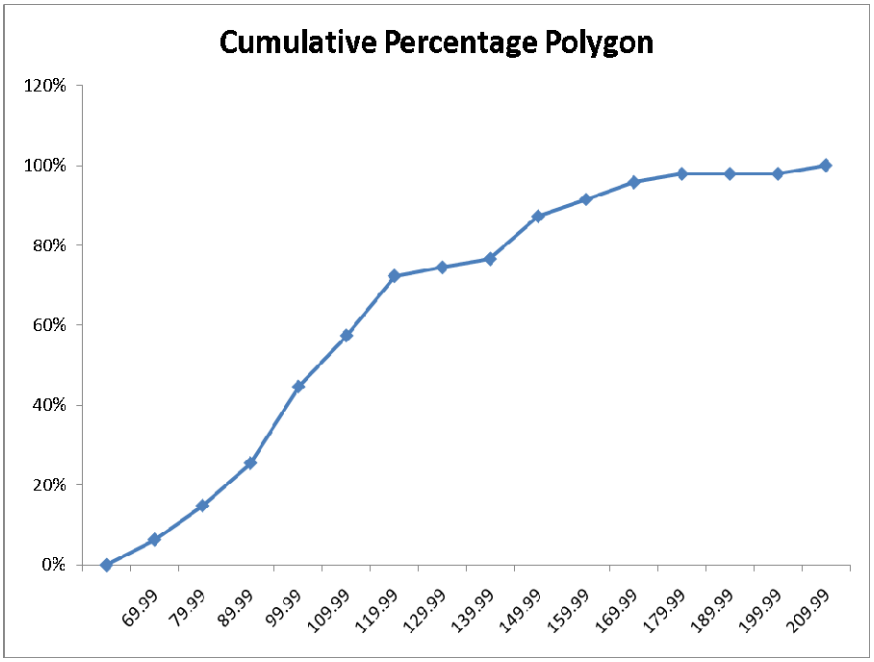


2.91 cont. (c)

Three-star:

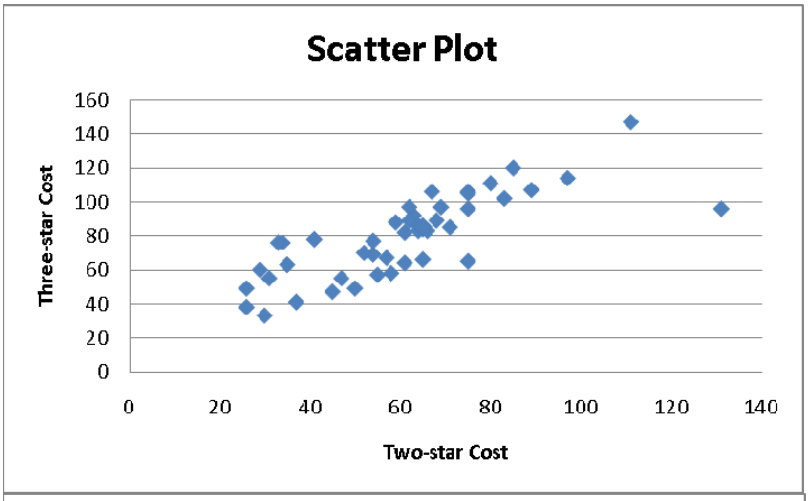


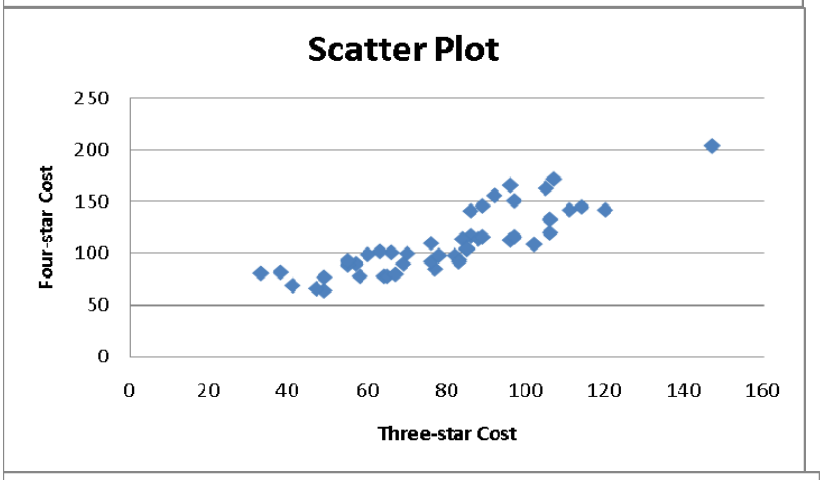
Four-star:

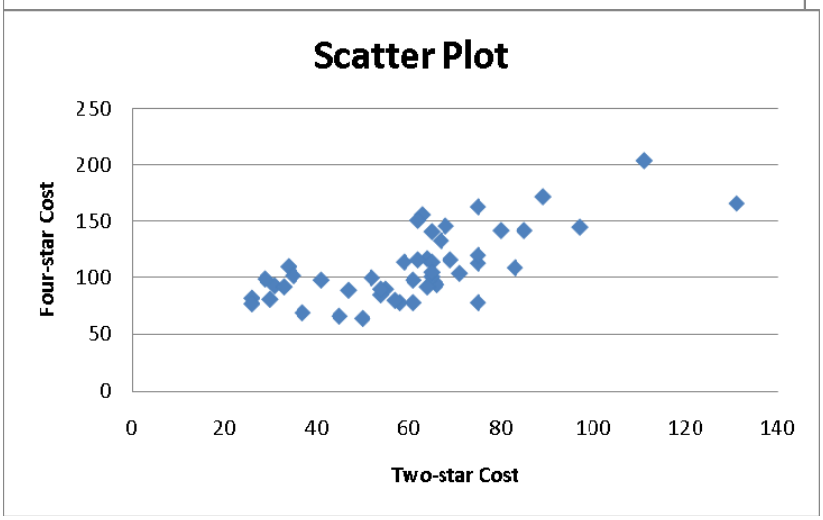


(d) The price of two-star, three-star and four-star hotels are all right-skewed. The median price of two-star , three-star and four-star hotels is around 62, 82, and 102 English pounds, respectively.

2.91 cont. (e)



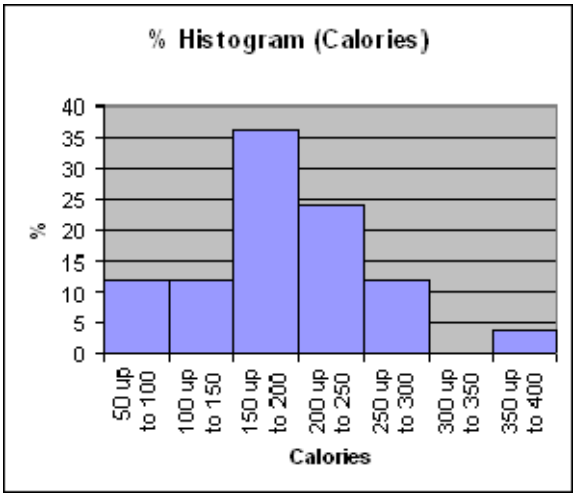




(f) The relationship of the price between two-star and three-star, three-star and four-star, and two-star and four-star hotels are all positve.

2.92 (a)

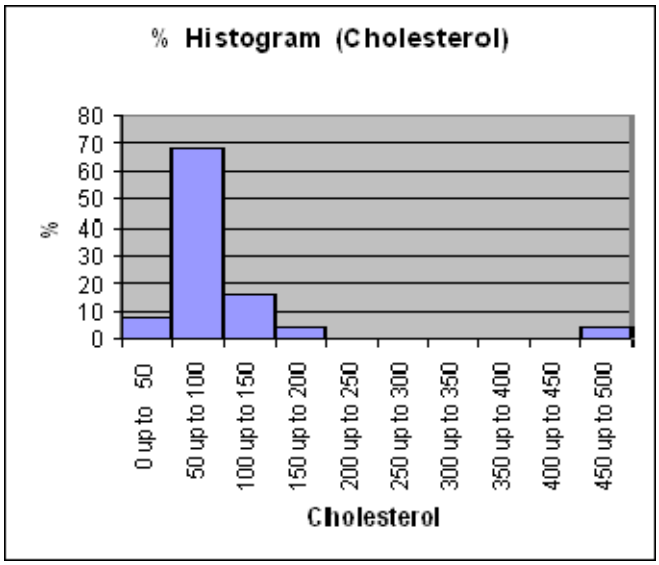
Calories	Frequency	Percentage	Percentage Less Than
50 up to 100	3	12%	12%
100 up to 150	3	12	24
150 up to 200	9	36	60
200 up to 250	6	24	84
250 up to 300	3	12	96
300 up to 350	0	0	96
350 up to 400	1	4	100



(b)

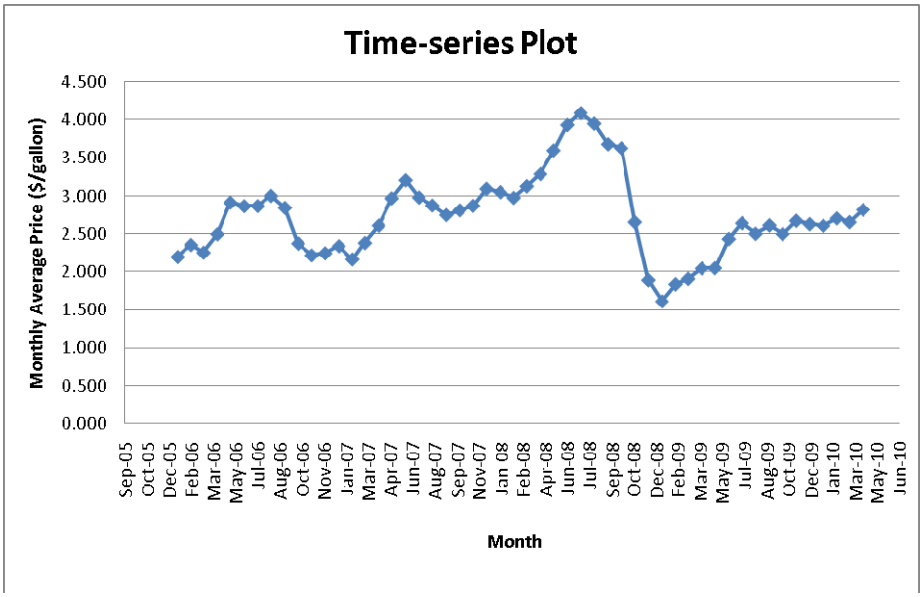
Cholesterol	Frequency	Percentage	Percentage Less Than
0 up to 50	2	8	8%
50 up to 100	17	68	76
100 up to 150	4	16	92
150 up to 200	1	4	96
200 up to 250	0	0	96
250 up to 300	0	0	96
300 up to 350	0	0	96
350 up to 400	0	0	96
400 up to 450	0	0	96
450 up to 500	1	4	100

2.92 cont. (b)



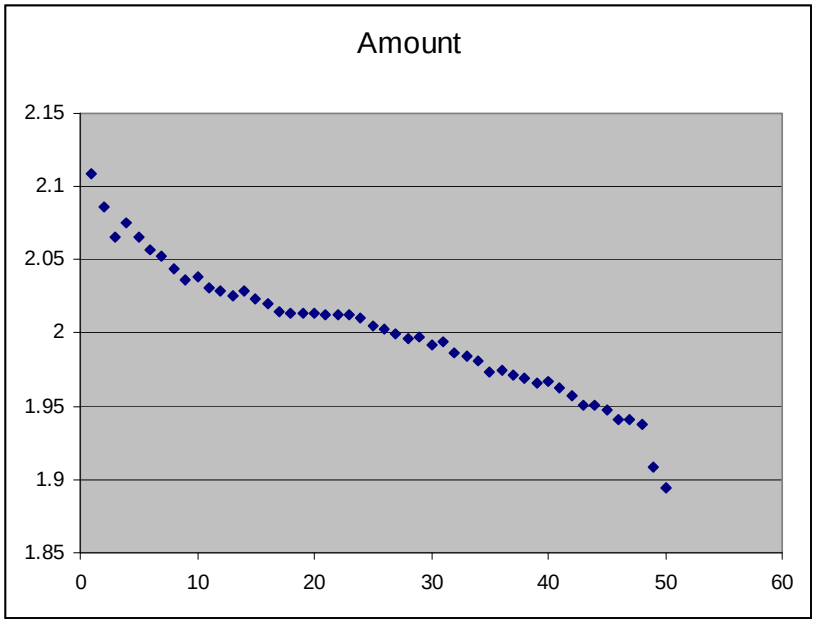
(c) The sampled fresh red meats, poultry, and fish vary from 98 to 397 calories per serving, with the highest concentration between 150 to 200 calories. One protein source, spareribs, with 397 calories, is more than 100 calories above the next highest caloric food. The protein content of the sampled 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.

2.93 (a)



(b) The average price of gasoline in the United States is higher in the summer in general and seems to peak in June.

2.94 (a)

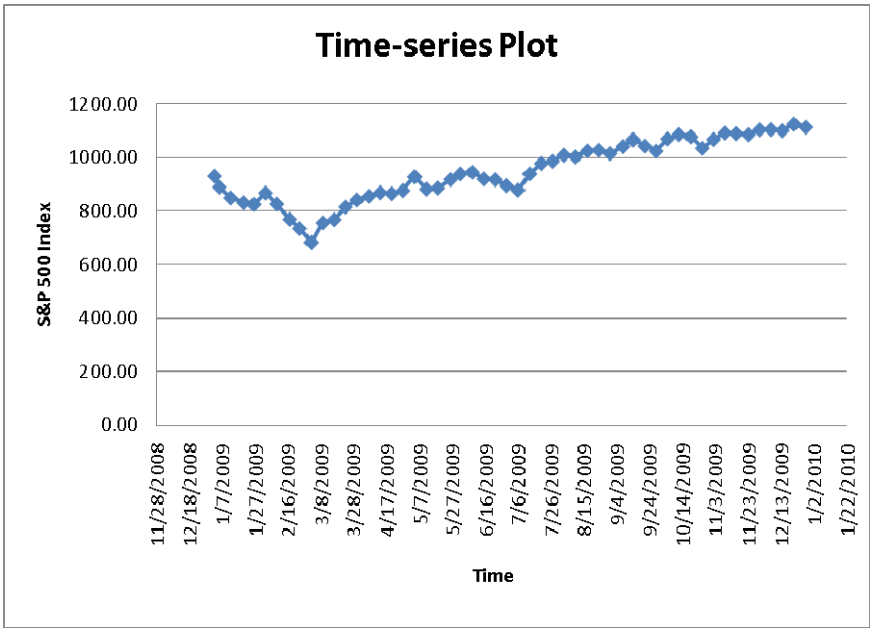


(b) There is a downward trend in the amount filled.

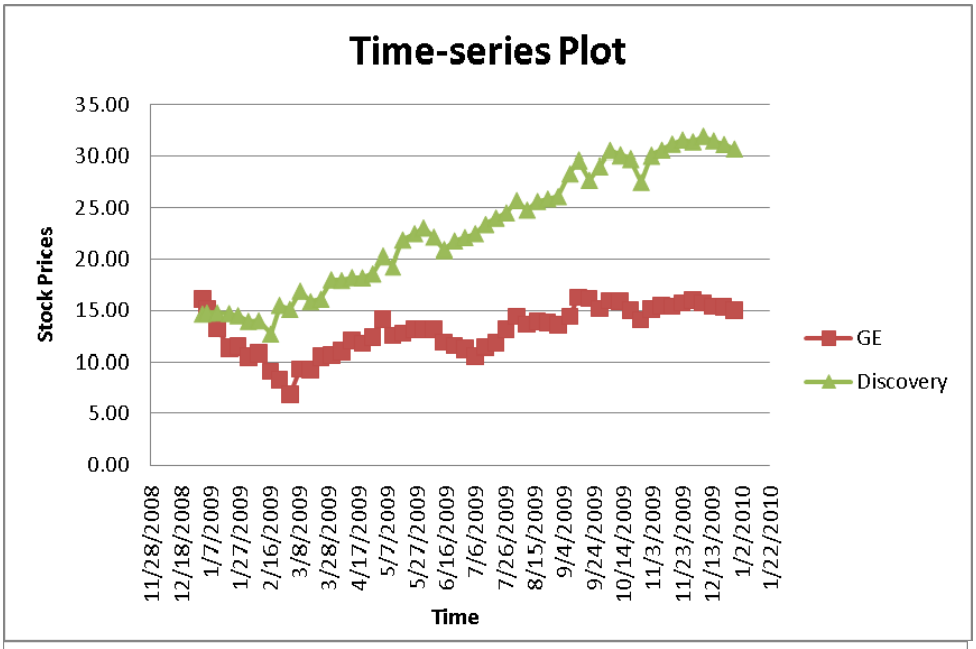
(c) The amount filled in the next bottle will most likely be below 1.894 liter.

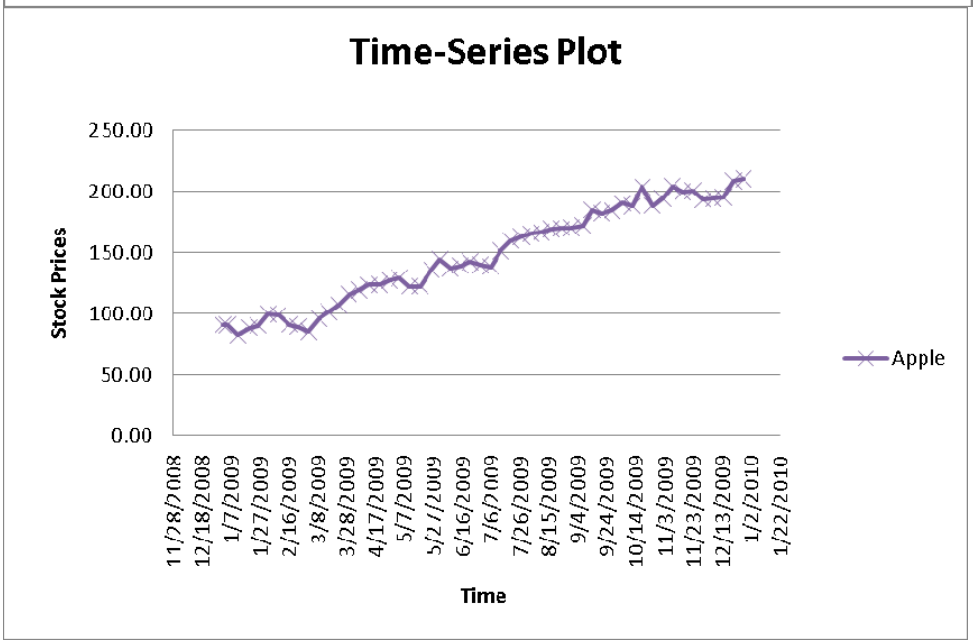
(d) The scatter plot of the amount of soft drink filled against time reveals the trend of the data, whereas a histogram only provides information on the distribution of the data.

2.95 (a)



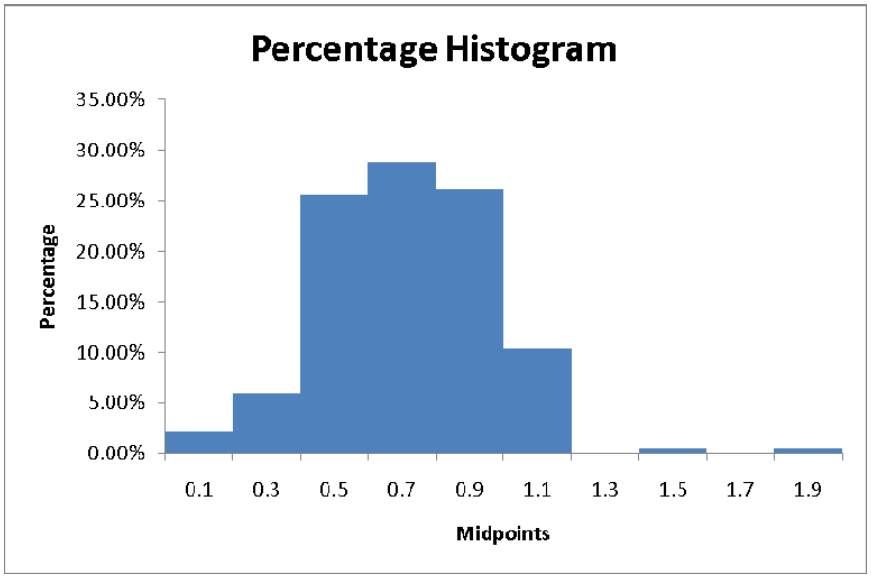
2.95 cont. (a)



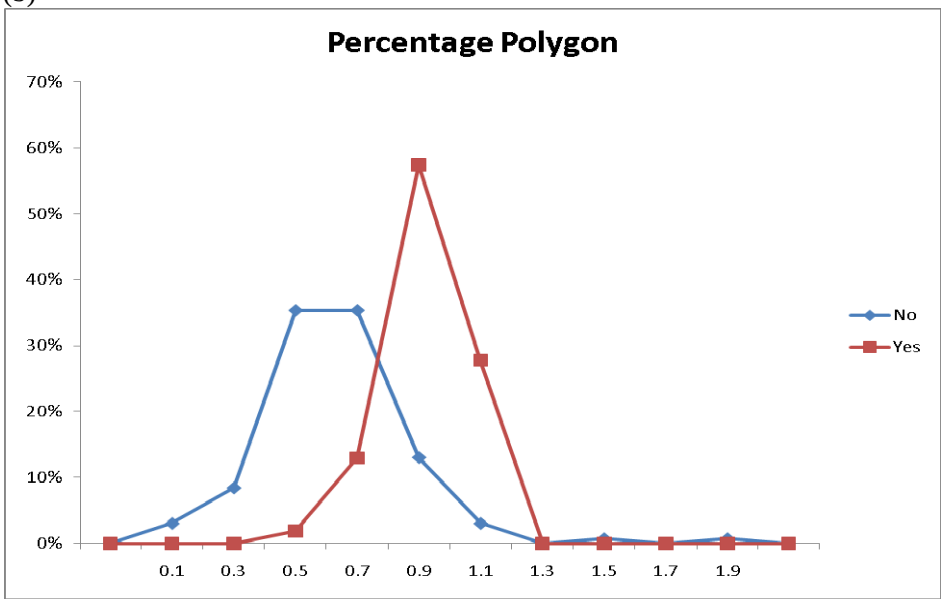


(b) Even though there appeared to be cyclical pattern in the S&P index, there was a general upward trend with a big drop that took place for the week of 3/2/2009. The stock price of Apple fluctuated between $120 and $210 with a general upward trend. The stock price of GE trended downward from about $15 during the week of 1/2/2009 to about $6 during the week of 3/2/2009 but turned around and trended upward back to about $15 during the week of 12/28/2009. Discovery's stock price trended upward from about $15 to about $30.

2.100 (a) **Expense Ratio**

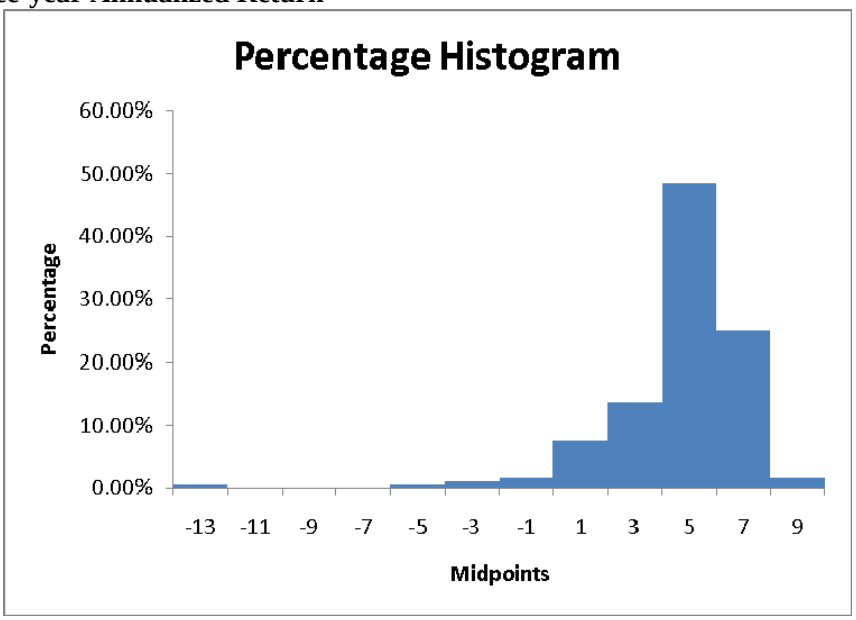


(b)

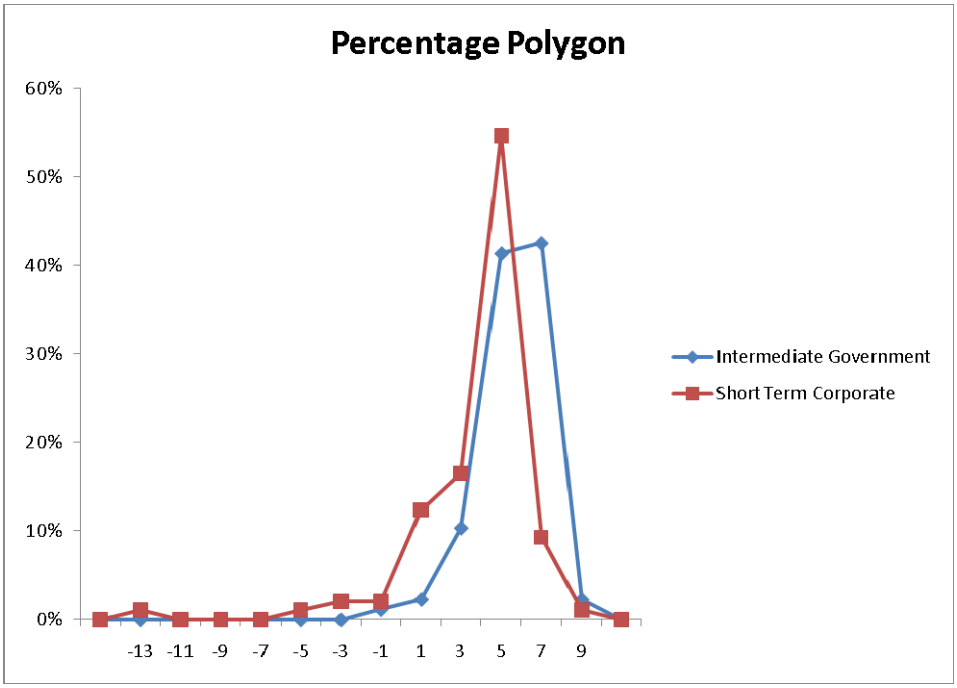


(c) The expense ratio of all bond funds is scattered around 0.75. Bond funds with fees have expense ratios scattered around 0.9 while bond funds without fees have expense ratios scattered around 0.6.

2.101 (a) **Three-year Annualized Return**

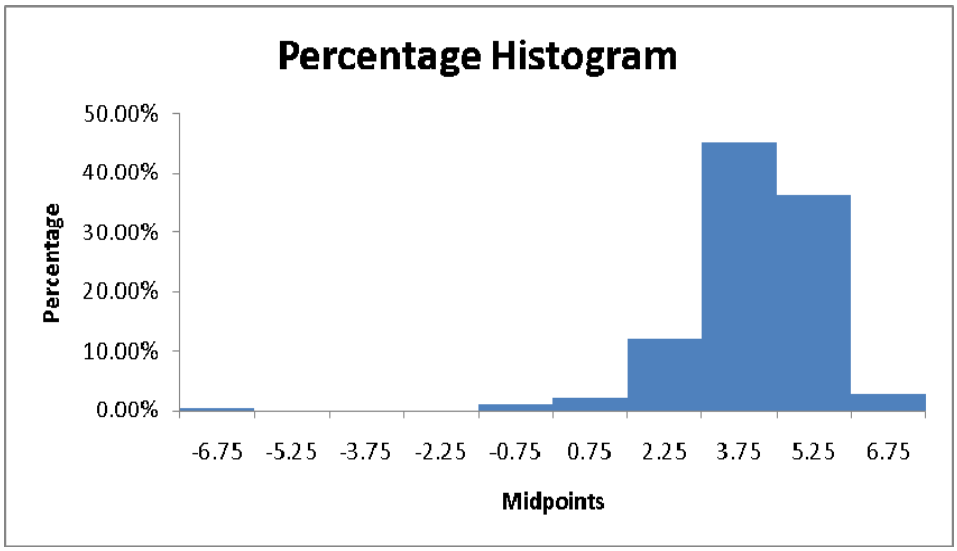


(b)

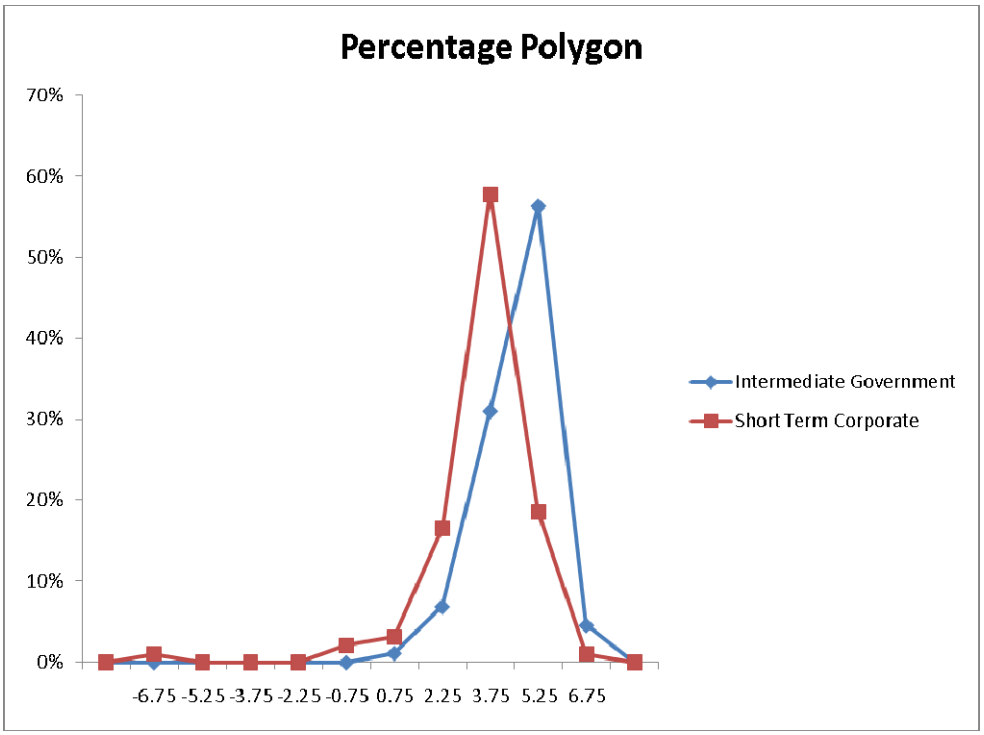


(c) The three-year annualized return of all the bond funds is left-skewed with majority of them (about 87%) scattered between 2% and 8%. About 3.8% of the bond funds have a negative three-year annualized return while about 1.6% of them have a return higher than 8%. In general, the intermediate government funds have higher three-year annualized returns than short term corporate funds. Both types of bond funds have three-year annualized returns skewed to the left.

2.102 (a) **Five-year Annualized Return**



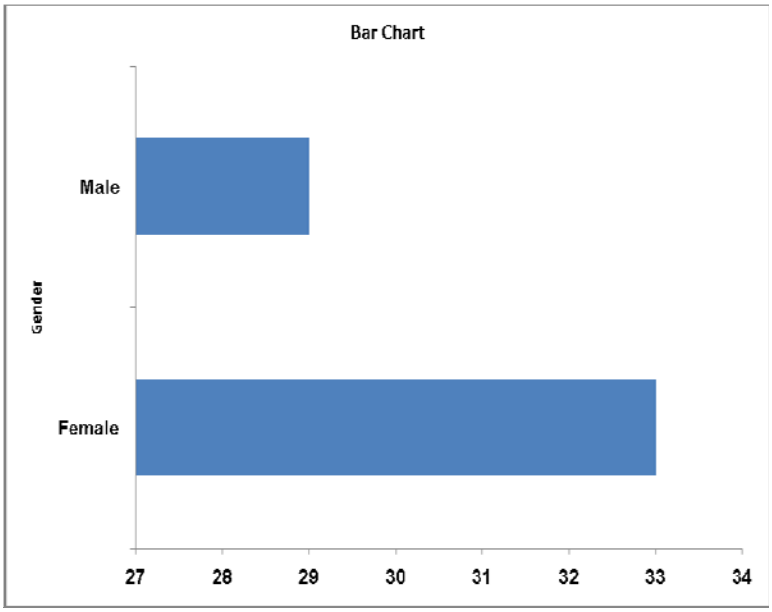
(b)

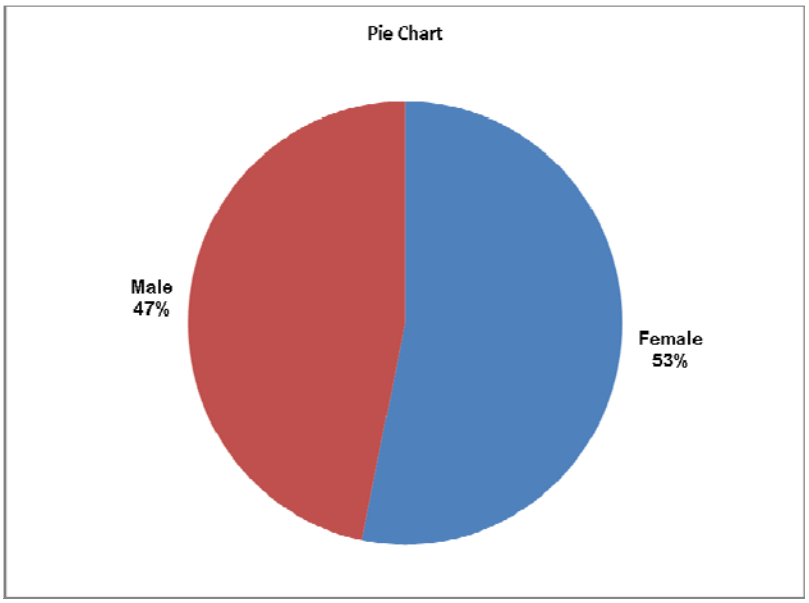


(c) The five-year annualized return of all the bond funds is left-skewed with majority of them (about 93%) scattered between 1.5% and 6%. About 1.6% of the bond funds have a negative five-year annualized return while about 2.7% of them have a return higher than 6%. In general, the intermediate government funds have higher five-year annualized returns than short term corporate funds. Both types of mutual funds have five-year annualized returns skewed to the left.

2.103

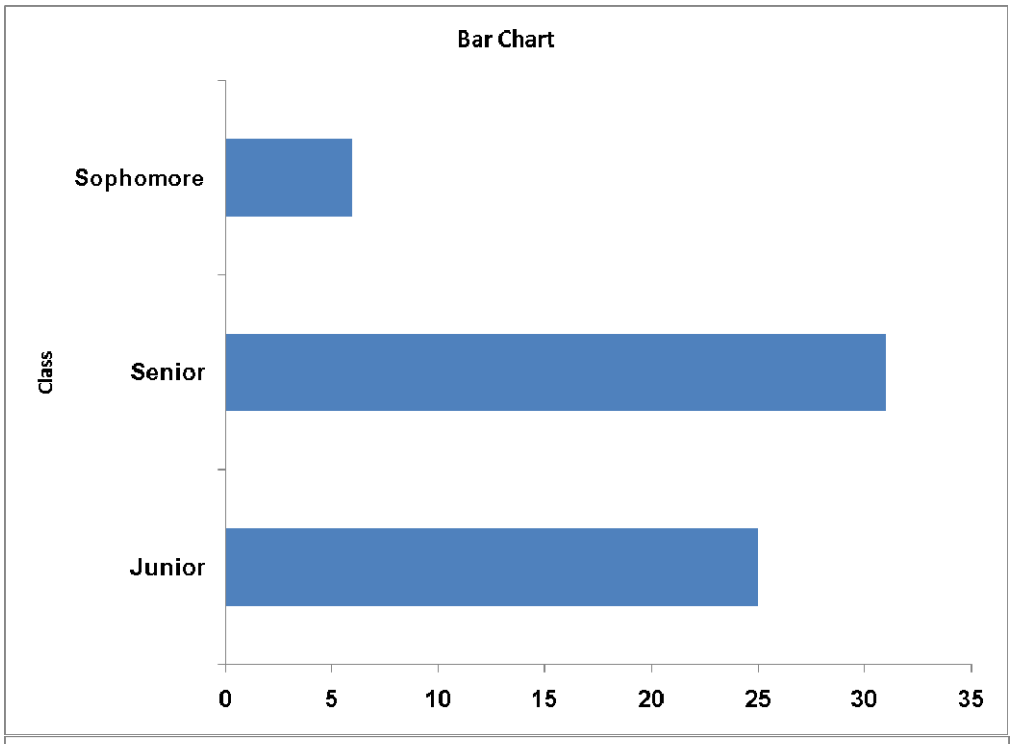
Gender:

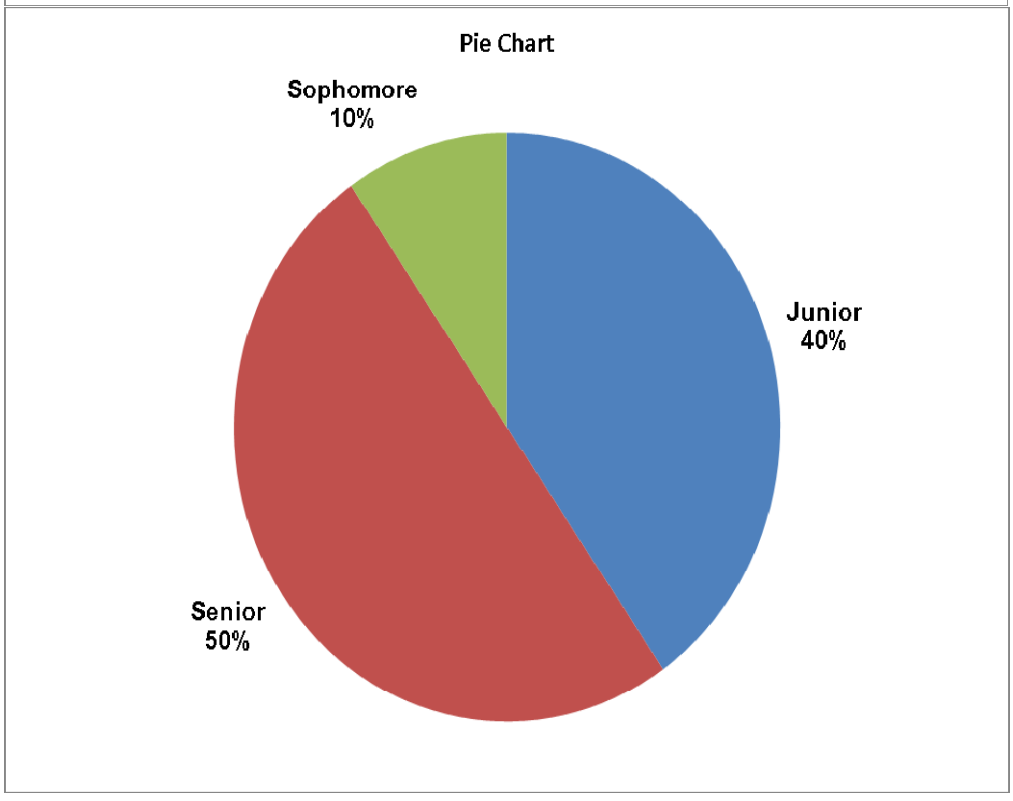




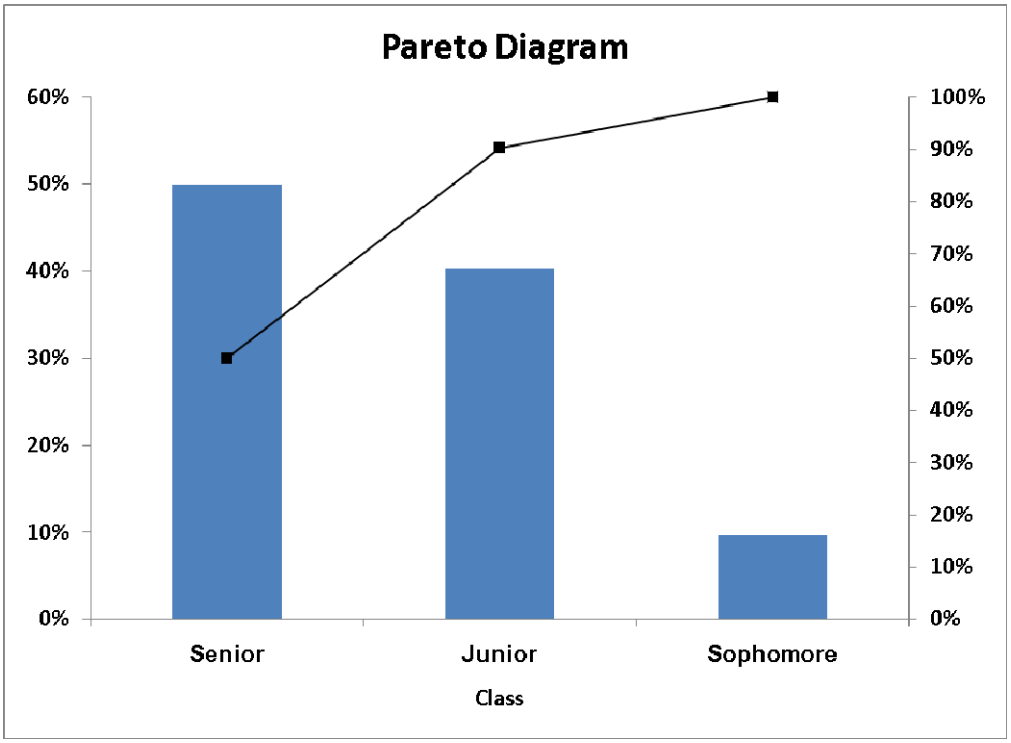
There are more females than males in the survey.

2.103 **Class:**
cont.



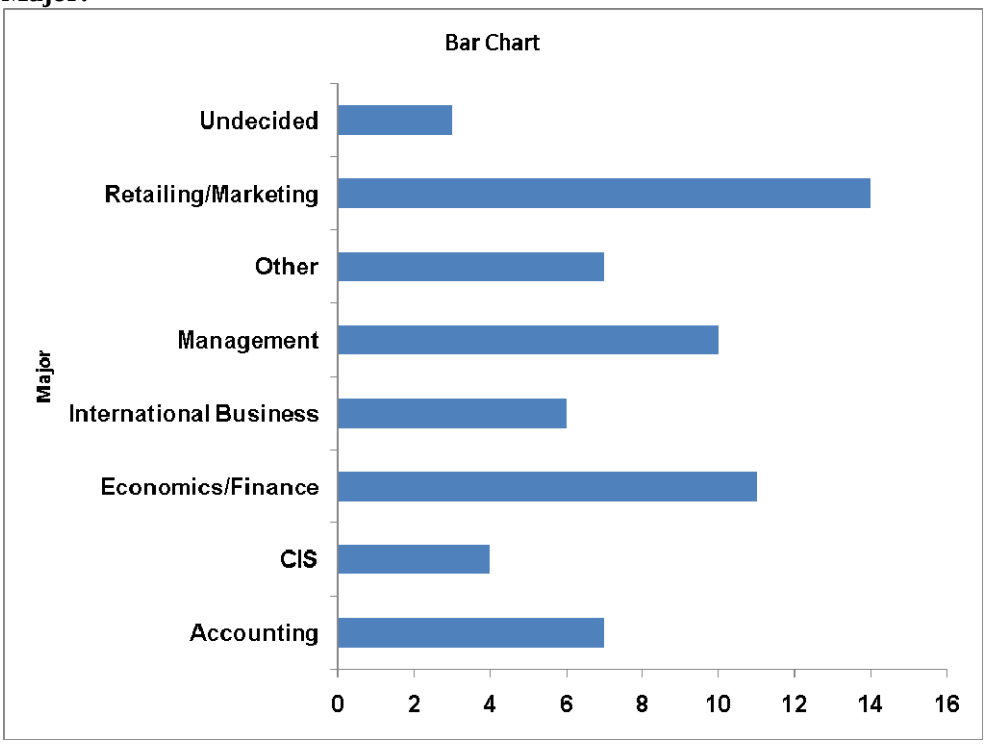


2.103
cont.

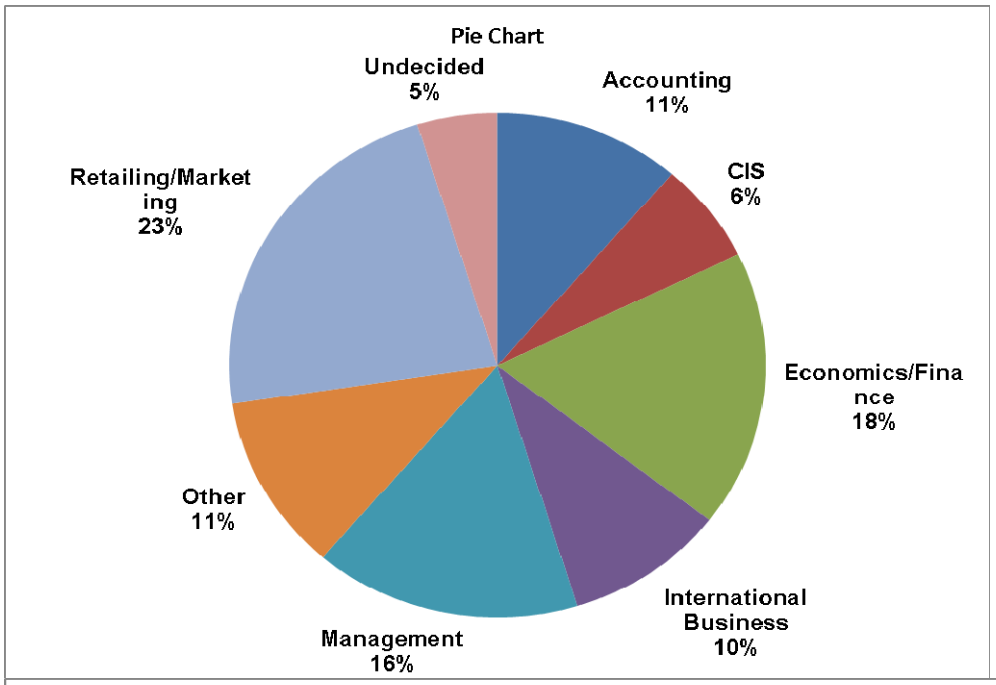


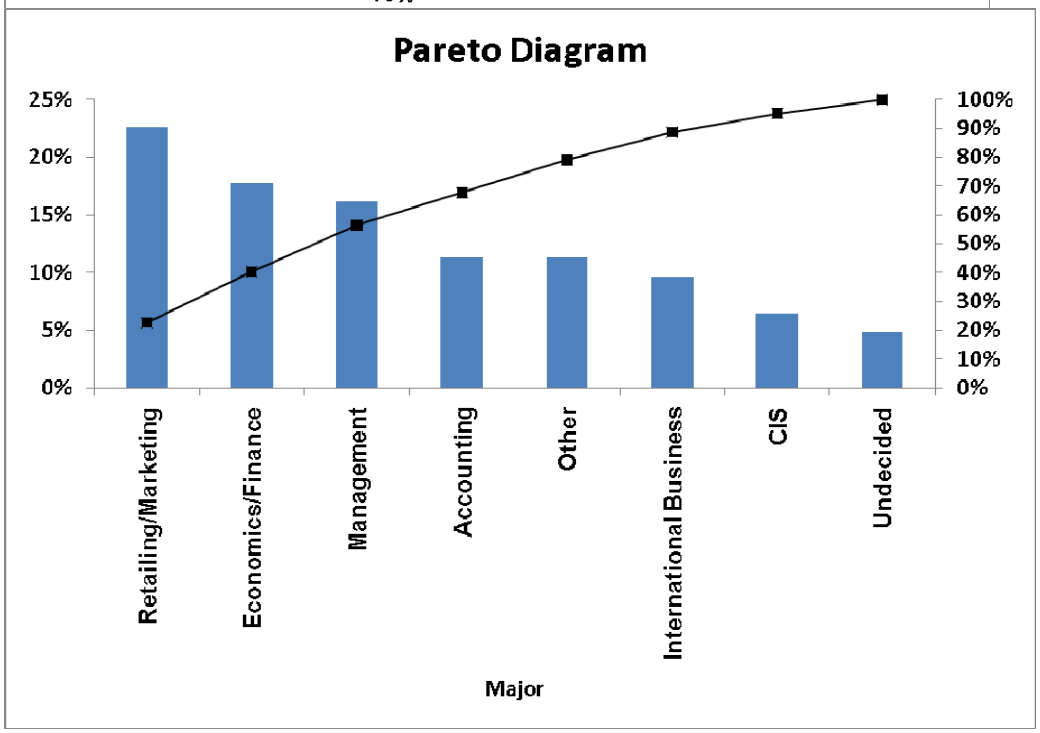
There are more senior and junior students than sophomore.

Major:



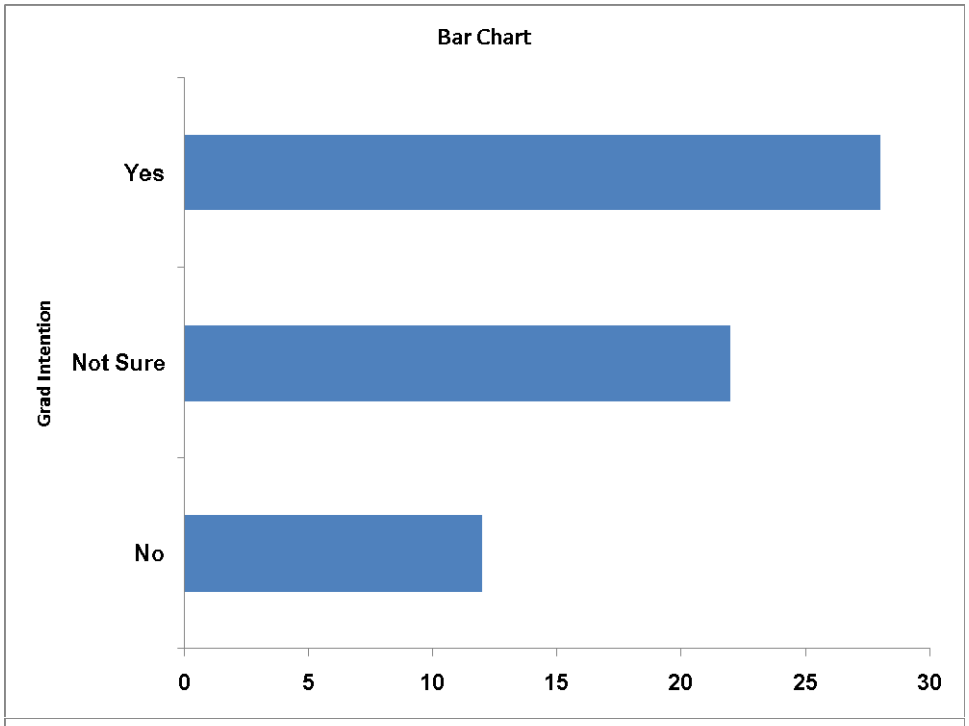
2.103
cont.

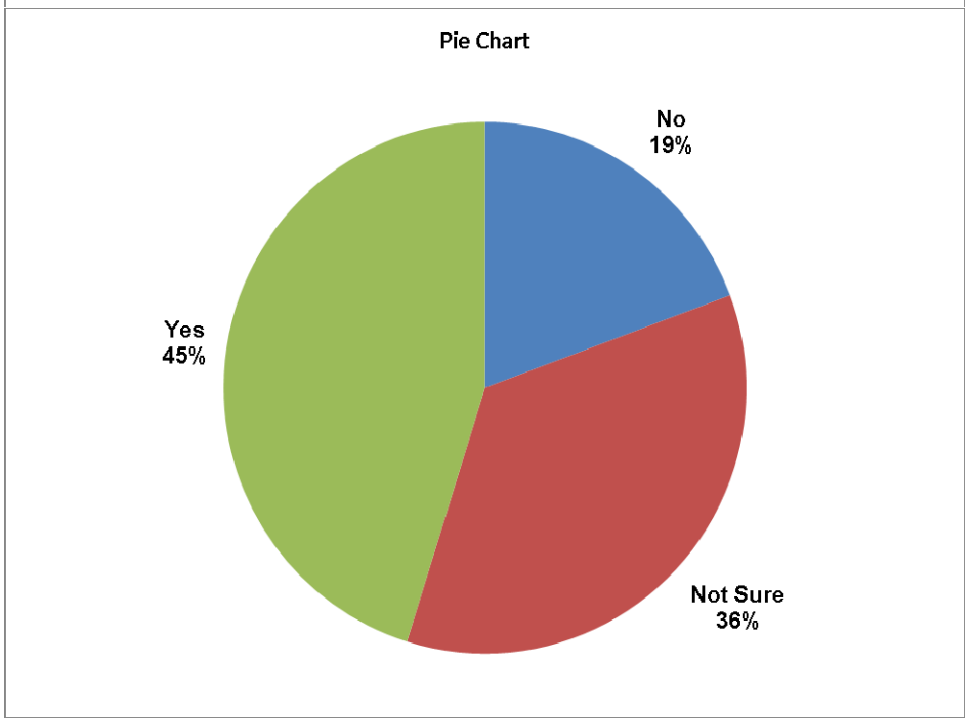




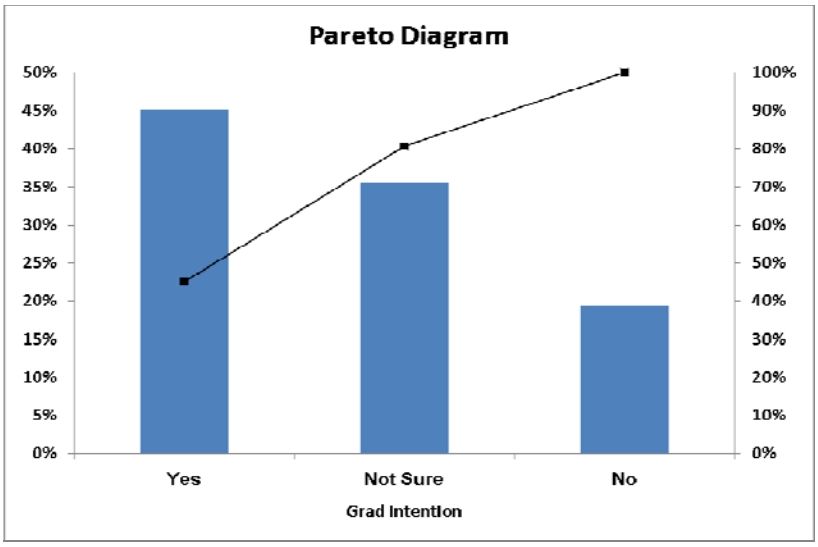
Retailing/marketing, economics/finance and management constituted the "vital few" while the rest of the majors make up the "trivial many".

2.103
cont. **Grad Intention:**



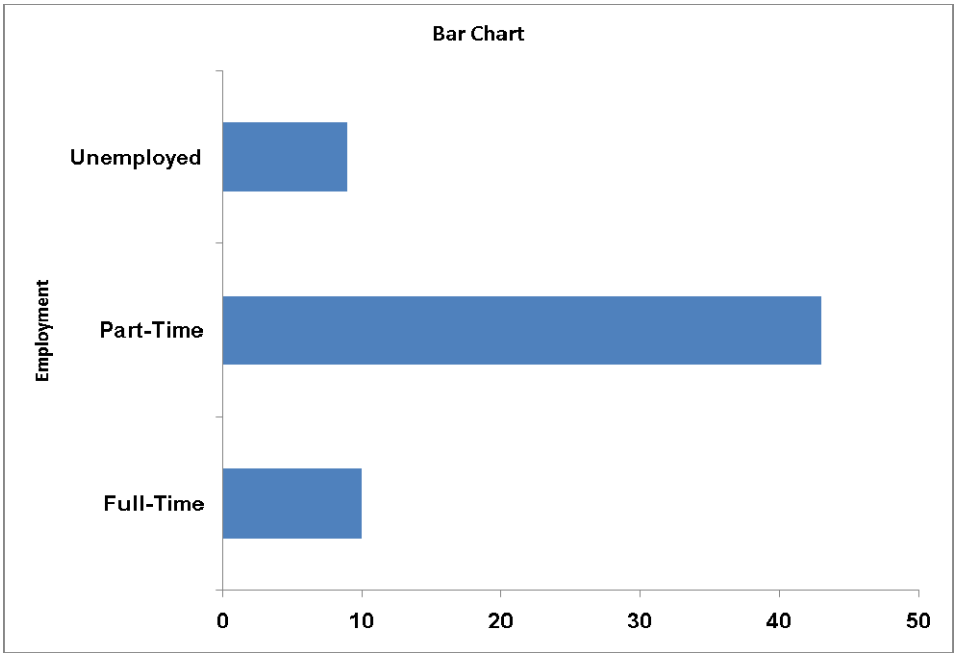


2.103
cont.

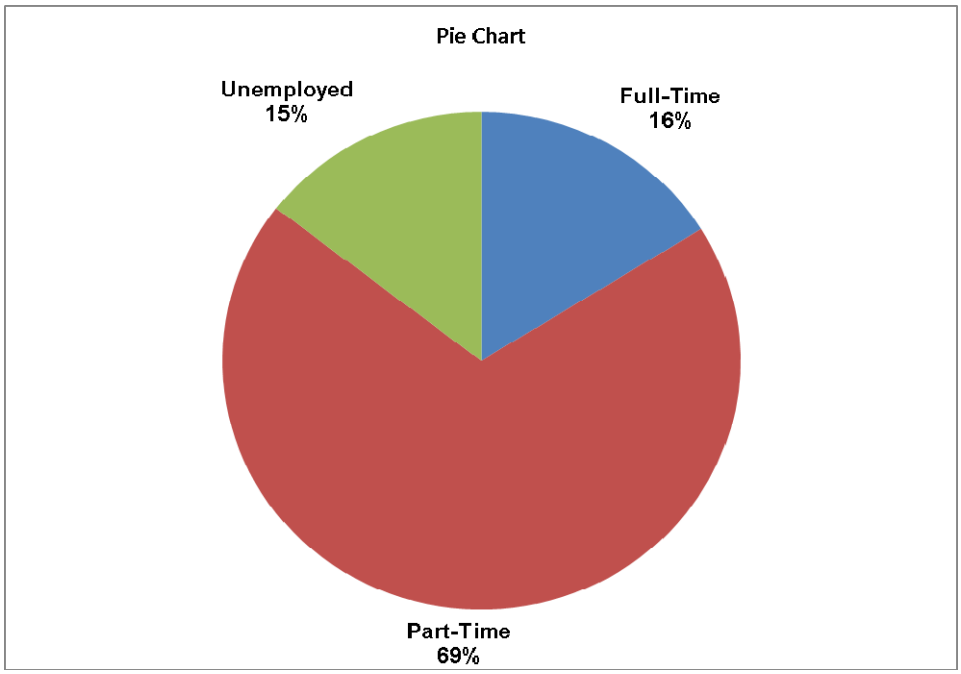


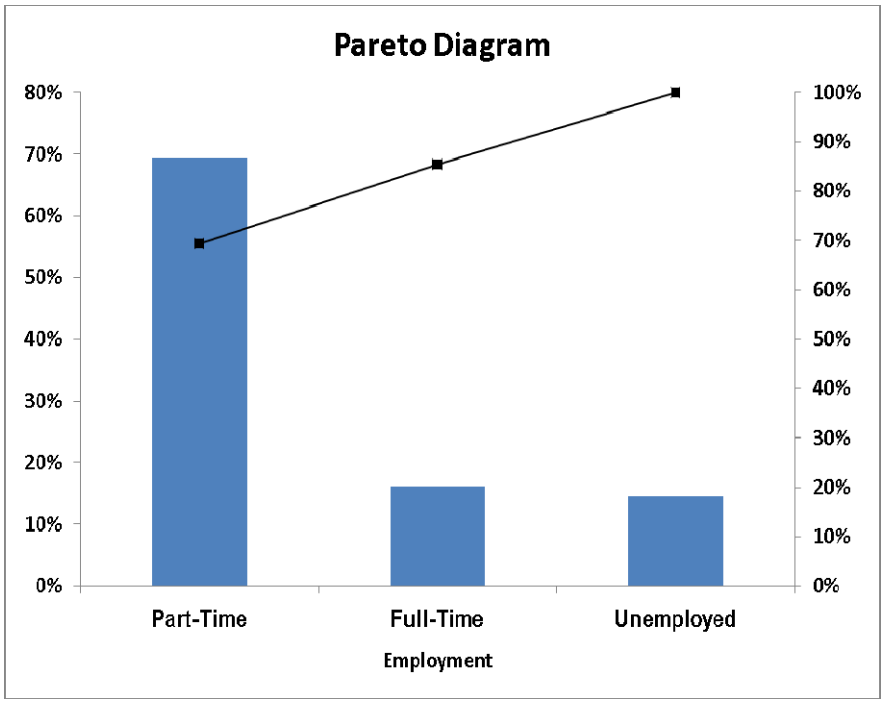
There are more students with a grad intention than either of the other categories.

Employment:



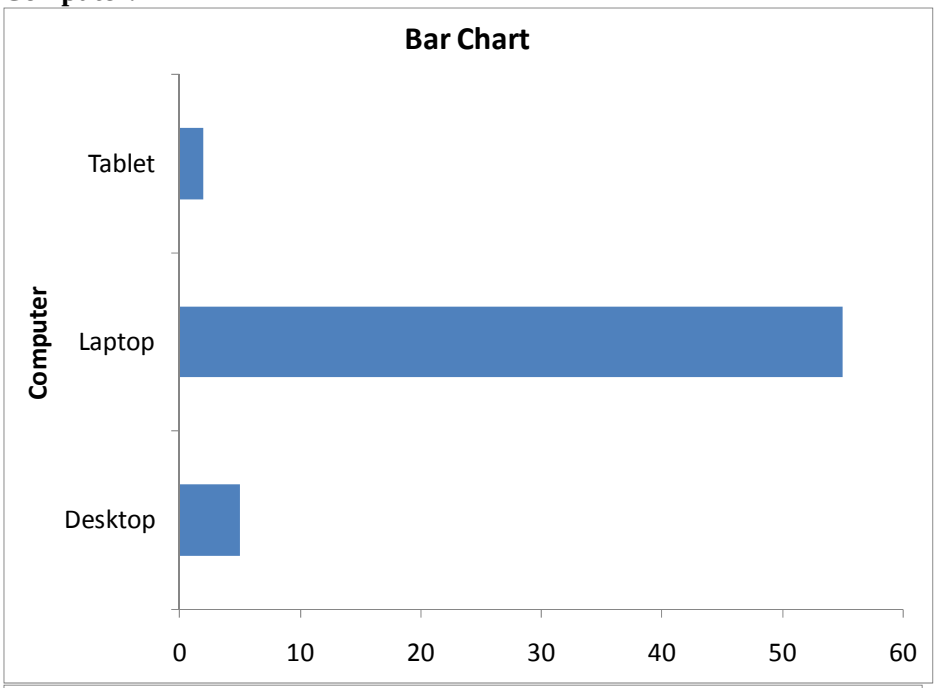
2.103
cont.

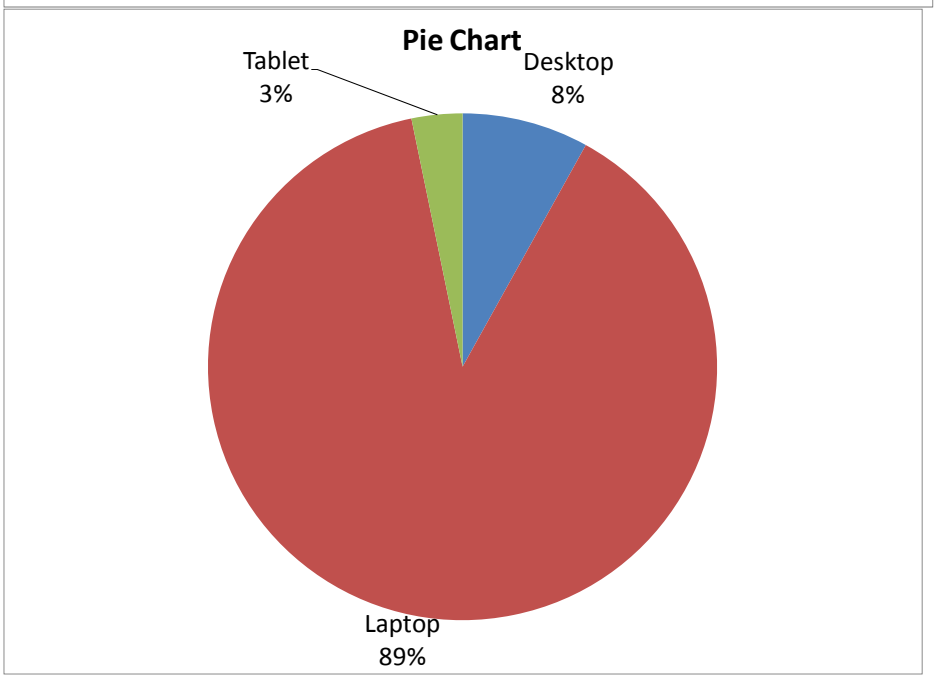




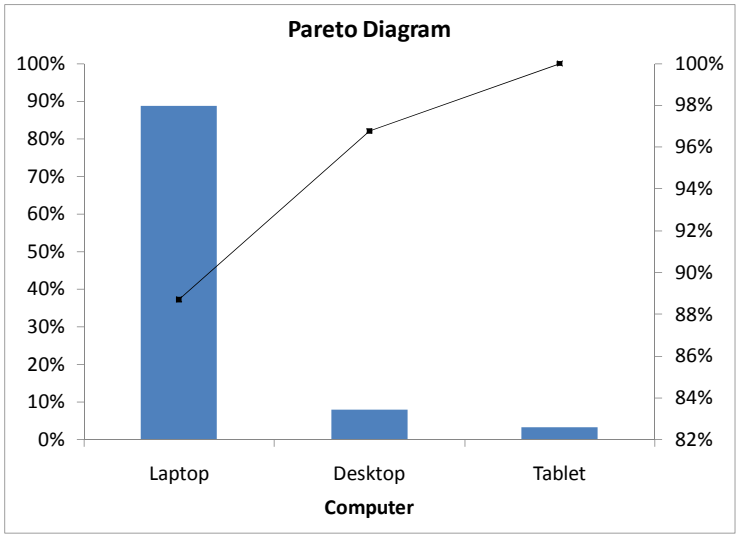
Most of the students have part-time employment.

2.103
cont. **Computer:**





2.103 cont.



Nearly 90% of the students use laptops.

Age:

Stem-and-Leaf Display	
Stem unit	1
18	0
19	00000
20	00000000000000
21	000000000000000000000
22	00000000000
23	00000
24	000
25	
26	0

Majority of the students surveyed are between 20 and 22 year old.

Social Networking:

Stem-and-Leaf Display	
Stem unit	1
0	000
1	00000000000000000000000000000000
2	00000000000000000
3	00000
4	00

Majority of the students are registered at between 1 and 2 social networking sites.

2.103
cont.

Satisfaction:

Stem-and-Leaf Display	
Stem unit	1
1	00000
2	00
3	0000000000000000
4	00000000000000000000000000
5	0000000000
6	0000

Majority of the satisfaction ratings fall between 3.0 and 5.0.

Text Messages:

Stem-and-Leaf Display	
Stem unit	100
0	013445555555567
1	00000000114555558
2	00000055
3	000000000055
4	00
5	000
6	00
7	005
8	0
9	0

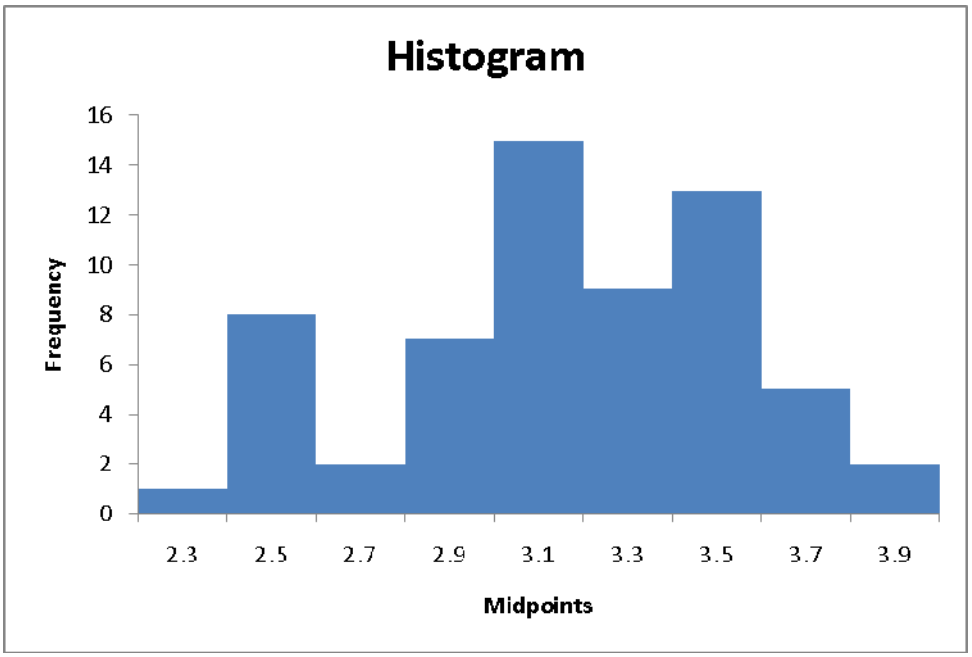
Majority of the students sent less than 400 messages in a typical week.

2.103 cont. **Spending:**

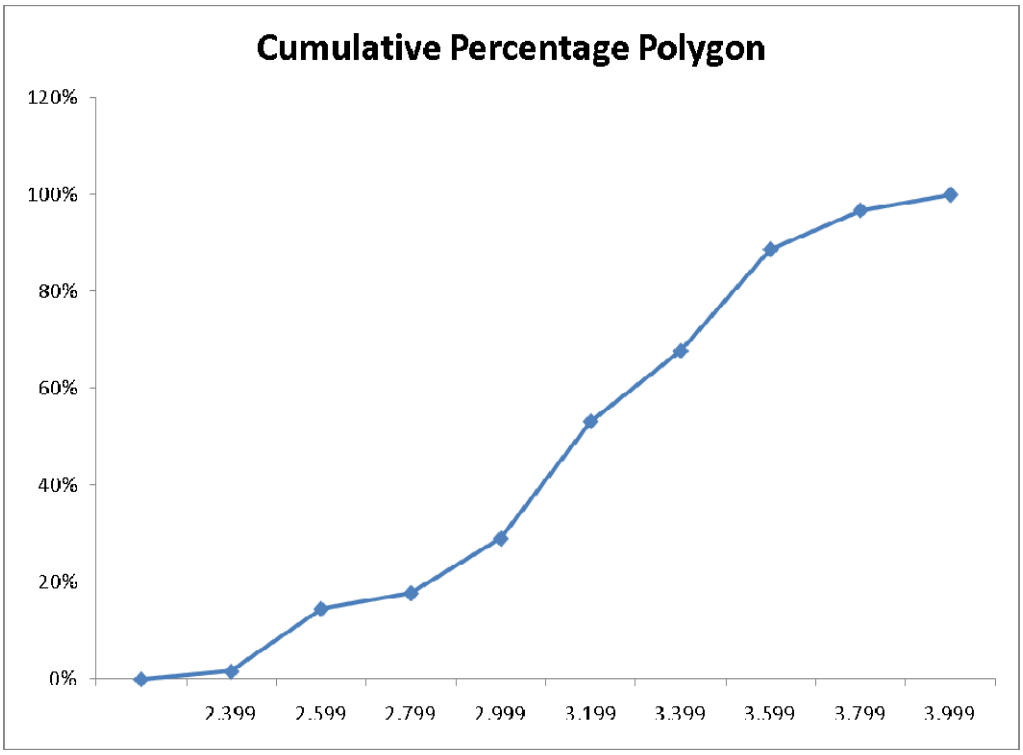
Stem-and-Leaf Display	
Stem unit	100
1	0
2	0000255
3	0000000055568
4	00000559
5	000000000000002
6	0000000005589
7	0
8	
9	0
10	0
11	0
12	
13	
14	0

Majority of the students spend between $200 and $700 for textbooks and supplies.

GPA:

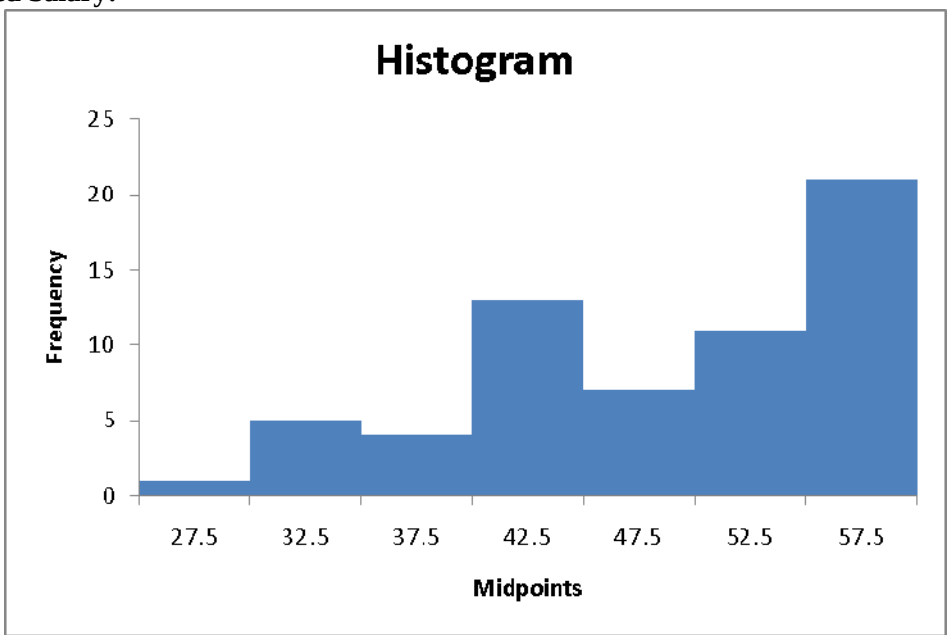


2.103
cont.

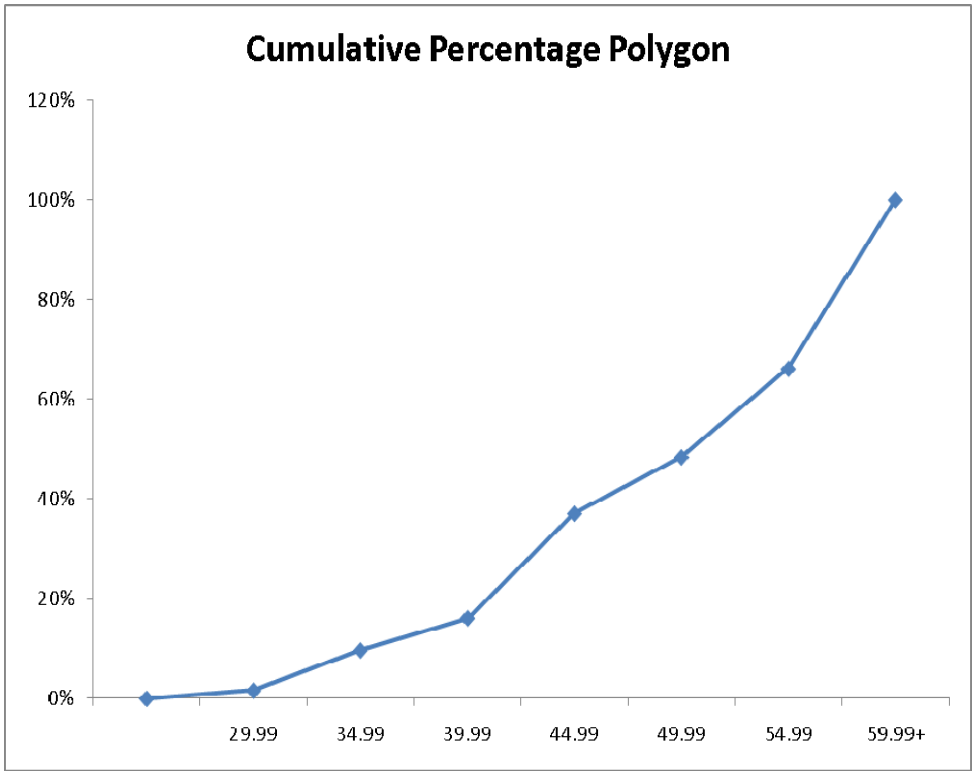


GPA is slightly left-skewed.

Expected Salary:

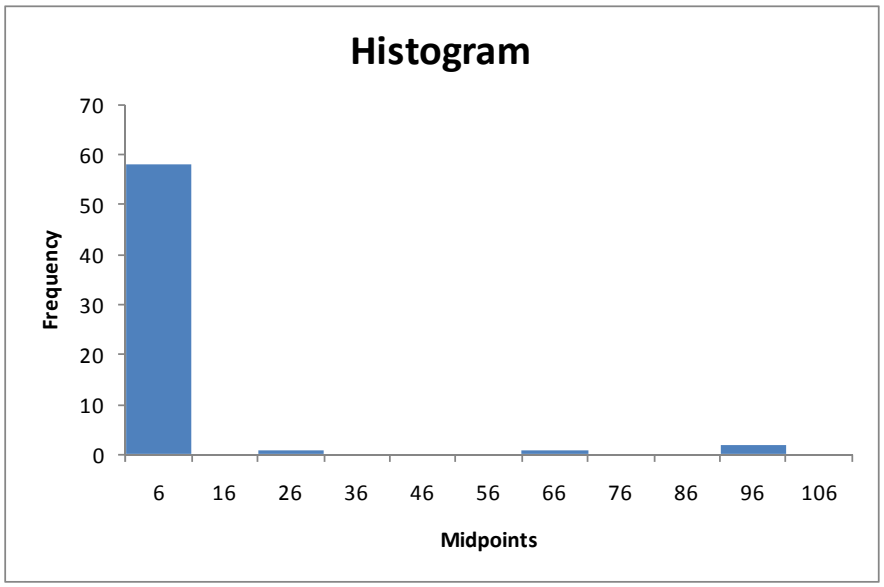


2.103
cont.

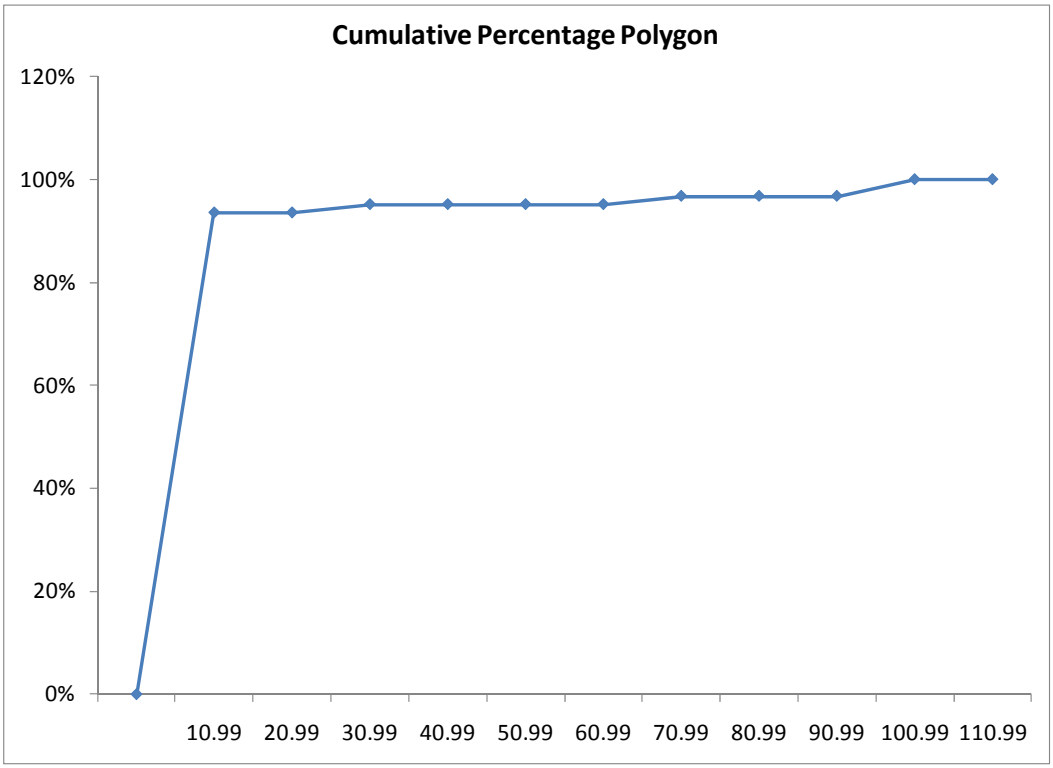


Expected salary is left-skewed.

Wealth:



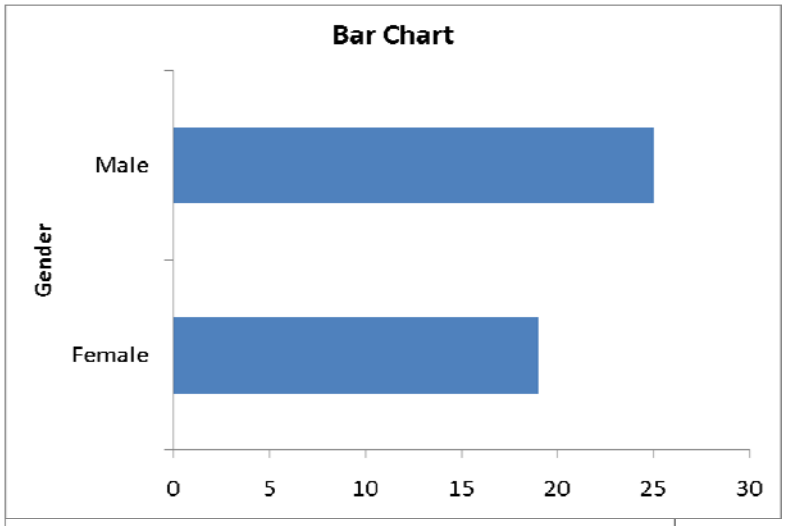
2.103
cont.

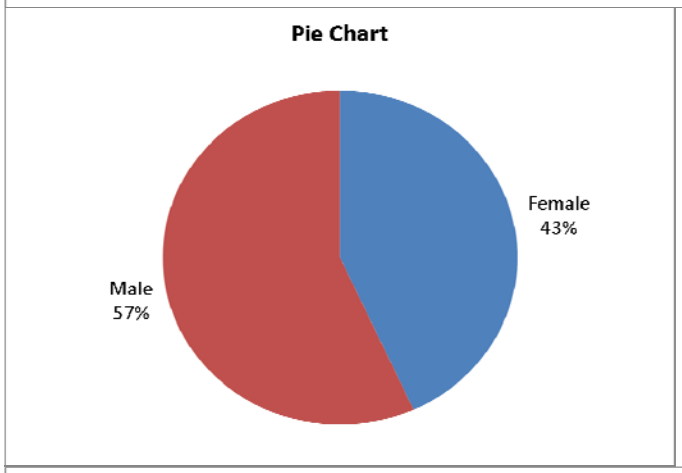


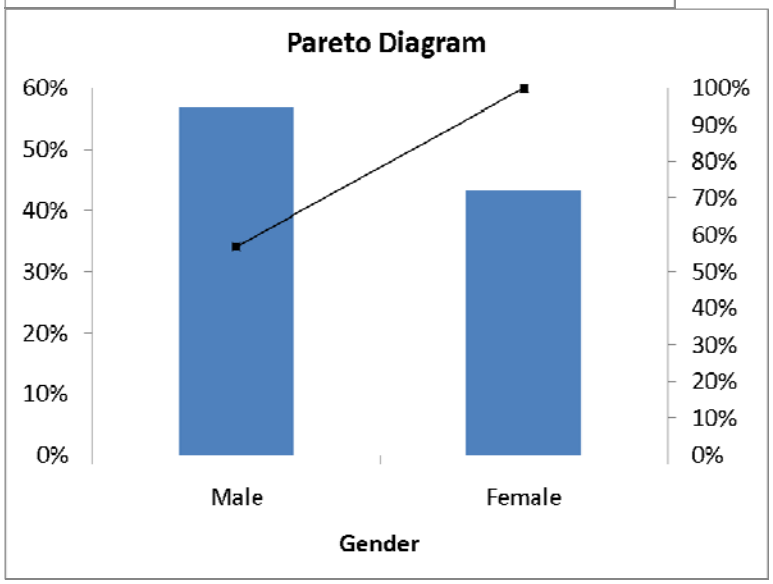
Wealth is right-skewed.

2.105

Gender:



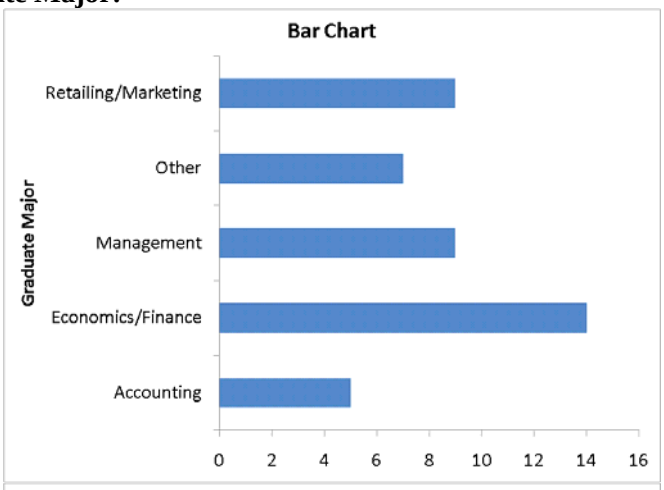


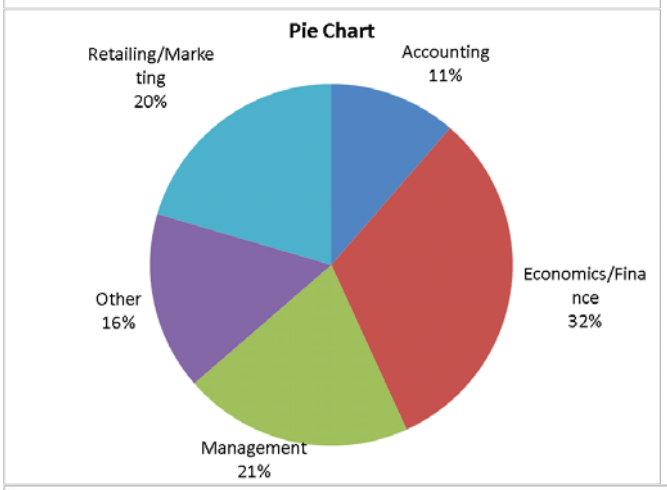


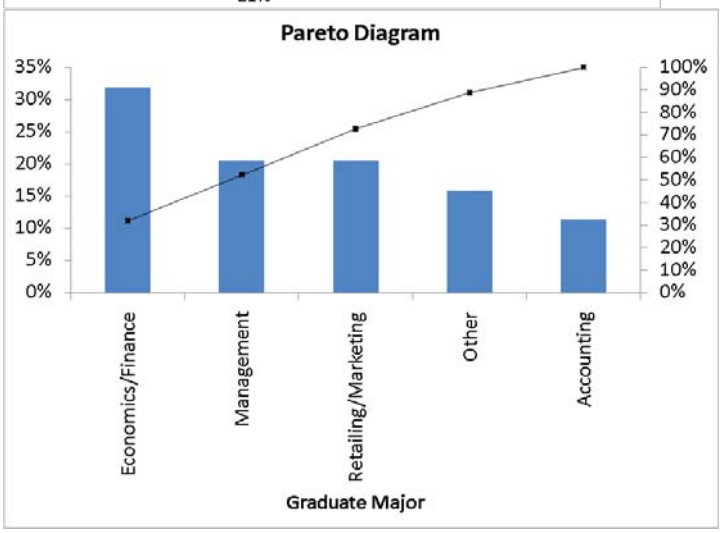
There are more males than females in the survey.

2.105
cont.

Graduate Major:

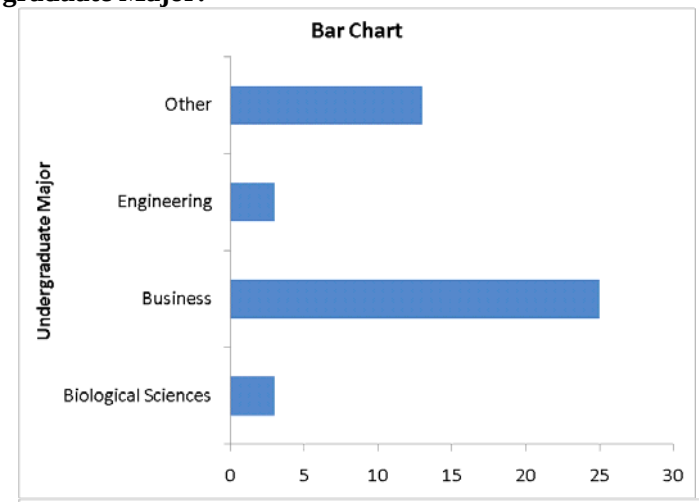


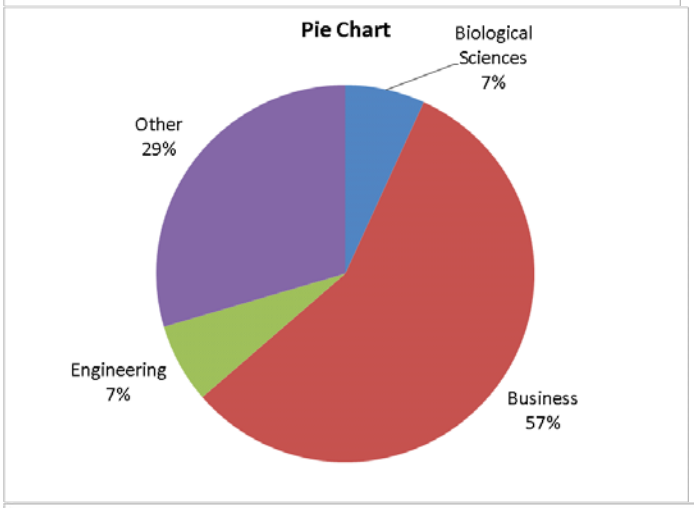


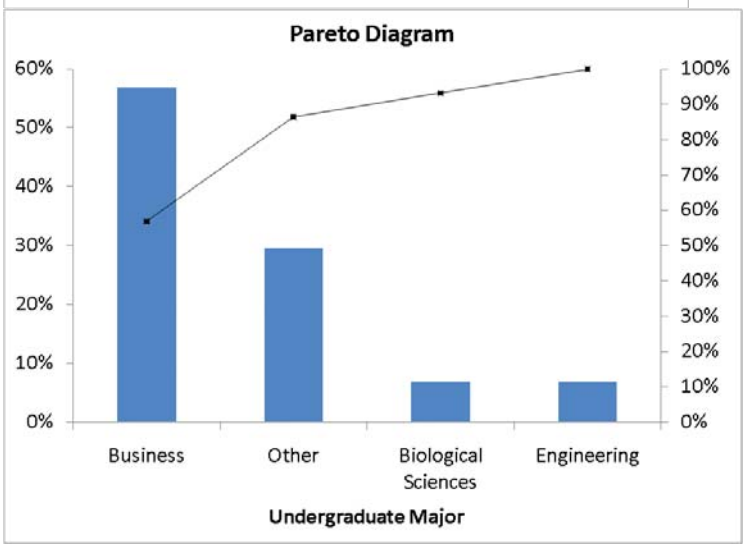


The "vital few" of economics/finance, management, and marketing/retailing account for more than 70% of the graduate majors.

2. 105
cont. **Undergraduate Major:**

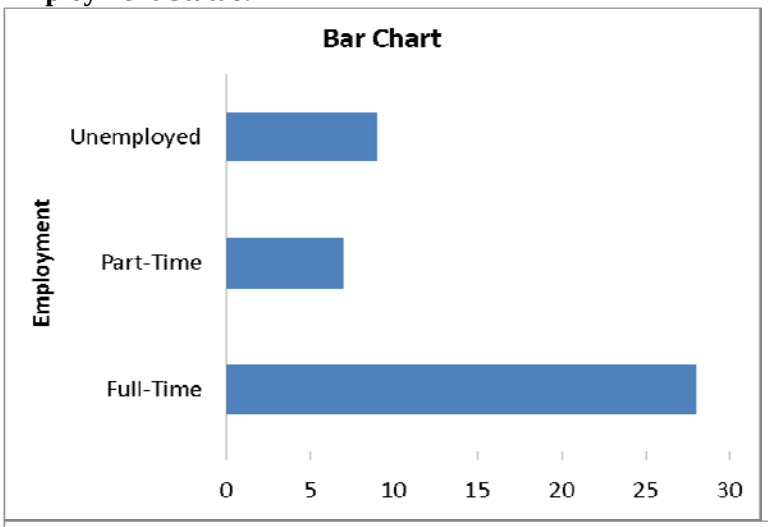


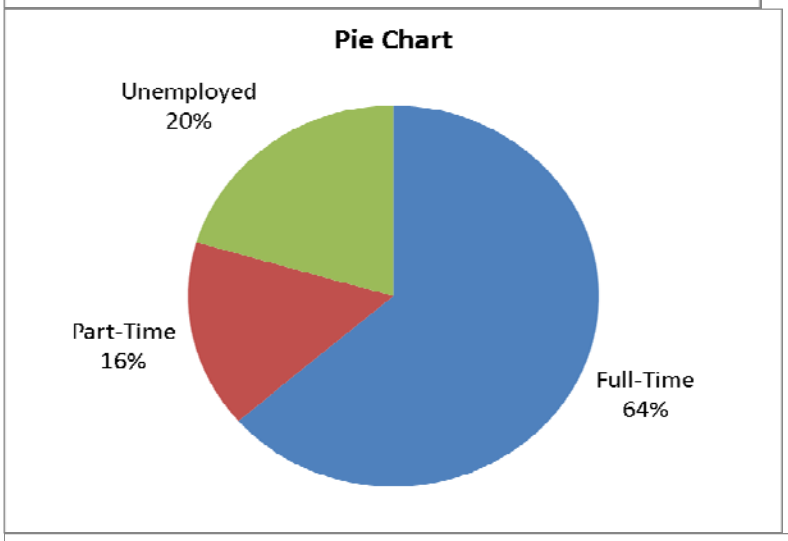


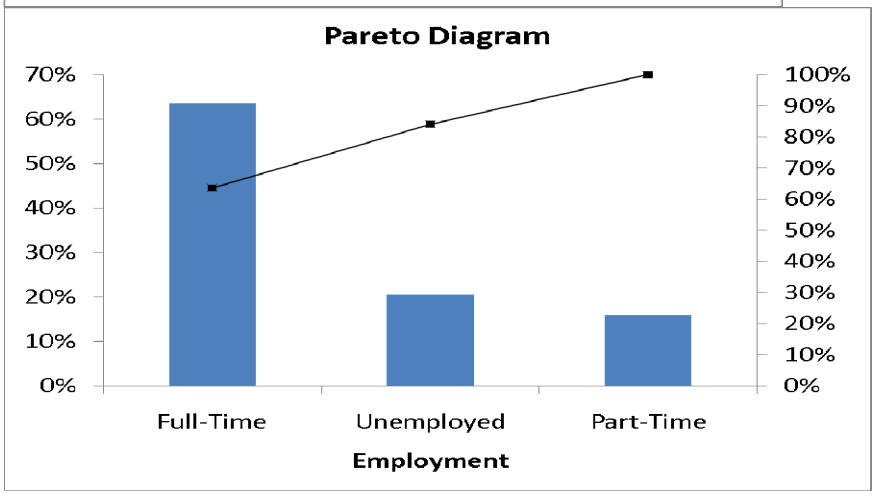


The "vital few" of business administration and other account for more than 80% of the undergraduate majors.

2.105
cont. **Employment Status:**



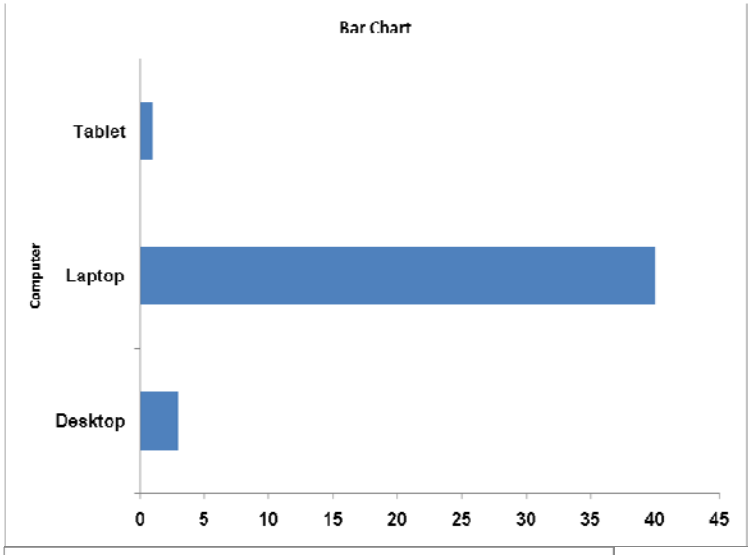


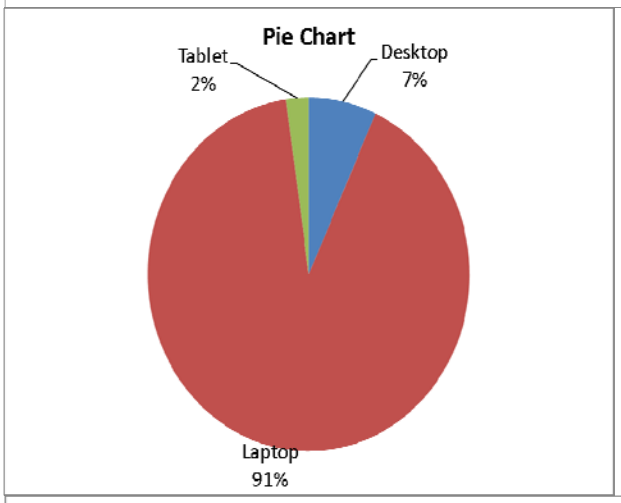


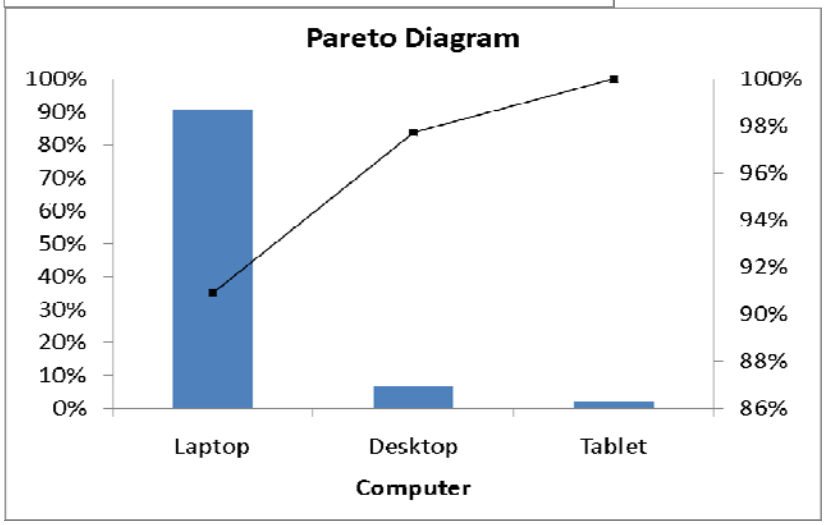
"Full-time" employment status accounts for more than 60% of the students.

2.105 cont.

. **Computers:**







More than 90% of the students use laptop computer for their studies.

2.105
cont.

Age:

Stem-and-Leaf Display	
Stem unit	1
21	0 0
22	0 0 0 0 0 0
23	0 0 0
24	0 0 0 0 0 0 0
25	0 0 0 0
26	0 0 0 0 0 0 0 0
27	0 0 0 0 0
28	
29	0 0 0
30	0
31	0
32	0
33	
34	
35	
36	
37	
38	
39	
40	
41	0
42	0
43	
44	
45	
46	
47	0

Majority of the students are between 22 and 27 years of age and the distribution of age is right-skewed.

2.105
cont.

Full-time Jobs:

Stem-and-Leaf Display	
Stem unit	1
0	0000
1	0000000000000000
2	00000000000000000
3	00000
4	000

Majority of the students have held between 1 and 2 jobs in the past 10 years.

Spending:

Stem-and-Leaf Display	
Stem unit	100
0	78
1	4555789
2	000235555
3	0000000000000556
4	00000
5	06
6	08
7	
8	
9	
10	
11	
12	
13	
14	
15	0
16	
17	
18	
19	
20	
21	
22	0

Majority of the students spend between $100 and $500 on textbooks and supplies.

2.105
cont.

Advisory Rating:

Stem-and-Leaf Display	
Stem unit	1
2	0
3	0 0 0
4	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6	0 0 0 0 0 0
7	0

Majority of the advisory service ratings is between 4 and 5.

Text Message:

Stem-and-Leaf Display	
Stem unit	100
0	0 0 0 0 0 0 1 1 1 1 2 3 5 5 6 6 6 8 8
1	0 0 0 0 0
2	0 0 0 1 1 5
3	0 5
4	0 0 0
5	0 0 3
6	0 5
7	
8	
9	
10	0 0
11	
12	5

Majority of the students sent less than 300 text message in a typical week.

2.105
cont.

Graduate GPA:

Stem-and-Leaf Display	
Stem unit	0.1
30	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
31	0
32	0 0
33	0
34	0
35	0
36	0
37	0 0 0 0
38	0 0
39	0 0
40	0 0 0 0 0 0

Majority of the students have a graduate GPA of 3.0.

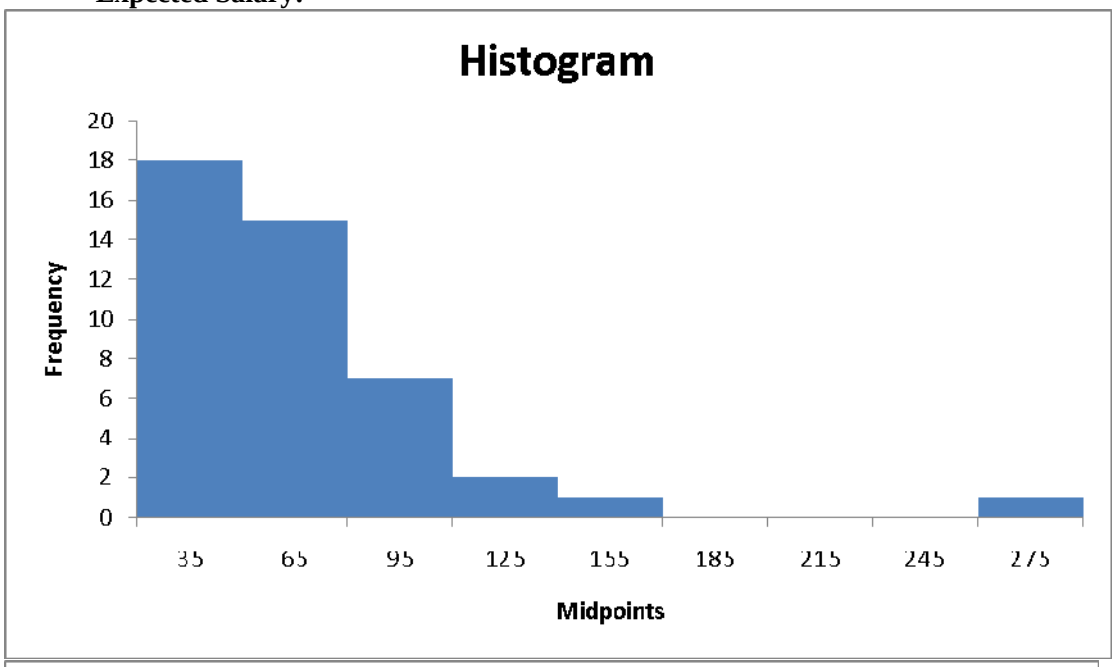
Undergraduate GPA:

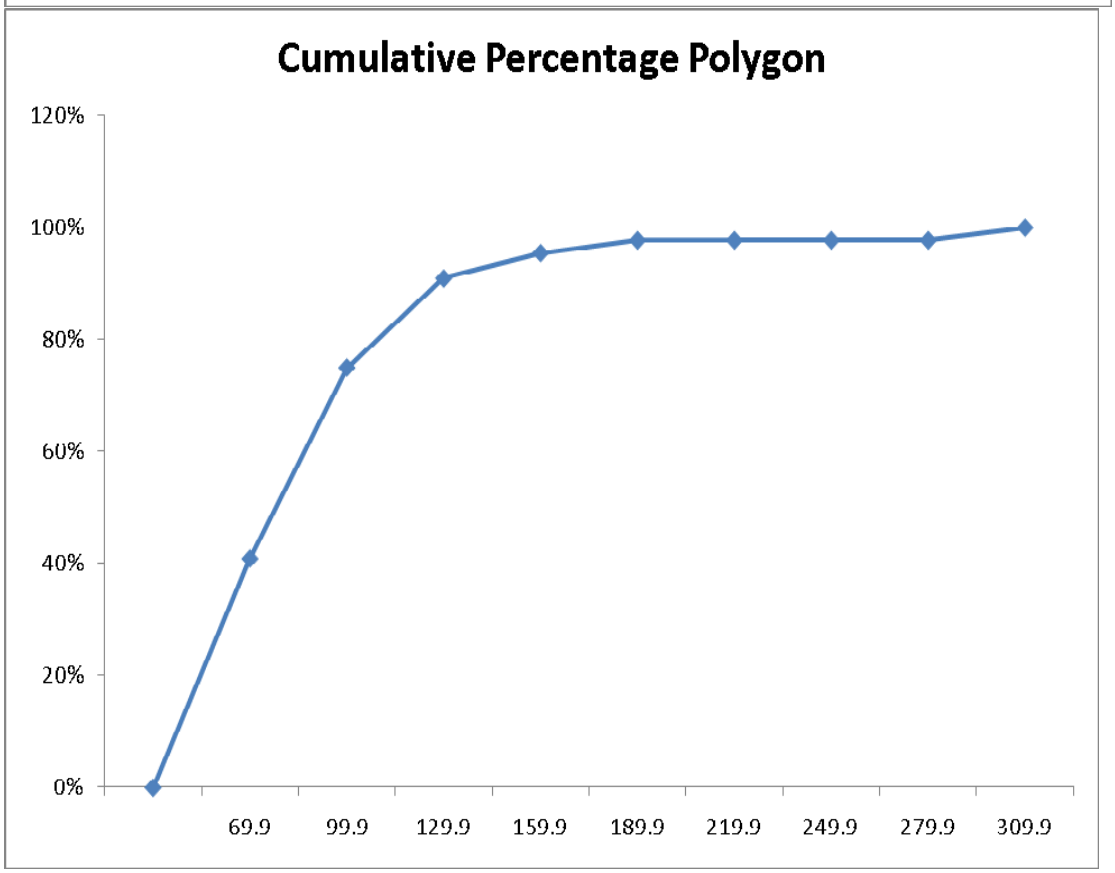
Stem-and-Leaf Displa	
Stem unit	0.1
28	0
29	0 0 0 0 0
30	0 0 0 0 0
31	0 0
32	0 0 0
33	0 0 0 0
34	0 0 0 0 0
35	0 0
36	0 0 0 0 0 0
37	0 0 0 0 0 0
38	0 0 0 0
39	0

The distribution of undergraduate GPA is quite symmetrical around 3.35.

2.105
cont.

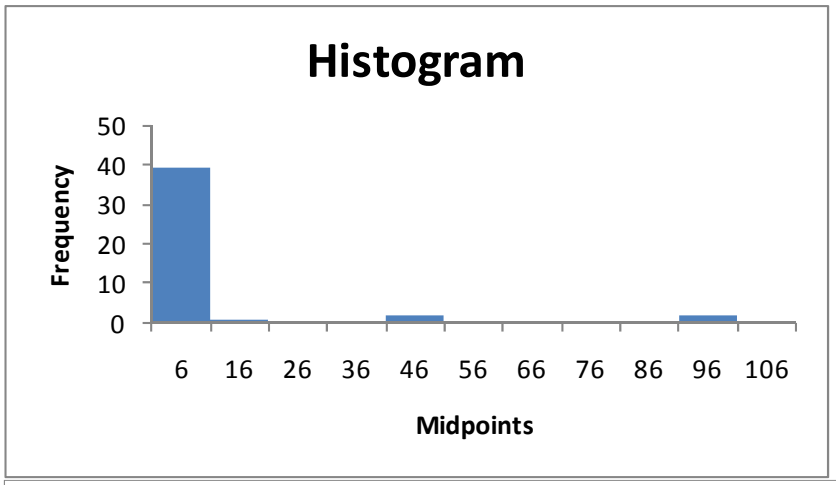
Expected Salary:

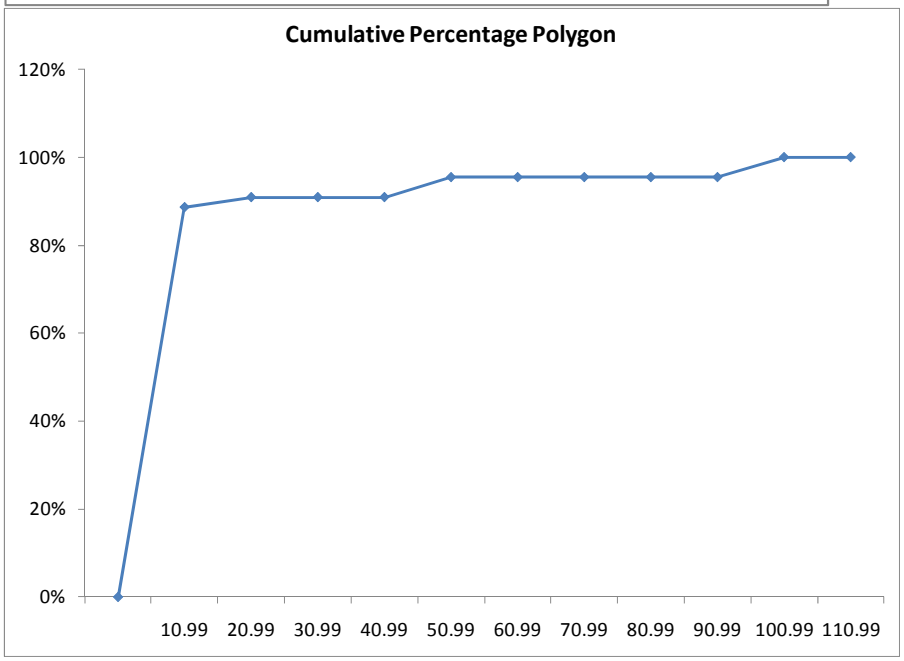




Expected salary is right-skewed.

2.105 cont. **Wealth:**





Wealth is right-skewed.