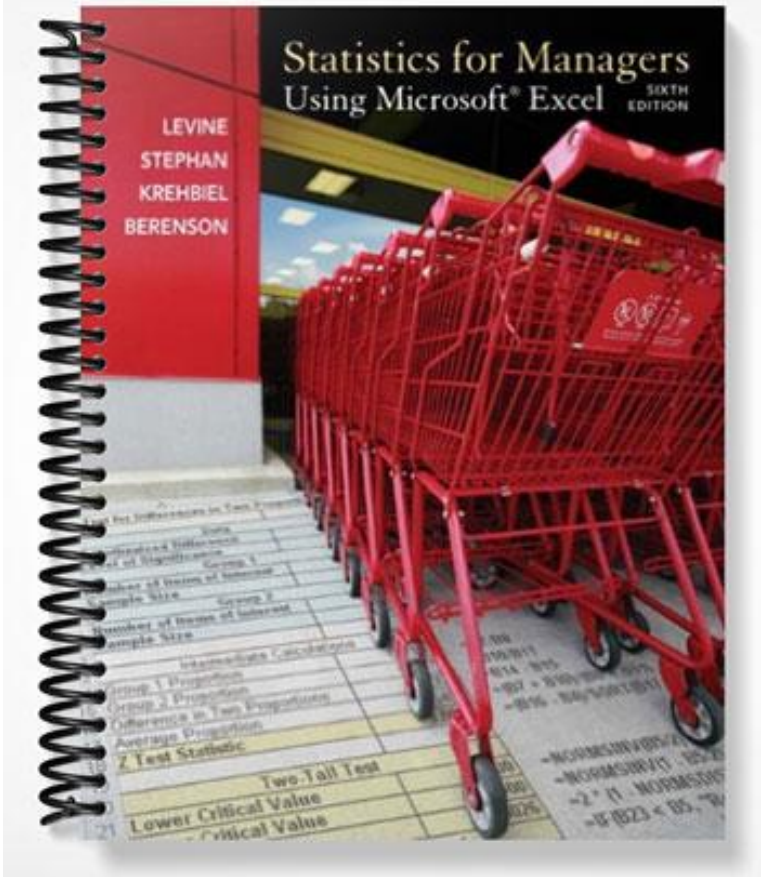
SOLUTIONS MANUAL



CHAPTER 2

2.1 (a) The types of beverages sold yield categorical or "qualitative" responses.
(b) The types of beverages sold yield distinct categories in which no ordering is implied.

2.2 Three sizes of soft drink are classified into distinct categories—small, medium, and large—in which order is implied.

2.3 (a) The time it takes to download a video from the Internet yields numerical or "quantitative" responses.
(b) The download time is a ratio scaled variable because the true zero point in the measurement is zero units of time.

2.4 (a) The number of telephones is a numerical variable that is discrete because the outcome is a count. It is ratio scaled because it has a true zero point.
(b) The length of the longest long-distance call is a numerical variable that is continuous because any value within a range of values can occur. It is ratio scaled because it has a true zero point.
(c) Whether there is a cell phone in the household is a categorical variable because the answer can be only yes or no. This also makes it a nominal-scaled variable.
(d) Same answer as in (c).

2.5 (a) numerical, continuous, ratio scale
(b) numerical, discrete, ratio scale
(c) categorical, nominal scale
(d) categorical, nominal scale

2.6 (a) categorical, nominal scale
(b) numerical, continuous, ratio scale
(c) numerical, discrete, ratio scale
(d) numerical, discrete, ratio scale

2.7 (a) numerical, continuous, ratio scale *
(b) categorical, nominal scale
(c) categorical, nominal scale
(d) numerical, discrete, ratio scale
*Some researchers consider money as a discrete numerical variable because it can be "counted."

2.8 (a) numerical, continuous, ratio scale *
(b) numerical, discrete, ratio scale
(c) numerical, continuous, ratio scale *
(d) categorical, nominal
*Some researchers consider money as a discrete numerical variable because it can be "counted."

2.9 (a) Income may be considered discrete if we "count" our money. It may be considered continuous if we "measure" our money; we are only limited by the way a country's monetary system treats its currency.

(b) The first format is preferred because the responses represent data measured on a higher scale.

2.10 The underlying variable, ability of the students, may be continuous, but the measuring device, the test, does not have enough precision to distinguish between the two students.

2.11 (a) The population is "all working women from the metropolitan area." A systematic or random sample could be taken of women from the metropolitan area. The director might wish to collect both numerical and categorical data.

(b) Three categorical questions might be occupation, marital status, type of clothing. Numerical questions might be age, average monthly hours shopping for clothing, income.

2.12 (a) (i) categorical, (ii) categorical, (iii) numerical discrete, (iv) categorical

(b) e.g. "Where do you usually purchase your cat food?"

(c) e.g. "How many cats do you have in your household?"

2.13 (a) (i) numerical, (ii) categorical, (iii) categorical, (iv) numerical

(b) (i) discrete, (iv) continuous*

*Some researchers consider money as a discrete numerical variable because it can be "counted."

2.14 The answers to this question depend on which data set is being selected.

2.15 The answers to this question depend on which data set is being selected.

2.16 The answer to this question depends on which top story is being selected.

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

2.18 The information presented here is based on data distributed by an organization, i.e., the U.S. Census Bureau.

2.19 (a)

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

(b) Category "B" is the majority.

2.20 (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%

2.21 (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.

2.22 (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 19.61%.

2.23 (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.

2.24 (a) Table of total percentages

	Gender		
Enjoy Shopping for Clothing	Male	Female	Total
Yes	27%	45%	72%
No	21%	7%	28%
Total	48%	52%	100%

Table of row percentages

	Gender		
Enjoy Shopping for Clothing	Male	Female	Total
Yes	38%	62%	100%
No	74%	26%	100%
Total	48%	52%	100%

Table of column percentages

	Gender		
Enjoy Shopping for Clothing	Male	Female	Total
Yes	57%	86%	72%
No	43%	14%	28%
Total	100%	100%	100%

(b) The percentage of shoppers who enjoy shopping for clothing is higher among females than males.

2.25 (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.

2.26 The percentage of MBA and undergraduate students who choose the lowest cost fund and the second-lowest cost fund is about the same. A higher percentage of MBA students choose the third-lowest cost fund while a higher percentage of undergraduate students choose the highest cost fund.

2.27 Ordered array: 63 64 68 71 75 88 94

2.28 Ordered array: 73 78 78 78 85 88 91

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

2.30 (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.

2.31 (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.

2.32 (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.

2.33 (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.

2.34 (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.

2.35 (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.

2.36 (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

2.36 (b) cont.

Bulb Life (hrs)	A		B	
	Percentage	Cumulative %	Percentage	Cumulative %
650 – 749	7.50%	7.50%	.00%	0.00%
750 – 849	12.50%	20.00%	.00%	5.00%
850 – 949	50.00%	70.00%	0.00%	25.00%
950 – 1049	22.50%	92.50%	0.00%	65.00%
1050 – 1149	7.50%	100.00%	2.50%	87.50%
1150 – 1249	0.00%	100.00%	2.50%	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

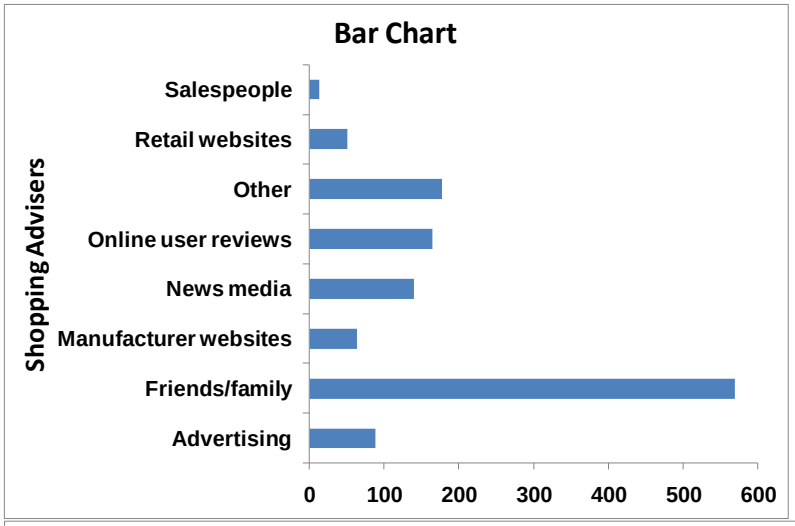
2.37 (a)

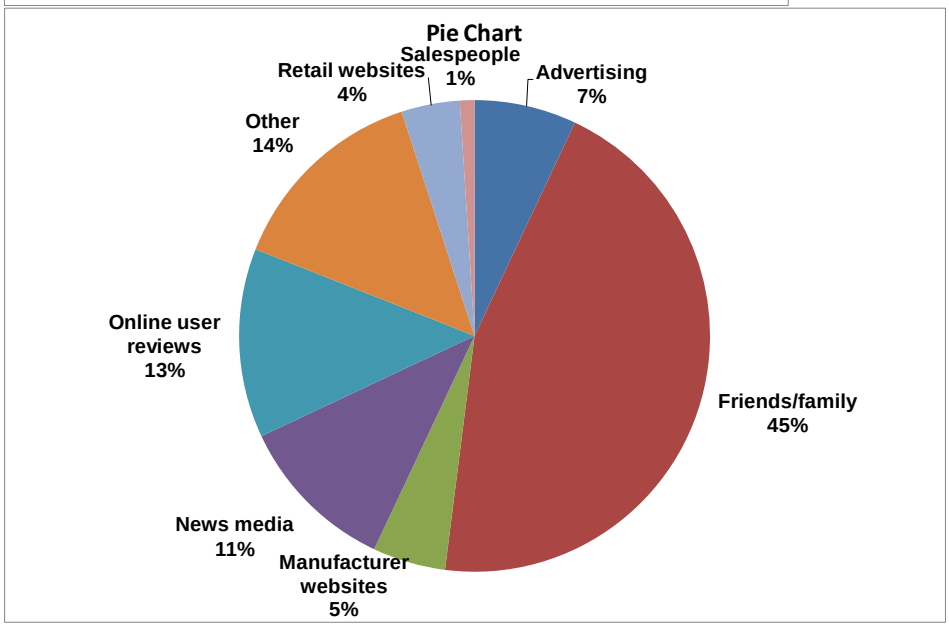
Amount of Soft Drink	Frequency	Percentage
1.850 – 1.899	1	2%
1.900 – 1.949	5	10
1.950 – 1.999	18	36
2.000 – 2.049	19	38
2.050 – 2.099	6	12
2.100 – 2.149	1	2

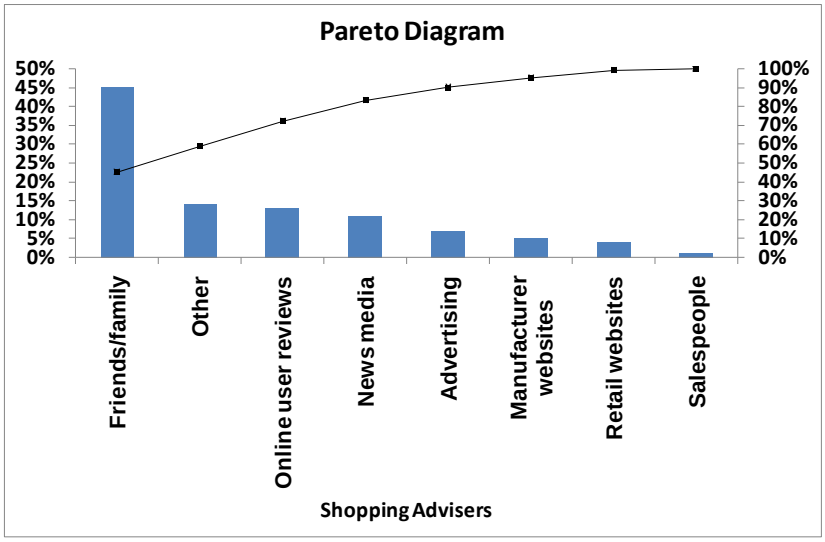
Amount of Soft Drink	Frequency Less Than	Percentage Less Than
1.85 – 1.89	1	2%
1.90 – 1.94	6	12
1.95 – 1.99	24	48
2.00 – 2.04	43	86
2.05 – 2.09	49	98
2.10 – 2.14	50	100

(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.

2.38 (a)



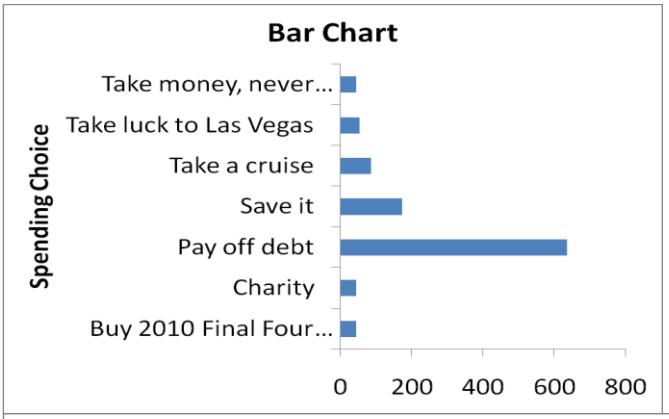


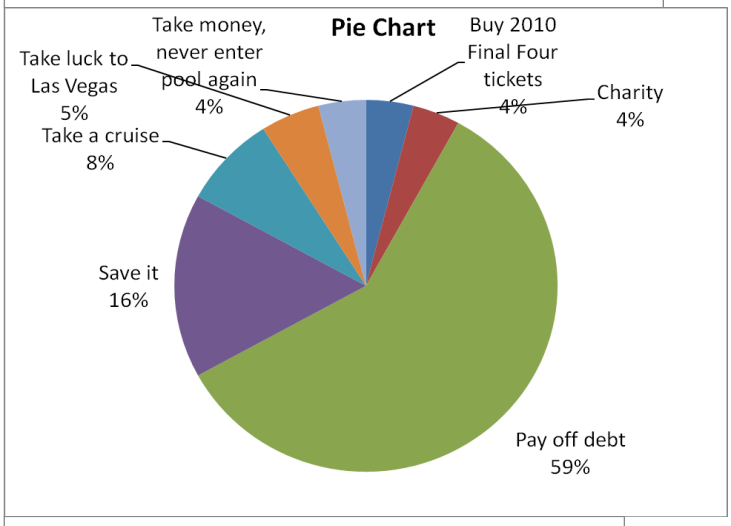


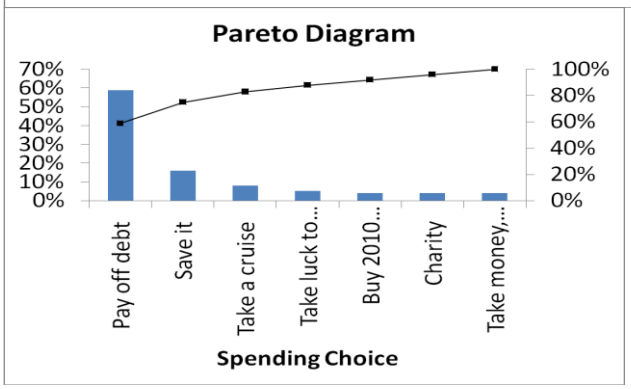
2.38 cont. (b) The Pareto diagram is better than the pie chart to portray these data because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

(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%.

2.39 (a)



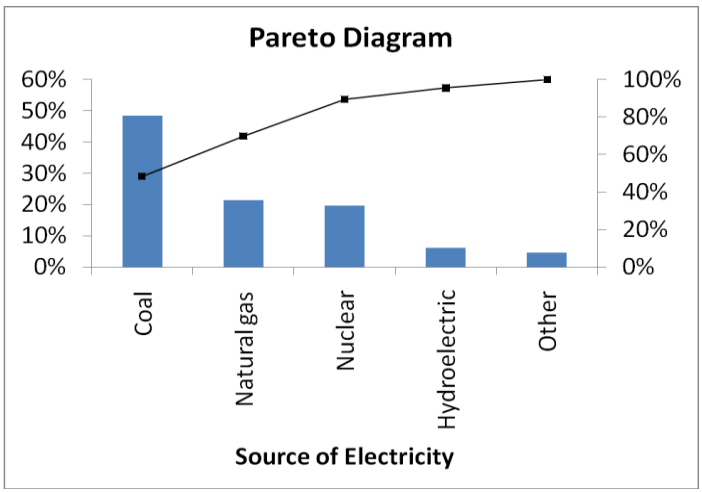




(b) The Pareto diagram is better than the pie chart or the bar chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

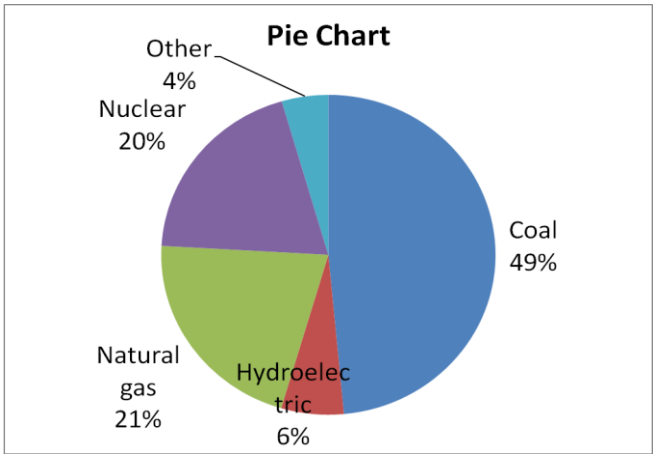
(c) From the Pareto diagram, it is obvious that more than 50% would pay off their debt with $1 million.

2.40 (a)



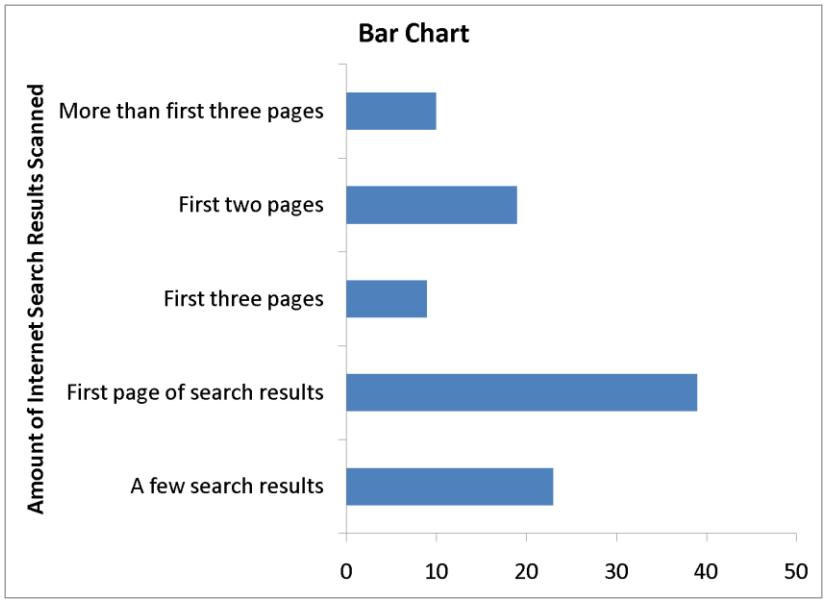
(b) From the Pareto chart, about 90% of power is derived from coal, nuclear, or natural gas.

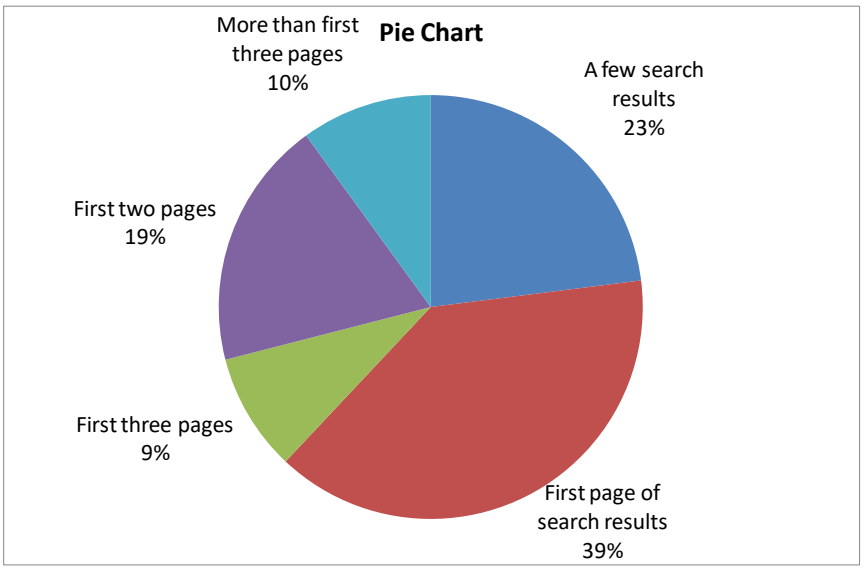
(c)



(d) The Pareto chart allows you to see which sources account for most of the electricity.

2.41 (a)



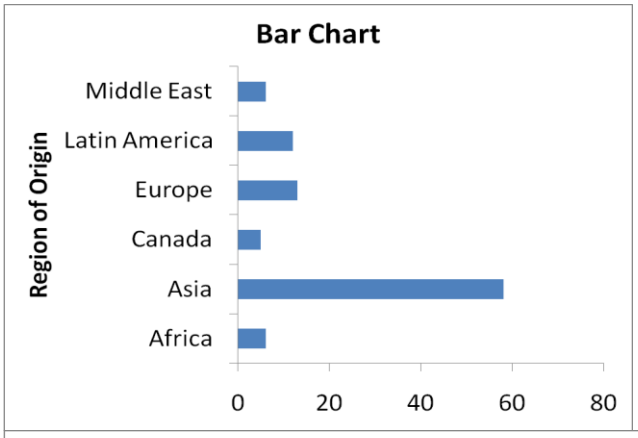


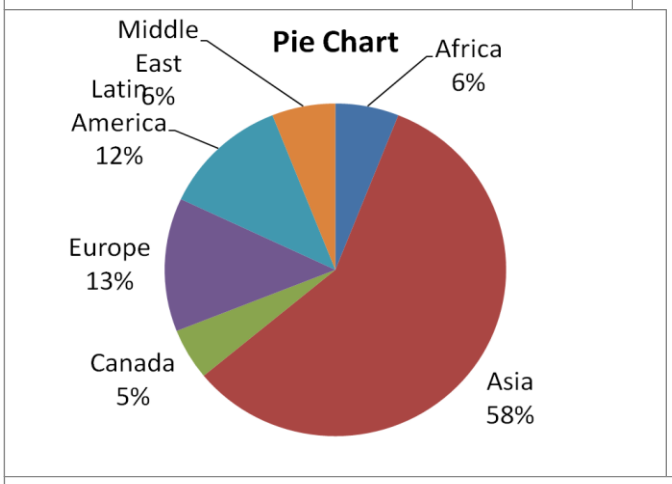
(b) The bar chart is more suitable if the purpose is to compare the categories. The pie chart is more suitable if the main objective is to investigate the portion of the whole that is in a particular category. *

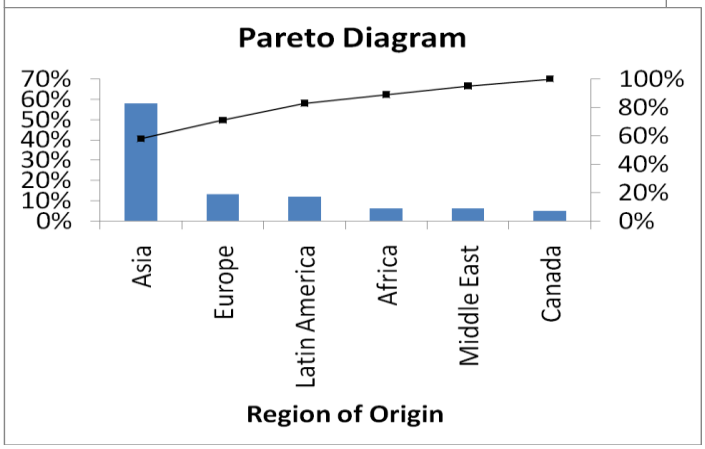
* Note: This is one of the many possible solutions for the question.

(c) You can conclude that most of the people (39%) scan Internet search results according to the "first page of search results", followed by "a few search results" (23%) and "first two pages" (19%).

2.42 (a)



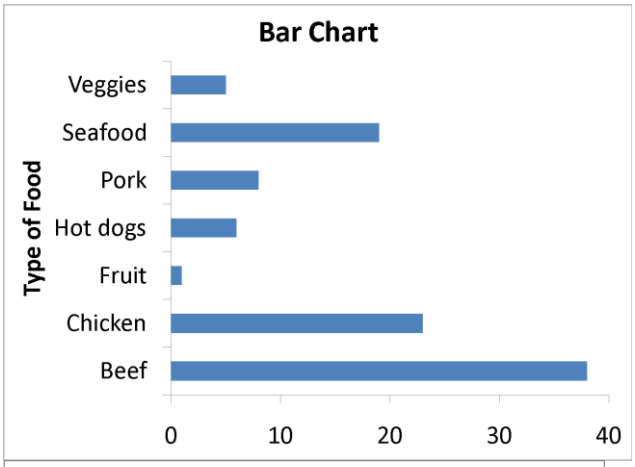


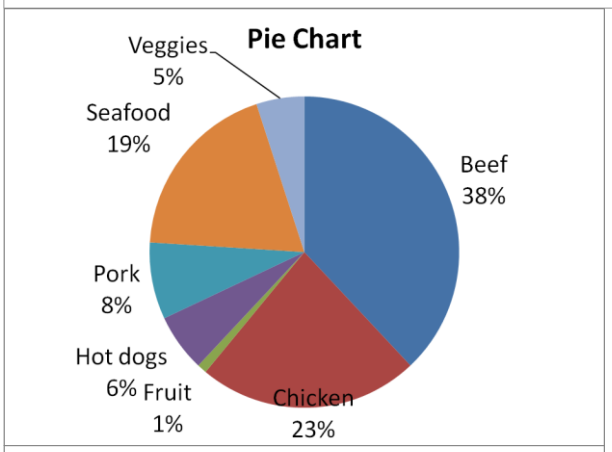


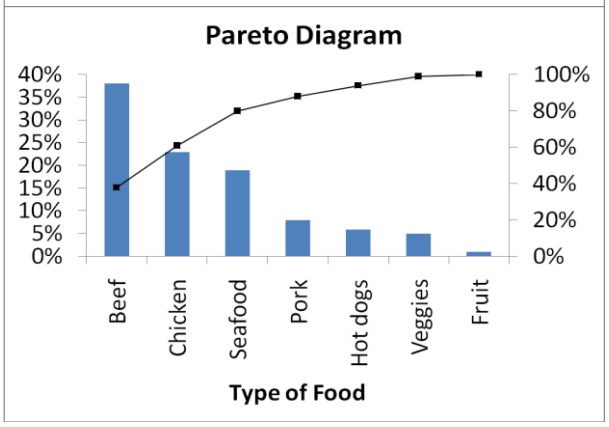
(b) Because a large percentage of students are from Asia, the Pareto chart allows you to focus on this dominant group. Almost sixty percent are from Asia. Including Asia, Europe and the Latin America represents 83% of all the foreign students.

(c) From the Pareto chart, almost half of the foreign students studying at the U.S. colleges are from Asia.

2.43 (a)



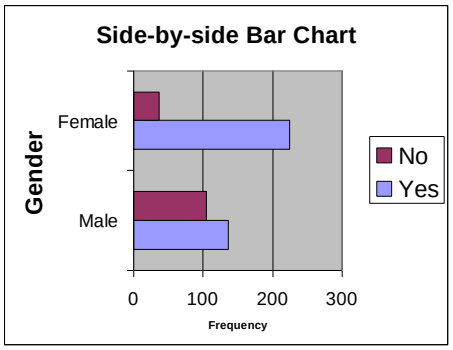




(b) The Pareto diagram is better than the pie chart because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

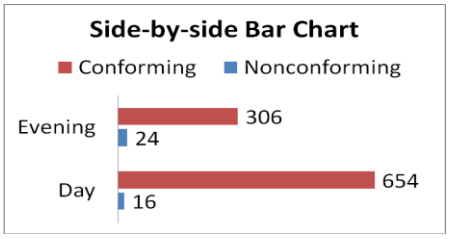
(c) From the Pareto chart, beef, chicken and seafood make up 80% of what folks want sizzling on the grill during barbecue season.

2.44 (a)



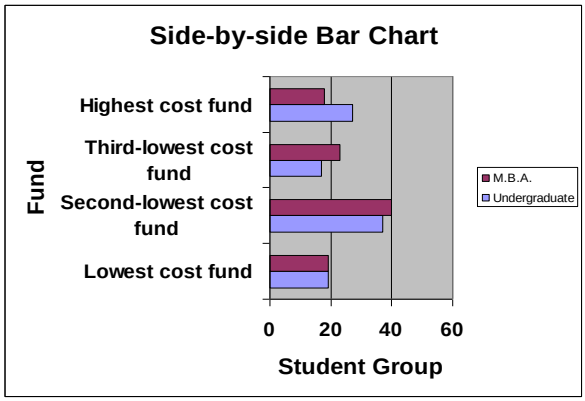
(b) A higher percentage of females enjoy shopping for clothing.

2.45 (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.

2.46 (a)



(b) The percentage of MBA and undergraduate students who choose the lowest-cost fund and the second-lowest-cost fund is about the same. A higher percentage of MBA students chose the third-lowest cost fund whereas a higher percentage of undergraduate students chose the highest cost fund.

2.47 Stem-and-leaf of Finance Scores

5	34
6	9
7	4
8	0
9	38

2.48 Ordered array: 50 74 74 76 81 89 92

2.49 (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 11.9 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.

2.50 (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.

2.51 (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

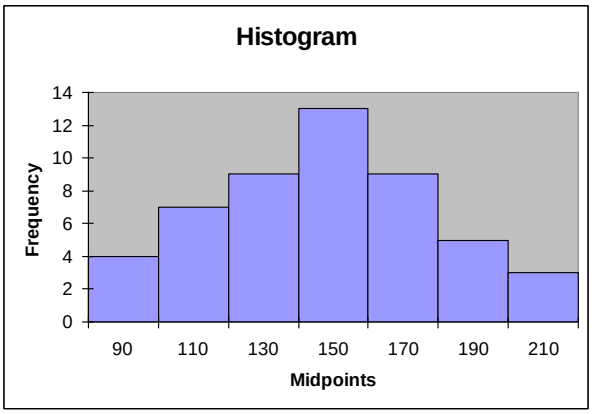
Stem unit: 0.1

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**

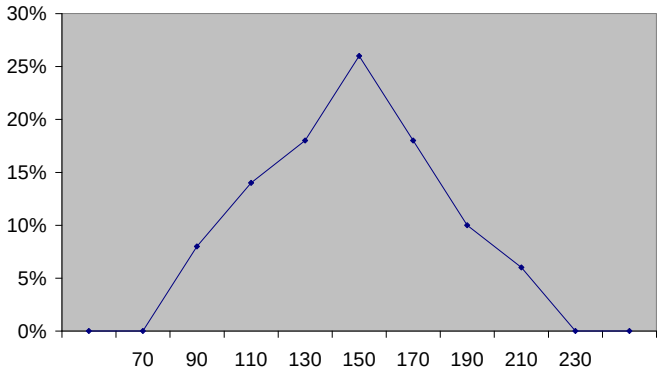
(c) The stem-and-leaf display conveys more information than the ordered array. We can more readily determine the arrangement of the data from the stem-and-leaf display than we can from the ordered array. We can also obtain a sense of the distribution of the data from the stem-and-leaf display.

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

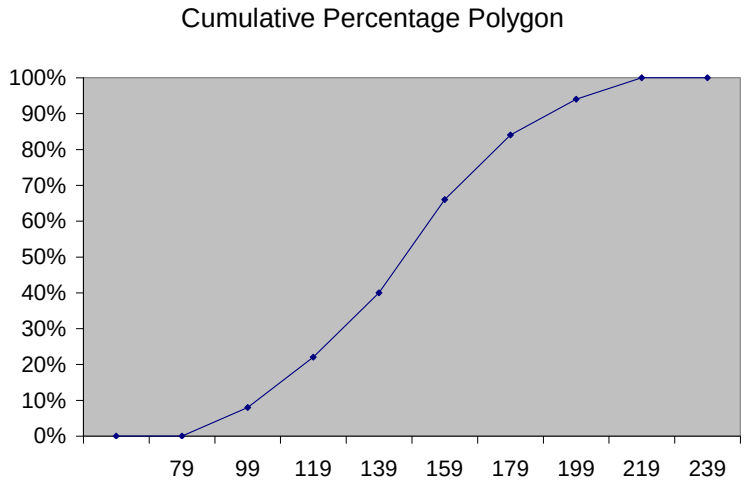
2.52 (a)



Percentage Polygon



2.52 (b)
cont.

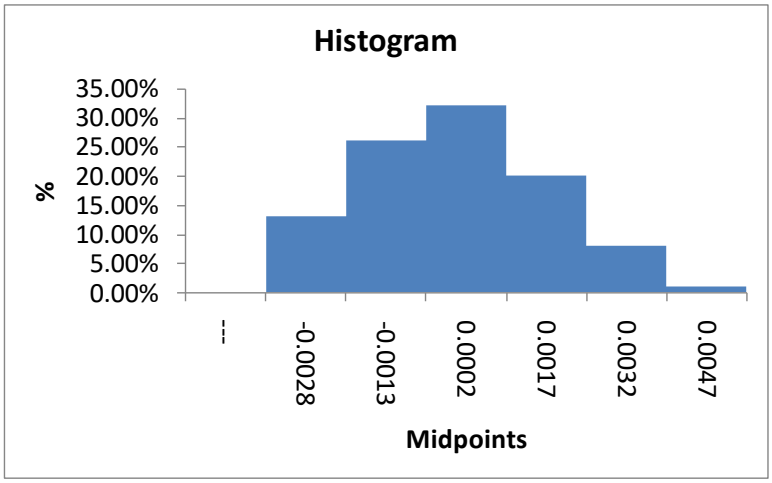


(c) The majority of utility charges are clustered between $120 and $180.

2.53 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.

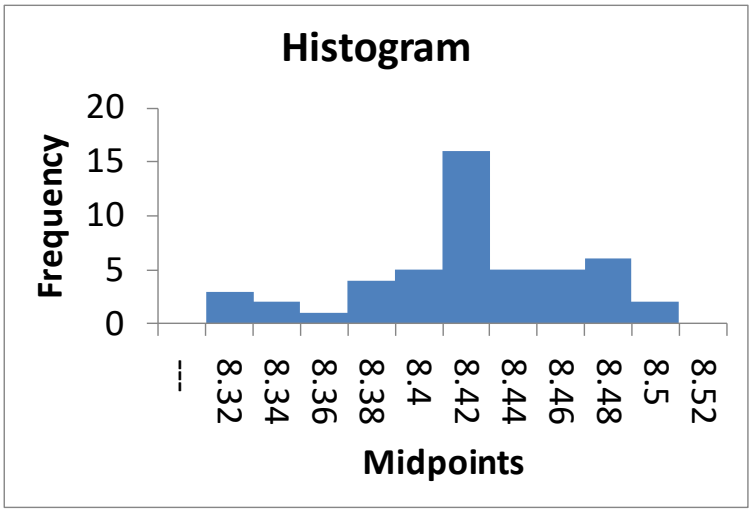
2.54 The property taxes per capita appear to be right-skewed with approximately 90% falling between $399 and $1,700, and the remaining 10% fall between $1,700 and $2,100. The center is at about $1,000.

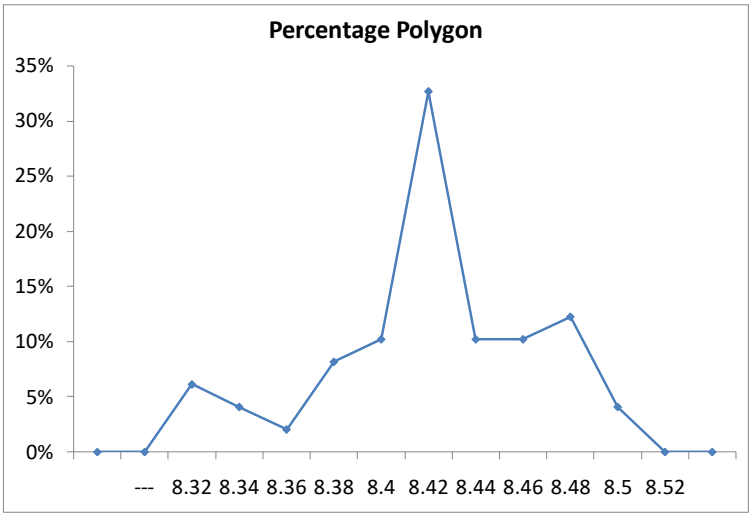
2.55 (a)



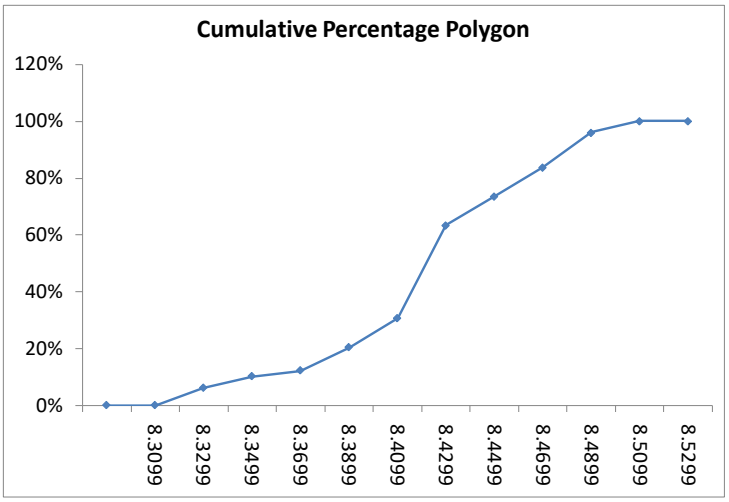
(b) Yes, the steel mill is doing a good job at meeting the requirement as there is only one steel part out of a sample of 100 that is as much as 0.005 inches longer than the specified requirement.

2.56 (a)



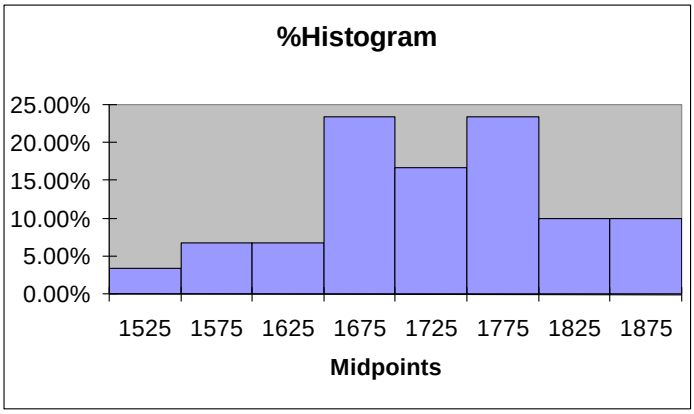


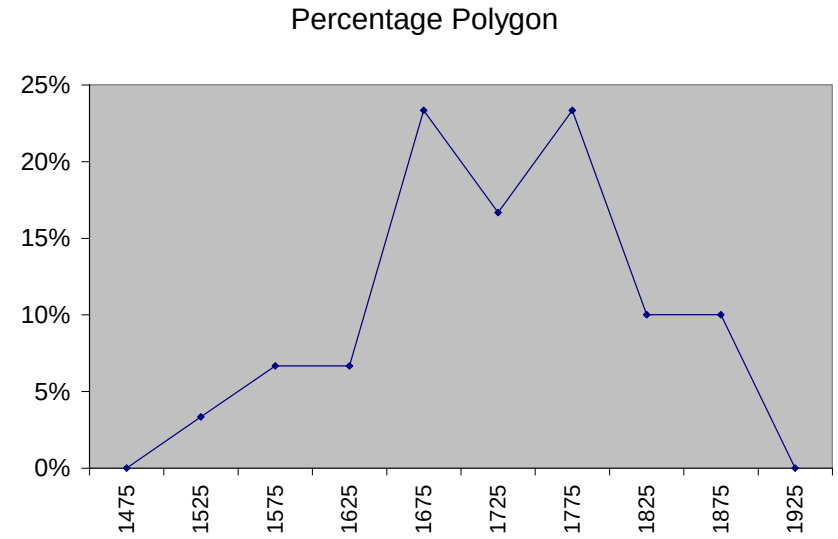
(b)



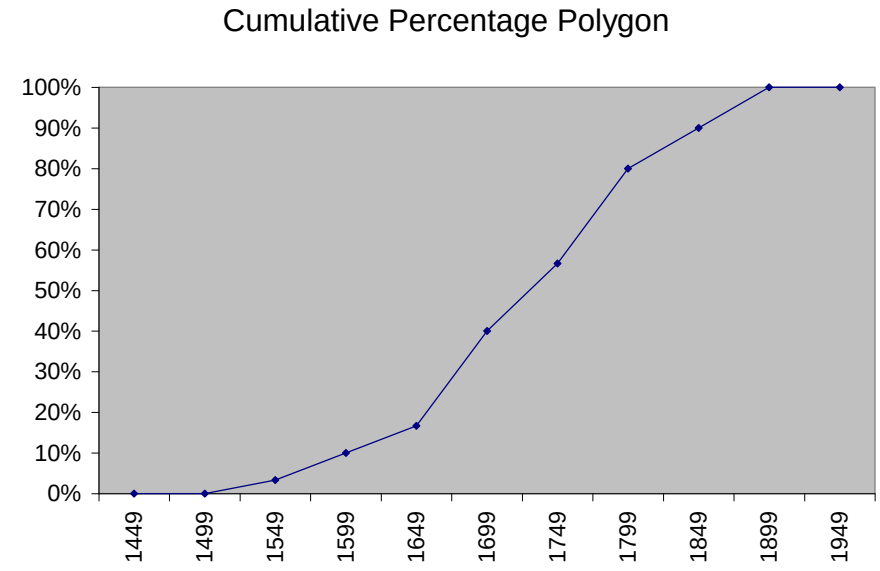
(c) All the troughs will meet the company's requirements of between 8.31 and 8.61 inches wide.

2.57 (a)



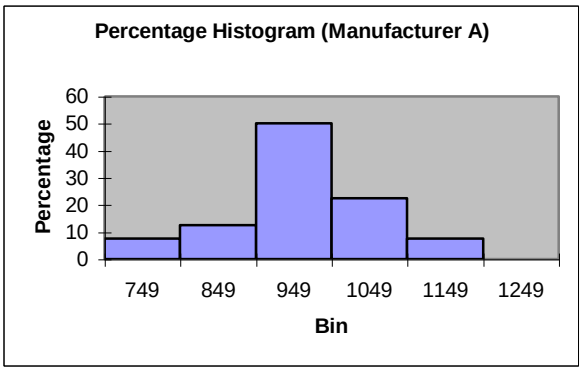


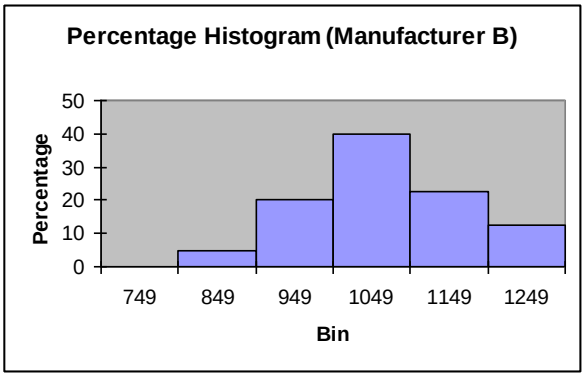
(b)

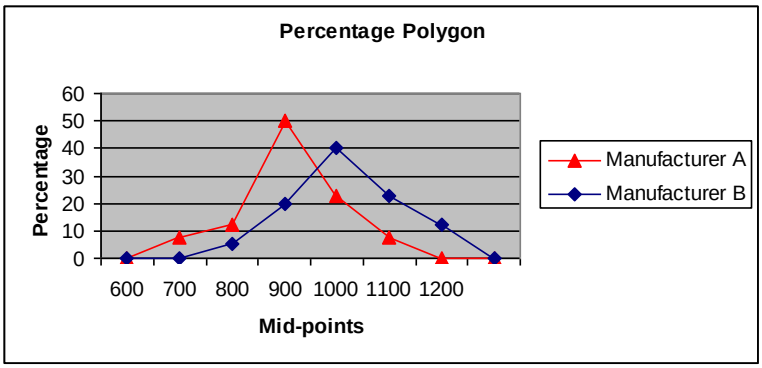


(c) The strength of all the insulators meets the company's requirement of at least 1500.

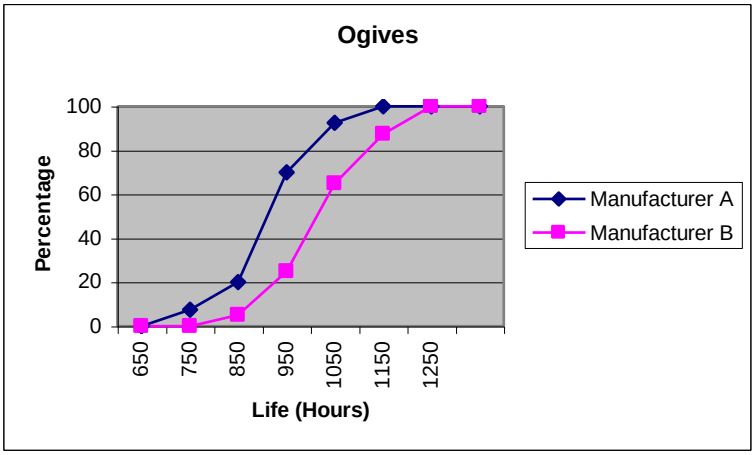
2.58 (a)





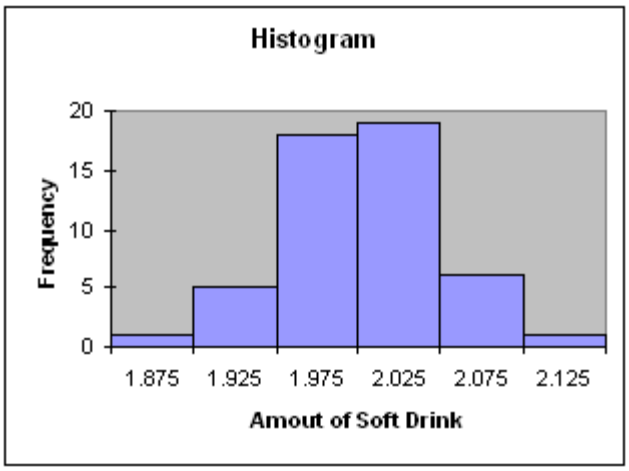


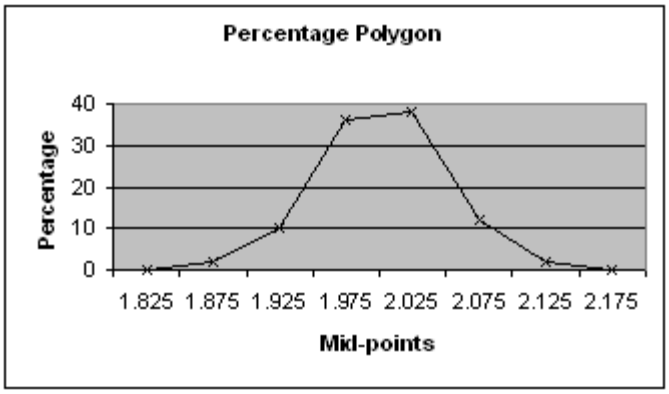
(b)



2.58 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.

2.59 (a)

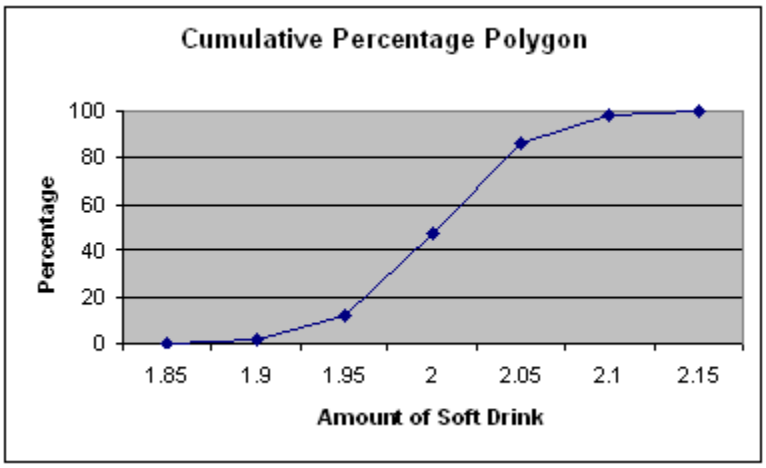




(b)

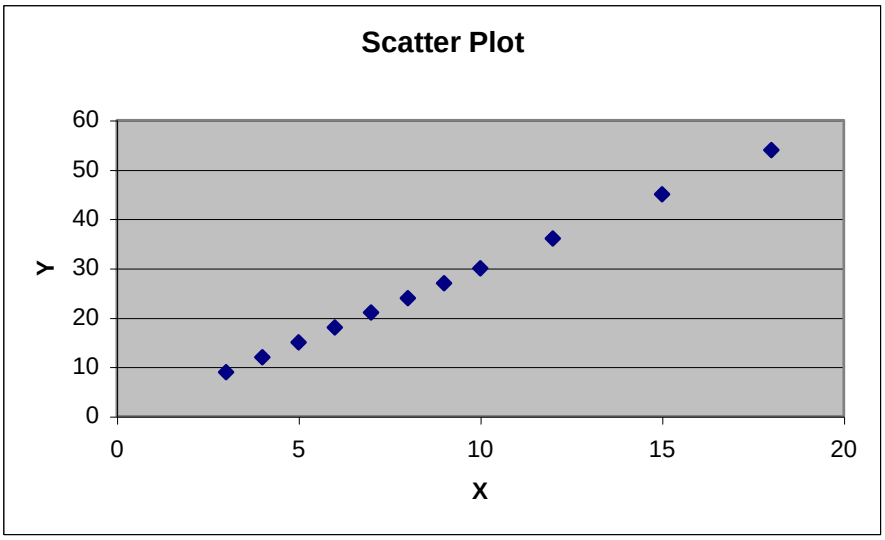
Amount of Soft Drink	Frequency Less Than	Percentage Less Than
1.85 – 1.89	1	2%
1.90 – 1.94	6	12
1.95 – 1.99	24	48
2.00 – 2.04	43	86
2.05 – 2.09	49	98
2.10 – 2.14	50	100

2.59 cont. (b)



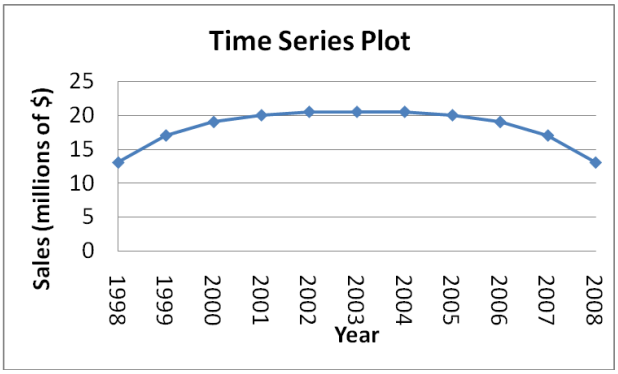
(c) The amount of soft drink filled in the two liter bottles is most concentrated in two intervals on either side of the two-liter mark, from 1.950 to 1.999 and from 2.000 to 2.049 liters. Almost three-fourths of the 50 bottles sampled contained between 1.950 liters and 2.049 liters.

2.60 (a)



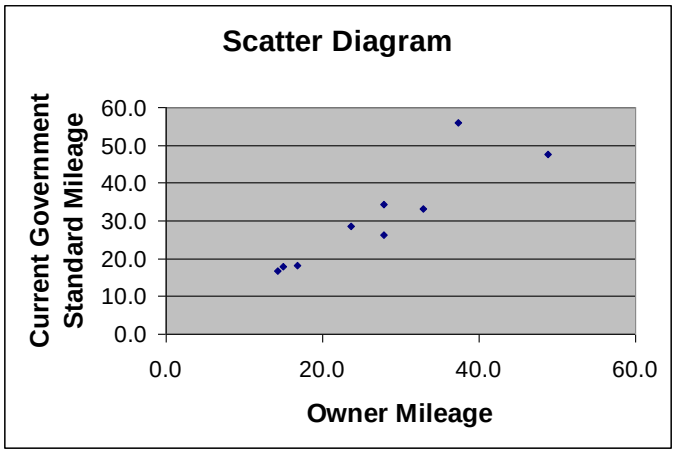
(b) Yes, there is a strong positive relationship between *X* and *Y*. As *X* increases, so does *Y*.

2.61 (a)



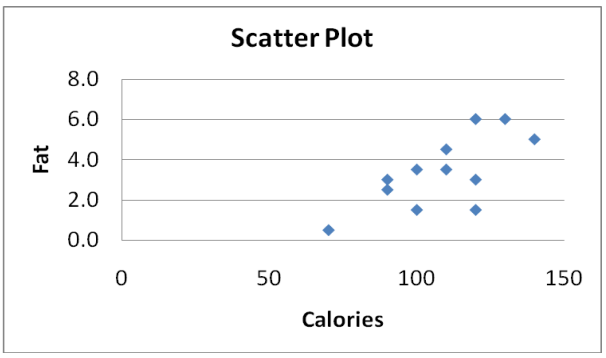
(b) Annual sales appear to be increasing in the earlier years before 2002 but start to decline after 2004.

2.62 (a)



(b) There is a positive relationship between owner mileage and current government standard mileage.

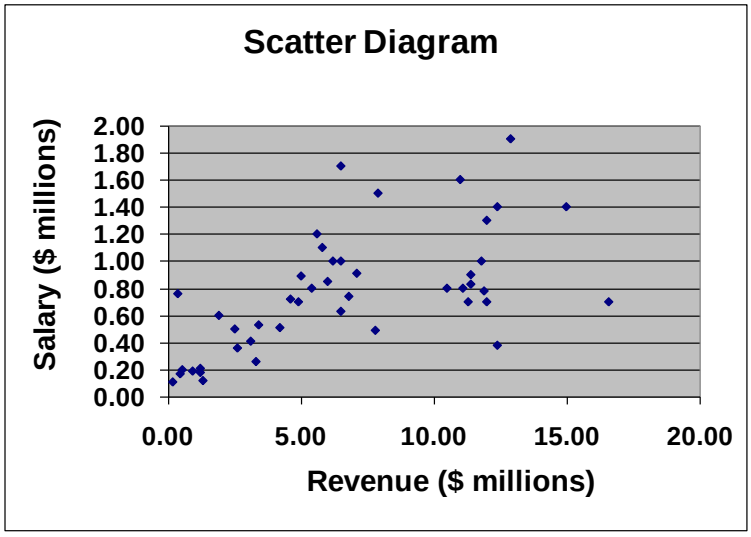
2.63 (a)



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

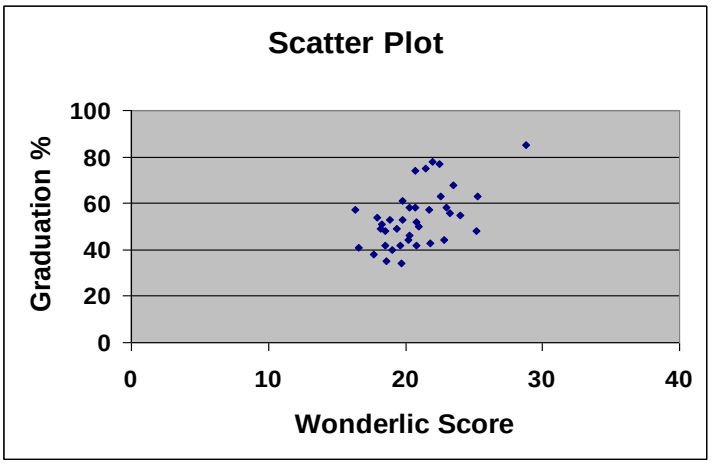
2.64 (a) Yes, schools with higher revenues will also have higher coach's salaries.

(b)



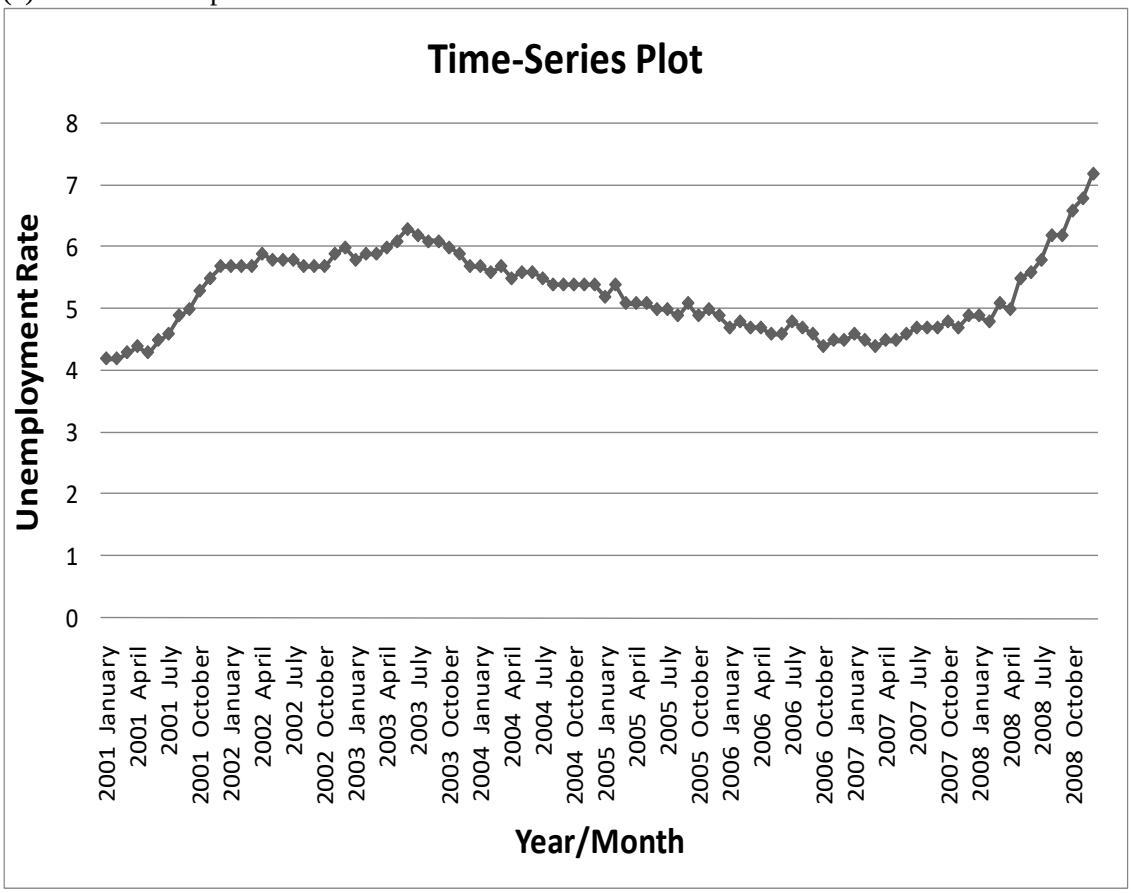
(c) There appears to be a positive relationship between coaches' salary and revenue. Yes, this is borne out by the data.

2.65 (a)



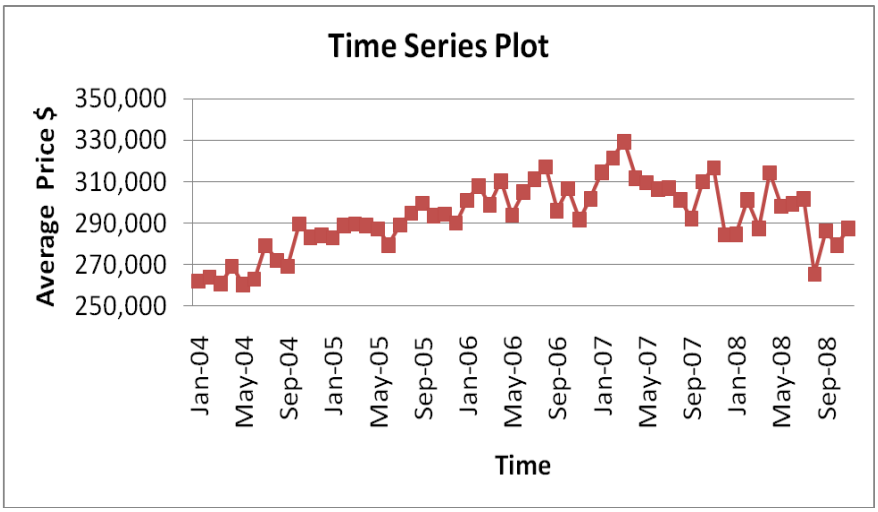
(b) There is a positive relationship between Wonderlic score and graduation rate.

2.66 (a) Excel output:



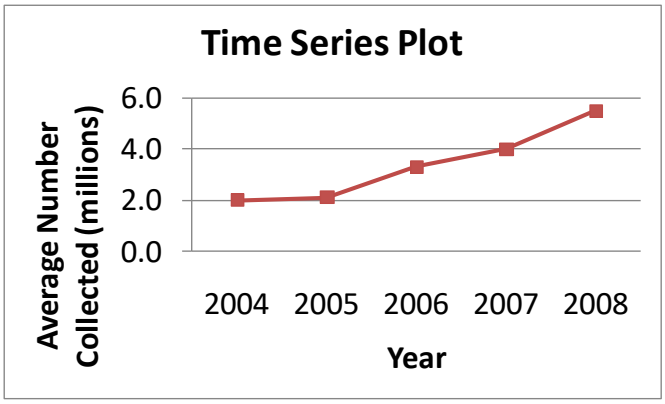
(b) The unemployment rate trended upward and leveled off at around 6% by December 2001. Around October 2003, it started to trend downward and reached about 4.5% by December 2006 before staying between 4.5% and 5% in 2007. It then trended upward and reached 7.2% in December 2008.

2.67 (a)



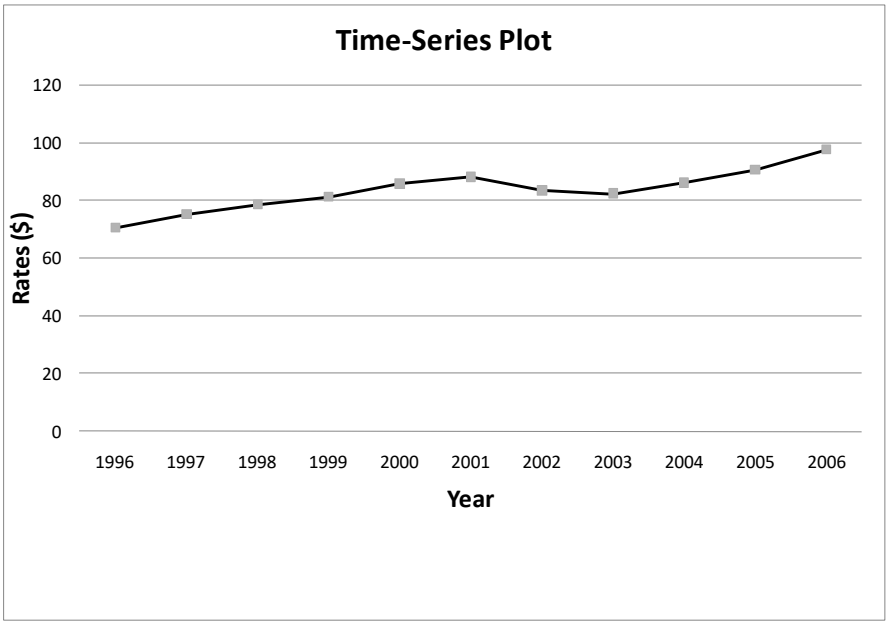
(b) There does not appear to be any obvious pattern present in the data.

2.68 (a)



(b) There is an obvious increasing trend from 2004 to 2008 with a sharp increase in 2008.
(c) You would predict about 6.5 to 7 million in 2009.

2.69 (a)



(b) The rates have a cyclical component and appear to be on the upswing in 2006.
(c) You would predict that the rate in 2007 will be around $110.

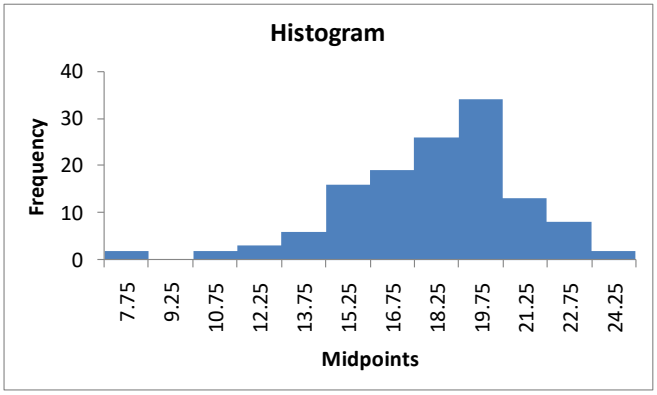
2.70 (a)

Count of Risk		Objective		
Category	Risk	Growth	Value	Grand Total
Large Cap	Average	97	77	174
	High	116	11	127
	Low	18	131	149
Large Cap Total		231	219	450
Mid Cap	Average	32	23	55
	High	76	10	86
	Low	3	30	33
Mid Cap Total		111	63	174
Small Cap	Average	11	71	82
	High	110	32	142
	Low	1	19	20
Small Cap Total		122	122	244
Grand Total		464	404	868

Count of Obj		Objective		
Category	Risk	Growth	Value	Grand Total
Large Cap	Average	11.18%	8.87%	20.05%
	High	13.36%	1.27%	14.63%
	Low	2.07%	15.09%	17.17%
Large Cap Total		26.61%	25.23%	51.84%
Mid Cap	Average	3.69%	2.65%	6.34%
	High	8.76%	1.15%	9.91%
	Low	0.35%	3.46%	3.80%
Mid Cap Total		12.79%	7.26%	20.05%
Small Cap	Average	1.27%	8.18%	9.45%
	High	12.67%	3.69%	16.36%
	Low	0.12%	2.19%	2.30%
Small Cap Total		14.06%	14.06%	28.11%
Grand Total		53.46%	46.54%	100.00%

(b) Large cap growth funds are very likely to be high risk while large cap value funds are very likely to be low risk. Mid cap growth funds are very likely to be high risk while mid cap value funds are very likely to be average or low risk. Small cap growth funds are very likely to be high risk while small cap value funds are likely to be high or average risk.

(c)



2.70 cont. (d) The 2006 return of the large cap, value, and low risk mutual funds is left-skewed with most of the returns concentrated around 19%. A few of the funds have a return as low as around 7.75%.

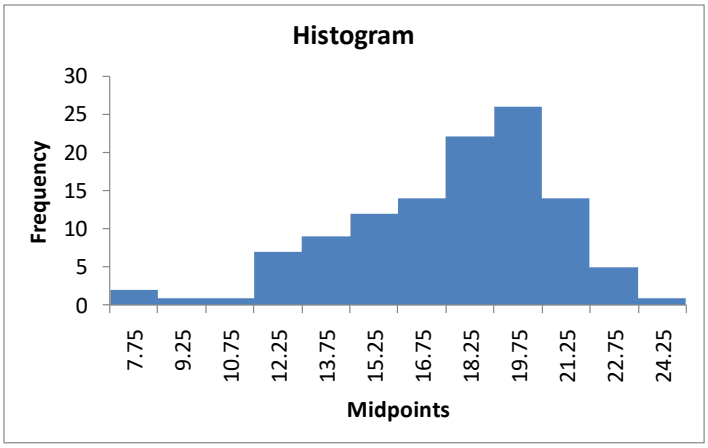
2.71 (a)

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.

(c)



(d) The 2006 return of the large cap, value, and no fee mutual funds is left-skewed with most of the returns concentrated around 19%. A few of the funds have a return as low as around 7.75%.

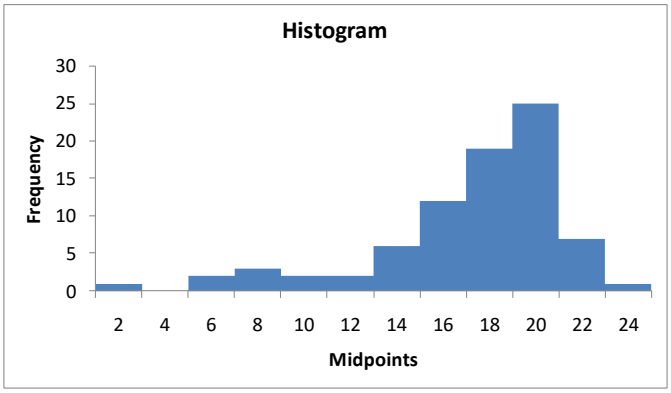
2.72 (a)

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.

(c)



(d) The 2006 return of the large cap, no fee and low risk mutual funds is left-skewed with most of the returns concentrated around 19%. A few of the funds have a return as low as around 2%.

2.73 (a)

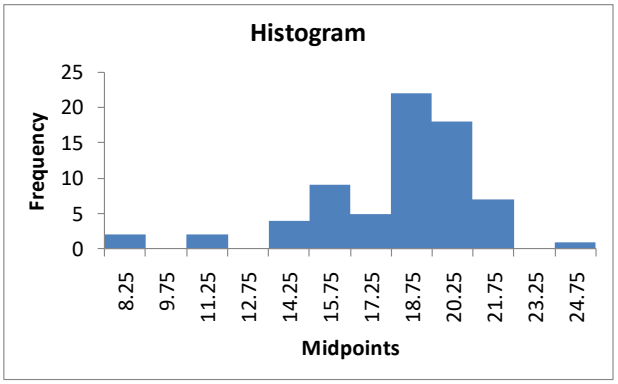
Count of Ris		Objective	Risk							
		Growth			Growth Total	Value			Value Total	Grand Total
Category	Fees	Average	High	Low		Average	High	Low		
Large Cap	No	59	68	10	137	36	8	70	114	251
	Yes	38	48	8	94	41	3	61	105	199
Large Cap Total		97	116	18	231	77	11	131	219	450
Mid Cap	No	22	36		58	11	5	23	39	97
	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 Ris		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 no fee; high risk, value and fee combinations that are dominated by the small cap.

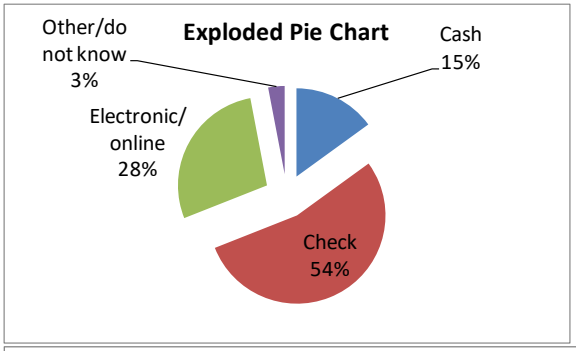
(c) The Pivot Tables in Problems 2.70-2.72 are easier to interpret because there are fewer combinations.

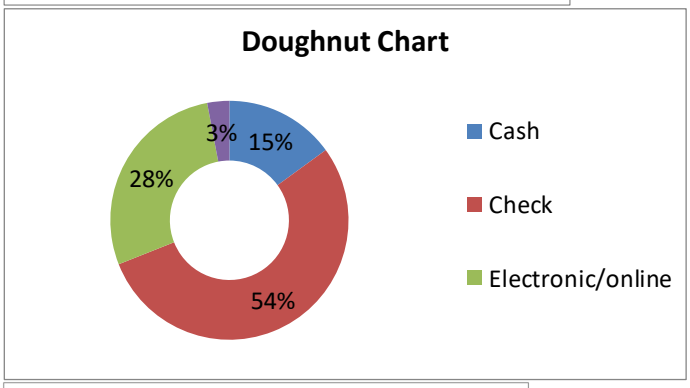
(d)

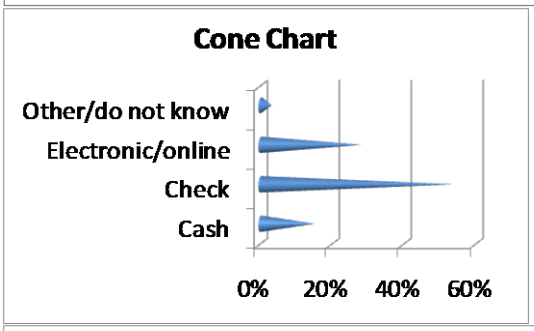


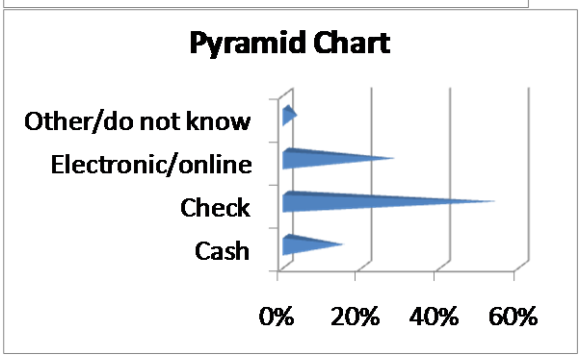
(e) The 2006 return of the large cap, value, no fee and low risk mutual funds is left-skewed with most of the returns concentrated around 19%. A few of the funds have a return as low as around 8.25% while two has a return as high as around 24.75%.

2.81 (a)



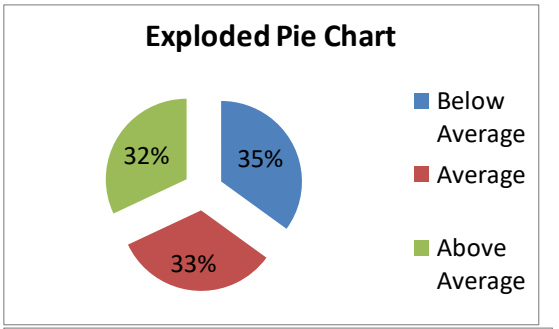


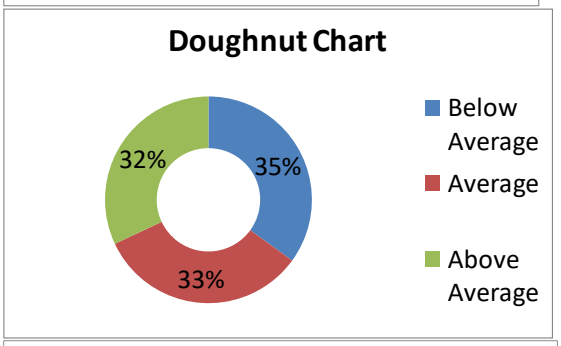


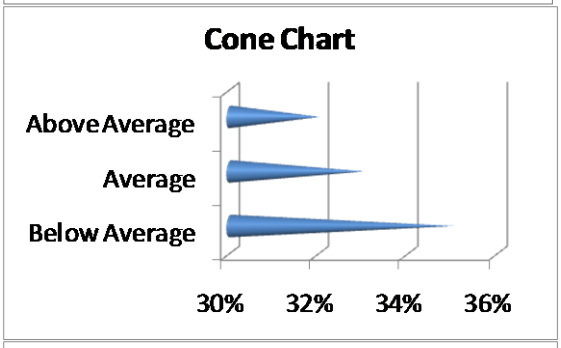


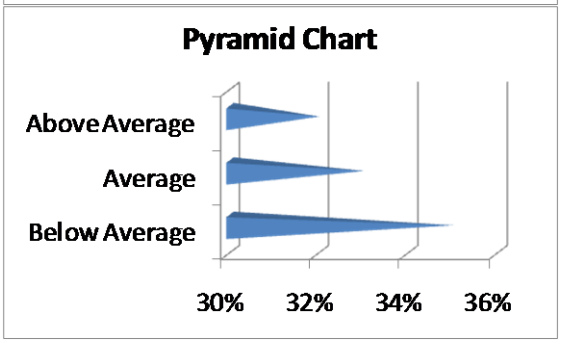
(b) The bar chart and the pie chart should be preferred over the exploded pie chart, doughnut chart, the cone chart and the pyramid chart since the former set is simpler and easier to interpret.

2.82 (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.

2.83 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.

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

2.85 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.

2.86 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.

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

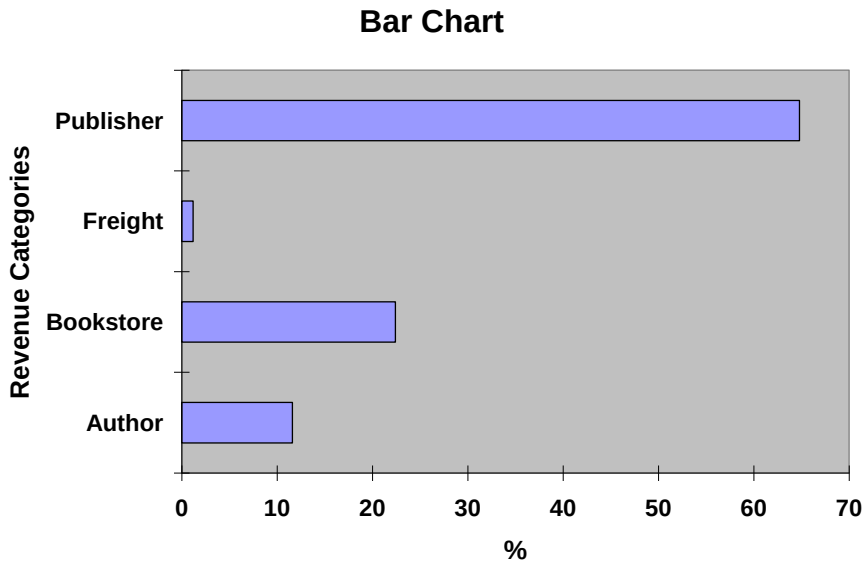
2.88 Because the categories are arranged according to frequency or importance, it allows the user to focus attention on the categories that have the greatest frequency or importance.

2.89 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.

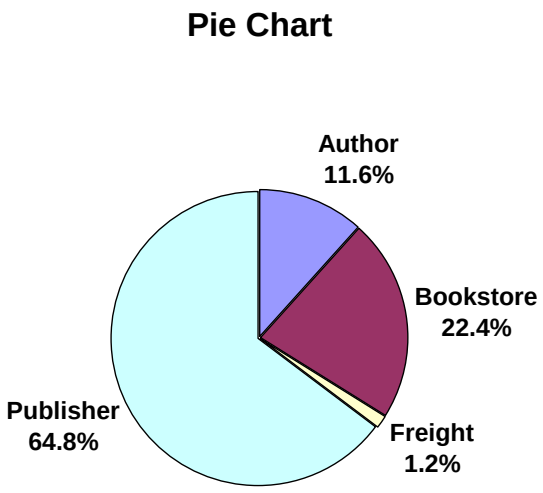
2.90 A contingency table contains information on two categorical variables whereas a Pivot Table can display information on more than two categorical variables.

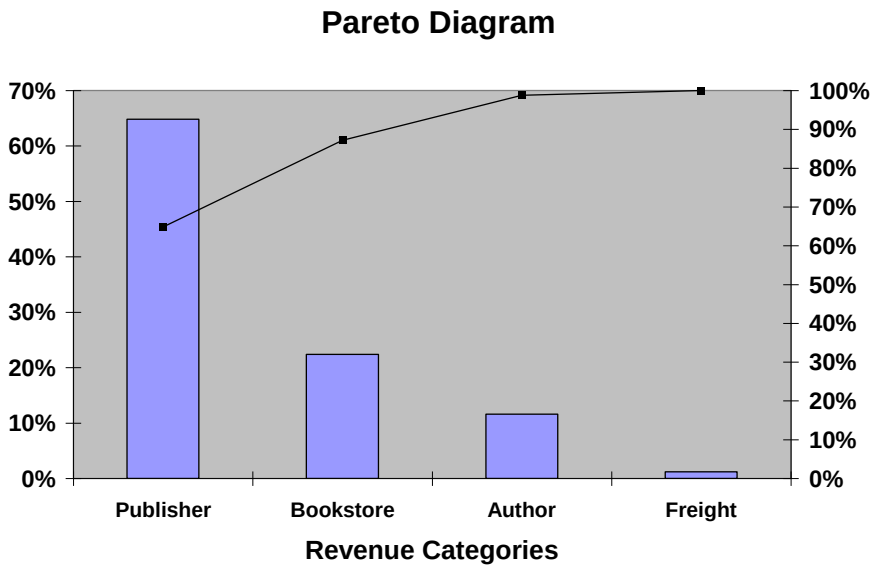
2.91 The multidimensional PivotTable can reveal additional patterns that cannot be seen in the a contingency table. One can also change the statistic displayed and compute descriptive statistics which can add insight into the data.

2.92 (a)

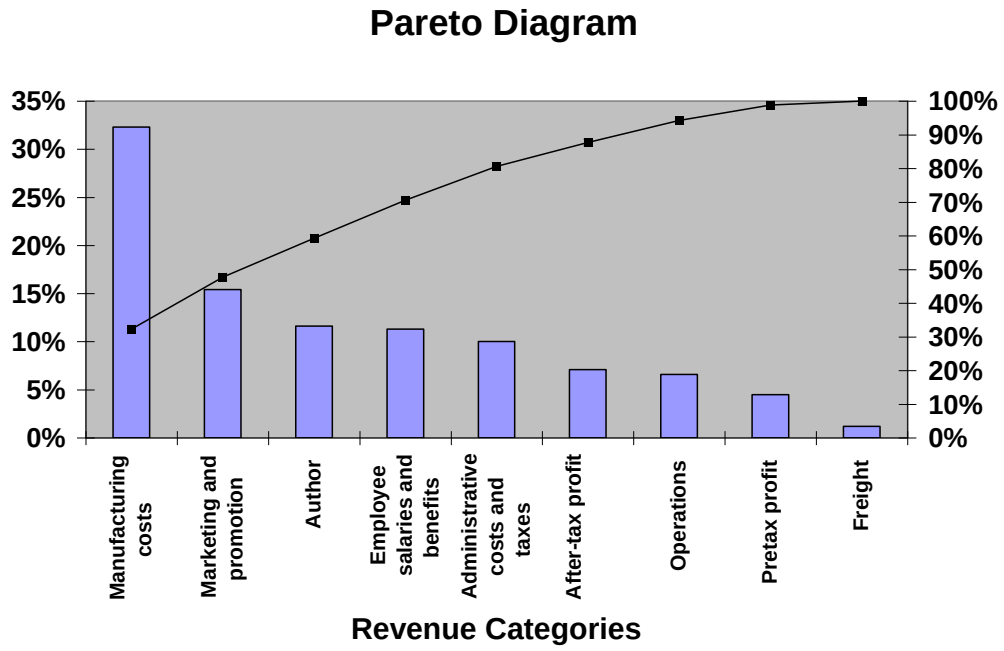


2.92 (a) cont.



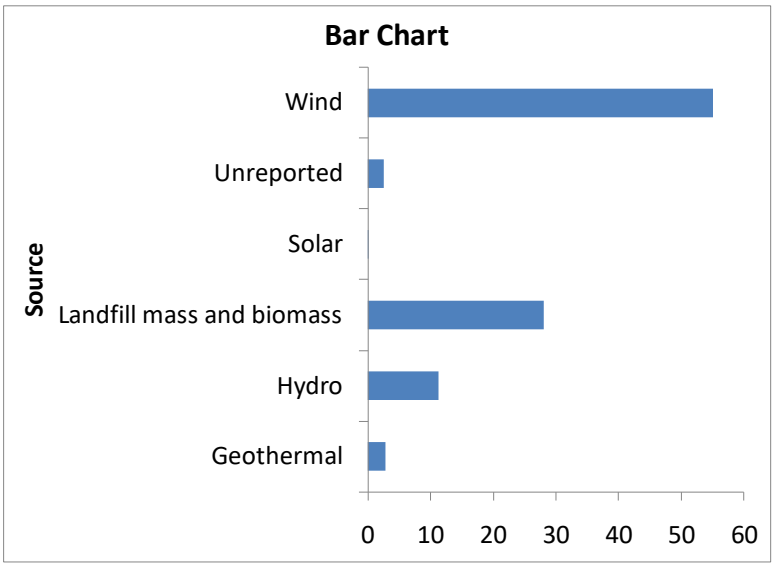


2.92 (b)
cont.

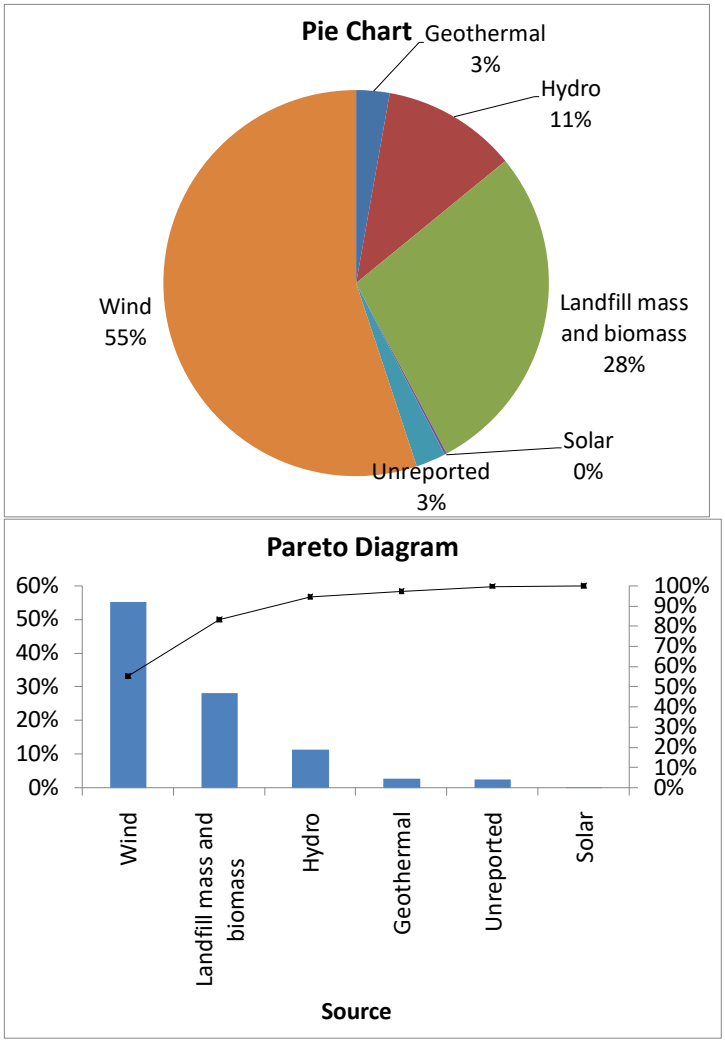


(c) The publisher gets the largest portion (64.8%) of the revenue. About half (32.3%) of the revenue received by the publisher covers manufacturing costs. The publisher's marketing and promotion account for the next largest share of the revenue, at 15.4%. Author, bookstore employee salaries and benefits, 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.

2.93 (a)

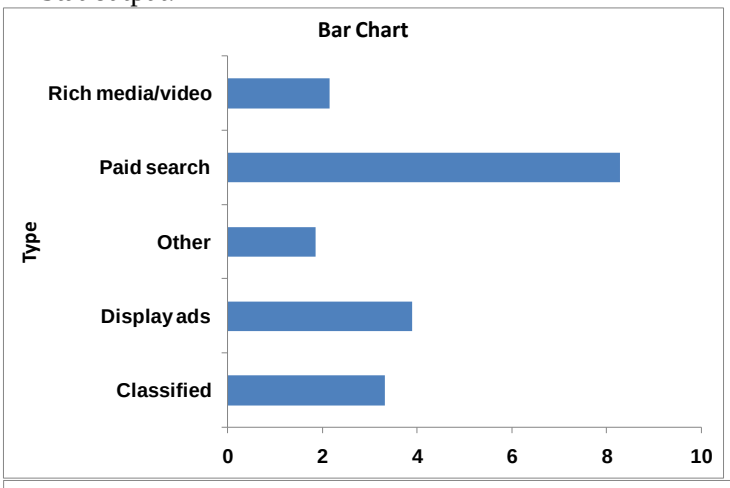


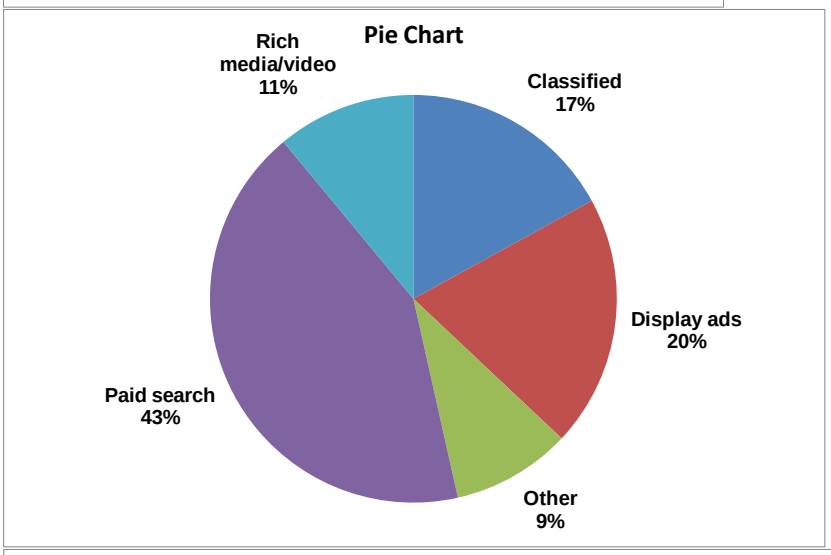
2.93 cont. (a)

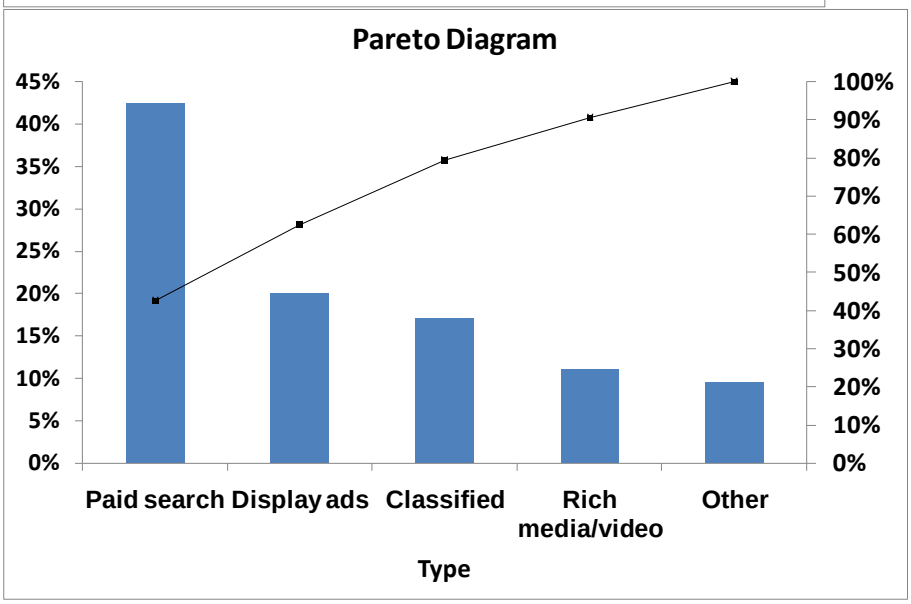


(b) Majority of the green power comes from wind power at over 50% while more than 80% of the green power is derived from wind, and landfill mass and biomass.

2.94 (a) PHStat output:

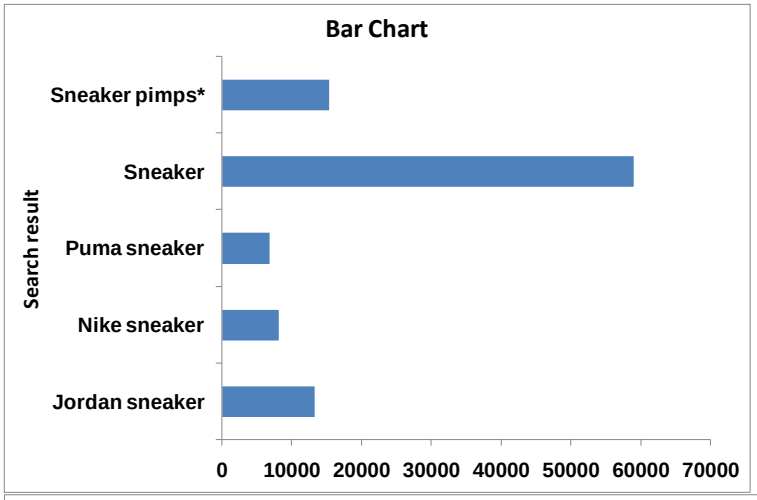


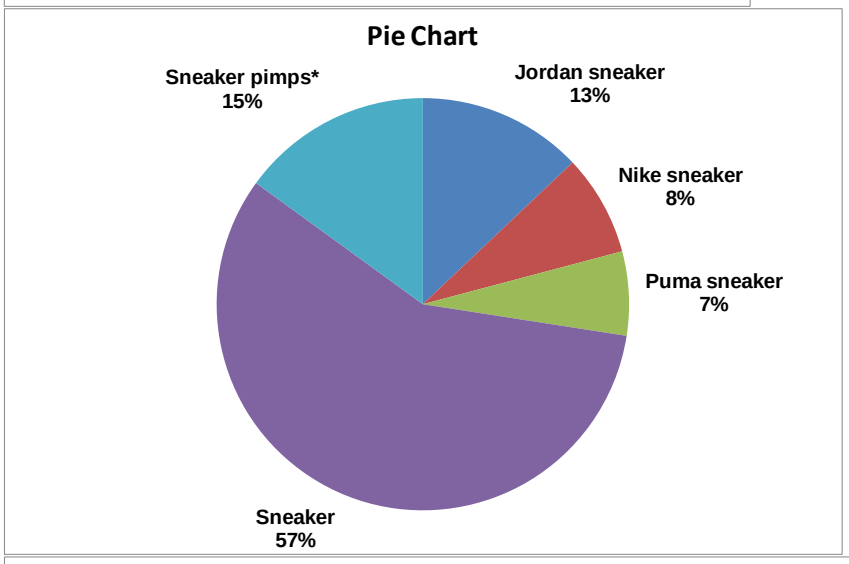


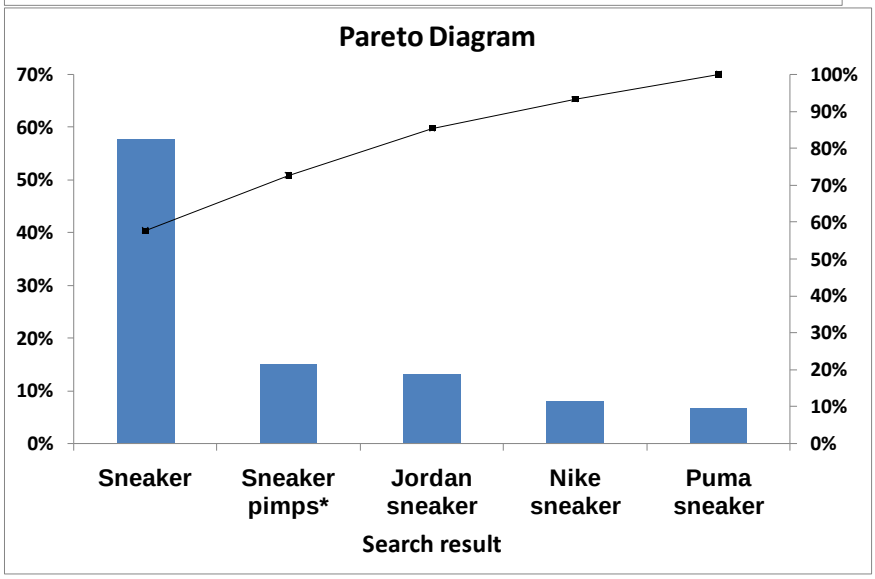


(b) The Pareto plot is most appropriate because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

2.94 cont. (c)







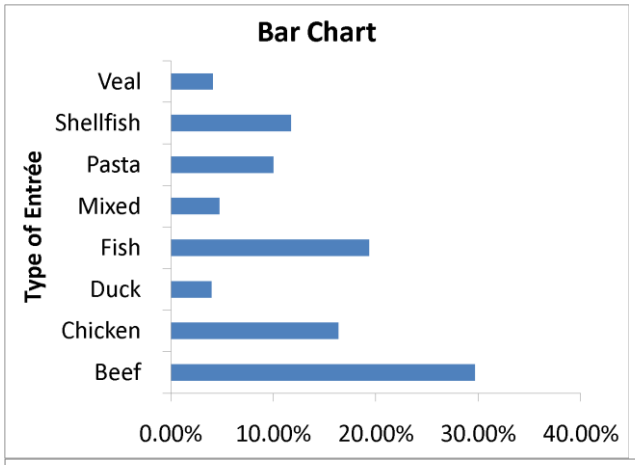
(d) The Pareto plot is most appropriate because it not only sorts the frequencies in descending order, it also provides the cumulative polygon on the same scale.

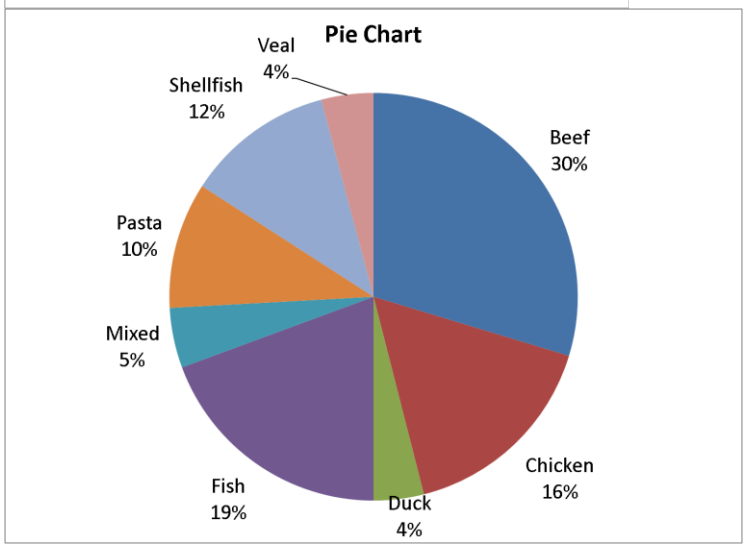
2.94 cont. (e) "Paid search" constitutes the largest category on US online ad spending at 43%. Excluding the generic keyword "sneaker", searches using the keywords "sneaker pimps" and "Jordan sneaker" make up majority of the search for sneakers on specific brands.

2.95 (a)

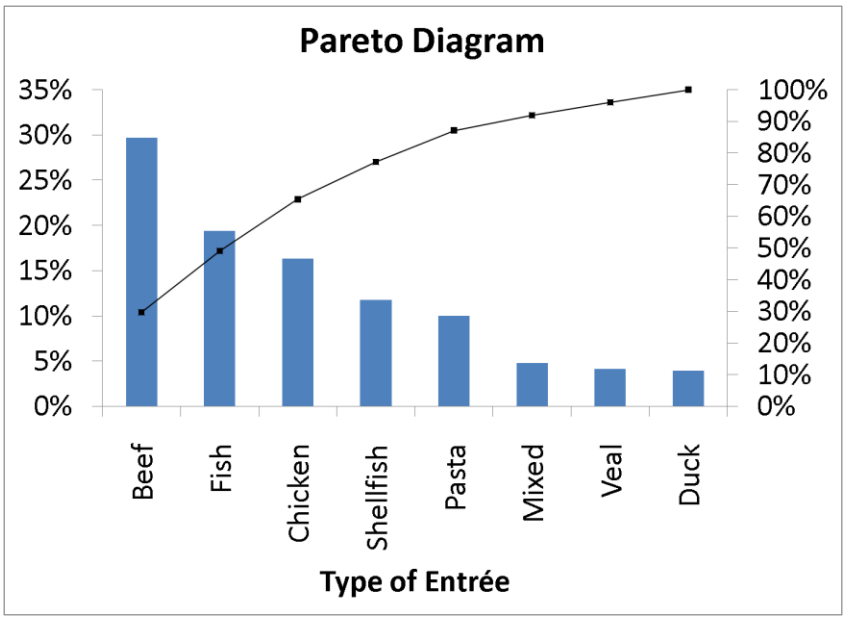
Type of Entrée	%	Number S
Beef	29.68%	187
Chicken	16.35%	103
Mixed	4.76%	30
Duck	3.97%	25
Fish	19.37%	122
Pasta	10.00%	63
Shellfish	11.75%	74
Veal	4.13%	26
Total	100.00%	630

(b)





2.95 (b) 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 more than 50% of all entrees ordered by weekend patrons of a continental restaurant. When chicken is included, better than two-thirds of the entrees are accounted for.

2.96 (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%

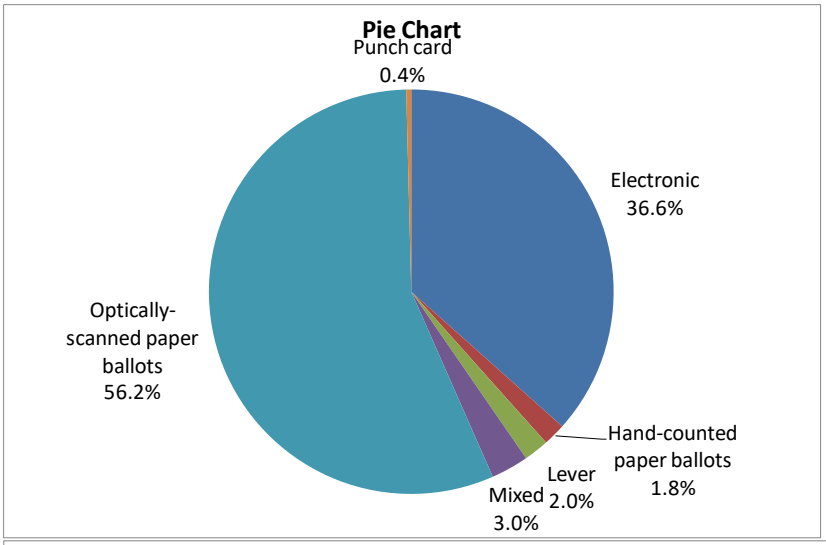
	Gender		
Dessert Ordered	**Male**	**Female**	**Total**
Yes	16%	7%	23%
No	37%	40%	77%
Total	53%	47%	100%

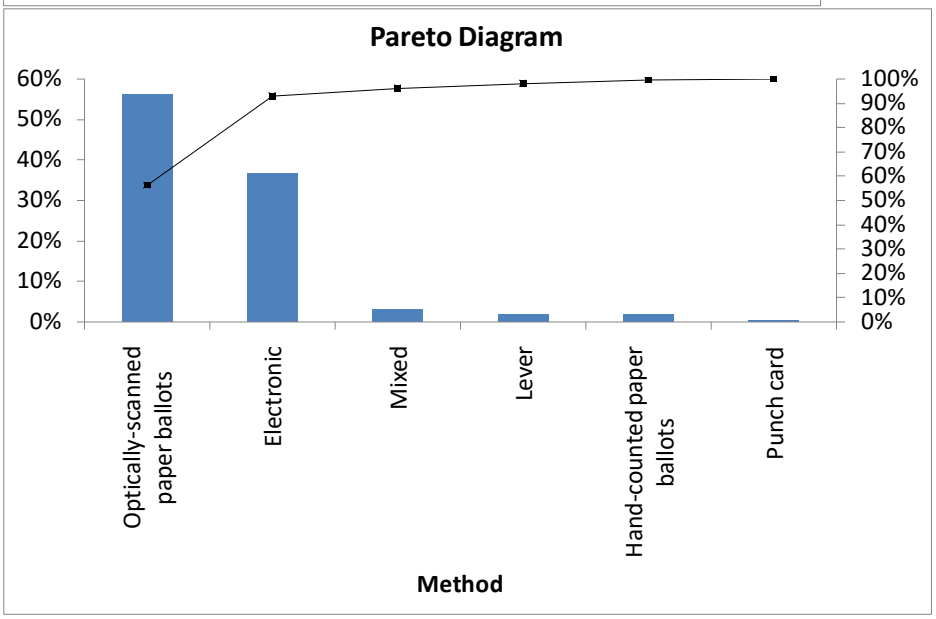
	Beef Entrée		
Dessert Ordered	**Yes**	**No**	**Total**
Yes	12%	11%	23%
No	19%	58%	77%
Total	31%	69%	100%

2.96 cont. (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.

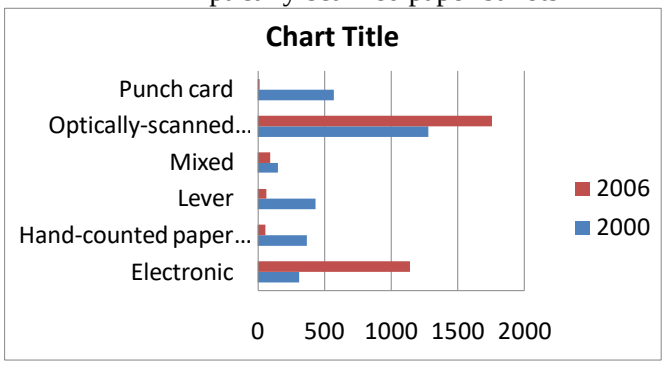
(c) 30% of the men sampled ordered desserts compared to 14% of the women. Men are more than twice as likely to order desserts as women. Almost 38% of the patrons ordering a beef entree ordered dessert compared to less than 16% of patrons ordering all other entrees. Patrons ordering beef are better than 2.3 times as likely to order dessert as patrons ordering any other entree.

2.97 (a)



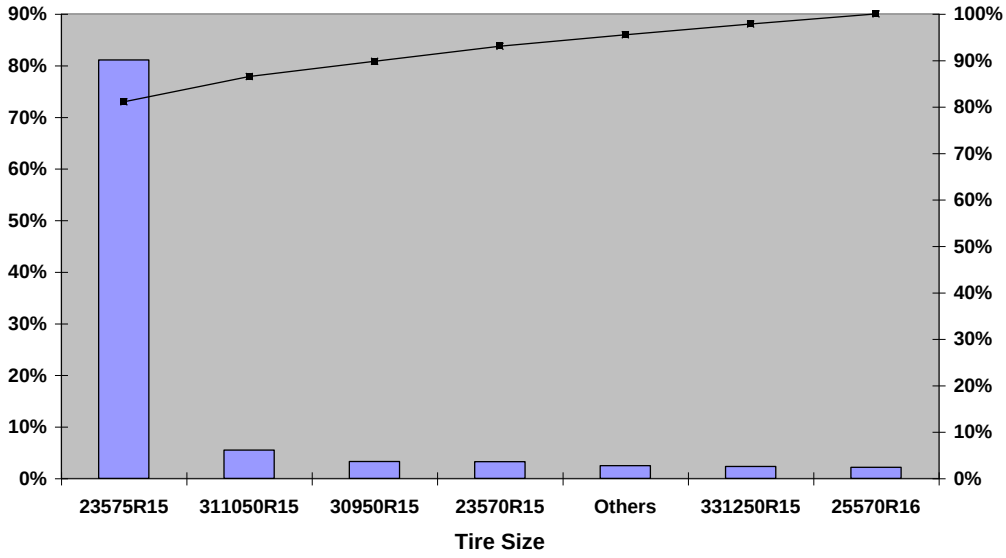


2.97 cont. (b) From the Pareto diagram in part (a), one can see that more than 90% of the counties used either the "optically-scanned paper ballots" or "electronic" method in 2006.



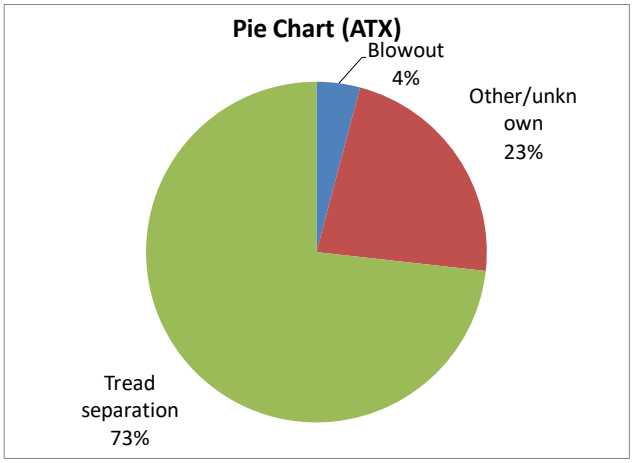
(c) More counties moved from the "punch card", "mixed", "level" or "hand-counted paper" methods to using the "optically-scanned paper ballots" or "electronic" methods in 2006 compared to 2000.

2.98 (a)

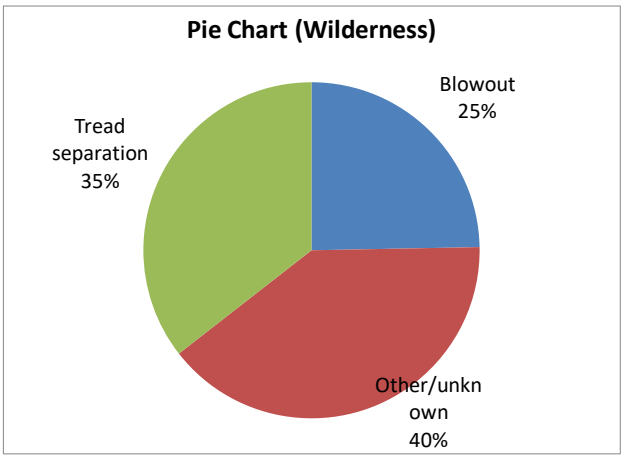


23575R15 accounts for over 80% of the warranty claims.

(b)

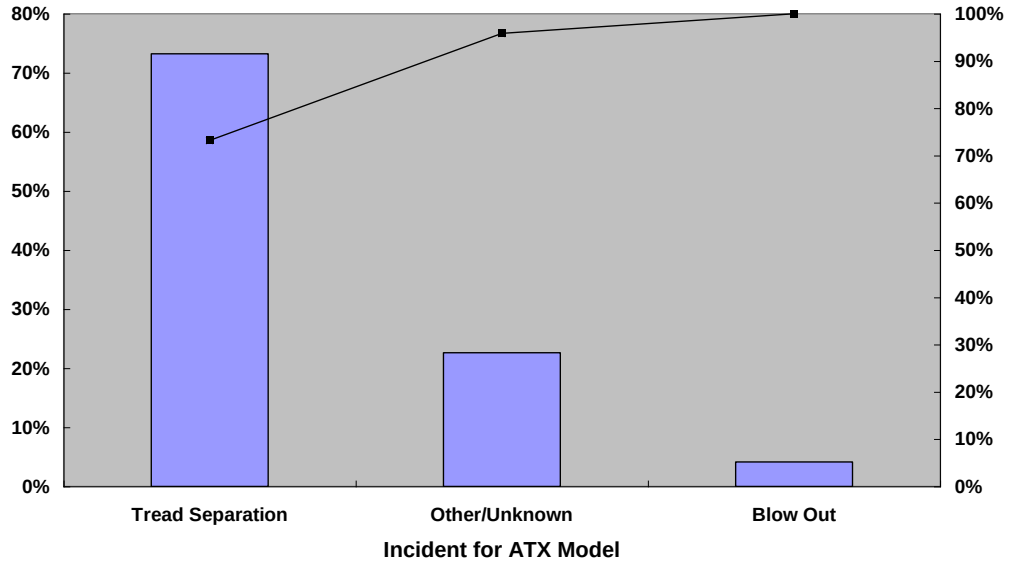


2.98 cont. (b)



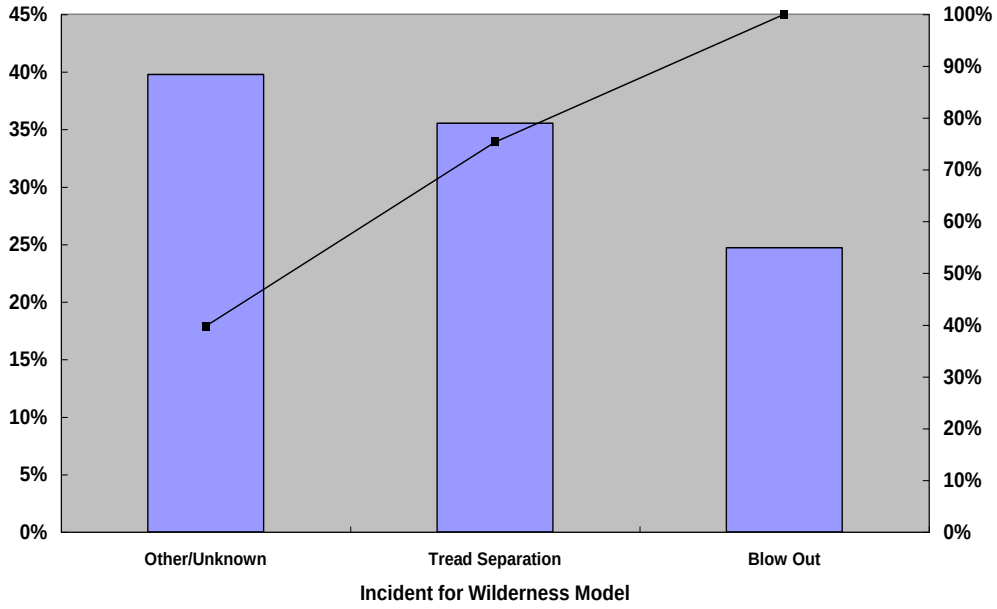
91.82% of the warranty claims are from the ATX model.

(c)



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

2.98 (d) cont.

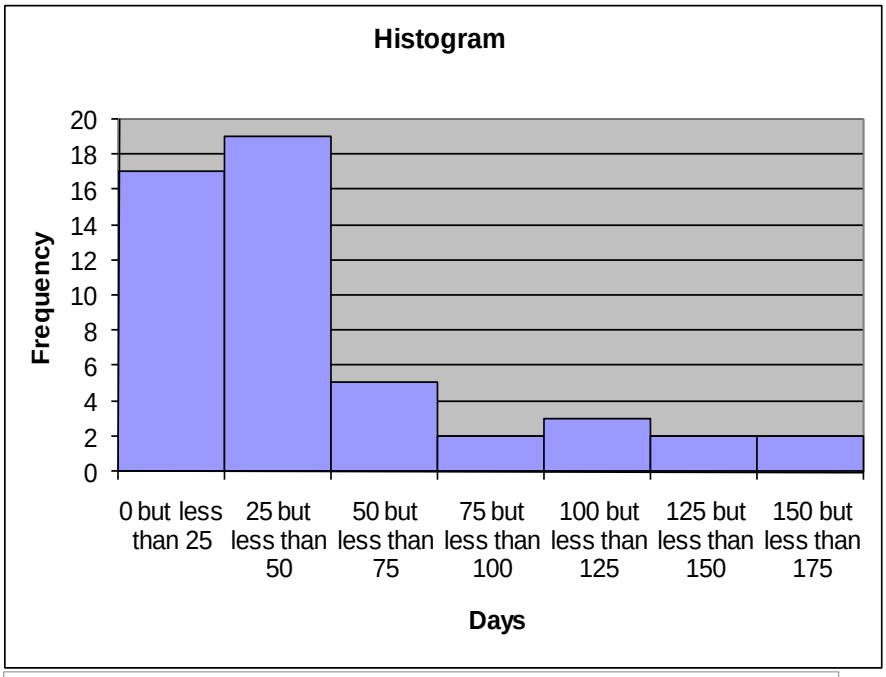


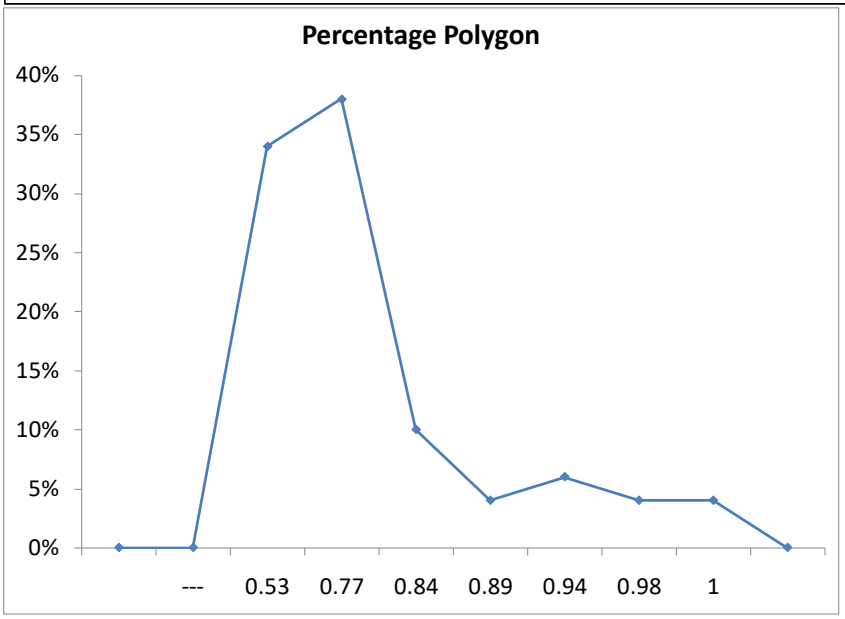
The number of claims is evenly distributed among the three incidents; other/unknown incidents account for almost 40% of the claims, tread separation accounts for about 35% of the claims, and blowout accounts for about 25% of the claims.

2.99 (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%

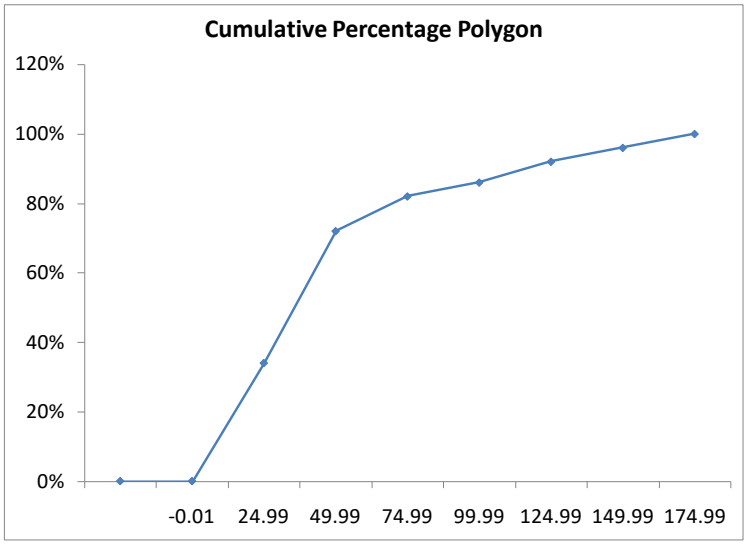
2.99 (b)
cont.





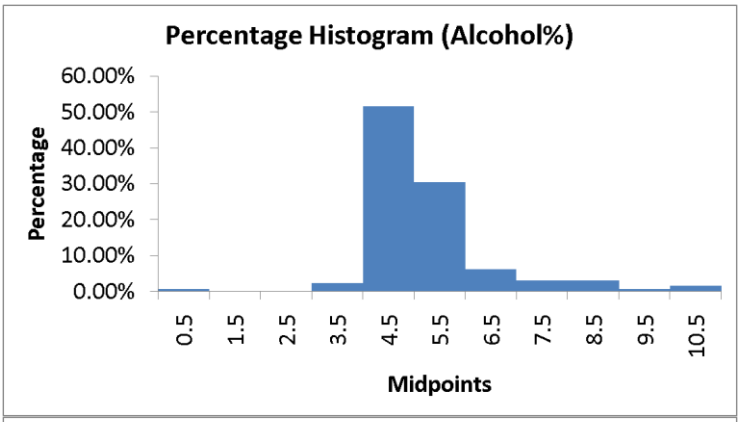
2.99 cont. (c)

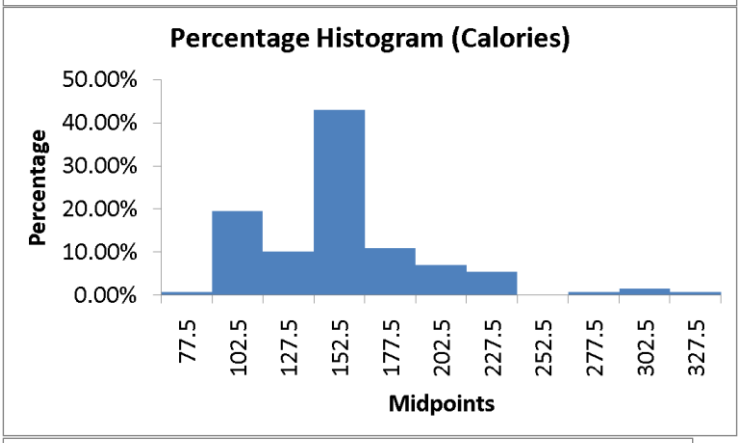
Range	*Cumulative %*
0 but less than 25	34%
25 but less than 50	72%
50 but less than 75	82%
75 but less than 100	86%
100 but less than 125	92%
125 but less than 150	96%
150 but less than 175	100%

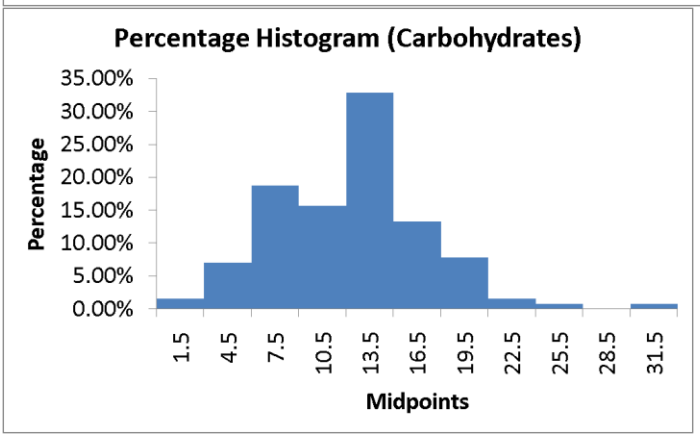


(d) You should tell the president of the company that over half of the complaints are resolved within a month, but point out that some complaints take as long as three or four months to settle.

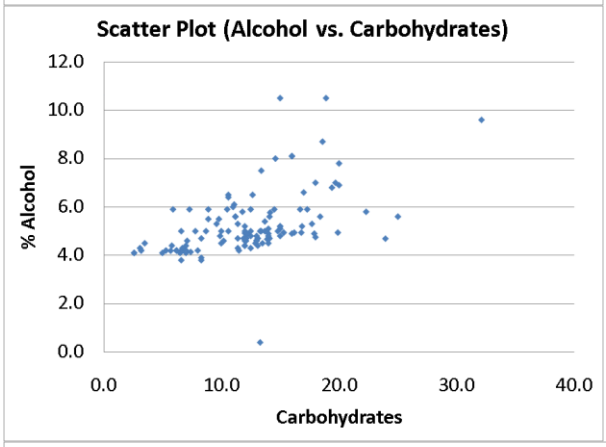
2.100 (a)

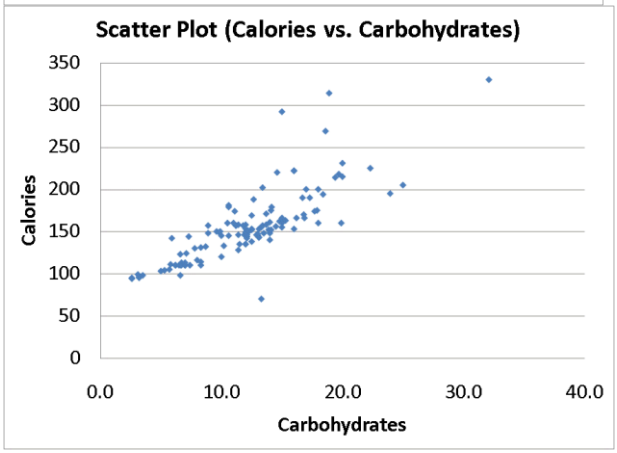






2.100 cont. (b)



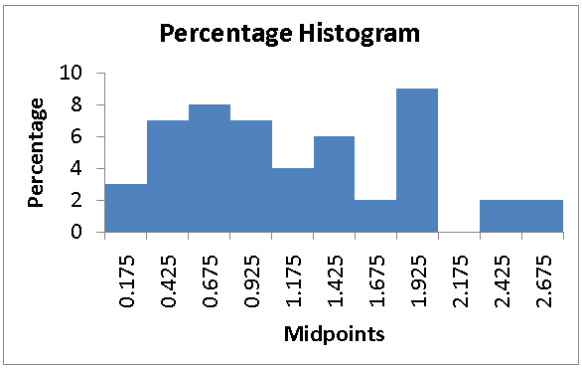


(c) The alcohol % is concentrated between 4 and 6, with more between 4 and 5. The calories are concentrated between 140 and 160. The carbohydrates are concentrated between 12 and 15. There are outliers in the percentage of alcohol in both tails. The outlier in the lower tail is due to the non-alcoholic 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.

2.101 (a) Ordered array:
0.070, 0.170, 0.180, 0.300, 0.339, 0.350, 0.360, 0.370, 0.425, 0.440, 0.550, 0.570, 0.600, 0.600, 0.620, 0.640, 0.695, 0.790, 0.800, 0.840, 0.870, 0.910, 0.980, 0.995, 1.030, 1.150, 1.150, 1.180, 1.250, 1.330, 1.350, 1.360, 1.410, 1.504, 1.530, 1.700, 1.770, 1.990, 2.000, 2.000, 2.000, 2.000, 2.000, 2.000, 2.000, 2.025, 2.460, 2.510, 2.575, 2.750.

(b)



(c) There is a 2.68% 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 2.6%. Majority of the states though have cigarette tax concentrated around 0.7%.

2.102 (a)

Statistics	
Sample Size	23
Mean	0.955217
Median	0.75
Std. Deviation	0.756312
Minimum	0.01
Maximum	2.12

Money Market

Stem unit: 0.1

0	1 2
1	
2	0 0 0 0 5 5
3	
4	
5	0
6	0 0
7	5
8	
9	
10	0
11	
12	0 0 5
13	
14	
15	
16	
17	5 5
18	
19	0
20	0 0 2
21	2

2.102 (a) cont.

One-Year CD

Stem unit: 0.1

Statistics	
Sample Size	23
Mean	1.693043
Median	2
Std. Deviation	0.550433
Minimum	0.25
Maximum	2.3

2	5
3	
4	
5	
6	
7	5
8	5
9	
10	0
11	
12	0
13	5
14	
15	0
16	5
17	5 5 5
18	
19	
20	0 0 0 0 0 0 5
21	0
22	0 4 5
23	0

2.102 (a)
cont.

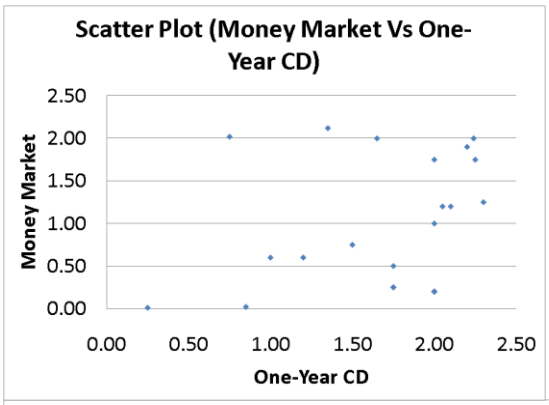
Five-Year CD

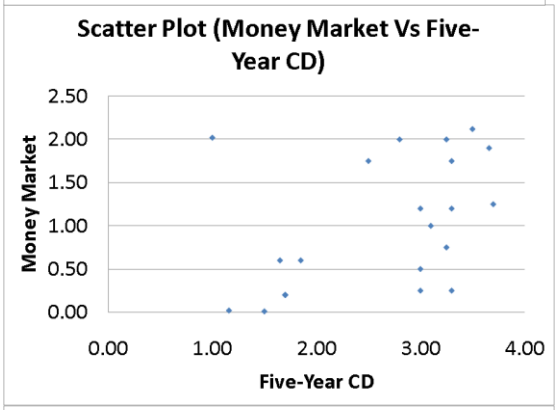
Stem unit: 0.1

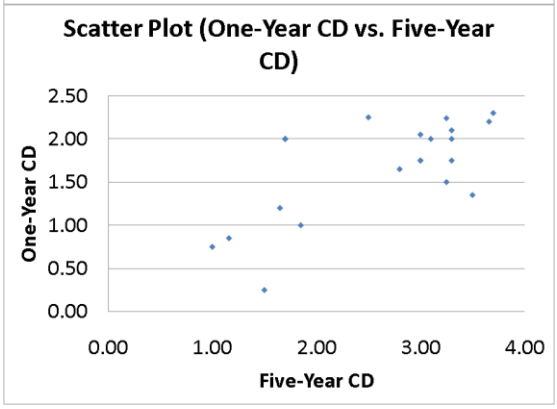
Statistics	
Sample Size	23
Mean	2.548696
Median	3
Std. Deviation	0.871953
Minimum	1
Maximum	3.7

10	0
11	6
12	
13	
14	
15	0
16	5
17	0 0 0 0
18	5
19	
20	
21	
22	
23	
24	
25	0
26	
27	
28	0
29	
30	0 0 0
31	0
32	5 5
33	0 0 0
34	
35	0
36	6
37	0

2.102 (b) cont.





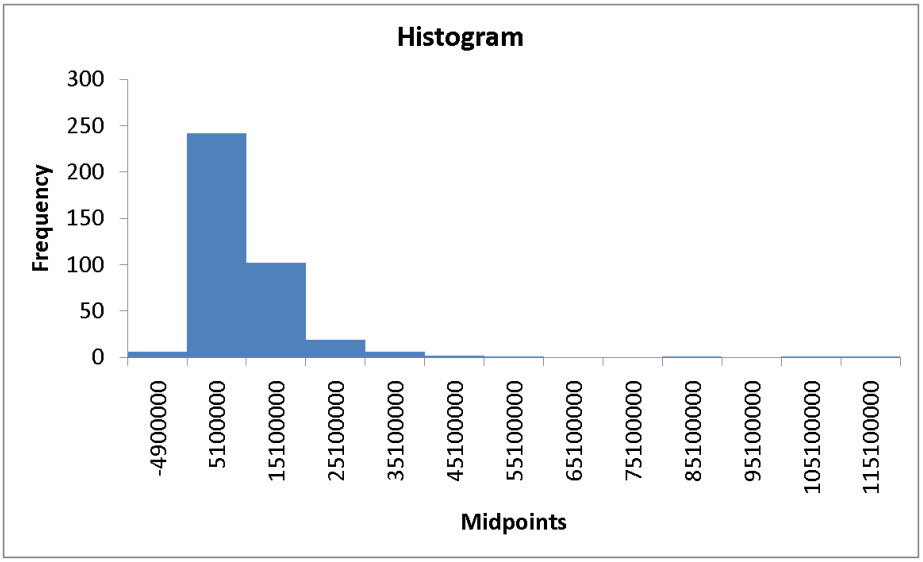


(c) The money market yield is concentrated between 0.2 and 0.3. The one-year CD is concentrated between 2 and 2.1. The five-year CD yield is concentrated between 1.7 and 1.8 and 3.0 and 3.4. In general, the five-year CD has the highest yield, followed by the one-year CD and then the money market.
There appears to be a positive relationship between the yield of the one-year CD and the five-year CD but no obvious relationship exists between the yield of the money market and the one-year CD, and the money market and the five-year CD.

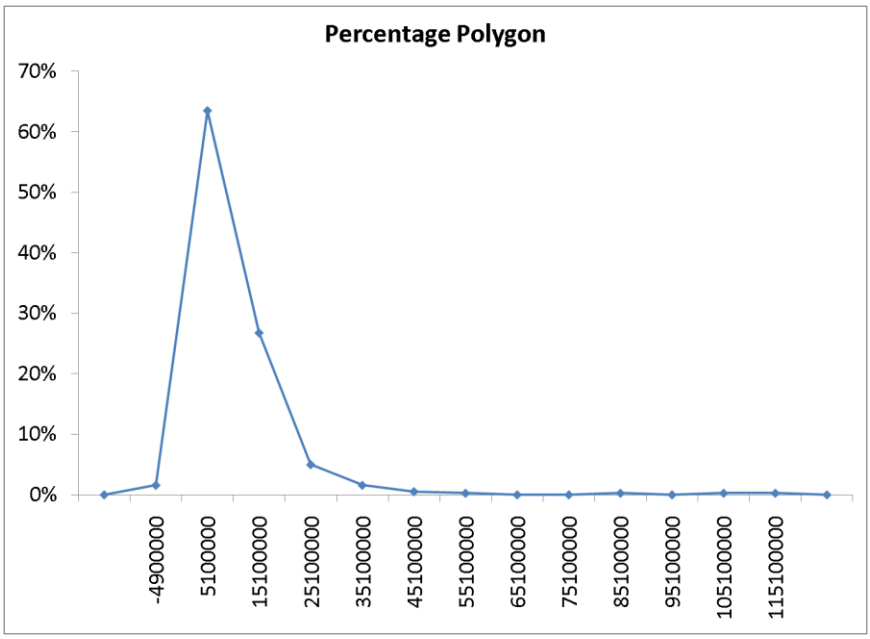
2.103 (a), (c)

bin	*Frequency*	*Percentage*	*Cumulative %*
-9,900,000 but less than 99,999	6	1.57%	1.57%
100,000 but less than 10,099,999	242	63.52%	65.09%
10,100,000 but less than 20,099,999	102	26.77%	91.86%
20,100,000 but less than 30,099,999	19	4.99%	96.85%
30,100,000 but less than 40,099,999	6	1.57%	98.43%
40,100,000 but less than 50,099,999	2	0.52%	98.95%
50,100,000 but less than 60,099,999	1	0.26%	99.21%
60,100,000 but less than 70,099,999	0	0.00%	99.21%
70,100,000 but less than 80,099,999	0	0.00%	99.21%
80,100,000 but less than 90,099,999	1	0.26%	99.48%
90,100,000 but less than 100,099,999	0	0.00%	99.48%
100,100,000 but less than 110,099,999	1	0.26%	99.74%
110,100,000 but less than 120,099,999	1	0.26%	100.00%

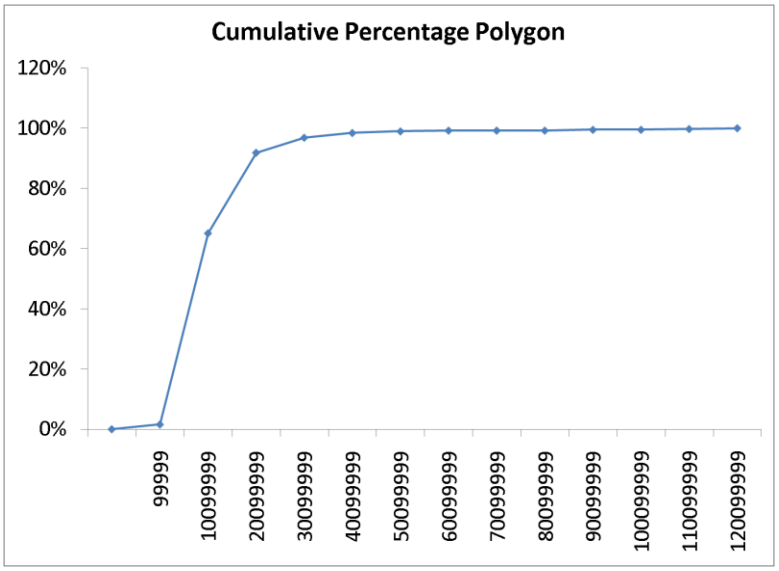
(b)



2.103 cont. (b)



(c)



(d) CEO compensation in 2008 is extremely right skewed. More than 90% of the CEOs have compensation lower than $20,100,000. On the other end, 1.57% of the CEOs have compensation lower than $100,000.

2.104 (a)

Frequencies (Boston)

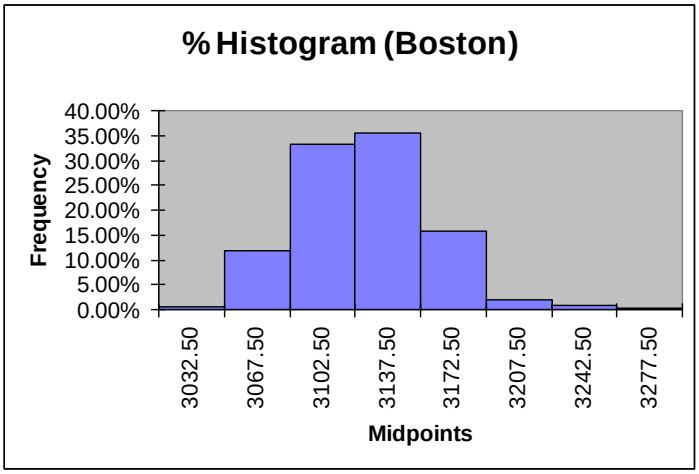
Weight (Boston)	*Frequency*	*Percentage*
3015 but less than 3050	2	0.54%
3050 but less than 3085	44	11.96%
3085 but less than 3120	122	33.15%
3120 but less than 3155	131	35.60%
3155 but less than 3190	58	15.76%
3190 but less than 3225	7	1.90%
3225 but less than 3260	3	0.82%
3260 but less than 3295	1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

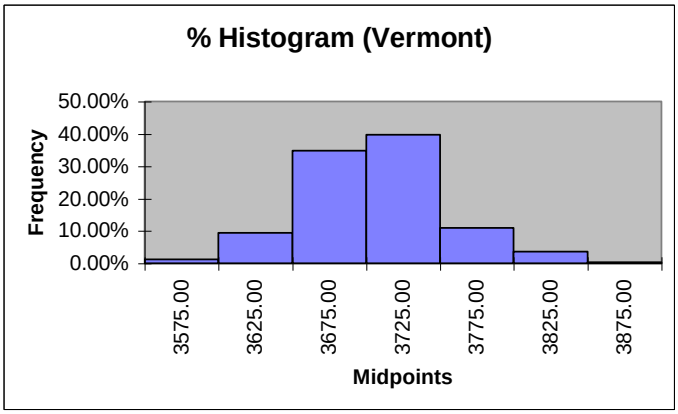
(b)

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%

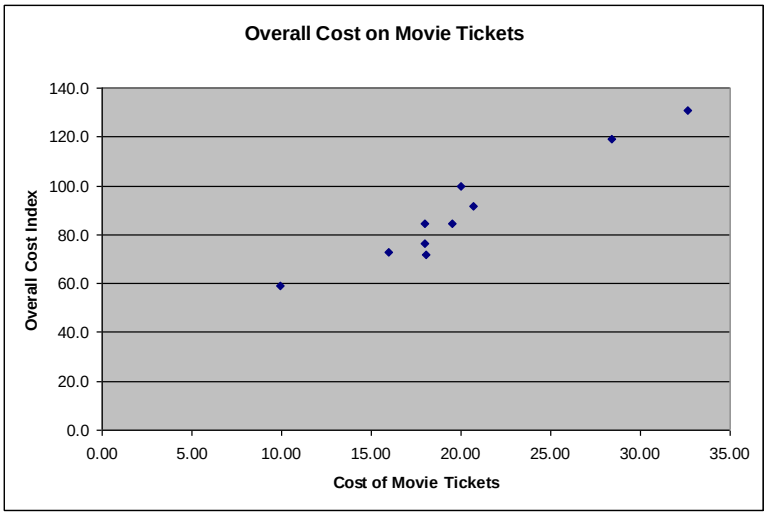
(c)

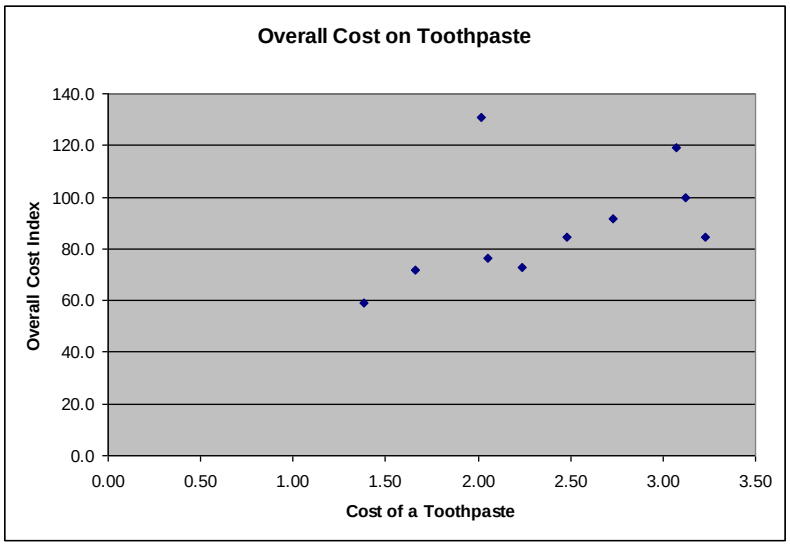


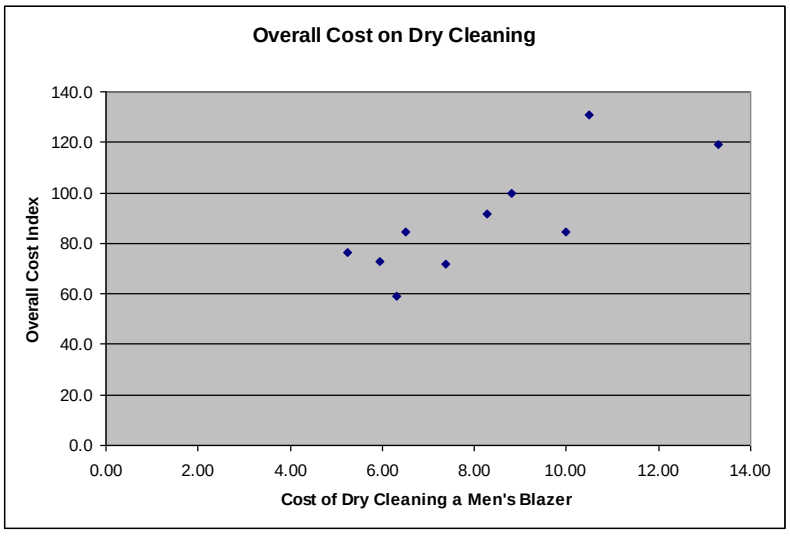


(d) 0.54% of the "Boston" shingles pallets are underweight while 0.27% are overweight. 1.21% of the "Vermont" shingles pallets are underweight while 3.94% are overweight.

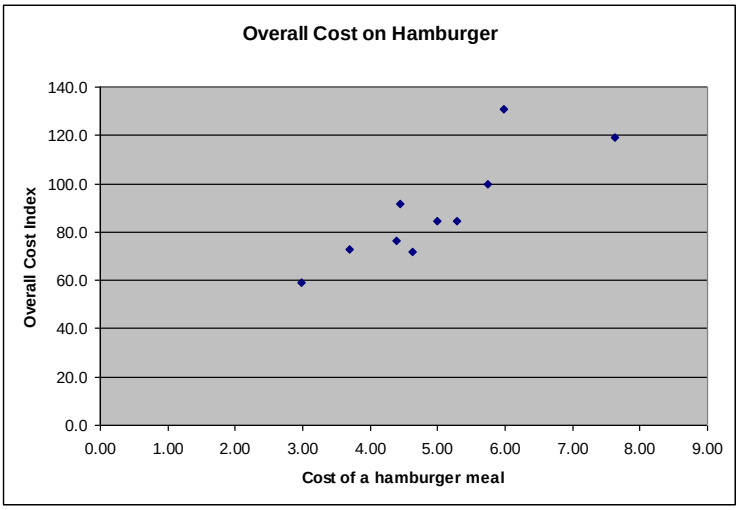
2.105 (a)

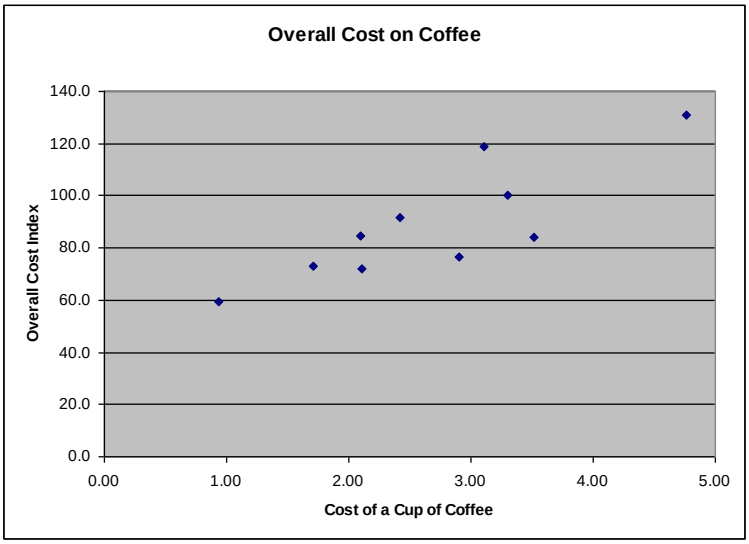


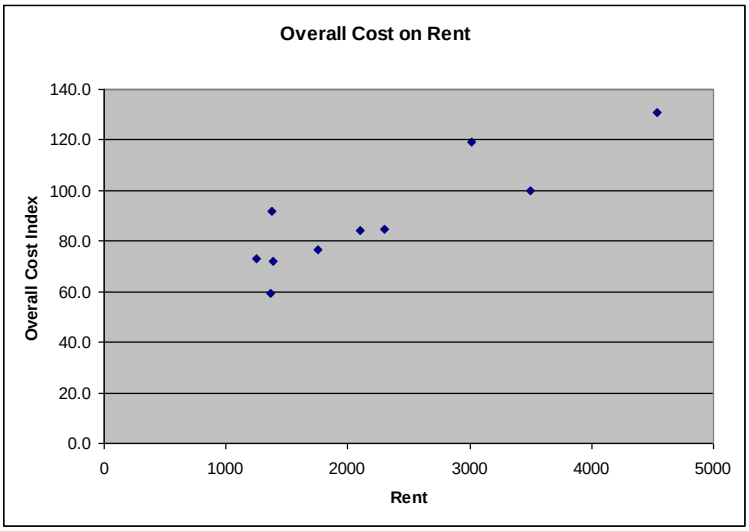




2.105 (a)
cont.



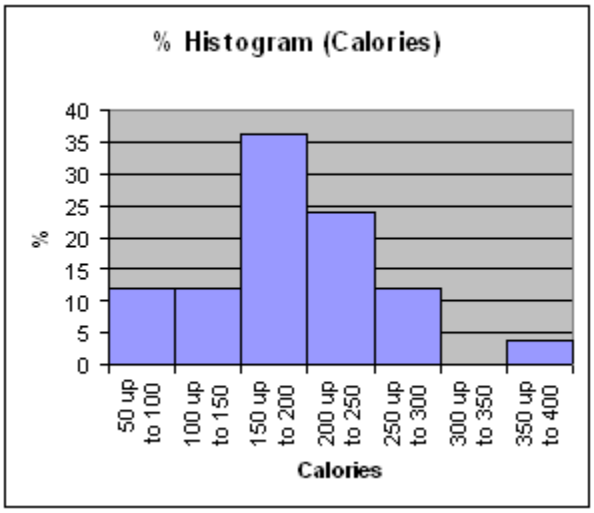




(b) There is a positive relationship between the overall cost index and each of these variables.

2.106 (a)

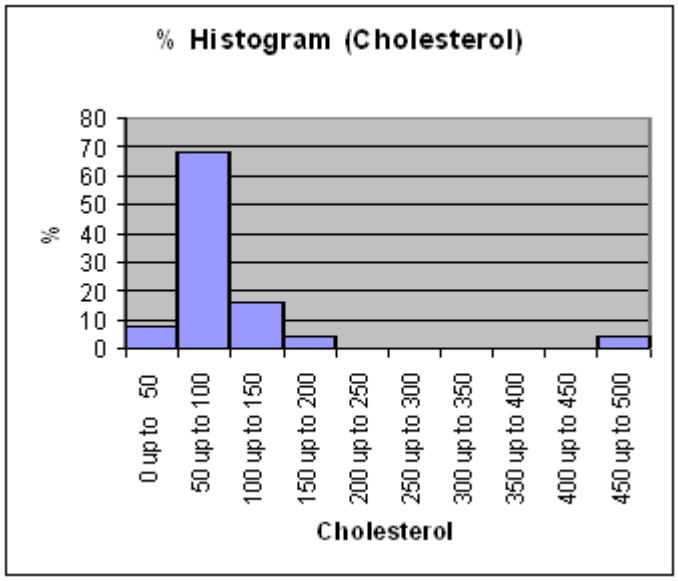
Calories	Frequency	Percentage	Percentage Less Than
50 up to 100	3	12%	12%
100 up to 150	3	12	24
150 up to 200	9	36	60
200 up to 250	6	24	84
250 up to 300	3	12	96
300 up to 350	0	0	96
350 up to 400	1	4	100



(b)

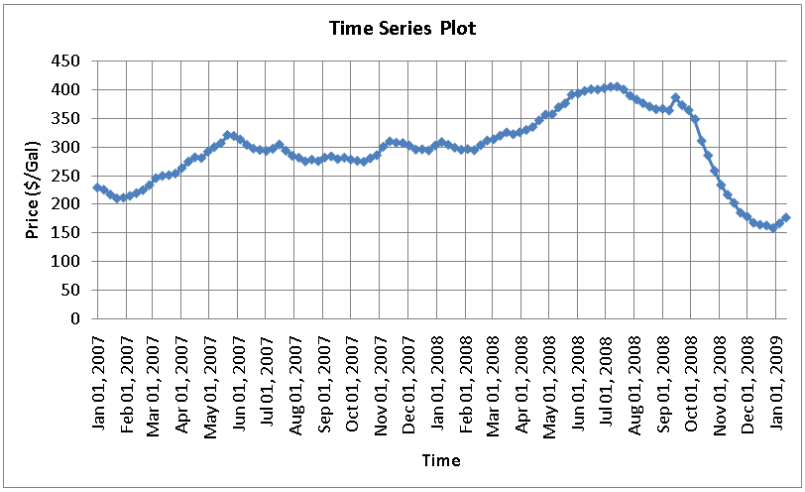
Cholesterol	Frequency	Percentage	Percentage Less Than
0 up to 50	2	8	8%
50 up to 100	17	68	76
100 up to 150	4	16	92
150 up to 200	1	4	96
200 up to 250	0	0	96
250 up to 300	0	0	96
300 up to 350	0	0	96
350 up to 400	0	0	96
400 up to 450	0	0	96
450 up to 500	1	4	100

2.106 (b) cont.



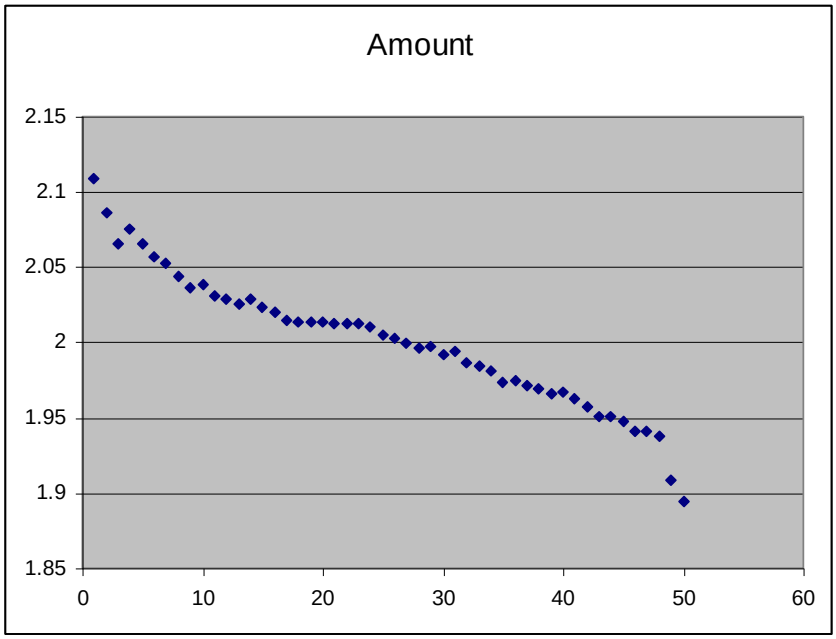
(c) The sampled fresh red meats, poultry, and fish vary from 98 to 397 calories per serving, with the highest concentration between 150 to 200 calories. One protein source, spareribs, with 397 calories, is more than 100 calories above the next highest caloric food. The protein content of the sampled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

2.107 (a)



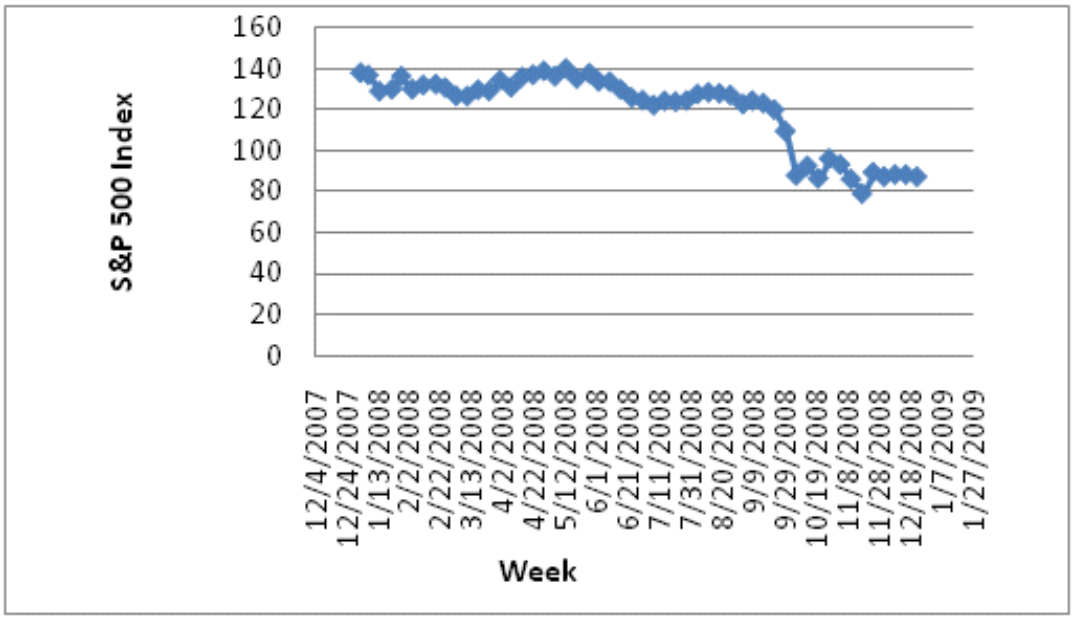
(b) The average price of gasoline in the United States is higher in the summer in general and seems to peak in June.

2.108 (a)

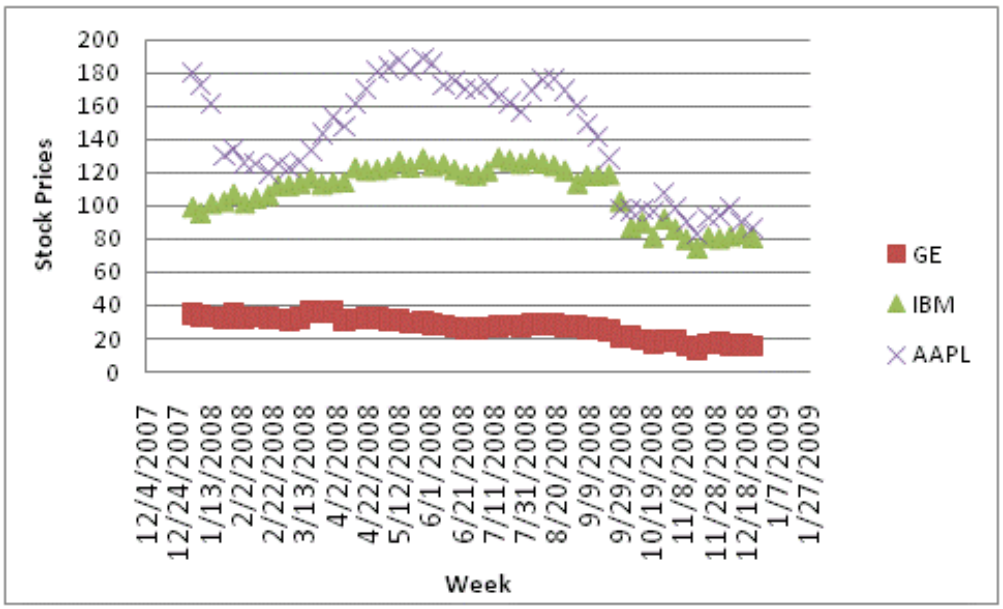


(b) There is a downward trend in the amount filled.
(c) The amount filled in next bottle will most likely be below 1.894 liter.
(d) The scatter plot of the amount of soft drink filled against time reveals the trend of the data, whereas a histogram only provides information on the distribution of the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

2.109 (a)

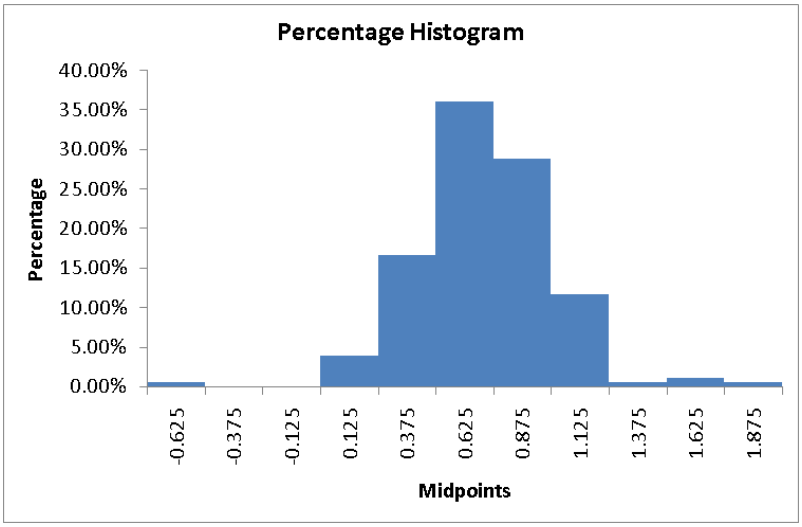


2.109 (a) cont.

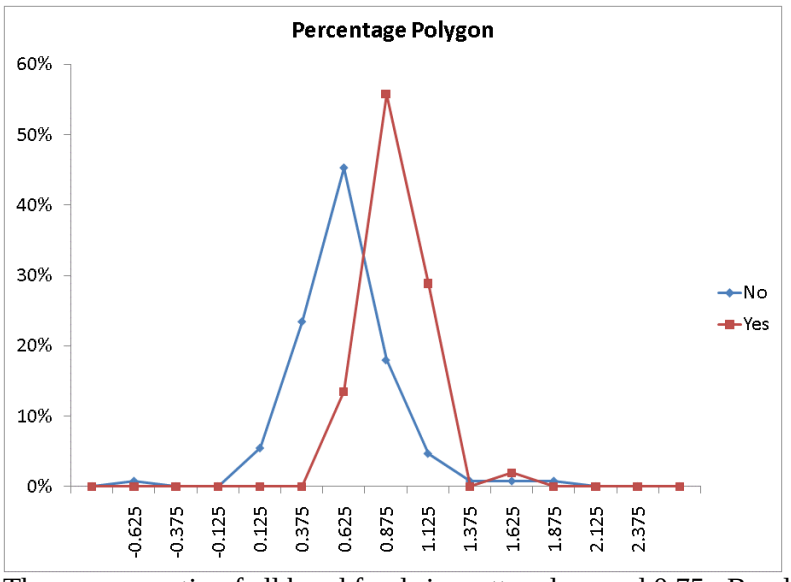


(b) Even though there appeared to be cyclical pattern in the S&P index, there was a general downward trend with a big drop that took place after 9/29/2009. The stock price of Apple fluctuated between $120 and $200 before 9/29/2009 and then dropped to around $90 after 9/29/2009. The stock price of GE trended downward from about $40 to about $20 while IBM's stock price was trending upward before the big drop that took place after 9/29/2009.

2.114 (a)

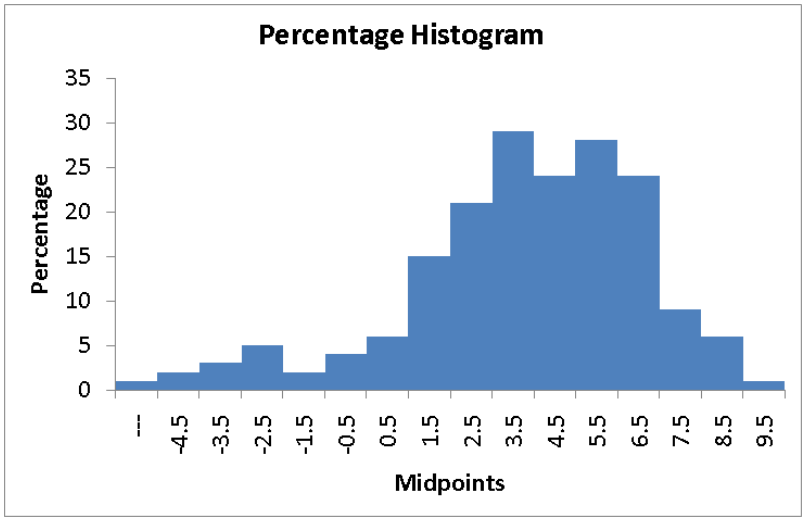


2.114 (b) cont.

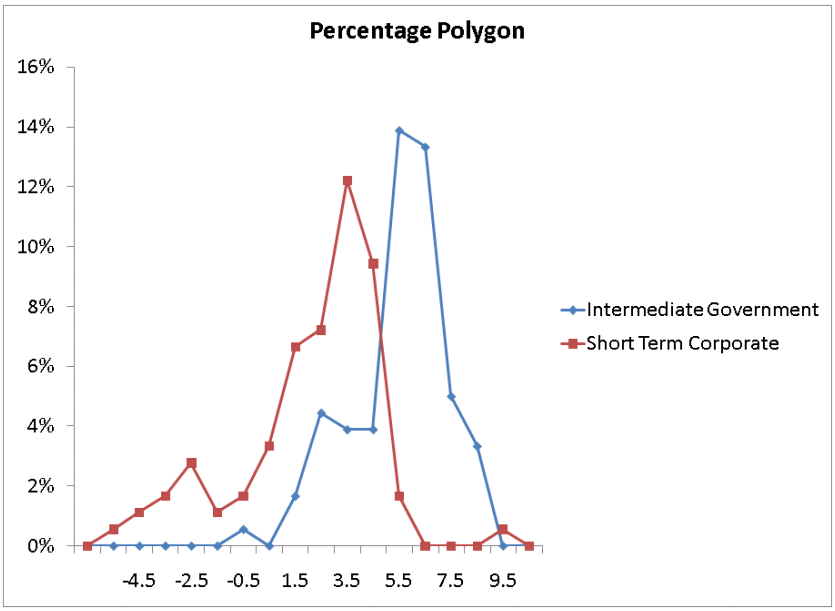


(c) The expense ratio of all bond funds is scattered around 0.75. Bond funds with fees have expense ratio that is scattered around 0.9 while bond funds without fees have expense ratio that is scattered around 0.625.

2.115 (a)

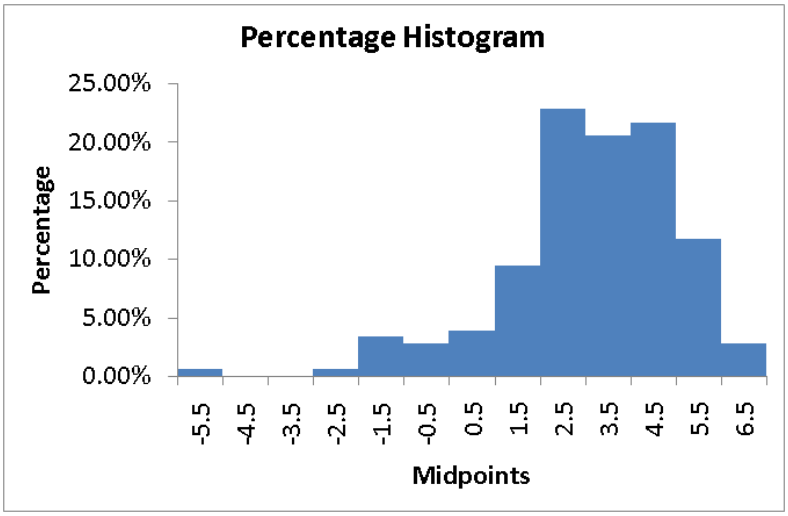


2.115 (b) cont.

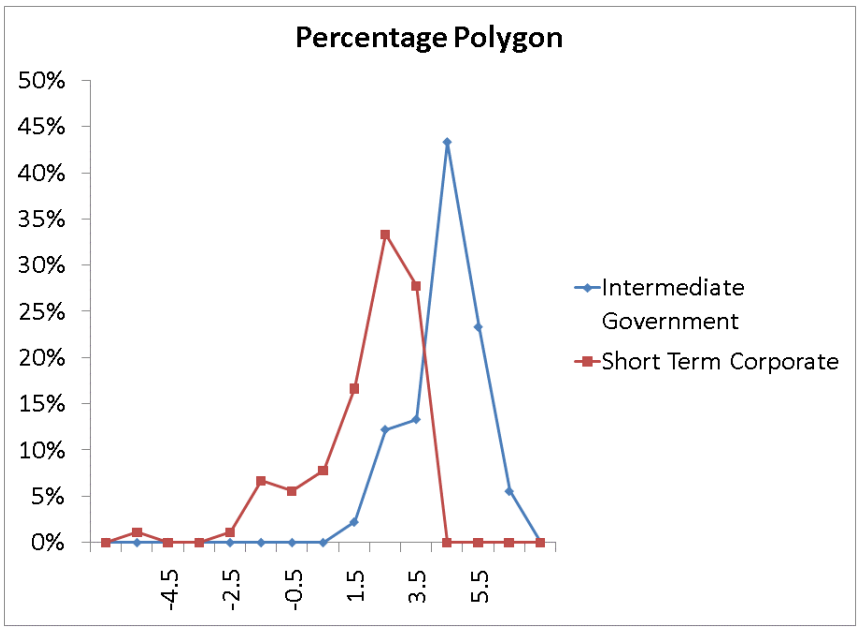


(c) The three-year annualized return of the 180 bond funds is left-skewed with majority of them (about 78%) scattered between 1% and 7%. About 9.5% of the mutual funds have a negative three-year annualized return while about 8.9% of them have return higher than 7%. In general, the intermediate government funds have higher three-year annualized return than short term corporate funds. Both types of mutual funds have three-year annualized return skewed to the left.

2.116 (a)



2.116 (b) cont.



(c) The five-year annualized return of the 181 bond funds is left-skewed with majority of them (about 65%) scattered between 2% and 5%. About 7% of the bond funds have a negative five-year annualized return while about 3% of them have return higher than 6%. In general, the intermediate government funds have higher five-year annualized return than short term corporate funds. Both types of mutual funds have five-year annualized return skewed to the left.

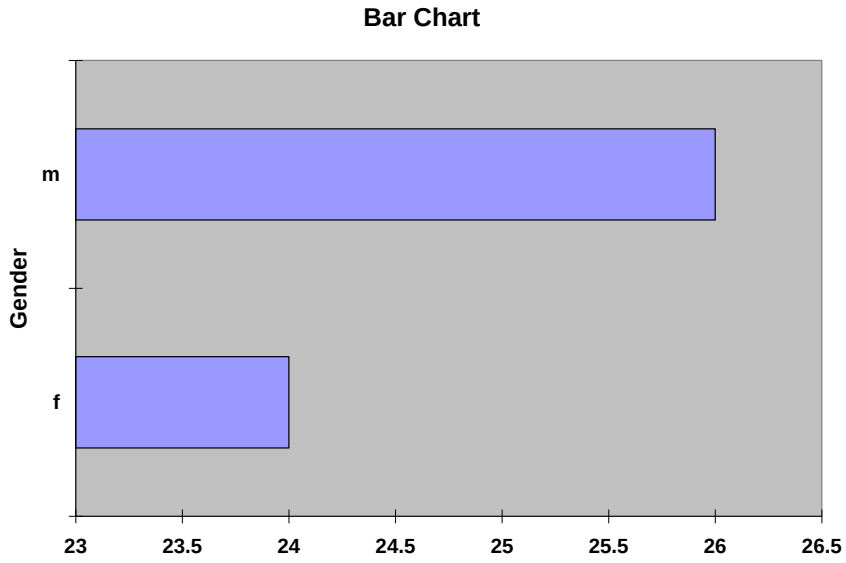
2.117 (a) Categorical variables: gender, class, major, grad school, employment status, satisfaction advisement.

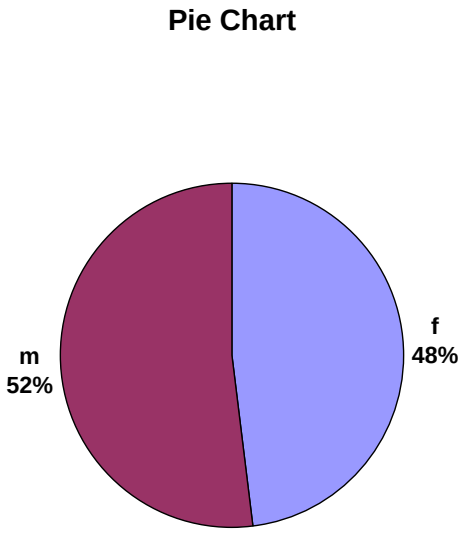
(b) Numerical variables: age, height, GPA, expected salary, annual salary in 5 years, number of affiliations, spending.

(c) Discrete numerical variables: age (in years), height(in inches), expected salary (in thousands of dollars), annual salary in 5 years (in thousands of dollars), number of affiliations.

(d)

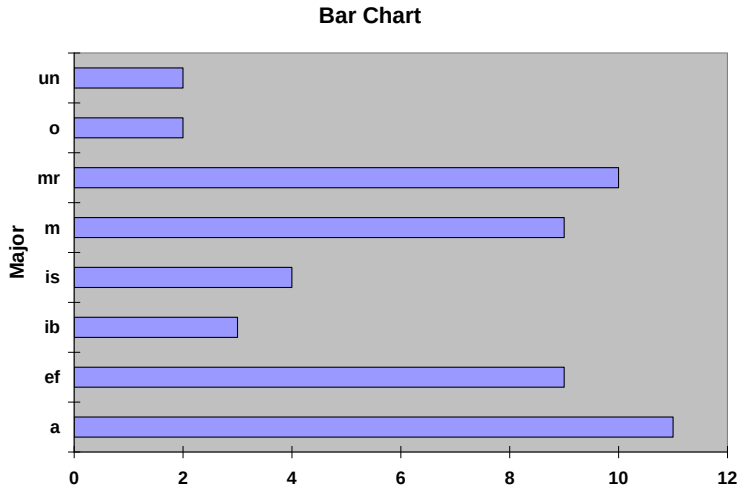
Gender:

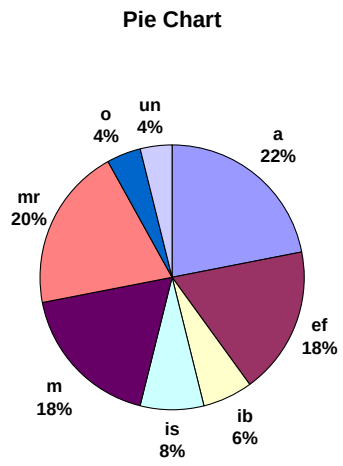


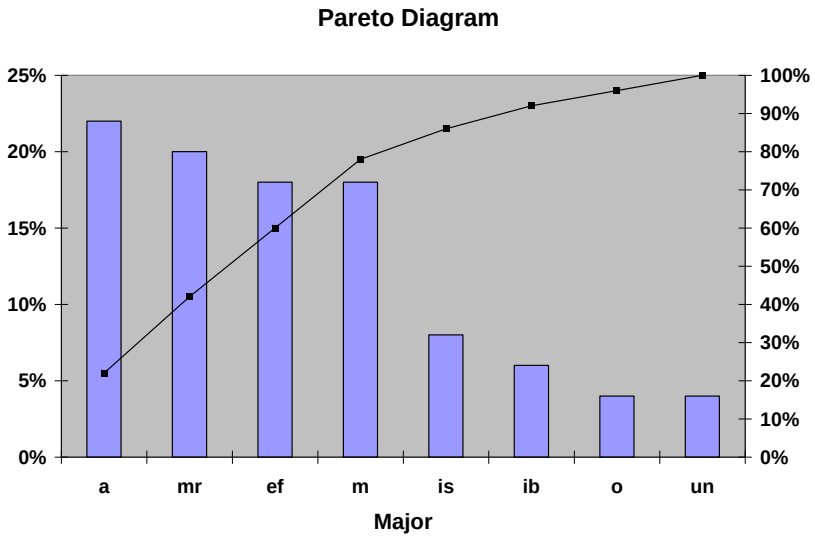


There are more males than females in the survey.

2.117
cont. **Major:**

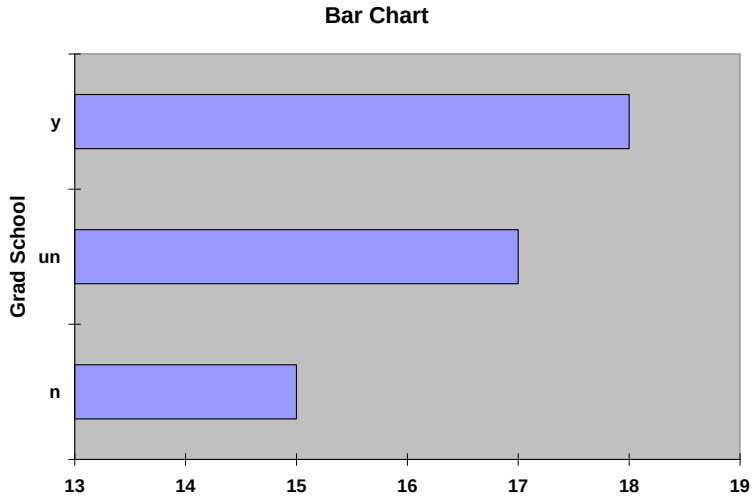


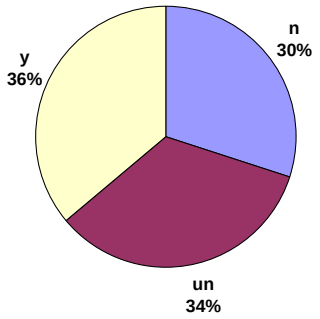


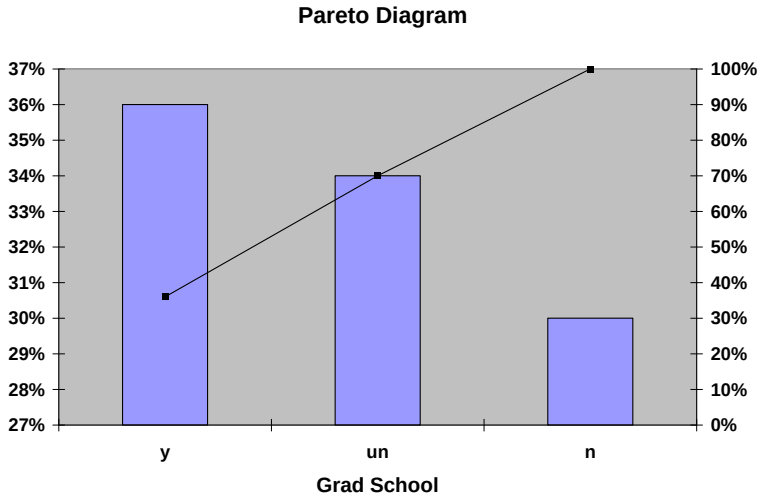


2.117 cont. Accounting, marketing/retailing, economics/finance and management constituted the "vital few" while the rest of the majors make up the "trivial many".

Grad School

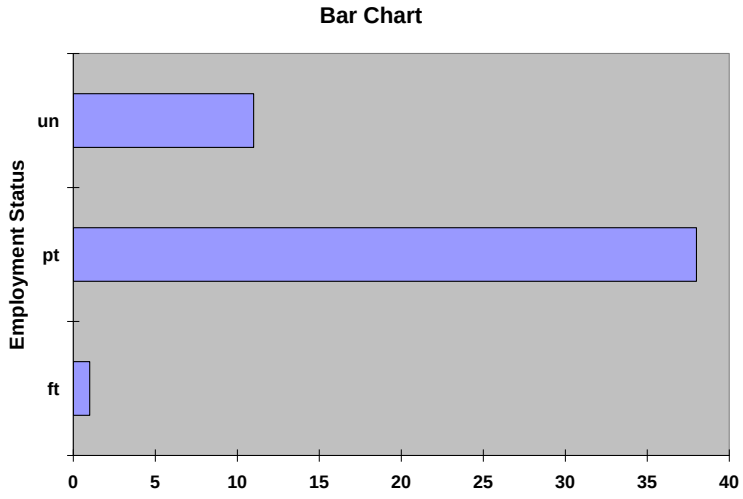


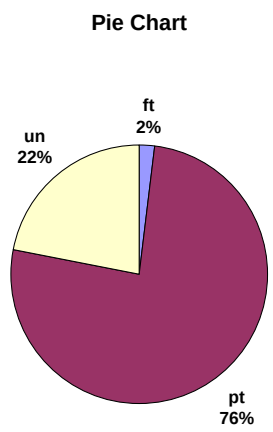


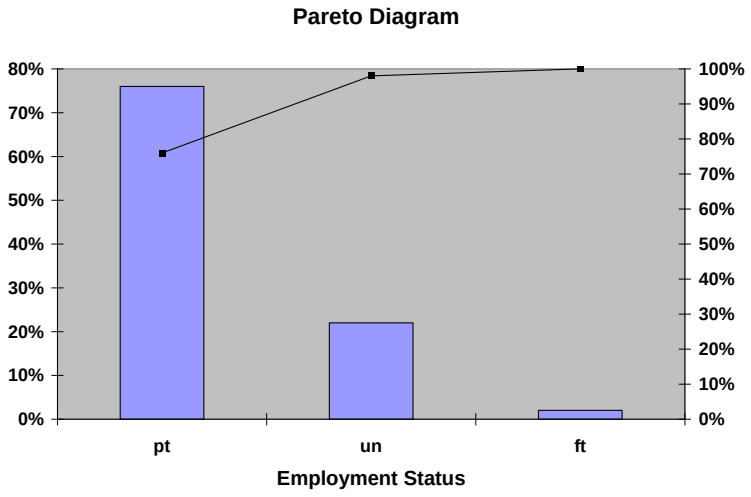


The percentages of students planning to attend graduate school are roughly evenly distributed among "Yes", "No" and "Not Sure".

2.117
cont. **Employment Status:**

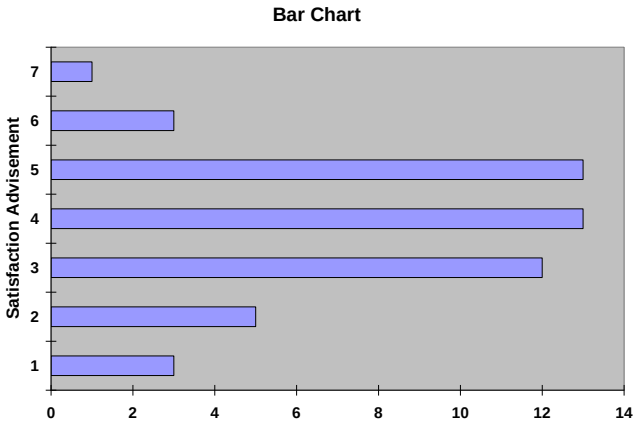


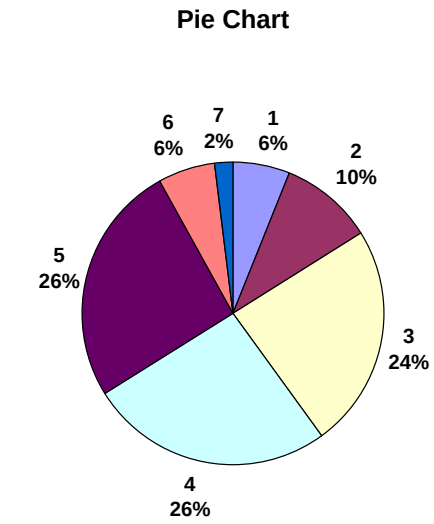


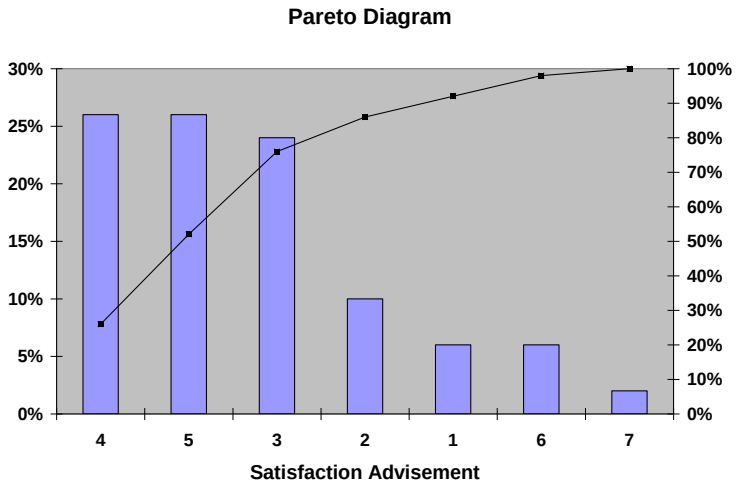


Part-time students constitute the "vital few" while full-time and unemployed students make up the "trivial many".

2.117
cont. **Satisfaction Advisement Services:**







About 80% of the students rated satisfaction with advisement services in the range between 3 to 5.

2.117
cont. **Number Affiliations:**

Stem-and-Leaf Display
for Number of Affiliations
Stem unit: 1

```
0 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  | 0 0 0 0 0
1 | 0 0 0 0 0 0 0 0 0 0 0
2 | 0 0 0 0 0
3 | 0 0
```

Majority of the students (64%) have no affiliation with clubs, groups, organizations or teams currently.

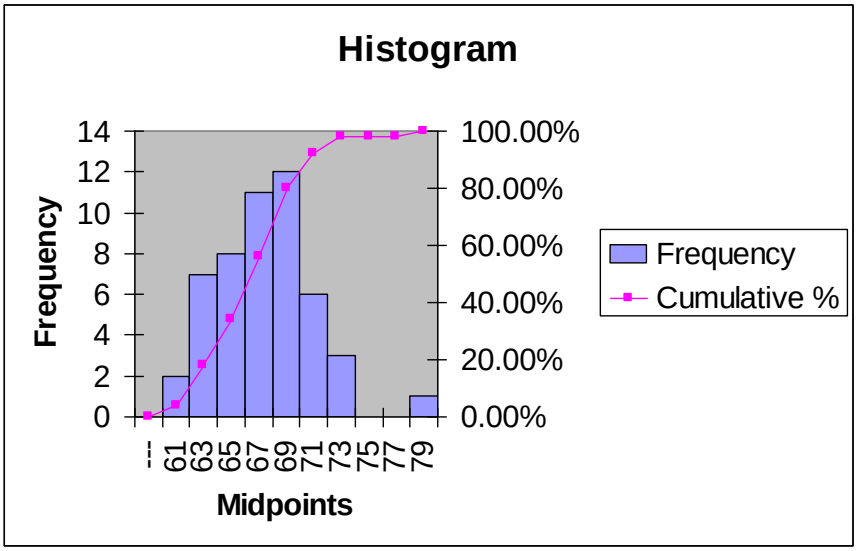
Age:

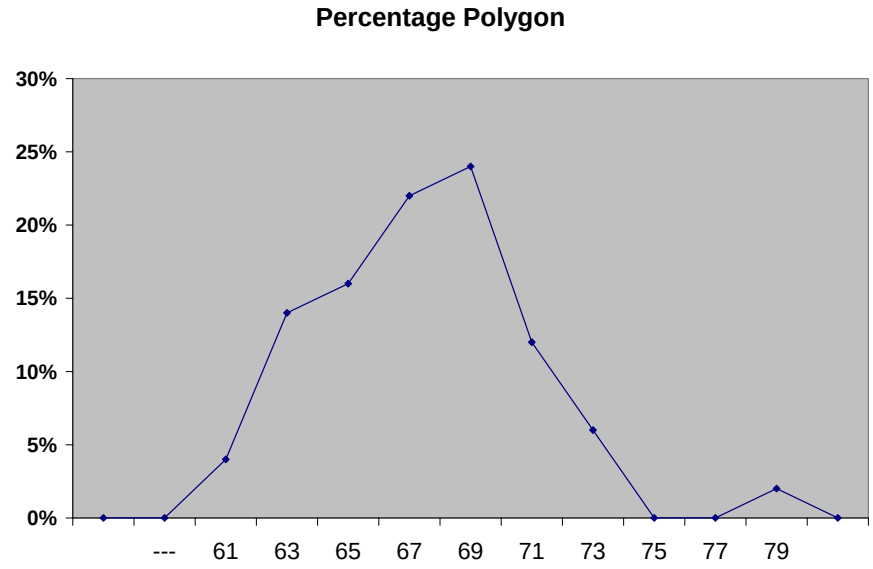
Stem-and-Leaf Display
for Age
Stem unit: 1

```
18 | 0
19 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0
20 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
21 | 0 0 0 0 0 0 0 0 0 0
22 | 0 0 0 0
23 | 0
24 | 0
25 |
26 |
27 |
28 |
29 |
30 | 0
31 |
32 |
33 | 0
34 |
35 |
36 | 0
```

Majority (88%) of the students surveyed are between 19 and 22 year old.

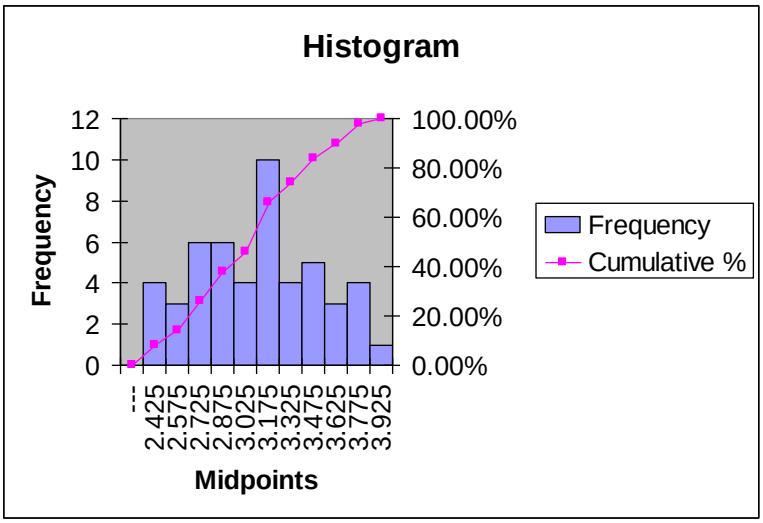
2.117
cont. **Height:**



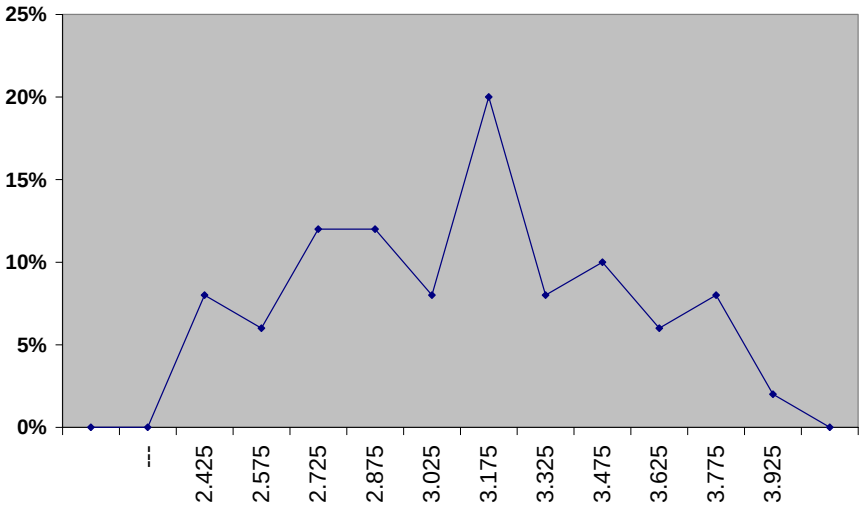


Height is right-skewed.

2.117
cont. **GPA:**



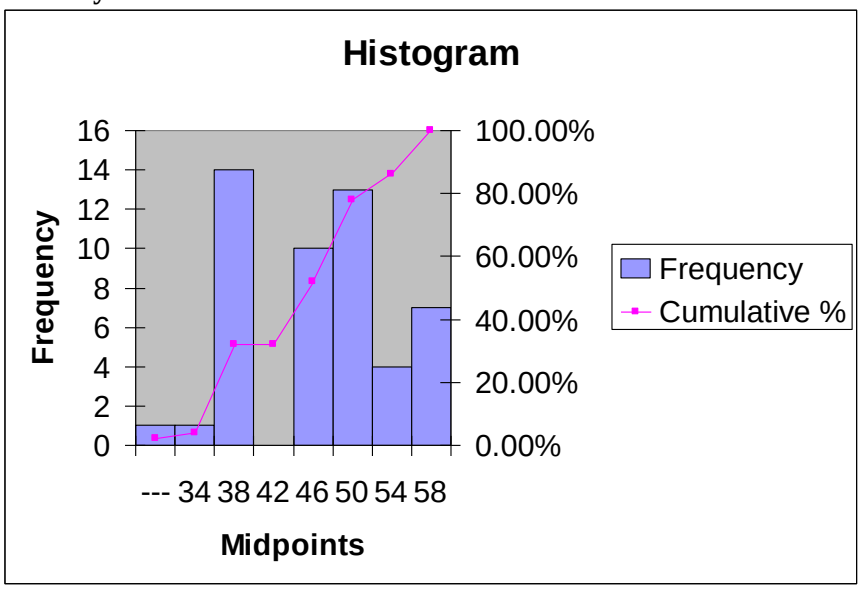
Percentage Polygon

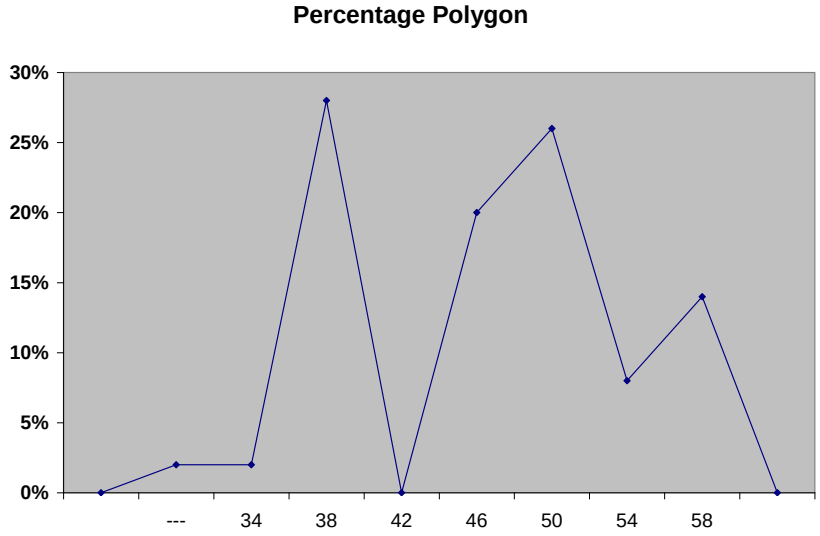


GPA is quite symmetrically distributed around 3.0.

2.117
cont.

Expected Salary:

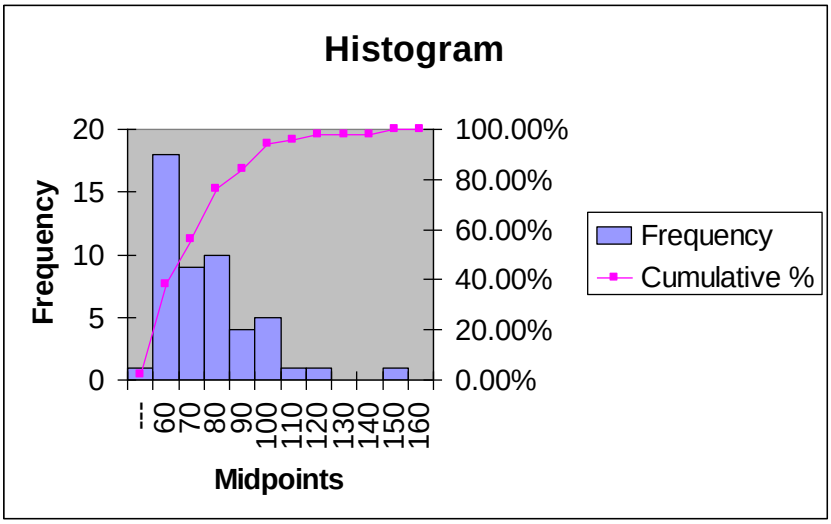


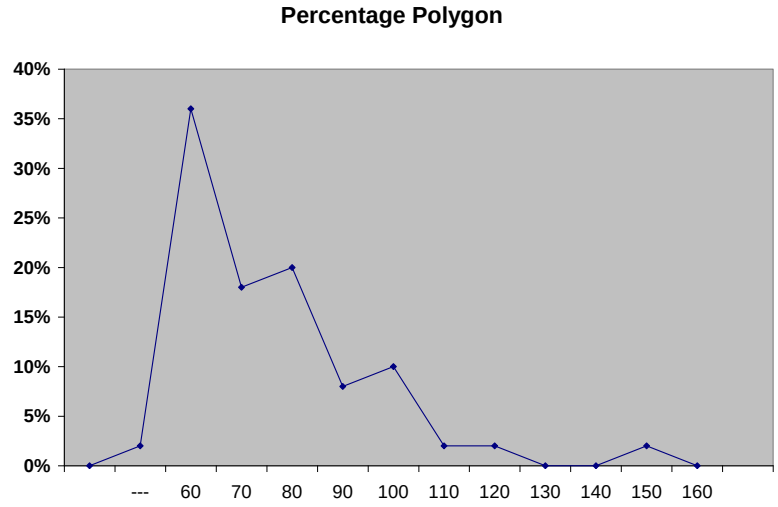


Expected salary is somewhat left-skewed.

2.117
cont.

Annual Salary in 5 Years:

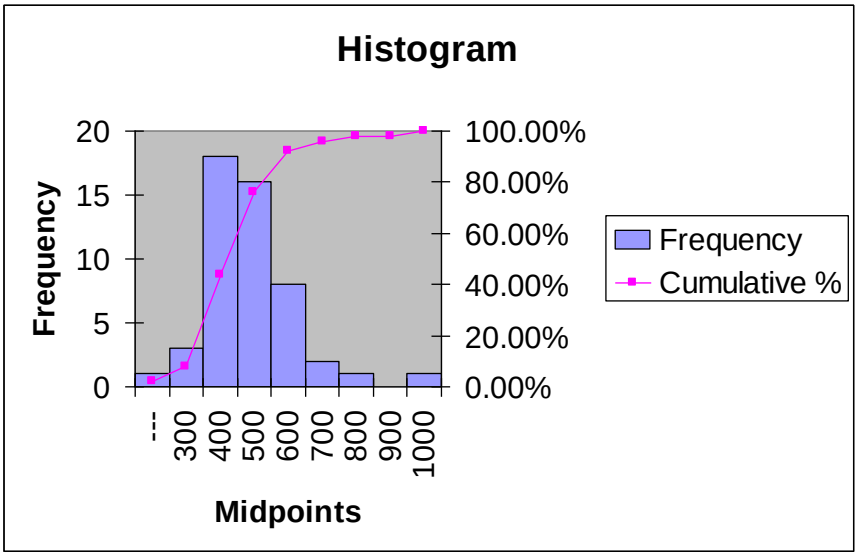


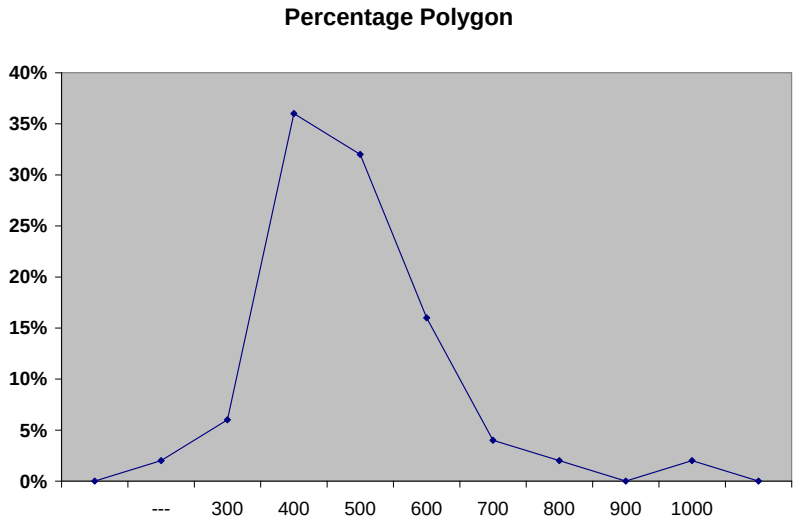


Annual salary in five years is right-skewed.

2.117
cont.

Spending:





Spending is also right-skewed.

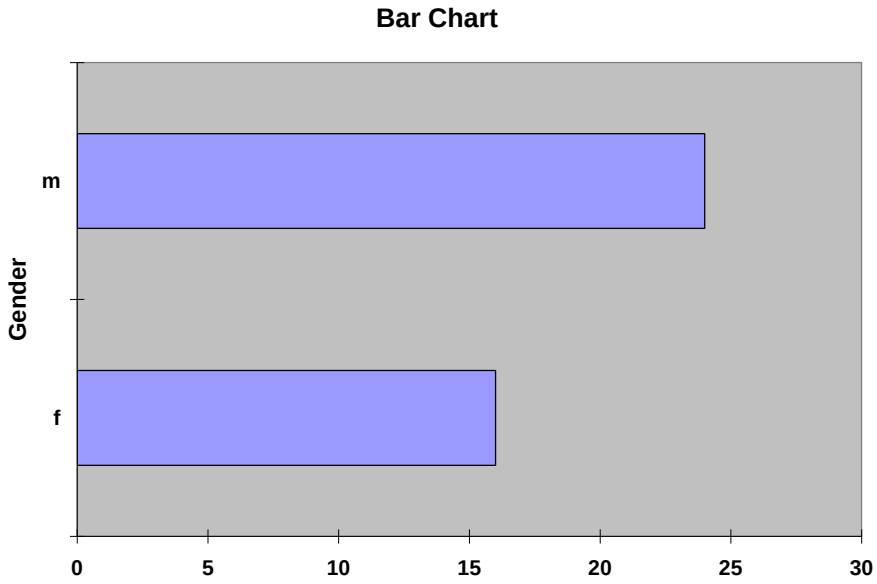
2.119 (a) Categorical variables: gender, major, undergrad specialization, employment status, satisfaction advisement.

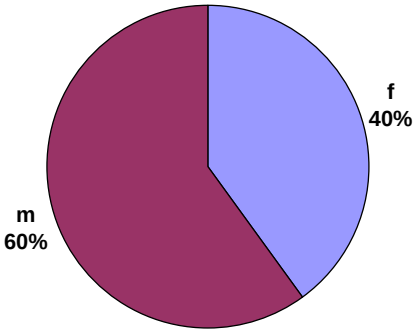
(b) Numerical variables: age, height, graduate GPA, undergrad GPA, GMAT, number of jobs, expected salary, anticipated salary in 5 years, spending.

(c) Discrete numerical variables: age (in years), height (in inches), GMAT, number of jobs, expected salary (in thousands of dollars), anticipated salary in 5 years (in thousands of dollars), spending (in dollars).

(d)

Gender:

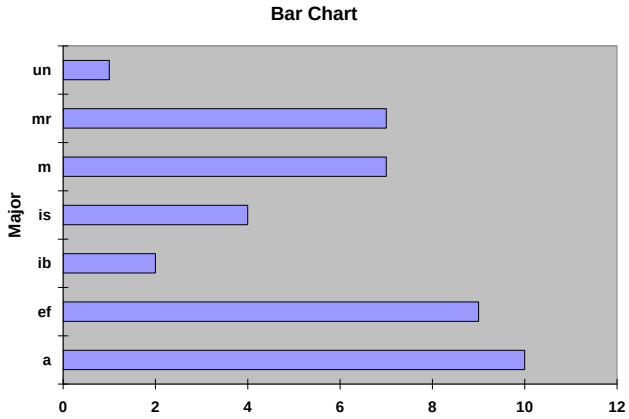


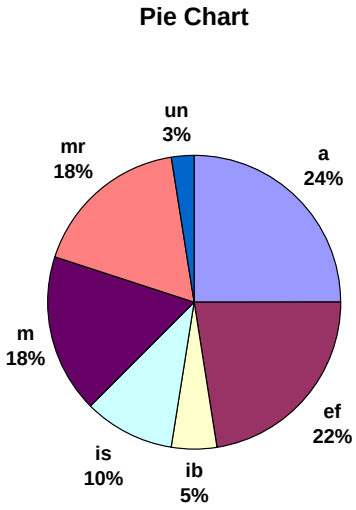


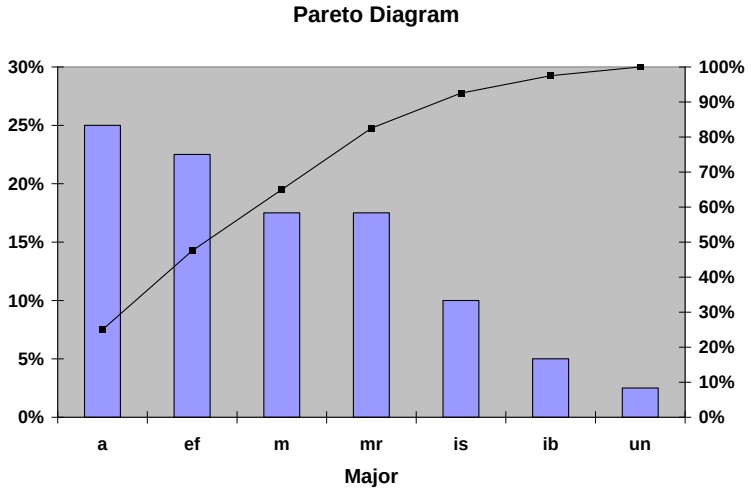
There are more males than females in the survey.

2.119
cont.

Major:

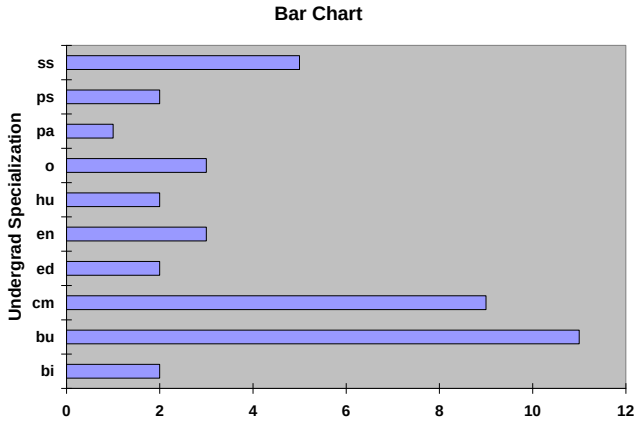


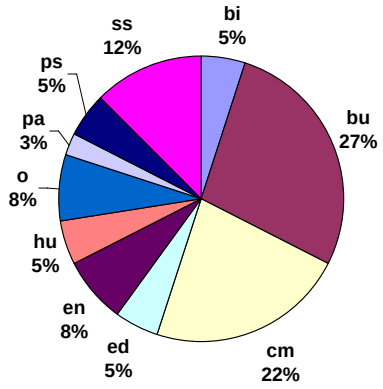


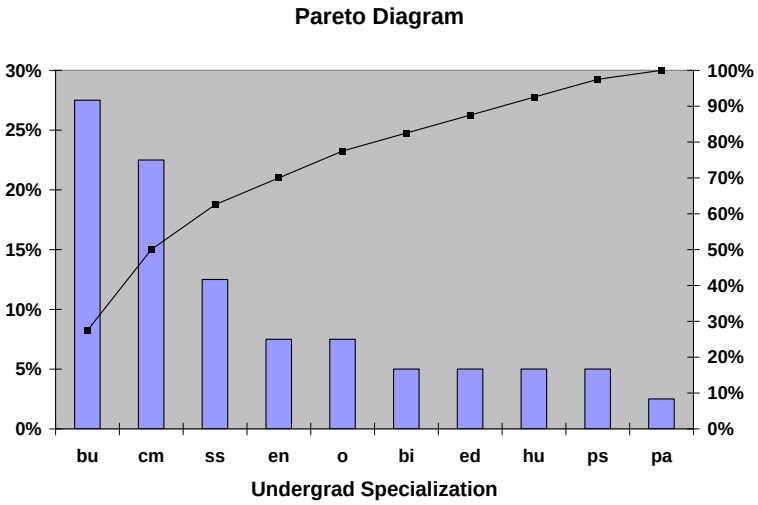


The "vital few" of accounting, economics/finance, management and marketing/retailing account for more than 80% of the majors.

2. 119
cont. **Undergraduate Specialization:**

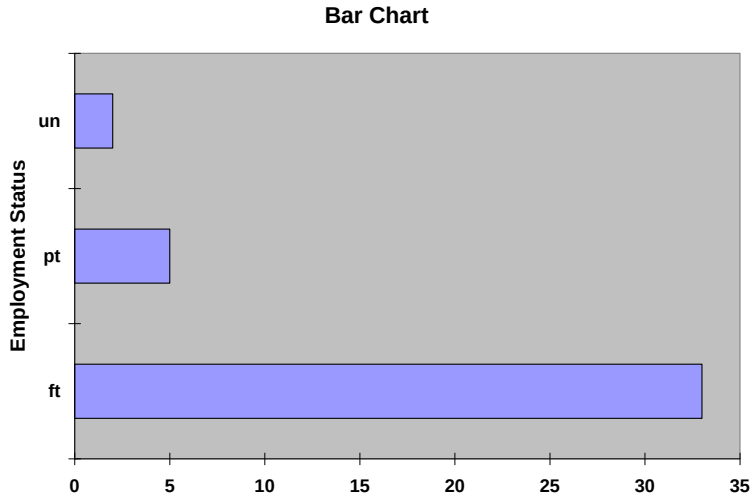


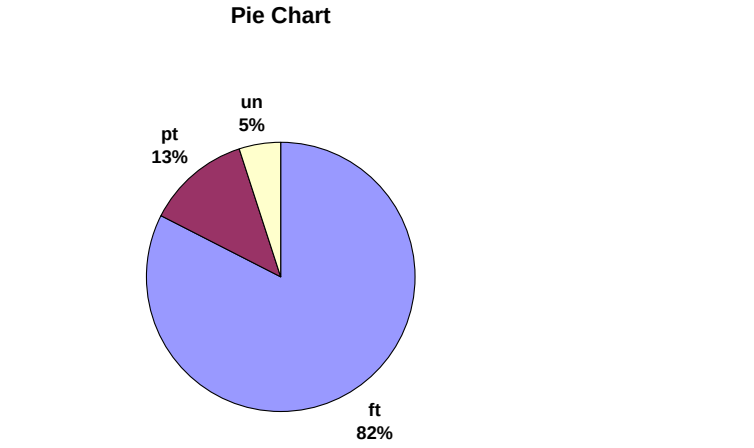


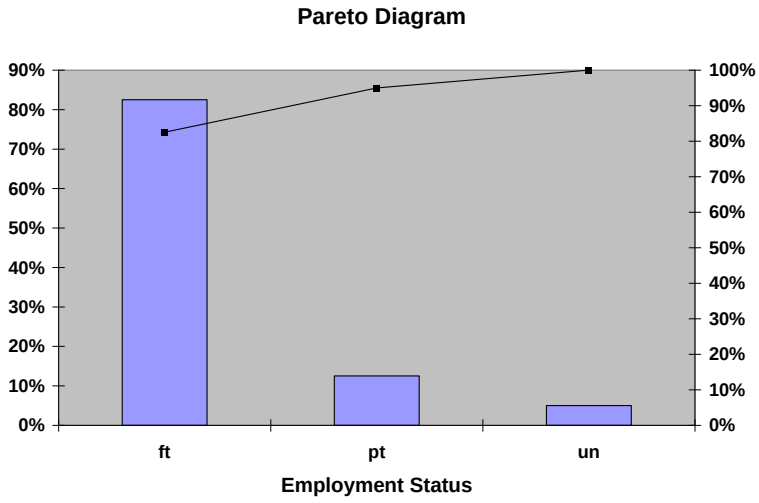


The "vital few" of business administration and computer/math account for half of the undergraduate specialization.

2. 119
cont. **Employment Status:**

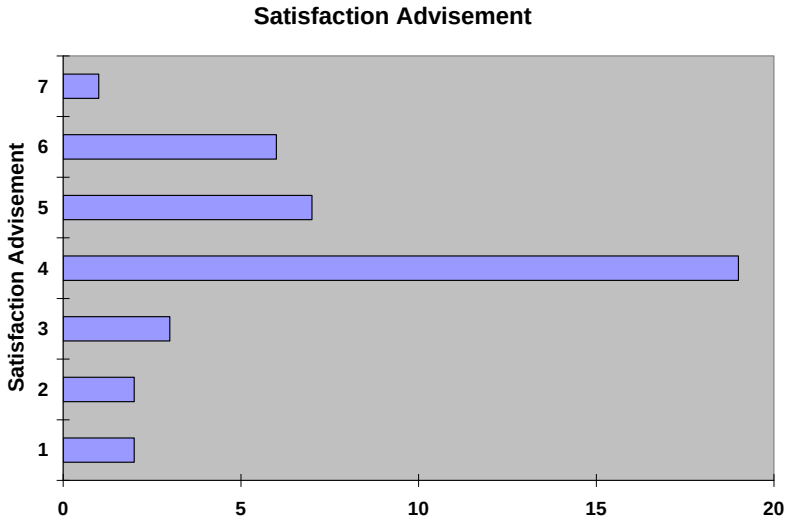


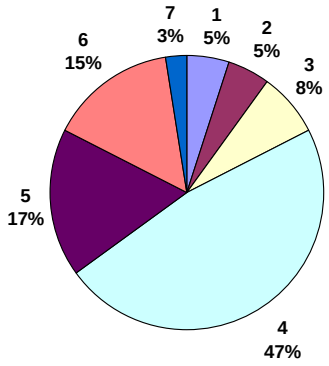


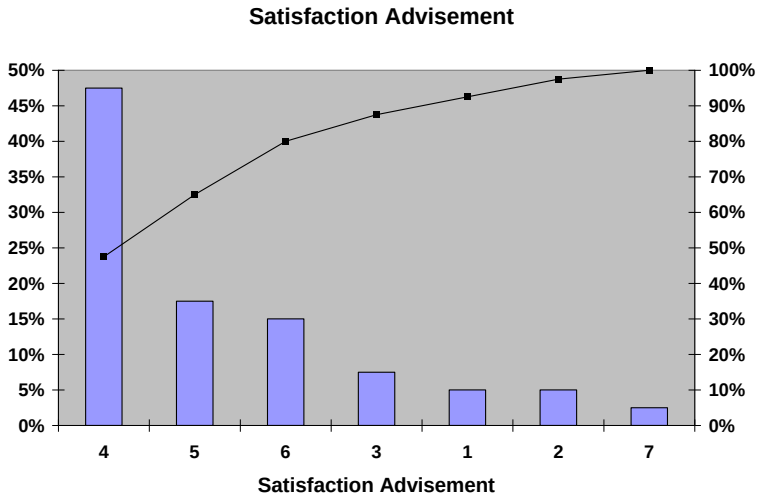


"Full-time" employment status accounts for more than 80% of the students.

2. 119
cont. **Satisfaction Advisement Services:**







80% of the students rated their satisfaction advisement services at between 4 and 6.

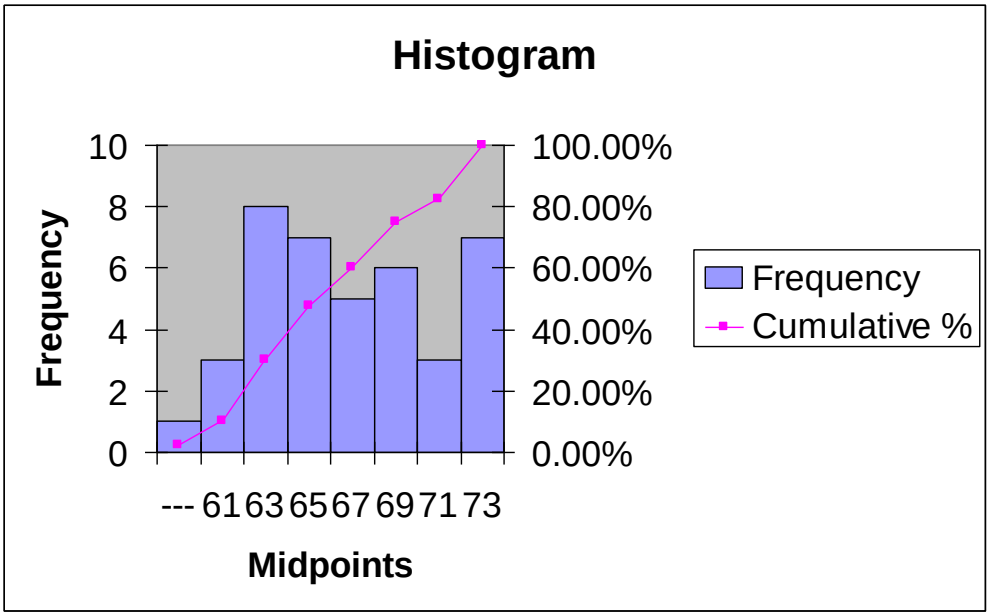
2. 119
cont. **Age:**

Stem-and-Leaf Display for Age
Stem unit: 1

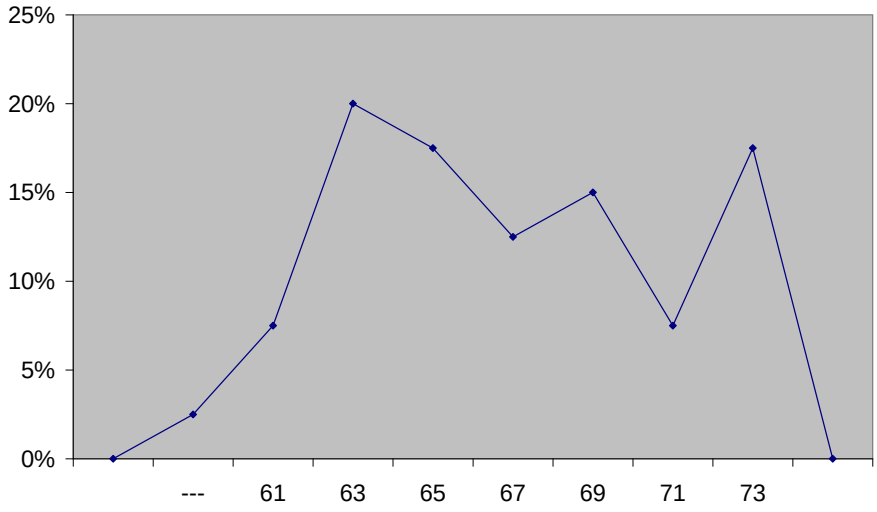
22	0
23	0
24	0 0
25	0 0 0 0 0
26	0 0 0
27	0 0 0
28	0 0 0
29	0 0
30	0 0 0 0
31	0 0 0
32	0 0 0
33	0 0
34	0
35	0 0
36	0
37	0
38	0
39	0
40	
41	0

Age is right-skewed.

2. 119 **Height:**
cont.



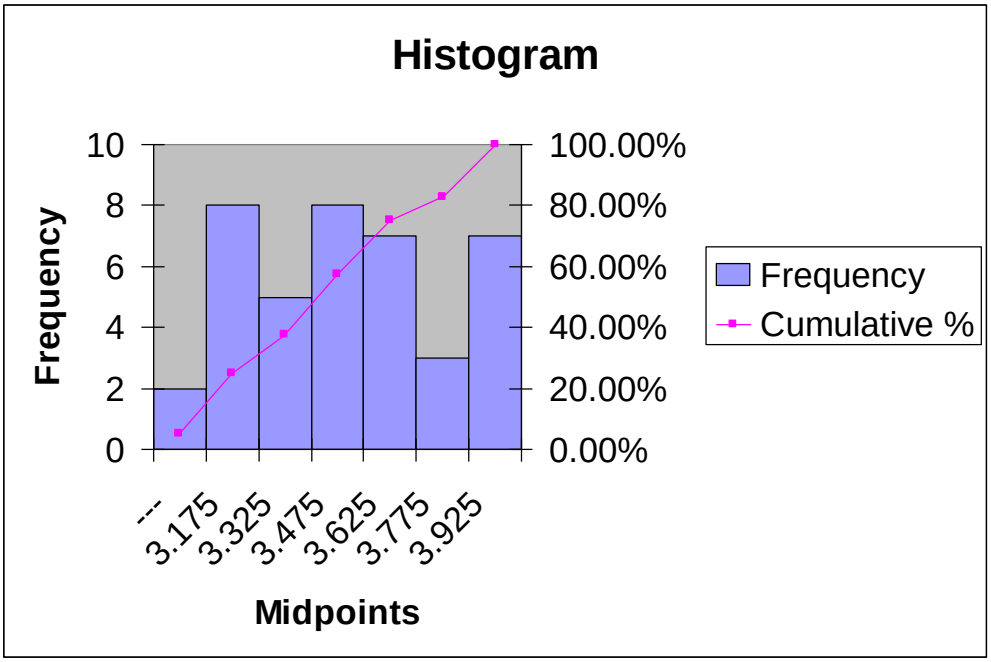
Percentage Polygon



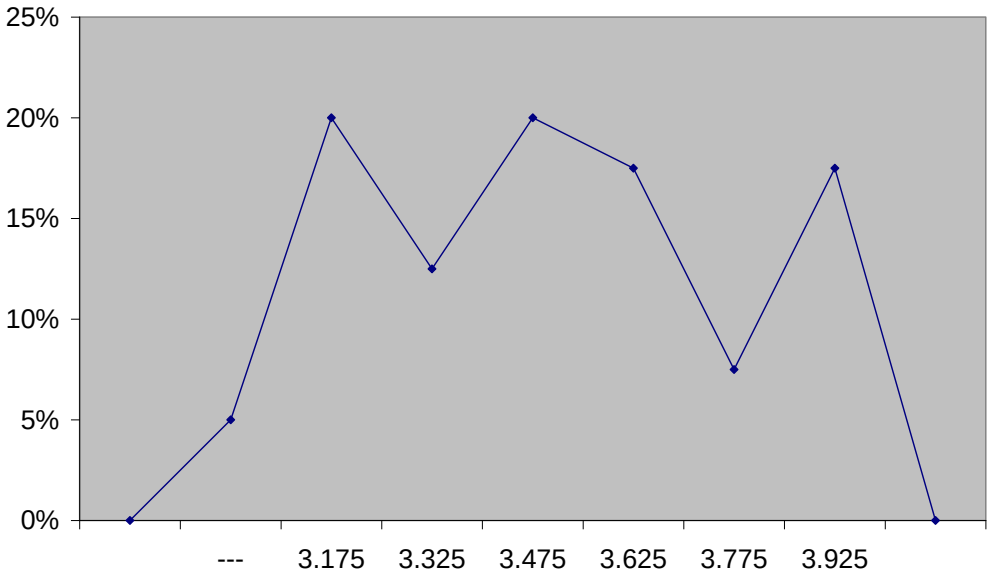
Height is left-skewed.

2. 119
cont.

Grad GPA:



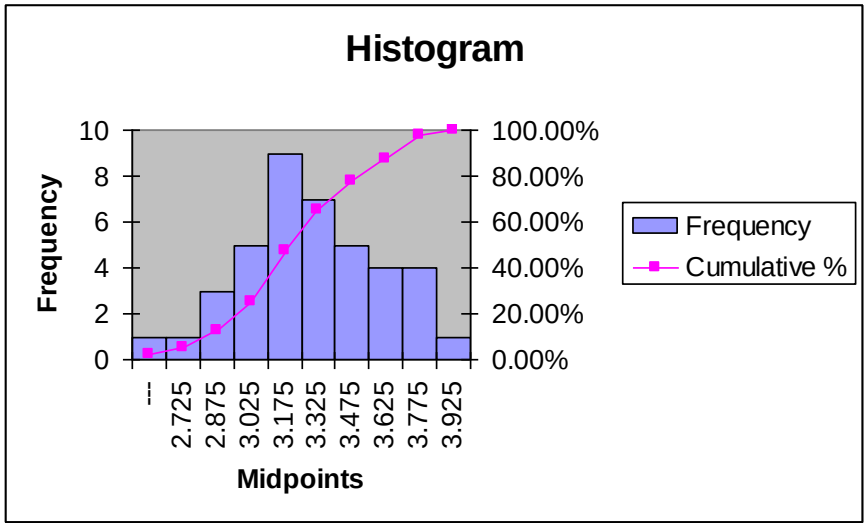
Percentage Polygon

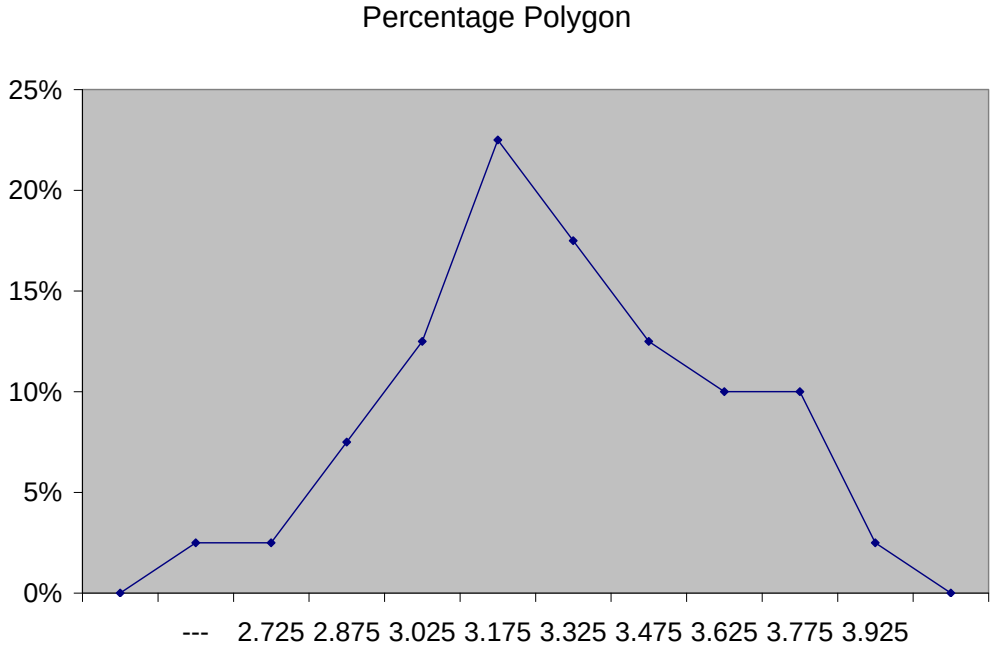


Graduate GPA is left-skewed.

2. 119
cont.

Undergraduate GPA:

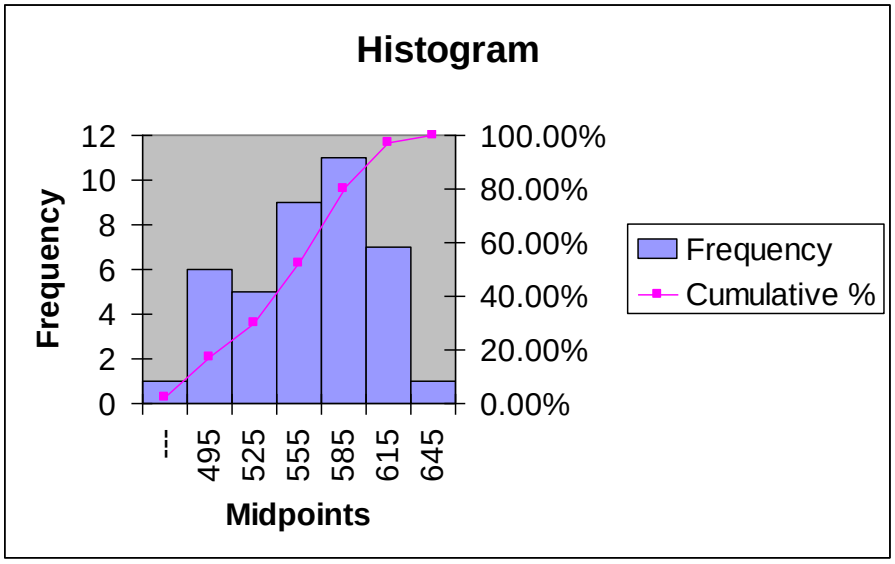




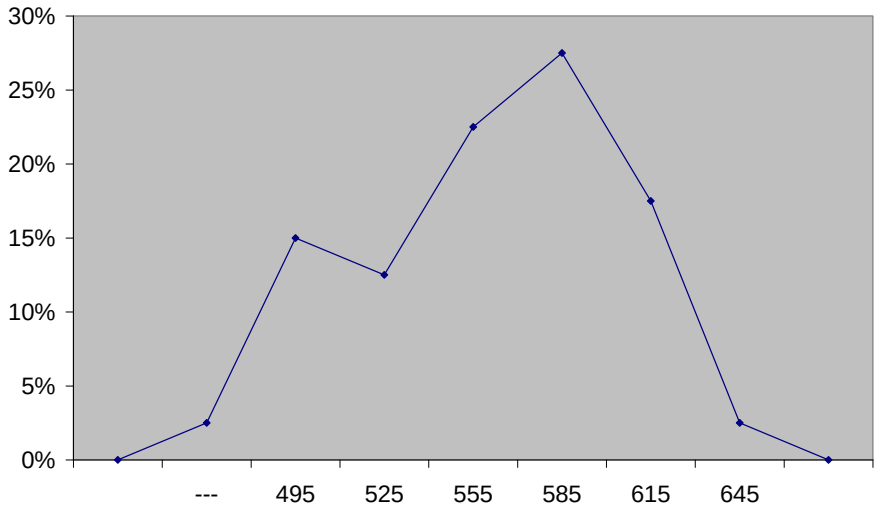
Undergraduate GPA is left-skewed.

2. 119
cont.

GMAT:



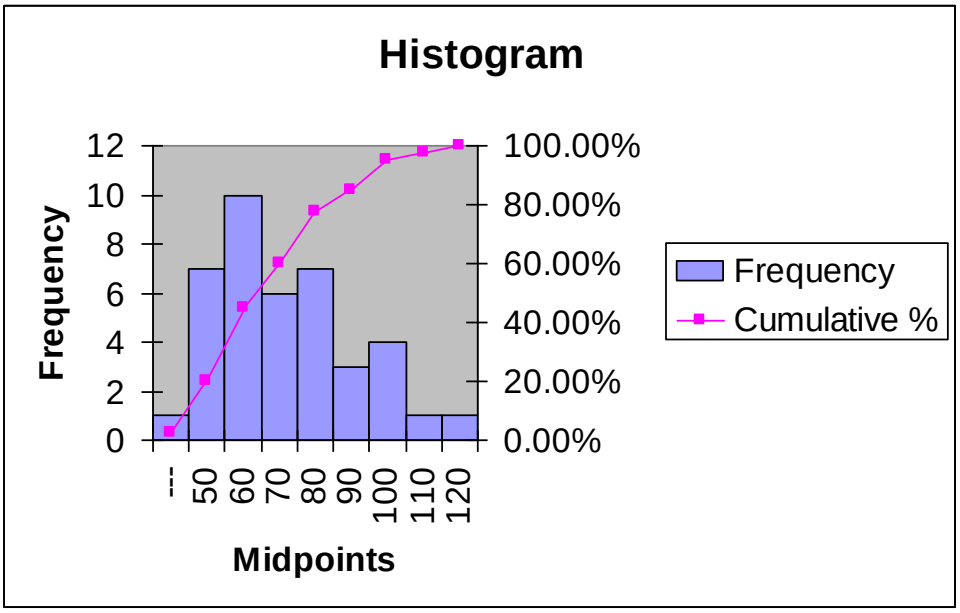
Percentage Polygon



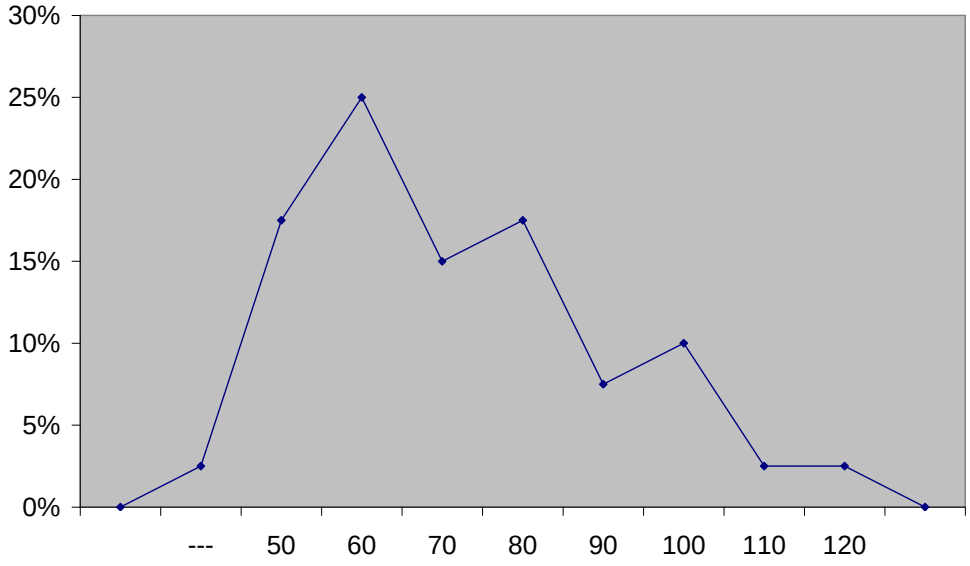
GMAT score is left-skewed.

2. 119
cont.

Expected Salary:



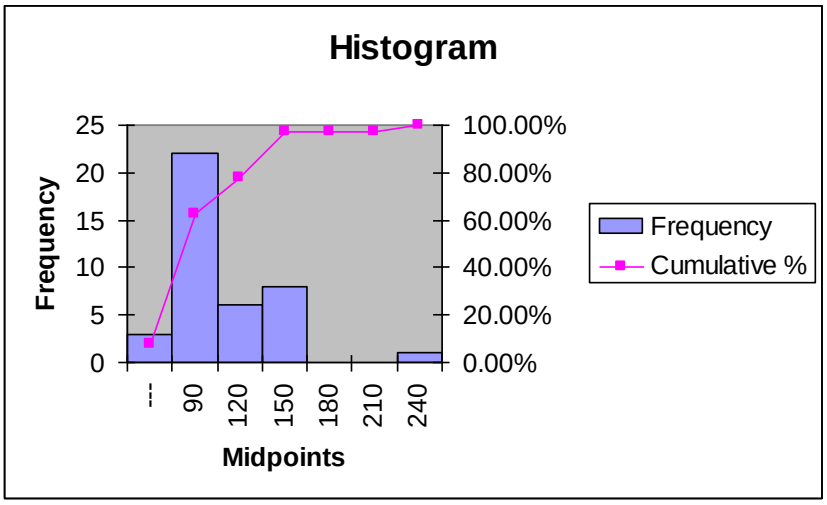
Percentage Polygon



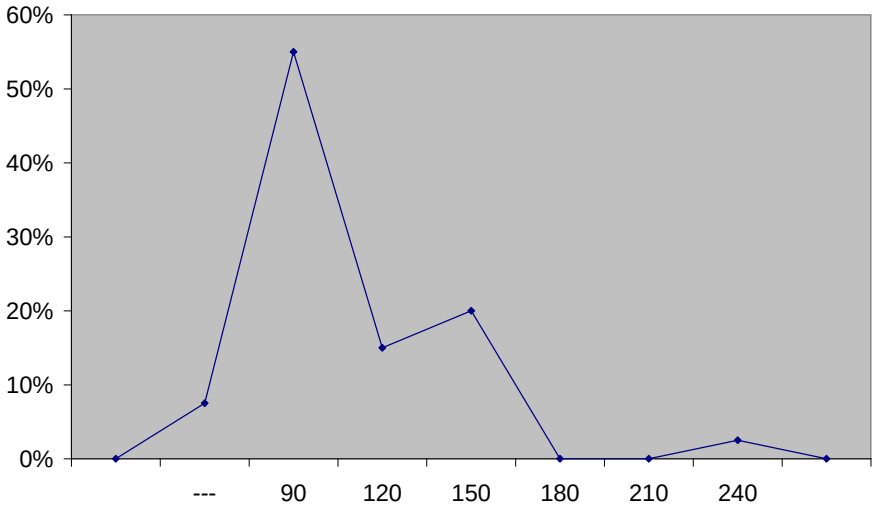
Expected salary is right-skewed.

2. 119
cont.

Anticipated Salary in 5 Years:



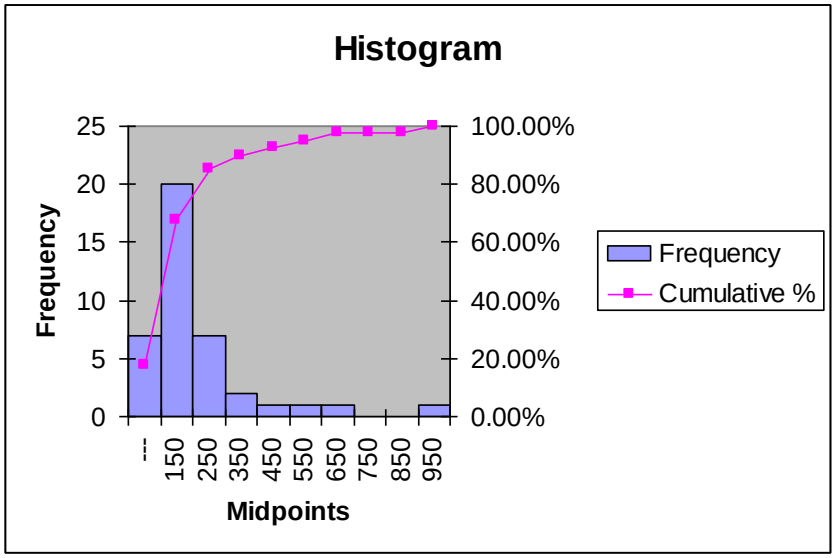
Percentage Polygon



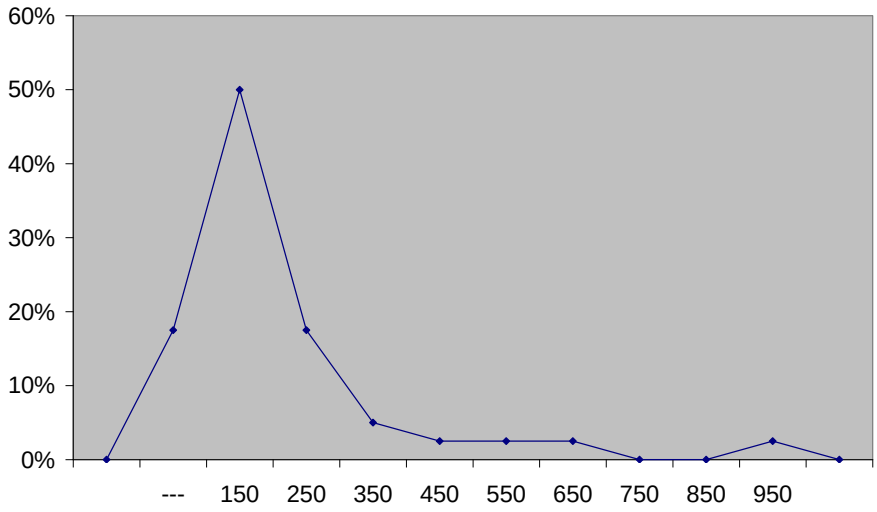
Anticipated salary is right-skewed.

2. 119
cont.

Spending:



Percentage Polygon



Spending is right-skewed.