Chapter 2
Descriptive Statistics: Tabular and Graphical Displays

Learning Objectives

1.   Learn how to construct and interpret summarization procedures for qualitative data such as: frequency and relative frequency distributions, bar graphs and pie charts.

2.   Learn how to construct and interpret tabular summarization procedures for quantitative data such as: frequency and relative frequency distributions, cumulative frequency and cumulative relative frequency distributions.

3.   Learn how to construct a dot plot and a histogram as graphical summaries of quantitative data.

4.   Learn how the shape of a data distribution is revealed by a histogram. Learn how to recognize when a data distribution is negatively skewed, symmetric, and positively skewed.

5.   Be able to use and interpret the exploratory data analysis technique of a stem-and-leaf display.

6.   Learn how to construct and interpret cross tabulations, scatter diagrams, side-by-side and stacked bar charts.

7.   Learn best practices for creating effective graphical displays and for choosing the appropriate type of display.

Solutions:

1.

Class	Frequency	Relative Frequency
A	60	60/120 = 0.50
B	24	24/120 = 0.20
C	36	36/120 = 0.30
	120	1.00

2. a. $1 - (.22 + .18 + .40) = .20$

   b. $.20(200) = 40$

   c/d.

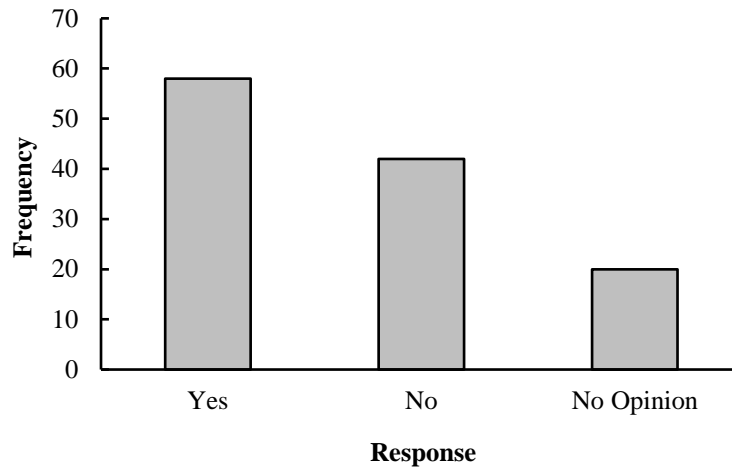| Class	Frequency	Percent Frequency
A	.22(200) = 44	22
B	.18(200) = 36	18
C	.40(200) = 80	40
D	.20(200) = 40	20
Total	200	100

3. a. $360° \times 58/120 = 174°$

   b. $360° \times 42/120 = 126°$

   c.
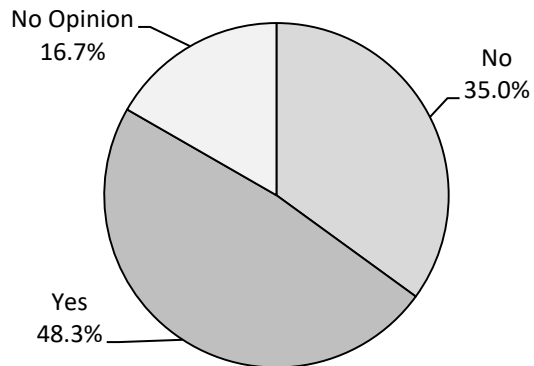
d.

4. a. These data are categorical.

b.

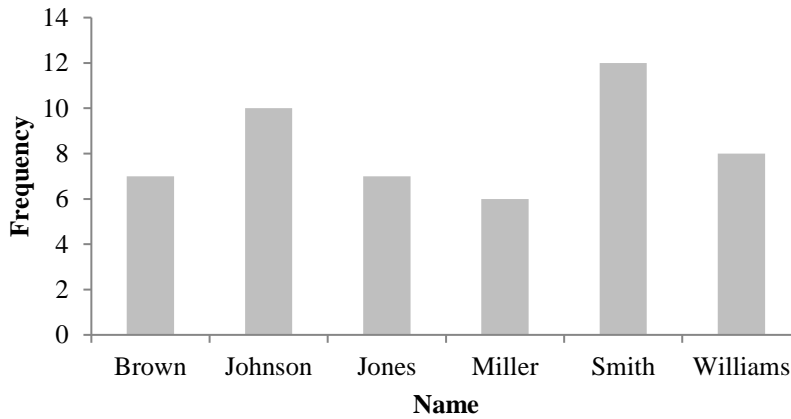| Show	Relative Frequency	% Frequency
Jep	10	20
JJ	8	16
OWS	7	14
THM	12	24
WoF	13	26
Total	50	100

c.
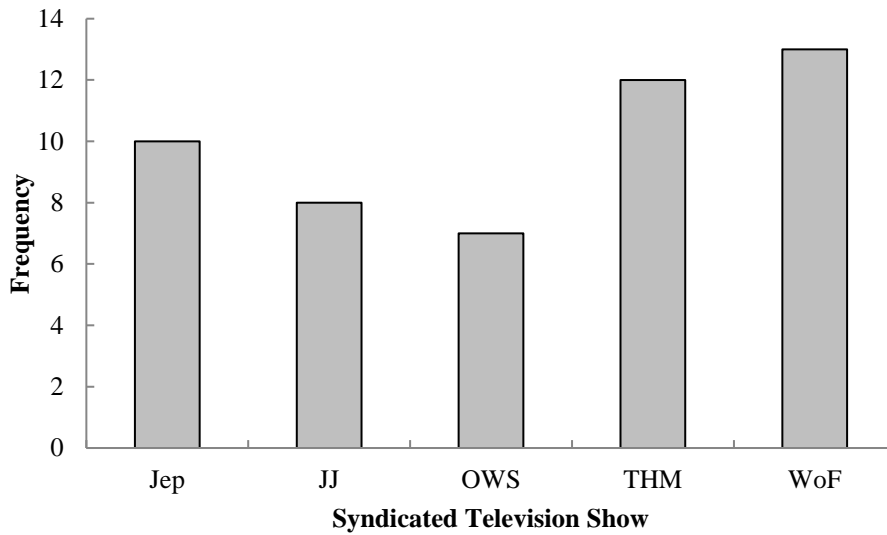
Syndicated Television Shows
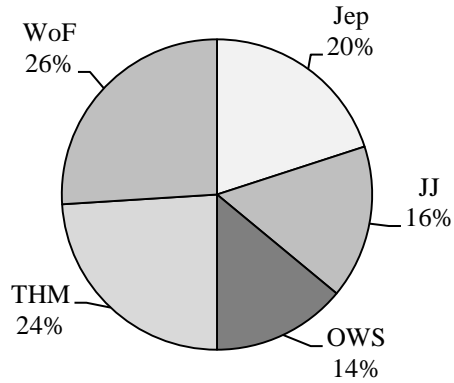
d. The largest viewing audience is for *Wheel of Fortune* and the second largest is for *Two and a Half Men*.

5. a.

Name	Frequency	Relative Frequency	Percent Frequency
Brown	7	0.14	14%
Johnson	10	0.20	20%
Jones	7	0.14	14%
Miller	6	0.12	12%
Smith	12	0.24	24%
Williams	8	0.16	16%
Total:	50	1	100%

b.

Common U.S. Last Names

c.

Common U.S. Last Names

Williams 16% — Brown 14% — Johnson 20% — Jones 14% — Miller 12% — Smith 24%
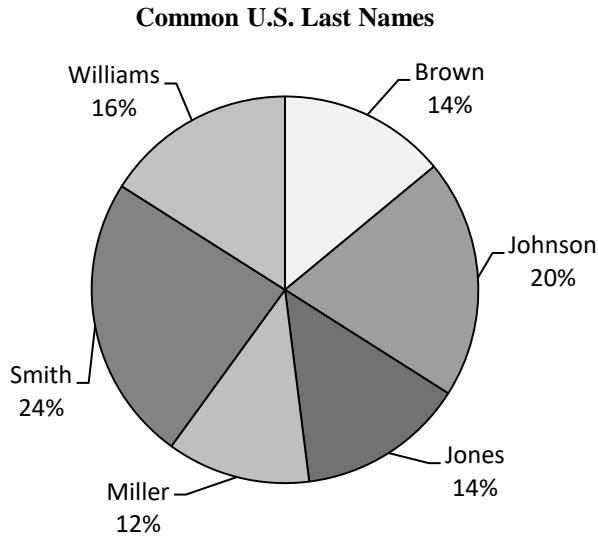
d. The three most common last names are Smith (24%), Johnson (20%), Williams (16%5)

6. a.

Network	Relative Frequency	% Frequency
ABC	6	24
CBS	9	36
FOX	1	4
NBC	9	36
Total:	25	100
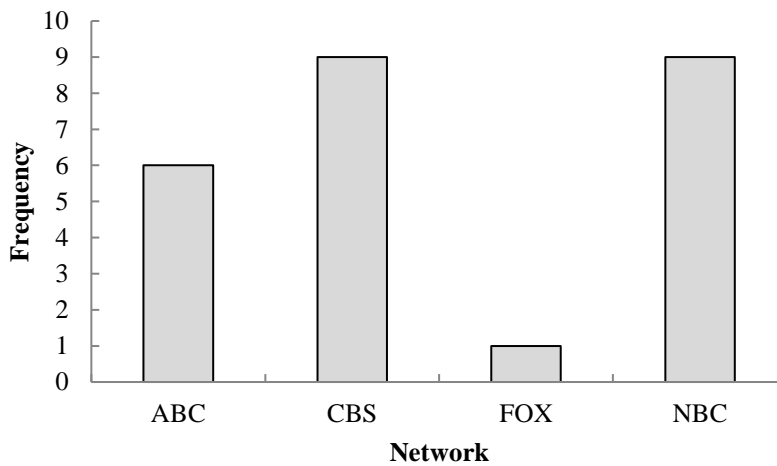
b. For these data, NBC and CBS tie for the number of top-rated shows. Each has 9 (36%) of the top 25. ABC is third with 6 (24%) and the much younger FOX network has 1(4%).

7. a.

Rating	Frequency	Percent Frequency
Excellent	20	40
Very Good	23	46
Good	4	8
Fair	1	2
Poor	2	4
	50	100

Management should be very pleased with the survey results. 40% + 46% = 86% of the ratings are very good to excellent. 94% of the ratings are good or better. This does not look to be a Delta flight where significant changes are needed to improve the overall customer satisfaction ratings.

b. While the overall ratings look fine, note that one customer (2%) rated the overall experience with the flight as Fair and two customers (4%) rated the overall experience with the flight as Poor. It might be insightful for the manager to review explanations from these customers as to how the flight failed to meet expectations. Perhaps, it was an experience with other passengers that Delta could do little to correct or perhaps it was an isolated incident that Delta could take steps to correct in the future.

8. a.

Position	Frequency	Relative Frequency
Pitcher	17	0.309
Catcher	4	0.073
1st Base	5	0.091
2nd Base	4	0.073
3rd Base	2	0.036
Shortstop	5	0.091
Left Field	6	0.109
Center Field	5	0.091
Right Field	7	0.127
	55	1.000

b. Pitchers (Almost 31%)

c. 3rd Base (3 – 4%)

d. Right Field (Almost 13%)

e. Infielders (16 or 29.1%) to Outfielders (18 or 32.7%)

9. a.

Living Area	Live Now	Ideal Community
City	32%	24%
Suburb	26%	25%
Small Town	26%	30%
Rural Area	16%	21%
Total	100%	100%

b. Where do you live now?
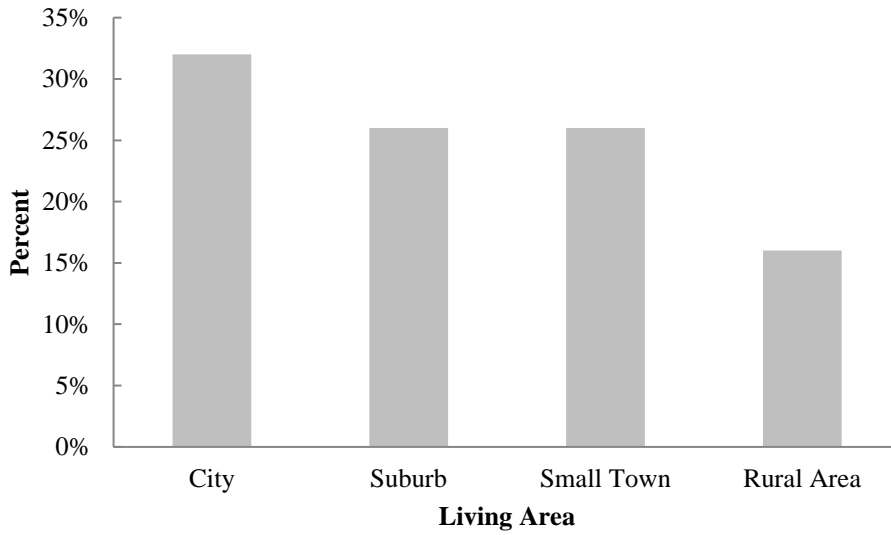
What do you consider the ideal community?
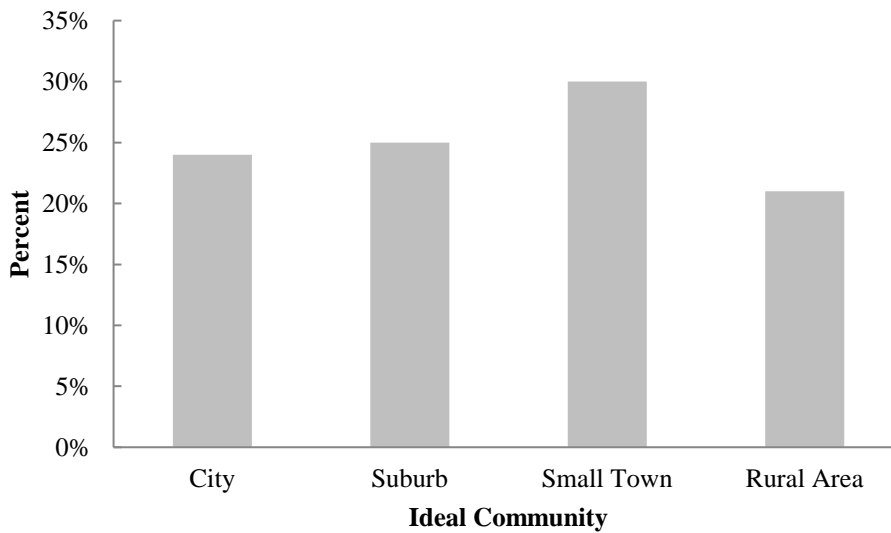
c.   Most adults are now living in a city (32%).

d.   Most adults consider the ideal community a small town (30%).

e.   Percent changes by living area: City –8%, Suburb –1%, Small Town +4%, and Rural Area +5%.
     Suburb living is steady, but the trend would be that living in the city would decline while
     living in small towns and rural areas would increase.

10. a.

Rating	Frequency
Excellent	20
Good	101
Fair	528
Bad	244
Terrible	122
Total	1015

b.

Rating	Percent Frequency
Excellent	2
Good	10
Fair	52
Bad	24
Terrible	12
Total	100

c.
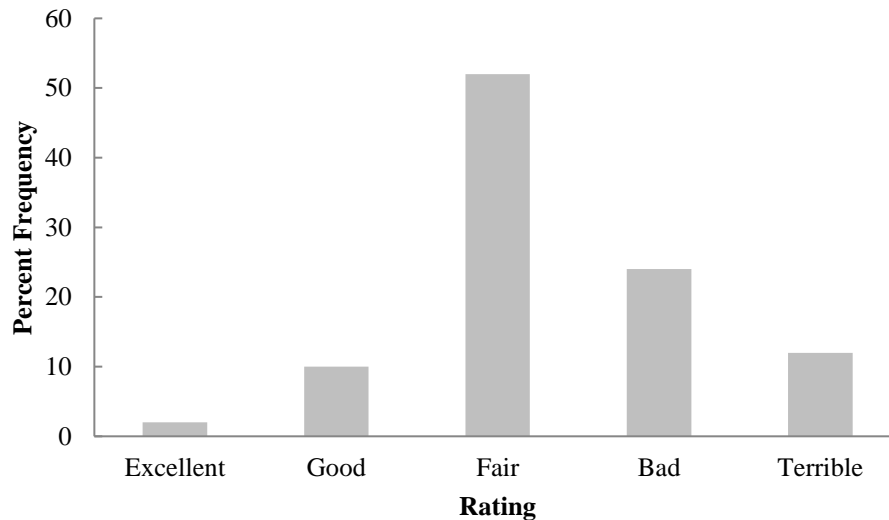
d.   24% + 12% = 36% of adults in the United Sates think the Federal Bank is doing a bad or a terrible
     job in handling the credit problems. Only 10% + 2% = 12% think the Federal Bank is doing a good
     or excellent job.

e. 40% + 10% = 50% of adults in Spain think the European Central Bank is doing a bad or terrible job in handling the credit problems. Only 4% of adults in Spain think the European Central Bank is doing a good or excellent job.

Both countries show pessimism and relatively low confidence in how the banks are handling the credit problems in the financial markets. But in comparing the two countries, adults in Spain show more concern and more pessimism about the bank's ability compared to adults in the United States.

11.

Class	Frequency	Relative Frequency	Percent Frequency
12–14	2	0.050	5.0
15–17	8	0.200	20.0
18–20	11	0.275	27.5
21–23	10	0.250	25.0
24–26	9	0.225	22.5
Total	40	1.000	100.0

12.

Class	Cumulative Frequency	Cumulative Relative Frequency
less than or equal to 19	10	.20
less than or equal to 29	24	.48
less than or equal to 39	41	.82
less than or equal to 49	48	.96
less than or equal to 59	50	1.00

13.
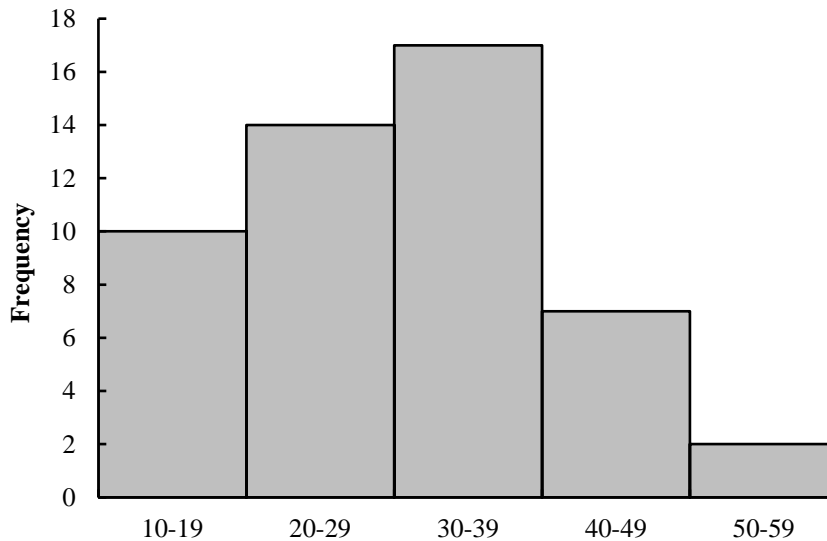
14. a.
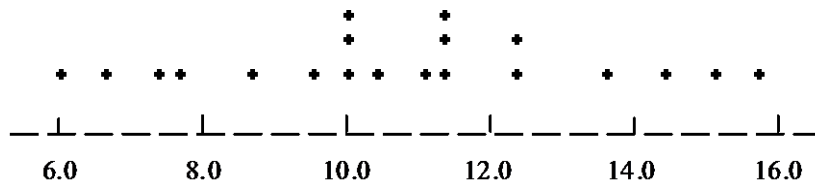
b/c.

Class	Frequency	Percent Frequency
6.0 – 7.9	4	20
8.0 – 9.9	2	10
10.0 – 11.9	8	40
12.0 – 13.9	3	15
14.0 – 15.9	3	15
	20	100

15.   Leaf Unit = .1

6	3
7	5 5 7
8	1 3 4 8
9	3 6
10	0 4 5
11	3

16.   Leaf Unit = 10

11	6
12	0 2
13	0 6 7
14	2 2 7
15	5
16	0 2 8
17	0 2 3

17. a/b.

Waiting Time	Frequency	Relative Frequency
0 – 4	4	0.20
5 – 9	8	0.40
10 – 14	5	0.25
15 – 19	2	0.10
20 – 24	1	0.05
Totals	20	1.00

c/d.

Waiting Time	Cumulative Frequency	Cumulative Relative Frequency
Less than or equal to 4	4	0.20
Less than or equal to 9	12	0.60
Less than or equal to 14	17	0.85
Less than or equal to 19	19	0.95
Less than or equal to 24	20	1.00

e.   12/20 = 0.60

18. a.

Salary	Frequency
150–159	1
160–169	3
170–179	7
180–189	5
190–199	1
200–209	2
210–219	1
Total	20

b.

Salary	Percent Frequency
150–159	5
160–169	15
170–179	35
180–189	25
190–199	5
200–209	10
210–219	5
Total	100

c.

Salary	Cumulative Percent Frequency
Less than or equal to 159	5
Less than or equal to 169	20
Less than or equal to 179	55
Less than or equal to 189	80
Less than or equal to 199	85
Less than or equal to 209	95
Less than or equal to 219	100
Total	100

d.
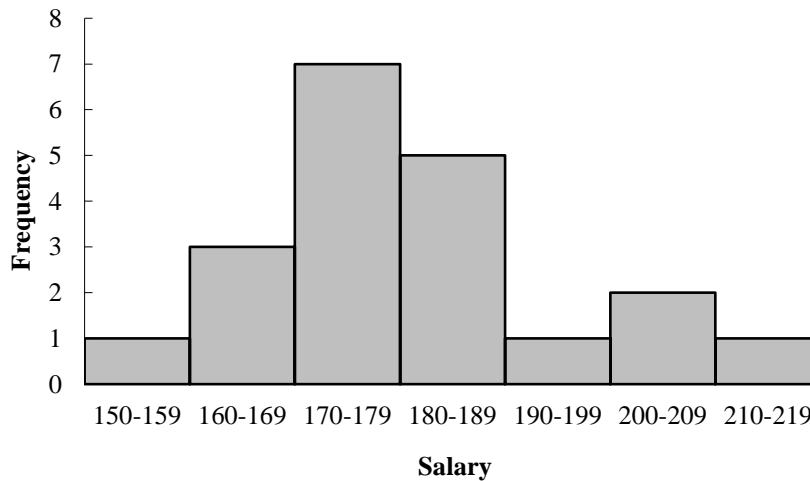
e.   There is skewness to the right.

f. $(3/20)(100) = 15\%$

19. a.   The highest price stock is for IBM with a price of $109 per share.
The lowest price stock is for Alcoa with a price of $11 per share.

b.   A class size of 10 results in 10 classes.

Price per Share	Frequency
$10–19	5
$20–29	10
$30–39	3
$40–49	2
$50–59	6
$60–69	2
$70–79	1
$80–89	0
$90–99	0
$100–109	1

c.
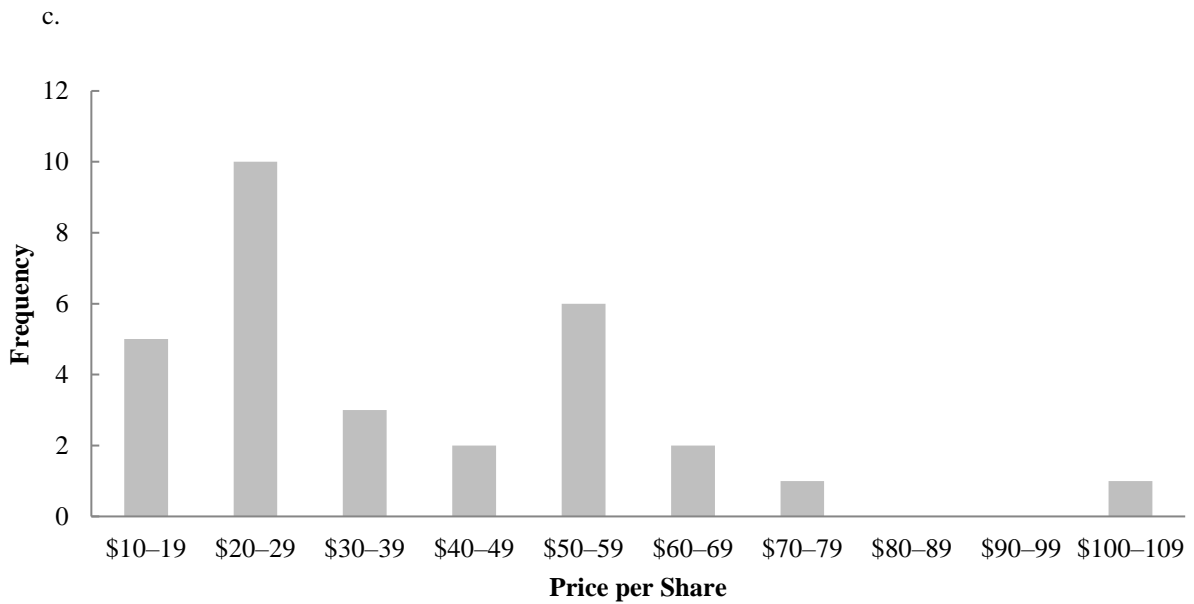
The general shape of the distribution is skewed to the right. Half of the companies (15) have a price per share less than $30. A mid-priced stock appears to be in the $30 to $49 range, while the most frequently priced stock is in the $20 to $29 range.

Five stocks are less than $20 per share (Alcoa, Bank of America, General Electric, Intel and Pfizer). Four stocks are $60 or more per share (3M, Chevron, ExxonMobil and IBM).
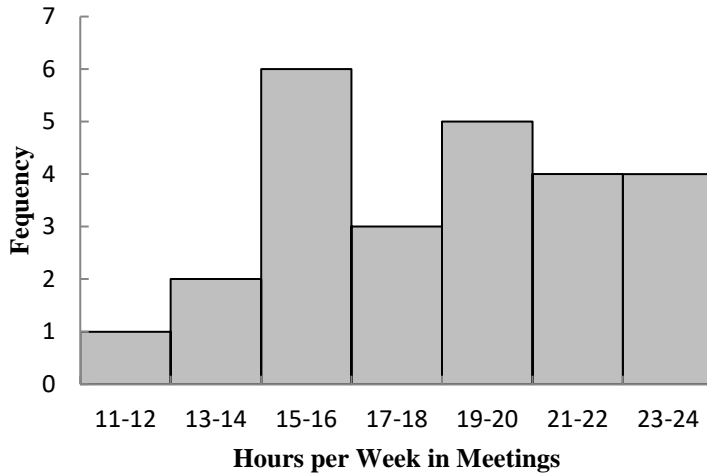
d.   A variety of comparisons are possible depending upon when the study is done.

20.   a.   Lowest = 12, Highest = 23

b.

Hours in Meetings per Week	Frequency	Percent Frequency
11-12	1	4%
13-14	2	8%
15-16	6	24%
17-18	3	12%
19-20	5	20%
21-22	4	16%
23-24	4	16%
	25	100%

c.

The distribution is slightly skewed to the left.

21.   a/b/c/d.

Revenue	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
0-49	6	.12	6	.12
50-99	29	.58	35	.70
100-149	11	.22	46	.92
150-199	0	.00	46	.92
200-249	1	.02	47	.94
250-299	1	.02	48	.96
300-349	0	.00	48	.96
350-399	0	.00	48	.96
400-449	2	.04	50	1.00
Total	50	1.00		

e. The majority of the large corporations (40) have revenues in the $50 billion to $149 billion range. Only 4 corporations have revenues of over $200 billion and only 2 corporations have revenues over $400 billion. .70, or 70%, of the corporations have revenues under $100 billion. .30, or 30%, of the corporations have revenues of $100 billion or more.

f.
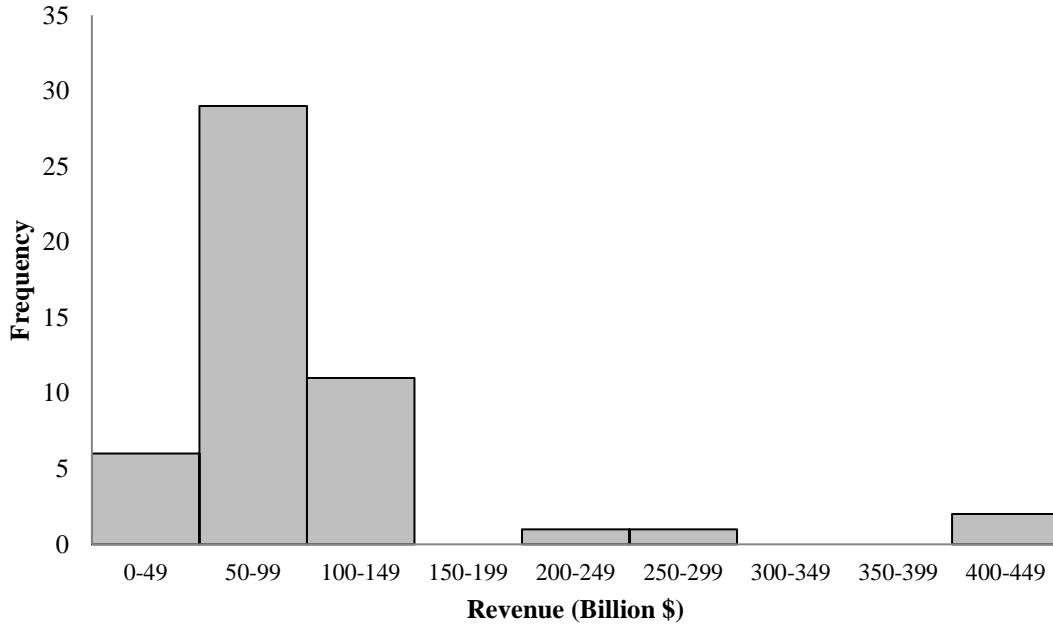
The histogram shows the distribution is skewed to the right with four corporations in the $200 to $449 billion range.

g. Exxon-Mobil is America's largest corporation with an annual revenue of $443 billion. Walmart is the second largest corporation with an annual revenue of $406 billion. All other corporations have annual revenues less than $300 billion. Most (92%) have annual revenues less than $150 billion.

22. a.

# U.S. Locations	Frequency	Percent Frequency
0-4999	10	50
5000-9999	3	15
10000-14999	2	10
15000-19999	1	5
20000-24999	0	0
25000-29999	1	5
30000-34999	2	10
35000-39999	1	5
Total:	20	100

b.

Number of U.S. Locations
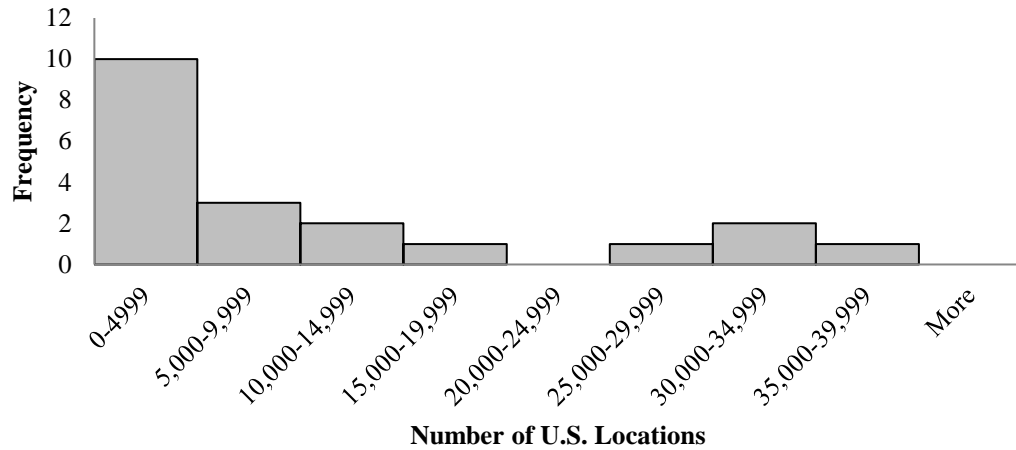
c. The distribution is skewed to the right. The majority of the franchises in this list have fewer than 20,000 locations (50% + 15% + 15% = 80%). McDonald's, Subway and 7-Eleven have the highest number of locations.

23. a/b.

Computer Usage (Hours)		Frequency	Relative Frequency
0.0 –	2.9	5	0.10
3.0 –	5.9	28	0.56
6.0 –	8.9	8	0.16
9.0 –	11.9	6	0.12
12.0 –	14.9	3	0.06
Total		50	1.00

c.

Computer Usage (Hours)
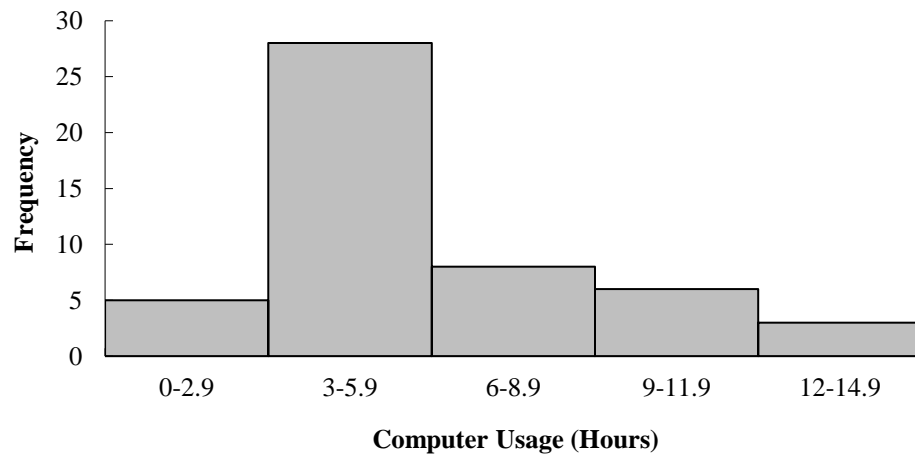
d. The majority of the computer users are in the 3 to 6 hour range. Usage is somewhat skewed toward the right with 3 users in the 12 to 14.9 hour range.

24.   Median Pay

```
 6 | 6  7  7
 7 | 2  4  6  7  7  8  9
 8 | 0  0  1  3  7
 9 | 9
10 | 0  6
11 | 0
12 | 1
```

The median pay for these careers is generally in the $70 and $80 thousands. Only four careers have a median pay of $100 thousand or more. The highest median pay is $121 thousand for a finance director.

Top Pay

```
10 | 0  6  9
11 | 1  6  9
12 | 2  5  6
13 | 0  5  8  8
14 | 0  6
15 | 2  5  7
16 |
17 |
18 |
19 |
20 |
21 | 4
22 | 1
```

The most frequent top pay is in the $130 thousand range. However, the top pay is rather evenly distributed between $100 and $160 thousand. Two unusually high top pay values occur at $214 thousand for a finance director and $221 thousand for an investment banker.   Also, note that the top pay has more variability than the median pay.

25.

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

26. a.

2	1 4
2	6 7
3	0 1 1 1 2 3
3	5 6 7 7
4	0 0 3 3 3 3 3 4 4
4	6 6 7 9
5	0 0 0 2 2
5	5 6 7 9
6	1 4
6	6
7	2

   b. Most frequent age group: 40-44 with 9 runners

   c. 43 was the most frequent age with 5 runners

27. a.

		y		
		1	2	Total
	A	5	0	5
x	B	11	2	13
	C	2	10	12
	Total	18	12	30

b.

	y		
	1	2	Total
A	100.0	0.0	100.0
x B	84.6	15.4	100.0
C	16.7	83.3	100.0

c.

	y	
	1	2
A	27.8	0.0
x B	61.1	16.7
C	11.1	83.3
Total	100.0	100.0

d. Category A values for *x* are always associated with category 1 values for *y*. Category B values for *x* are usually associated with category 1 values for y. Category C values for *x* are usually associated with category 2 values for *y*.

28. a.

		y			
		20-39	40-59	60-79	80-100	Grand Total
	10-29			1	4	5
x	30-49	2		4		6
	50-69	1	3	1		5
	70-90	4				4
	Grand Total	7	3	6	4	20

b.

		y			
		20-39	40-59	60-79	80-100	Grand Total
	10-29			20.0	80.0	100
x	30-49	33.3		66.7		100
	50-69	20.0	60.0	20.0		100
	70-90	100.0				100

c.

		y			
		20-39	40-59	60-79	80-100
	10-29	0.0	0.0	16.7	100.0
x	30-49	28.6	0.0	66.7	0.0
	50-69	14.3	100.0	16.7	0.0
	70-90	57.1	0.0	0.0	0.0
	Grand Total	100	100	100	100

d. Higher values of x are associated with lower values of y and vice versa
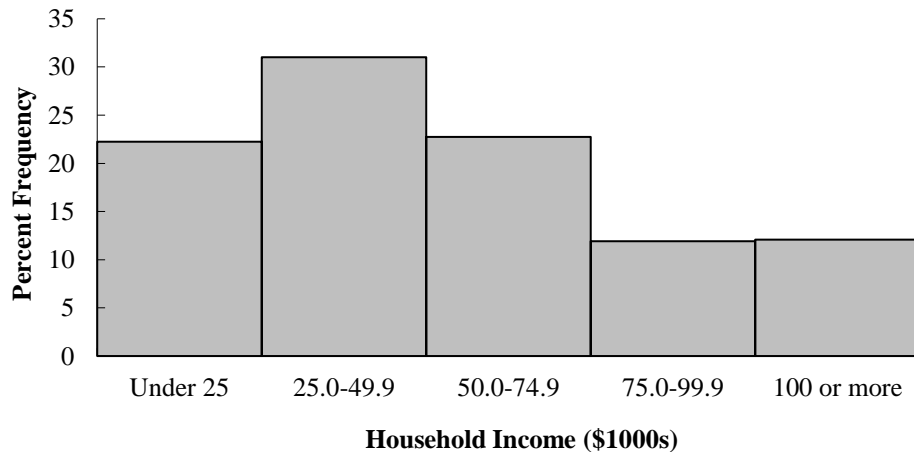
29. a. Row Percentages:

Education Level	Household Income ($1000s)					Total
	Under 25	25.0-49.9	50.0-74.9	75.0-99.9	100 or More	
Not H.S. Graduate	42.23	34.73	13.94	5.41	3.68	100.00
H.S. Graduate	22.25	31.00	22.75	11.93	12.07	100.00
Some College	13.99	26.20	23.31	16.20	20.30	100.00
Bachelor's Degree	6.42	15.19	20.66	18.72	39.02	100.00
Beyond Bach. Deg.	3.71	10.60	16.29	15.87	53.54	100.00
Total	17.77	25.08	20.64	13.90	22.62	100.00

There are six percent frequency distributions in this table with row percentages. The first five give the percent frequency distribution of income for each educational level. The total row provides an overall percent frequency distribution for household income.

The second row, labeled H.S. Graduate, is the percent frequency distribution for households headed by high school graduates. The fourth row, labeled Bachelor's Degree, is the percent frequency distribution for households headed by bachelor's degree recipients.

b. The percentage of households headed by high school graduates earning $75,000 or more is 11.93% + 12.07 = 24.00%. The percent of households headed by bachelor's degree recipients earning $75,000 or more is 18.72% + 39.02% = 57.74%.

c. The percent frequency histogram for high school graduates.

The percent frequency distribution for college graduates with a bachelor's degree.
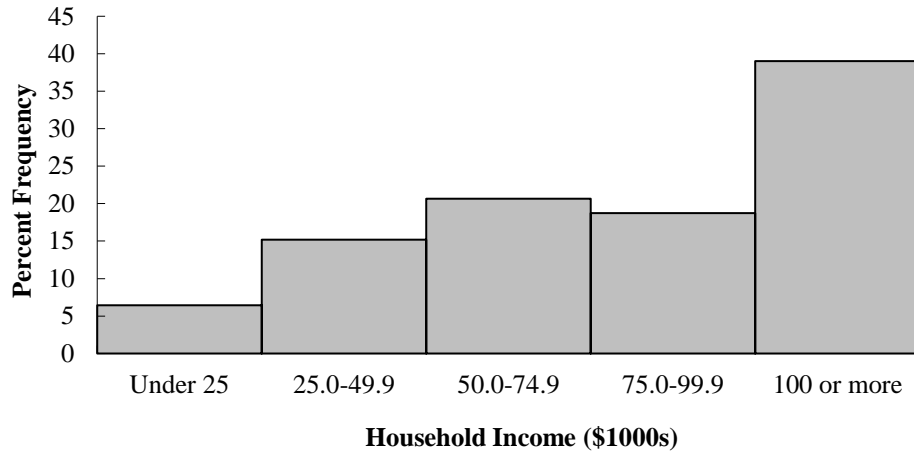
The histograms show that households headed by a college graduate with a bachelor's degree earn more than households headed by a high school graduate. Yes, there is a positive relationship between education level and income.

30. a. Column Percentages:

| Education Level | Household Income ($1000s) | | | | | Total |
	Under 25	25.0-49.9	50.0-74.9	75.0-99.9	100 or More	
Not H.S. Graduate	32.10	18.71	9.13	5.26	2.20	13.51
H.S. Graduate	37.52	37.05	33.04	25.73	16.00	29.97
Some College	21.42	28.44	30.74	31.71	24.43	27.21
Bachelor's Degree	6.75	11.33	18.72	25.19	32.26	18.70
Beyond Bach. Deg.	2.21	4.48	8.37	12.11	25.11	10.61
Total	100.00	100.00	100.00	100.00	100.00	100.00

There are six percent frequency distributions in this table of column percentages. The first five columns give the percent frequency distributions for each income level. The percent frequency distribution in the "Total" column gives the overall percent frequency distributions for educational level. From that percent frequency distribution we see that 13.51% of the heads of households did not graduate from high school.

b. The column percentages show that 25.11% of households earning $100,000 or more were headed by persons having schooling beyond a bachelor's degree. The row percentages show that 53.54% of the households headed by persons with schooling beyond a bachelor's degree earned $100,000 or more. These percentages are different because they came from different percent frequency distributions and provide different kinds of information.

c. Compare the "under 25" percent frequency distributions to the "Total" percent frequency distributions. We see that for this low income level the percentage with lower levels of education is higher than for the overall population and the percentage with higher levels of education is lower than for the overall population.

Compare the "100 or more" percent frequency distribution to "Total" percent frequency distribution. We see that for this high income level the percentage with lower levels of education is lower than for the overall population and the percentage with higher levels of education is higher than for the overall population.

From the comparisons it is clear that there is a positive relationship between household incomes and the education level of the head of the household.

31. a. The crosstabulation of condition of the greens by gender is below.

| | Green Condition | | |
Gender	Too Fast	Fine	Total
Male	35	65	100
Female	40	60	100
Total	75	125	200

The female golfers have the highest percentage saying the greens are too fast: 40/100 = 40%. Male golfers have 35/100 = 35% saying the greens are too fast.

b. Among low handicap golfers, 1/10 = 10% of the women think the greens are too fast and 10/50 = 20% of the men think the greens are too fast. So, for the low handicappers, the men show a higher percentage who think the greens are too fast.

c. Among the higher handicap golfers, 39/51 = 43% of the woman think the greens are too fast and 25/50 = 50% of the men think the greens are too fast. So, for the higher handicap golfers, the men show a higher percentage who think the greens are too fast.

d. This is an example of Simpson's Paradox. At each handicap level a smaller percentage of the women think the greens are too fast. But, when the crosstabulations are aggregated, the result is reversed and we find a higher percentage of women who think the greens are too fast.

The hidden variable explaining the reversal is handicap level. Fewer people with low handicaps think the greens are too fast, and there are more men with low handicaps than women.

32. a.

| Fund Type | 5 Year Average Return | | | | | | Total |
	0-9.99	10-19.99	20-29.99	30-39.99	40-49.99	50-59.99	
DE	1	25	1	0	0	0	27
FI	9	1	0	0	0	0	10
IE	0	2	3	2	0	1	8
Total	10	28	4	2	0	1	45

b.

5 Year Average Return	Frequency
0-9.99	10
10-19.99	28
20-29.99	4
30-39.99	2
40-49.99	0
50-59.99	1
Total	45

c.

Fund Type	Frequency
DE	27
FI	10
IE	8
Total	45

d. The right margin shows the frequency distribution for the fund type variable and the bottom margin shows the frequency distribution for the 5 year average return variable.

e. Higher returns are associated with International Equity funds and lower returns are associated with Fixed Income funds.

33. a.

Fund Type	Expense Ratio (%)						Total
	0-0.24	0.25-0.49	0.50-0.74	0.75-0.99	1.00-1.24	1.25-1.49	
DE	1	1	3	5	10	7	27
FI	2	4	3	0	0	1	10
IE	0	0	1	2	4	1	8
Total	3	5	7	7	14	9	45

b.

Expense Ratio (%)	Frequency	Percent
0-0.24	3	6.7
0.25-0.49	5	11.1
0.50-0.74	7	15.6
0.75-0.99	7	15.6
1.00-1.24	14	31.0
1.25-1.49	9	20.0
Total	45	100

c. Higher expense ratios are associated with Domestic Equity funds and lower expense ratios are associated with Fixed Income fund

34. a.

Net Asset Value	5-Year Average Return							Total
	0-5	5-10	10-15	15-20	20-25	30-35	50-55	
0-10	1							1
10-20	7	1	1	5		1		15
20-30			4	3	3	1		11
30-40			3	6				9
40-50		1	2	1	1			5
50-60			1	1			1	3
70-80				1				1
Total	8	2	11	17	4	2	1	45

b. There is some indication that higher 5-year returns are associated with higher net asset values.

35. a.

	Hwy MPG						
Size	15-19	20-24	25-29	30-34	35-39	40-44	Total
Compact	3	4	17	22	5	5	56
Large	2	10	7	3	2		24
Midsize	3	4	30	20	9	3	69
Total	8	18	54	45	16	8	149

b. Midsize and Compact seem to be more fuel efficient than Large.

c.

	City MPG						
Drive	10-14	15-19	20-24	25-29	30-34	40-44	Total
A	7	18	3				28
F		17	49	19	2	3	90
R	10	20		1			31
Total	17	55	52	20	2	3	149

d. Higher fuel efficiencies are associated with front wheel drive cars.

e.

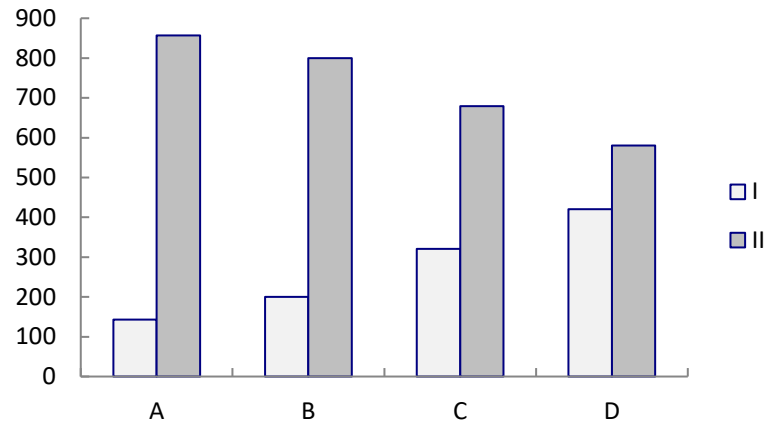|	City MPG						
Fuel Type	15-19	20-24	25-29	30-34	35-39	40-44	Total
P	8	16	20	12			56
R		2	34	33	16	8	93
Total	8	18	54	45	16	8	149

f. Higher fuel efficiencies are associated with cars that use regular gas.

36. a.
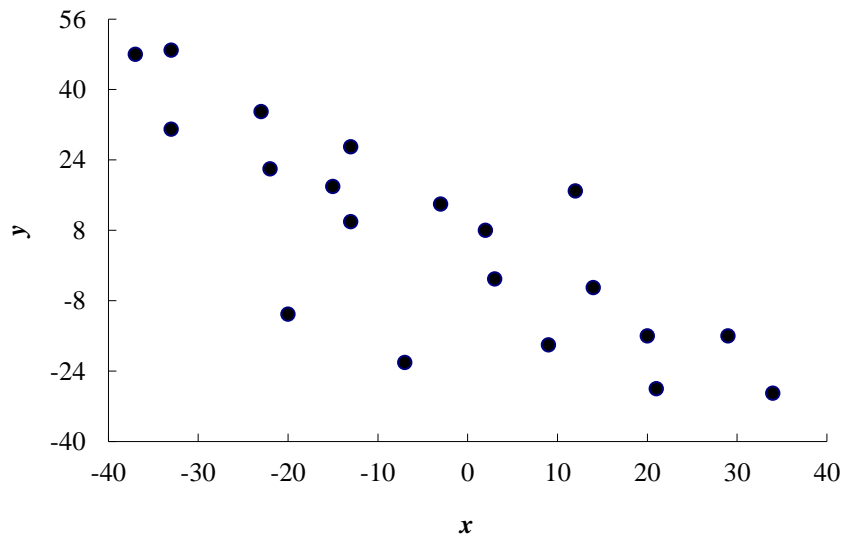
b. There is a negative relationship between x and y; y decreases as x increases.
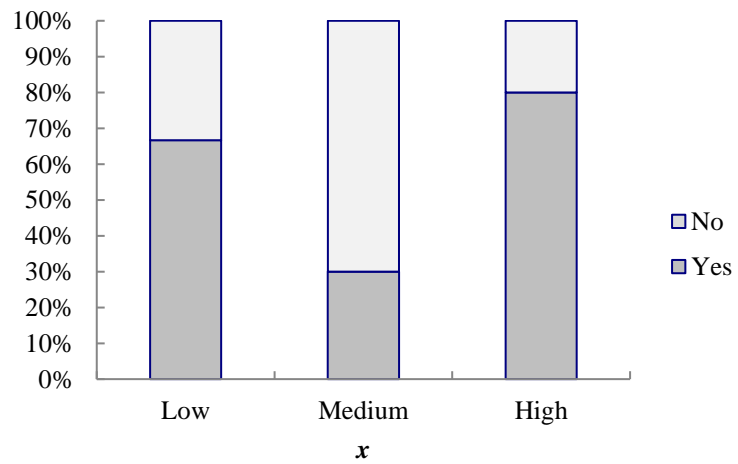
37. a.

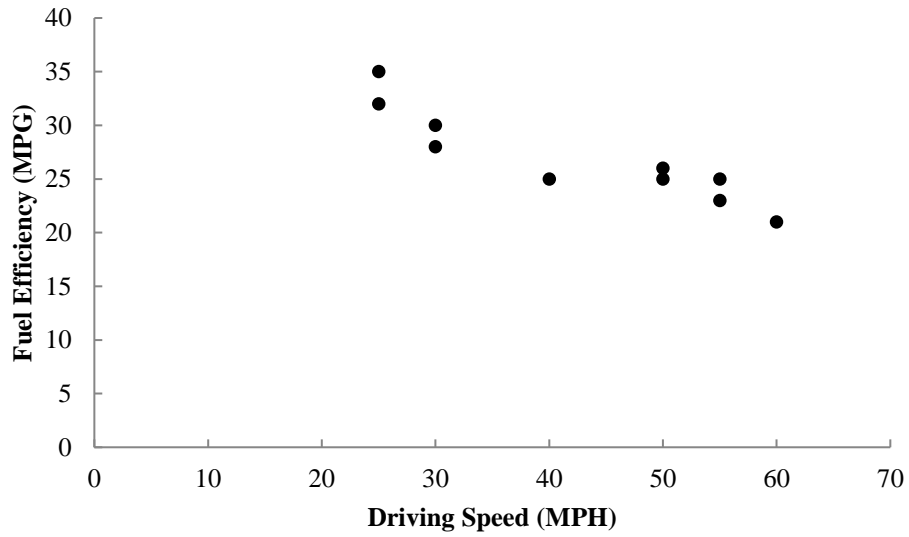b.   As X goes from A to D the frequency for I increases and the frequency of II decreases.

38. a.

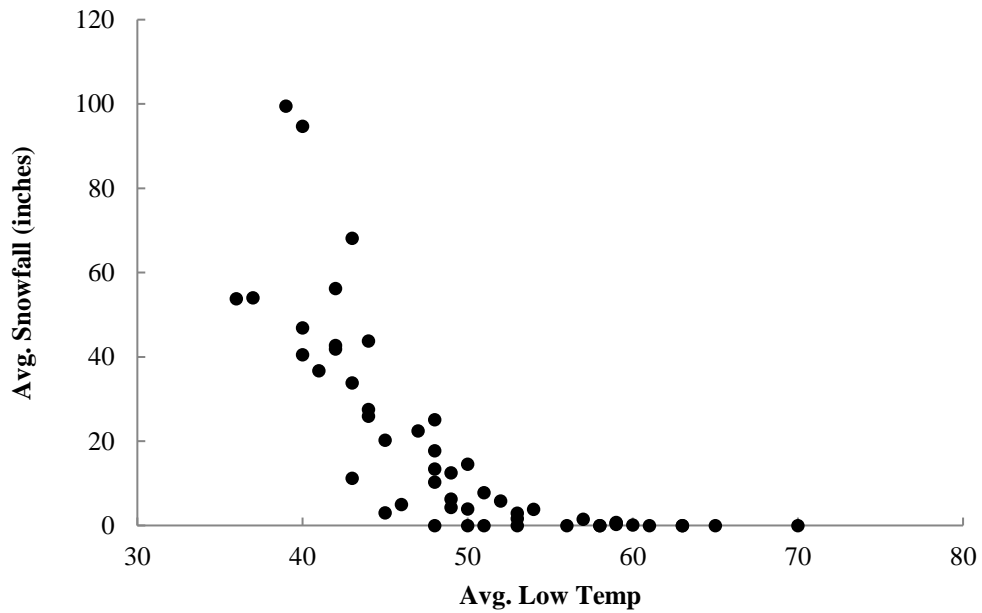|		y		
		Yes	No	
	Low	66.667	33.333	100
x	Medium	30.000	70.000	100
	High	80.000	20.000	100

b.

39. a.

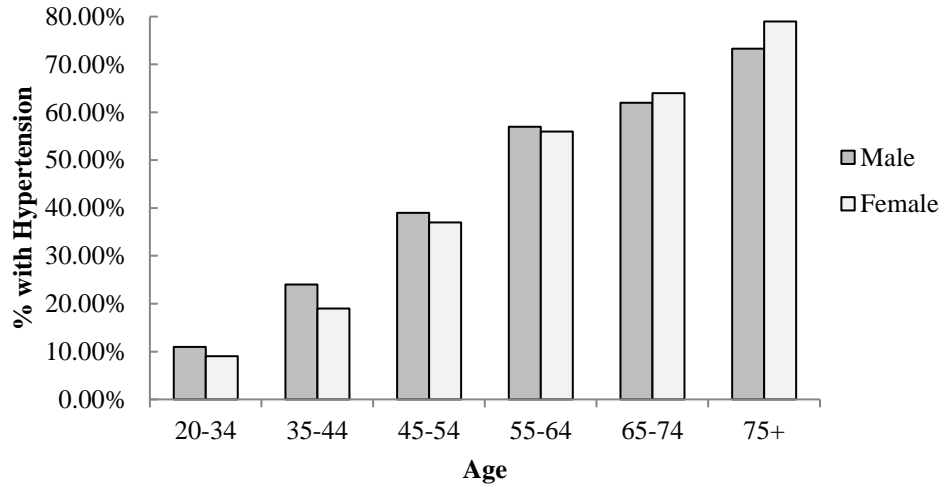b. For midsized cars, lower driving speeds seem to yield higher miles per gallon.

40. a.

b. Colder average low temperature seems to lead to higher amounts of snowfall.

c. Two cities have an average snowfall of nearly 100 inches of snowfall: Buffalo, N.Y and Rochester, NY. Both are located near large lakes in New York.

41. a.

b. The percentage of people with hypertension increases with age.

c. For ages earlier than 65, the percentage of males with hypertension is higher than that for females. After age 65, the percentage of females with hypertension is higher than that for males.

42. a.
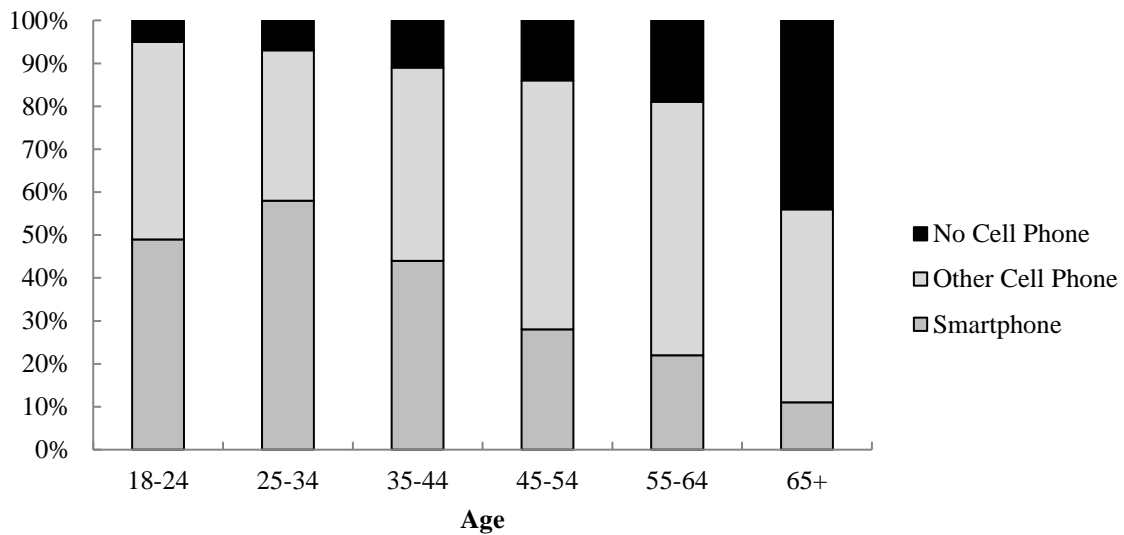
b. After an increase in age 25-34, smartphone ownership decreases as age increases. The percentage of people with no cell phone increases with age. There is less variation across age groups in the percentage who own other cell phones.

c. Unless a newer device replaces the smartphone, we would expect smartphone ownership would become less sensitive to age. This would be true because current users will become older and because the device will become to be seen more as a necessity than a luxury.
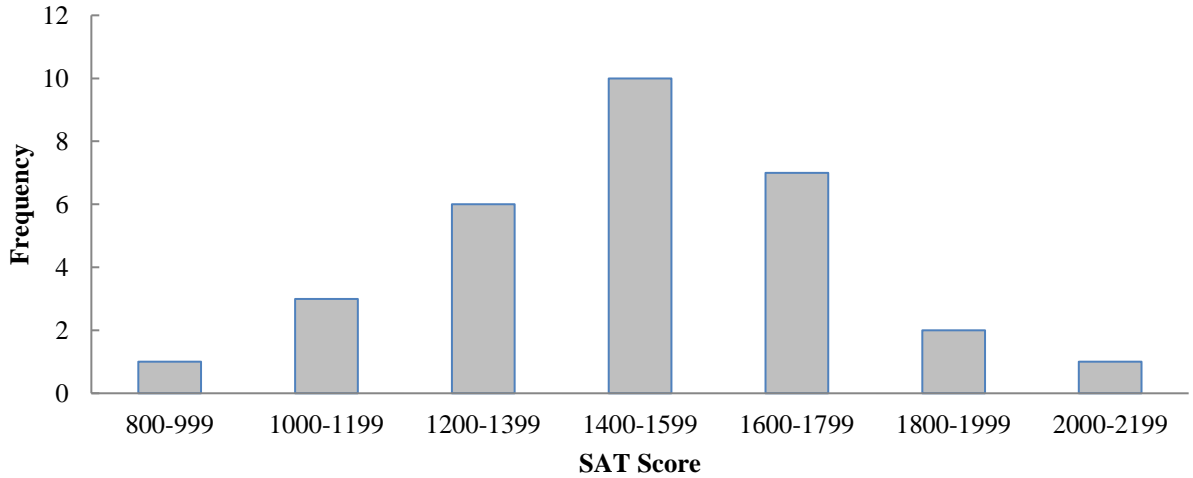
43. a.
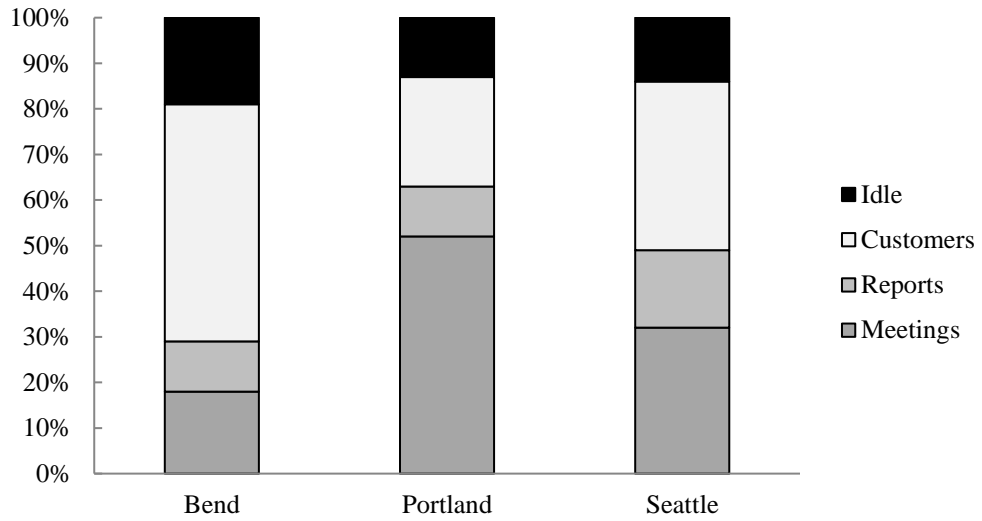
b.
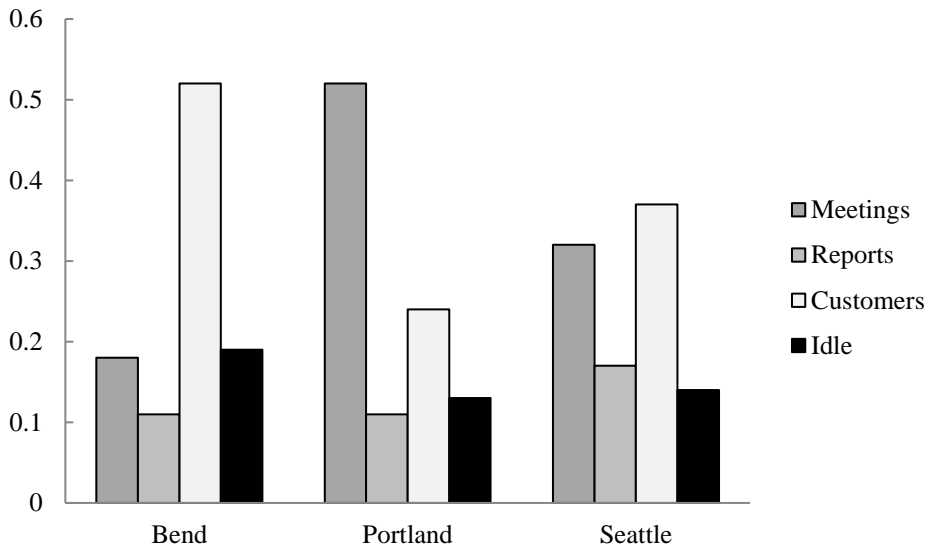
c. The stacked bar chart seems simpler than the side-by-side bar chart and more easily conveys the differences in store managers' use of time.

44. a.

Class	Frequency
800-999	1
1000-1199	3
1200-1399	6
1400-1599	10
1600-1799	7
1800-1999	2
2000-2199	1
Total	30

b. The distribution if nearly symmetrical. It could be approximated by a bell-shaped curve.

c. 10 of 30 or 33% of the scores are between 1400 and 1599. The average SAT score looks to be a little over 1500. Scores below 800 or above 2200 are unusual.

45. a.

State	Frequency
Arizona	2
California	11
Florida	15
Georgia	2
Louisiana	8
Michigan	2
Minnesota	1
Texas	2
Total	43
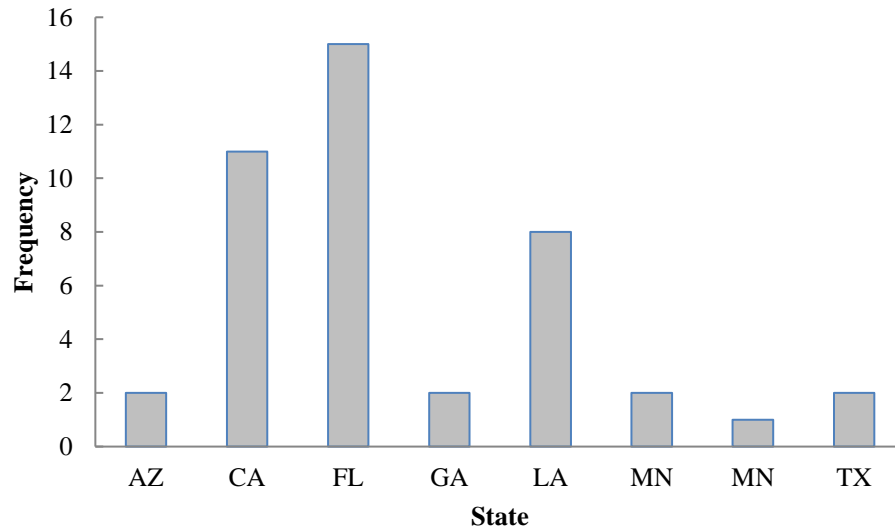
b. Florida has had the most Super Bowl with 15, or 15/43(100) = 35%. Florida and California have been the states with the most Super Bowls. A total of 15 + 11 = 26, or 26/43(100) = 60%. Only 3 Super Bowls, or 3/43(100) = 7%, have been played in the cold weather states of Michigan and Minnesota.

c.

```
0 | 1 3 3 3 3 3 4 4 4 4
0 | 5 7 7 7 9
1 | 0 0 0 1 2 2 3 4
1 | 5 6 7 7 7 7 8 9 9 9
2 | 1 2 3
2 | 5 7 7
3 | 2
3 | 5 6
4 |
4 | 5
```
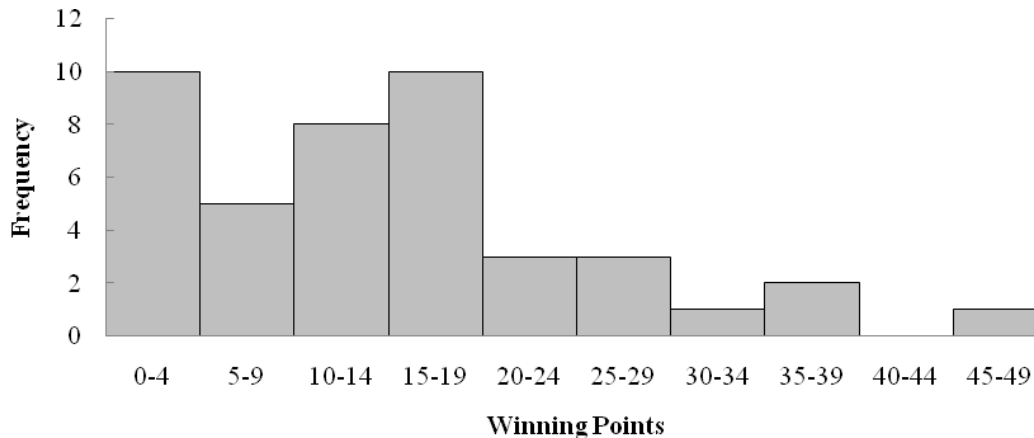

d. The most frequent winning points have been 0 to 4 points and 15 to 19 points. Both occurred in 10 Super Bowls. There were 10 close games with a margin of victory less than 5 points, 10/43(100) = 23% of the Super Bowls. There have also be 10 games, 23%, with a margin of victory more than 20 points.

e. The closest games was the 25th Super Bowl with a 1 point margin. It was played in Florida. The largest margin of victory occurred one year earlier in the 24th Super Bowl. It had a 45 point margin and was played in Louisiana. More detailed information not available from the text information.

   25th Super Bowl: 1991 New York Giants 20 Buffalo Bills 19, Tampa Stadium, Tampa, FL
   24th Super Bowl: 1990 San Francisco 49ers 55 Denver Broncos 10, Superdome, New Orleans, LA

   Note: The data set SuperBowl contains a list of the teams and the final scores of the 43 Super Bowls. This data set can be used in Chapter 2 and Chapter 3 to provide interesting data summaries about the points scored by the winning team and the points scored by the losing team in the Super Bowl. For example, using the median scores, the median Super Bowl score was 28 to 13.

46. a.

Population in Millions	Frequency	% Frequency
0.0 - 2.4	15	30.0%
2.5-4.9	13	26.0%
5.0-7.4	10	20.0%
7.5-9.9	5	10.0%
10.0-12.4	1	2.0%
12.5-14.9	2	4.0%
15.0-17.4	0	0.0%
17.5-19.9	2	4.0%
20.0-22.4	0	0.0%
22.5-24.9	0	0.0%
25.0-27.4	1	2.0%
27.5-29.9	0	0.0%
30.0-32.4	0	0.0%
32.5-34.9	0	0.0%
35.0-37.4	1	2.0%
37.5-39.9	0	0.0%
More	0	0.0%
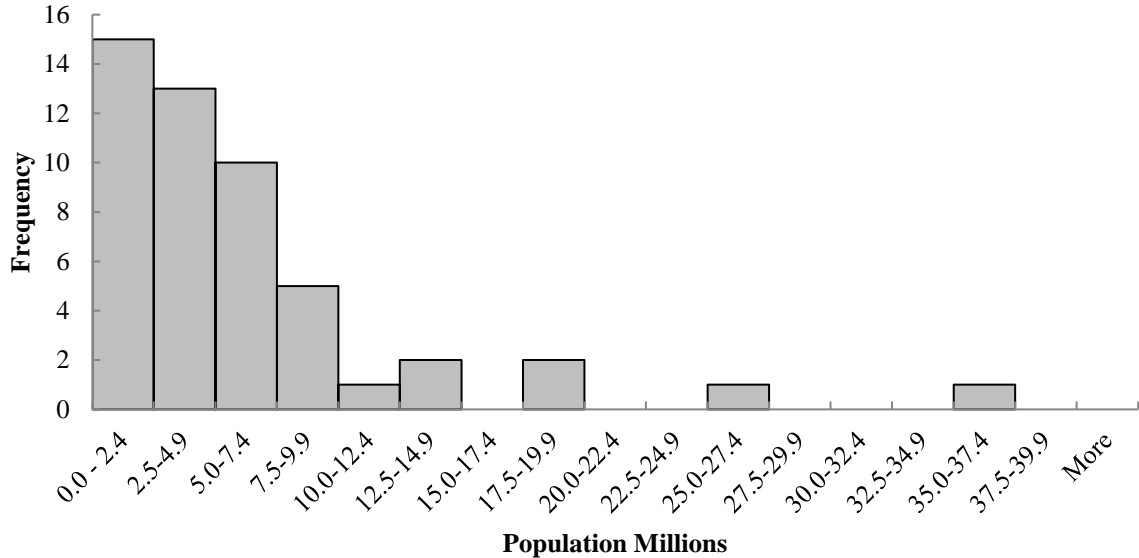
b. The distribution is skewed to the right.

c. 15 states (30%) have a population less than 2.5 million. Over half of the states have population less than 5 million (28 states – 56%). Only seven states have a population greater than 10 million (California, Florida, Illinois, New York, Ohio, Pennsylvania and Texas). The largest state is California (37.3 million) and the smallest states are Vermont and Wyoming (600 thousand).

47. a.

```
 1    8
 2    014
 3    18
 4    007899
 5    012444578
 6    00139
 7    237888
 8    011
 9    1
10    3
11    0289
12    9
13    01
14
15    46
16    68
17
18
19    2
20
21
22
23
24
25
26
27    2
```

b. The majority of the start-up companies in this set have less than $90 million in venture capital. Only 6 of the 50 (12%) have more than $150 million.

48. a.

Industry	Frequency	% Frequency
Bank	26	13%
Cable	44	22%
Car	42	21%
Cell	60	30%
Collection	28	14%
Total	200	100%

b.
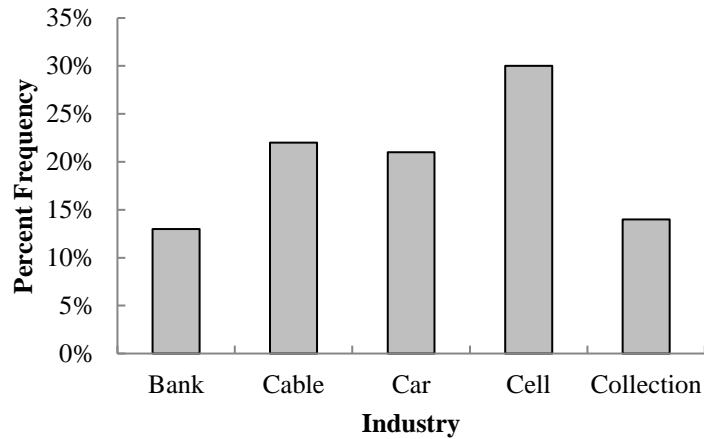
c. The cellular phone providers had the highest number of complaints.

d. The percentage frequency distribution shows that the two financial industries (banks and collection agencies) had about the same number of complaints. Also, new car dealers and cable and satellite television companies also had about the same number of complaints.

49. a.

Yield%	Frequency	Percent Frequency
0.0-0.9	4	13.3
1.0-1.9	2	6.7
2.0-2.9	6	20.0
3.0-3.9	10	33.3
4.0-4.9	3	10.0
5.0-5.9	2	6.7
6.0-6.9	2	6.7
7.0-7.9	0	0.0
8.0-8.9	0	0.0
9.0-9.9	1	3.3
Total	30	100.0

b.
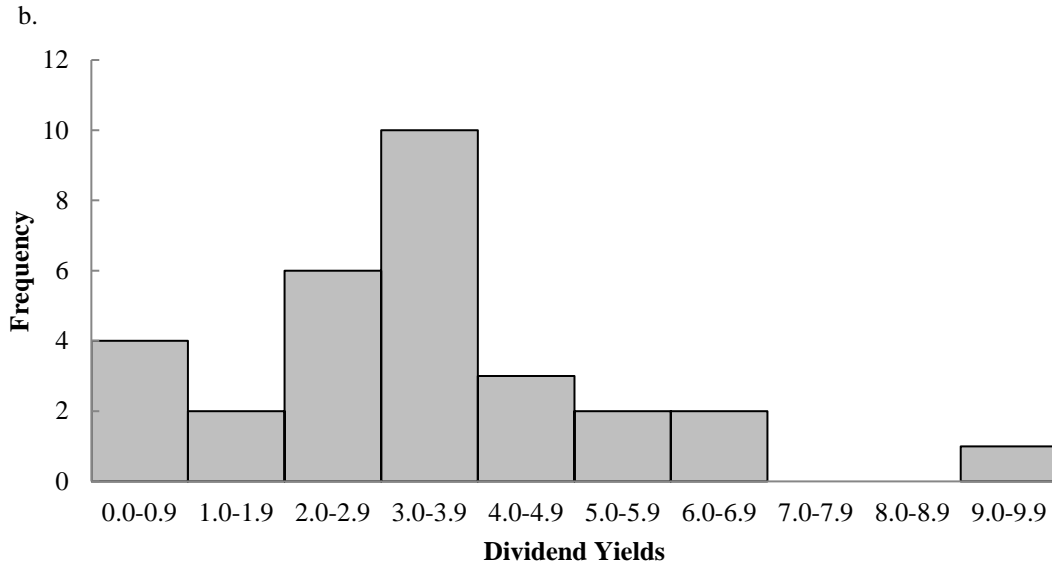
c. The distribution is skewed to the right.

d. Dividend yield ranges from 0% to over 9%. The most frequent range is 3.0% to 3.9%. Average dividend yields looks to be between 3% and 4%. Over 50% of the companies (16) pay from 2.0 % to 3.9%. Five companies (AT&T, DuPont, General Electric, Merck, and Verizon) pay 5.0% or more. Four companies (Bank of America, Cisco Systems, Hewlett-Packard, and J.P. Morgan Chase) pay less than 1%.

e. General Electric had an unusually high dividend yield of 9.2%. 500 shares at $14 per share is an investment of 500($14) = $7,000. A 9.2% dividend yield provides .092(7,000) = $644 of dividend income per year.

50.

Level of Support	Percent Frequency
Strongly favor	1617/5372 = 30.10
Favor more than oppose	1871/5372 = 34.83
Oppose more than favor	1135/5372 = 21.13
Strongly oppose	749/5372 = 13.94
Total	100.00

The results show support for a higher tax. Note that 30.10% + 34.83% = 64.93% of the respondents said they strongly favor or favor more than oppose a higher tax on higher carbon emission cars.

b.

Country	Percent Frequency
Great Britain	1087/5372 = 20.2
Italy	1045/5372 = 19.5
Spain	1109/5372 = 20.6
Germany	1111/5372 = 20.7
United States	1020/5372 = 19.0
Total	100.0

The poll had an approximately equal representation of the five countries with roughly 20% of the poll respondents coming from each country.

c. Converting the entries in the crosstabulation into column percentages provides the following results:

Support	Country Great Britain	Italy	Spain	Germany	United States
Strongly favor	31.00	31.96	45.99	19.98	20.98
Favor more than oppose	34.04	39.04	32.01	36.99	32.06
Oppose more than favor	23.00	17.99	13.98	24.03	26.96
Strongly oppose	11.96	11.01	8.03	18.99	20.00
Total	100.00	100.00	100.00	100.00	100.00

Considering the percentage of respondents who favor the higher tax by either saying "strongly favor" or "favor more than oppose", we have the following favorable support for the higher tax in each country.

Great Britain	31.00 + 34.04 = 65.04%
Italy	31.96 + 39.04 = 71.00%
Spain	45.99 + 32.01 = 78.00%
Germany	19.98 + 36.99 = 56.97%
United States	20.98 + 32.06 = 53.04%

More than 50% of the respondents favor the higher tax for the higher carbon emission cars in all five countries. But the support for the higher tax is greater in the European countries. Spain and Italy have the greatest support for the higher tax with 78% and 71% respectively. Germany is close in views to the United States with 56.97% expressing favor for the higher tax. United States shows the lowest level of support for the higher tax with 53.04%. Note that United States ranks first in terms of the response "strongly oppose" the higher tax with 20% of the respondents providing this opinion.

51. a. The batting averages for the junior and senior years for each player are as follows:

Junior year:

Allison Fealey	15/40 = .375
Emily Janson	70/200 = .350

Senior year:

Allison Fealey	75/250 = .300
Emily Janson	35/120 = .292

Because Allison Fealey had the higher batting average in both her junior year and senior year, Allison Fealey should receive the scholarship offer.

b. The combined or aggregated two-year crosstabulation is as follows:

Combined 2-Year Batting

Outcome	A. Fealey	E. Jansen
Hit	90	105
No Hit	200	215
Total At Bats	290	320

Based on this crosstabulation, the batting average for each player is as follows:

Combined Junior/Senior Years
  Allison Fealey   90/290 = .310
  Emily Janson   105/320 = .328

Because Emily Janson has the higher batting average over the combined junior and senior years, Emily Janson should receive the scholarship offer.

c. The recommendations in parts (a) and (b) are not consistent. This is an example of Simpson's Paradox. It shows that in interpreting the results based upon separate or un-aggregated crosstabulations, the conclusion can be reversed when the crosstabulations are grouped or aggregated. When Simpson's Paradox is present, the decision maker will have to decide whether the un-aggregated or the aggregated form of the crosstabulation is the most helpful in identifying the desired conclusion. Note: The authors prefer the recommendation to offer the scholarship to Emily Janson because it is based upon the aggregated performance for both players over a larger number of at-bats. But this is a judgment or personal preference decision. Others may prefer the conclusion based on using the un-aggregated approach in part (a).

52. a.

Year Constructed	Elec	Nat. Gas	Oil	Propane	Other	Total
			Fuel Type			
1973 or before	40	183	12	5	7	247
1974-1979	24	26	2	2	0	54
1980-1986	37	38	1	0	6	82
1987-1991	48	70	2	0	1	121
Total	149	317	17	7	14	504

b.

Year Constructed	Frequency	Fuel Type	Frequency
1973 or before	247	Electricity	149
1974-1979	54	Nat. Gas	317
1980-1986	82	Oil	17
1987-1991	121	Propane	7
Total	504	Other	14
		Total	504

c. Crosstabulation of Column Percentages

Year Constructed	Elec	Nat. Gas	Oil	Propane	Other
			Fuel Type		
1973 or before	26.9	57.7	70.5	71.4	50.0
1974-1979	16.1	8.2	11.8	28.6	0.0
1980-1986	24.8	12.0	5.9	0.0	42.9
1987-1991	32.2	22.1	11.8	0.0	7.1
Total	100.0	100.0	100.0	100.0	100.0

d. Crosstabulation of row percentages.

Year Constructed	Elec	Nat. Gas	Oil	Propane	Other	Total
			Fuel Type			
1973 or before	16.2	74.1	4.9	2.0	2.8	100.0
1974-1979	44.5	48.1	3.7	3.7	0.0	100.0
1980-1986	45.1	46.4	1.2	0.0	7.3	100.0
1987-1991	39.7	57.8	1.7	0.0	0.8	100.0

e. <u>Observations from the column percentages crosstabulation</u>

For those buildings using electricity, the percentage has not changed greatly over the years. For the buildings using natural gas, the majority were constructed in 1973 or before; the second largest percentage was constructed in 1987-1991. Most of the buildings using oil were constructed in 1973 or before. All of the buildings using propane are older.

<u>Observations from the row percentages crosstabulation</u>

Most of the buildings in the CG&E service area use electricity or natural gas. In the period 1973 or before most used natural gas. From 1974-1986, it is fairly evenly divided between electricity and natural gas. Since 1987 almost all new buildings are using electricity or natural gas with natural gas being the clear leader.

53. a.

Tuition & Fees ($)

Year Founded	1-5000	10001-15000	15001-20000	20001-25000	25001-30000	30001-35000	35001-40000	40001-45000	Total
1600-1649							1		1
1700-1749							2	1	3
1750-1799								4	4
1800-1849				1	3	3	6	8	21
1850-1899	1		2	2	13	14	13	4	49
1900-1949		1		2	3	4	8		18
1950-2000			2	4		1			7
Total	1	1	4	9	19	22	30	17	103

b.

Tuition & Fees ($)

Year Founded	1-5000	10001-15000	15001-20000	20001-25000	25001-30000	30001-35000	35001-40000	40001-45000	Grand Total
1600-1649							100.00		100
1700-1749							66.67	33.33	100
1750-1799								100.00	100
1800-1849				4.76	14.29	14.29	28.57	38.10	100
1850-1899	2.04		4.08	4.08	26.53	28.57	26.53	8.16	100
1900-1949		5.56		11.11	16.67	22.22	44.44		100
1950-2000			28.57	57.14		14.29			100

c. Colleges in this sample founded before 1800 tend to be expensive in terms of tuition.

54. a.

Year Founded	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100	Grand Total
1600-1649													1	1
1700-1749													3	3
1750-1799												1	3	4
1800-1849						1	2	4	2	3	4	3	2	21
1850-1899			1	2	4	3	11	5	9	6	3	4	1	49
1900-1949	1	1	1		1	3		3	2	4	1	1		18
1950-2000	1		1	3			2							7
Grand Total	2	1	3	5	5	7	15	12	13	13	8	9	10	103

(% Graduate — column headers span the 35-40 through 95-100 columns)

b.

Year Founded	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100	Grand Total
1600-1649													100.00	100
1700-1749													100.00	100
1750-1799												25.00	75.00	100
1800-1849						4.76	9.52	19.05	9.52	14.29	19.05	14.29	9.52	100
1850-1899			2.04	4.08	8.16	6.12	22.45	10.20	18.37	12.24	6.12	8.16	2.04	100
1900-1949	5.56	5.56	5.56		5.56	16.67		16.67	11.11	22.22	5.56	5.56		100
1950-2000	14.29		14.29	42.86			28.57							100

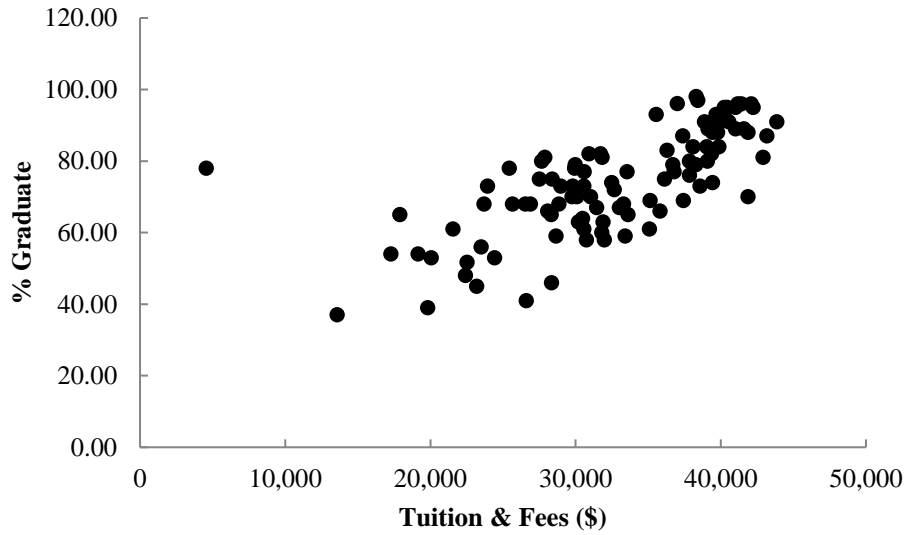(% Graduate — column headers span the 35-40 through 95-100 columns)

c. Older colleges and universities tend to have higher graduation rates.
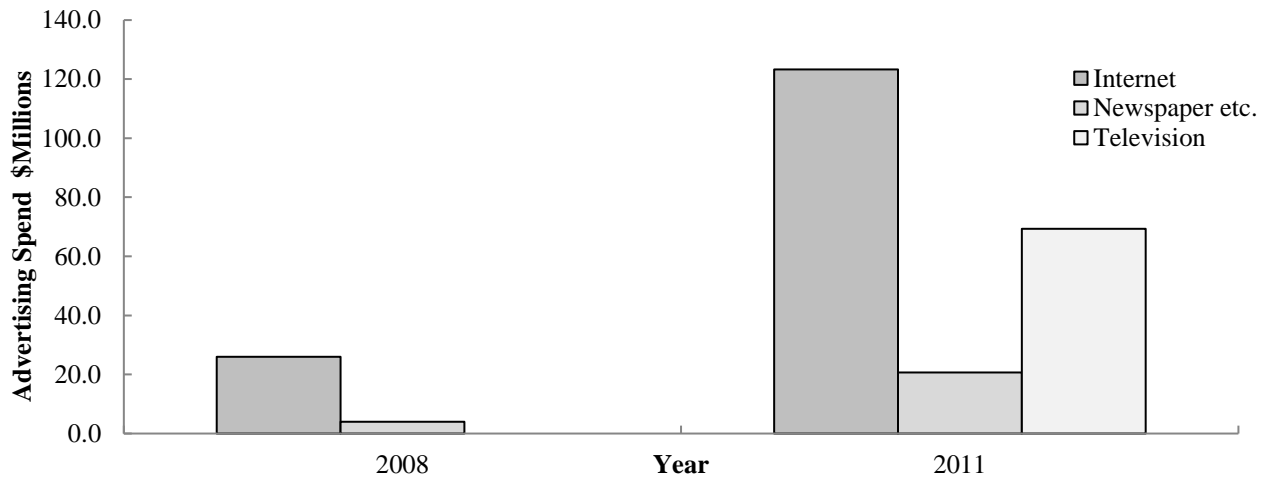
55. a.
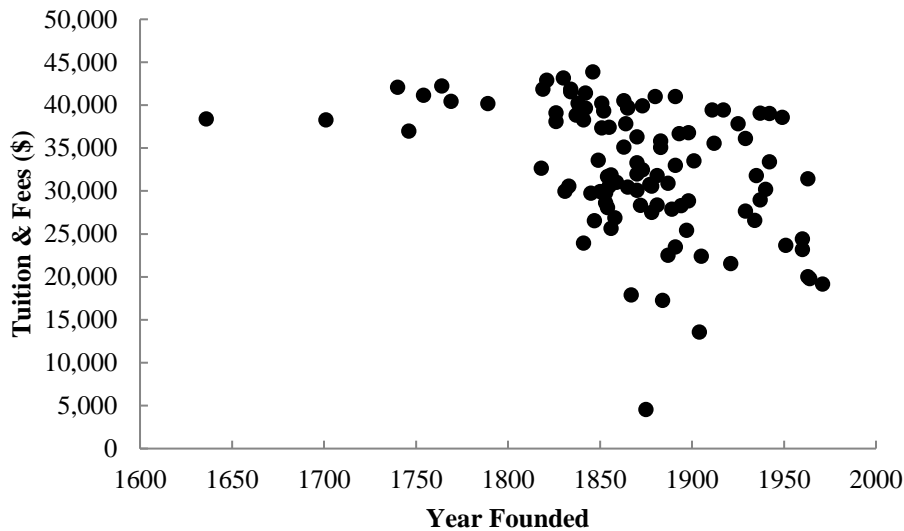
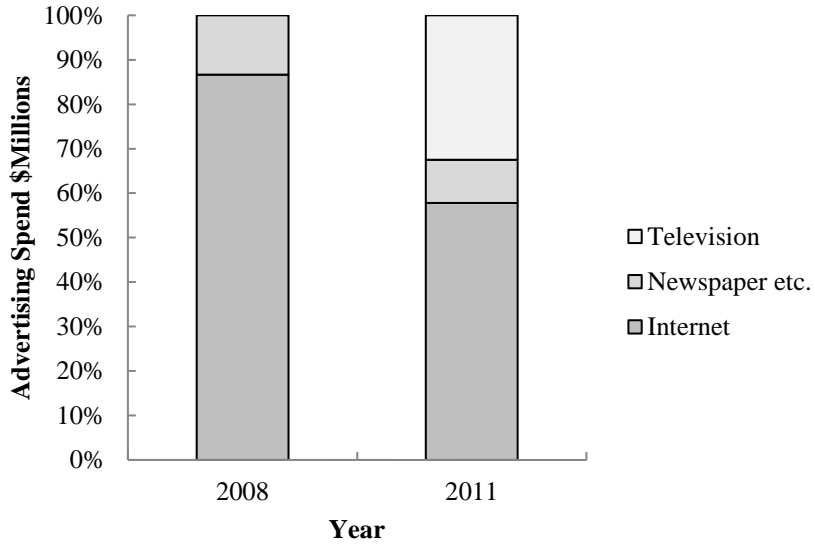b. Older colleges and universities tend to be more expensive.

56. a.

b.   There appears to be a strong positive relationship between Tuition & Fees and % Graduation.
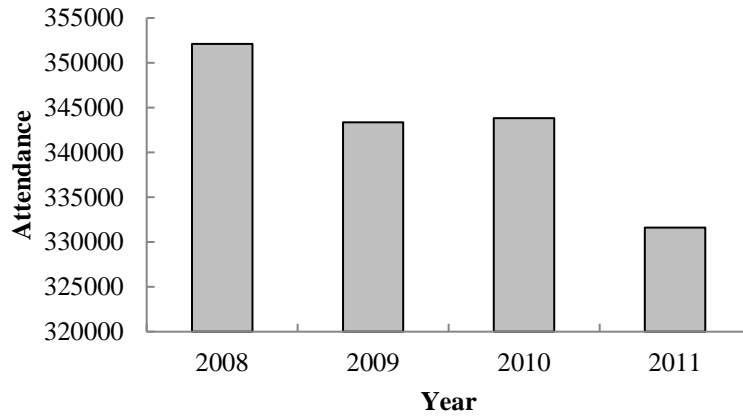
57. a.

b.

	2008	**2011**
Internet	86.7%	57.8%
Newspaper etc.	13.3%	9.7%
Television	0.0%	32.5%
Total	100.0%	100.0%

c. The graph is part a is more insightful because is shows the allocation of the budget across media, but also dramatic increase in the size of the budget.
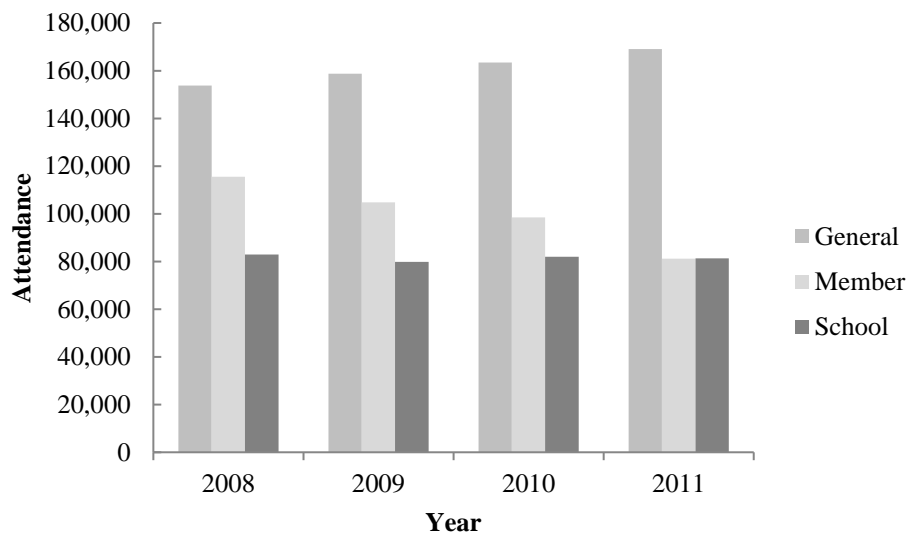
58. a.

Zoo attendance appears to be dropping over time.

b.

c.   General attendance is increasing, but not enough to offset the decrease in member attendance. School membership appears fairly stable.