

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	<u>36</u>	$36/120 = \underline{0.30}$
	120	1.00

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

b. $.20(200) = 40$

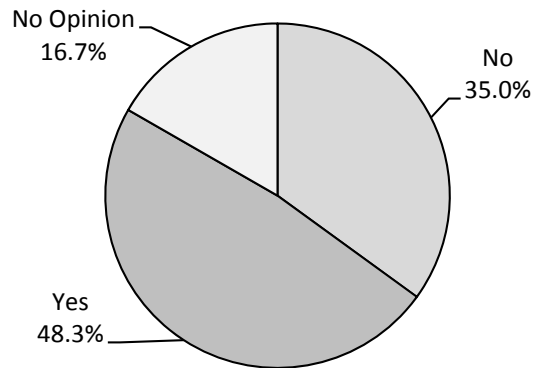
c/d.

Class	Frequency	Percent Frequency
A	$.22(200) = 44$	22
B	$.18(200) = 36$	18
C	$.40(200) = 80$	40
D	$.20(200) = \underline{40}$	<u>20</u>
Total	200	100

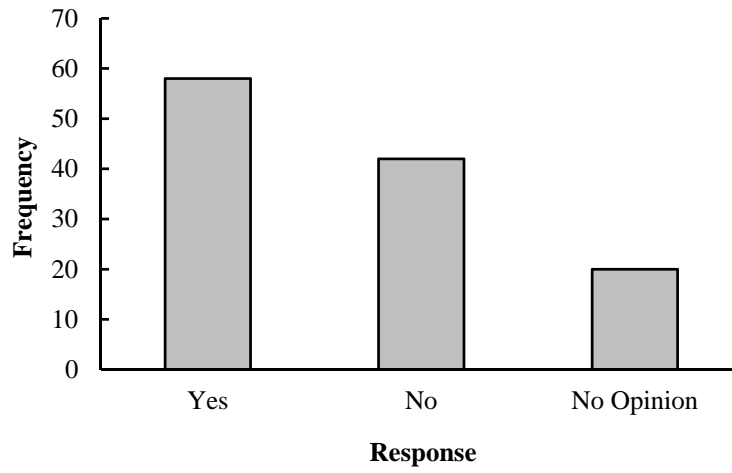
3. a. $360^\circ \times 58/120 = 174^\circ$

b. $360^\circ \times 42/120 = 126^\circ$

c.



d.

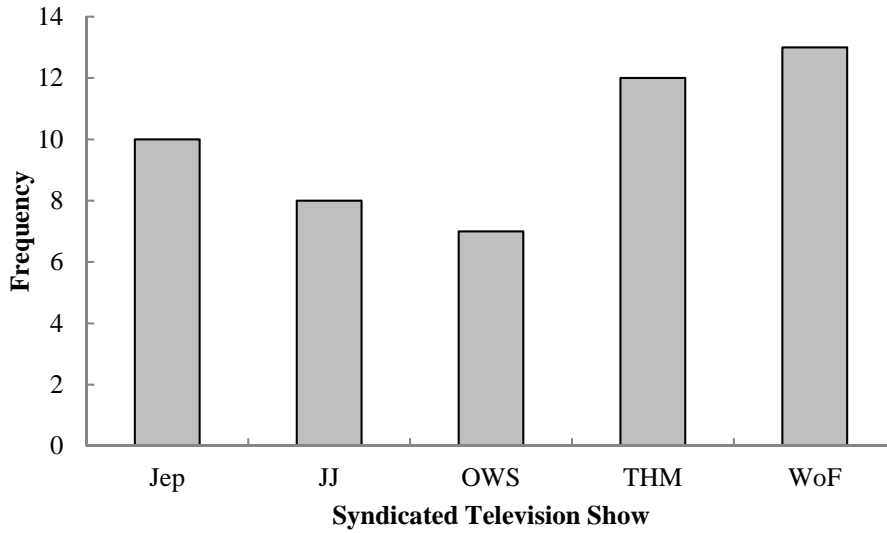


4. a. These data are categorical.

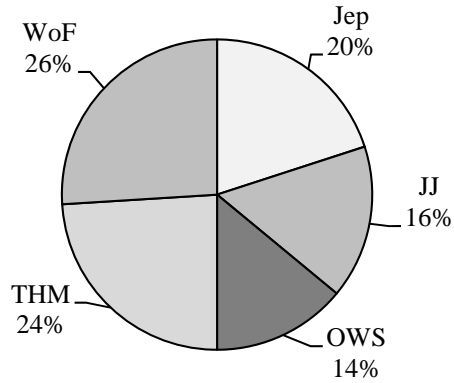
b.

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



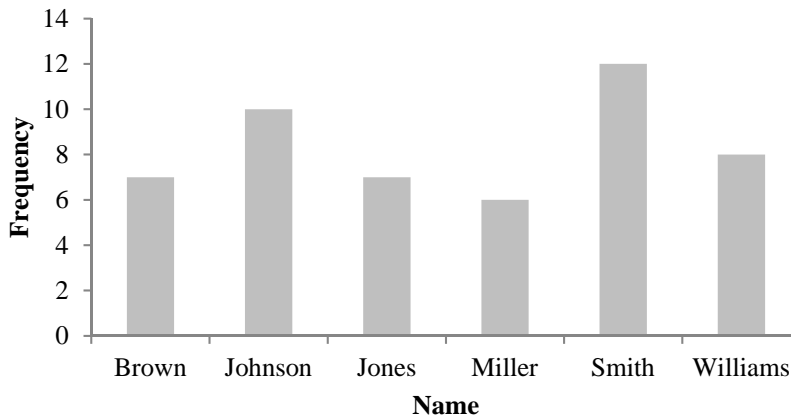
d. The largest viewing audience is for *Wheel of Fortune* and the second largest is for *Two and a Half Men*.

5. a.

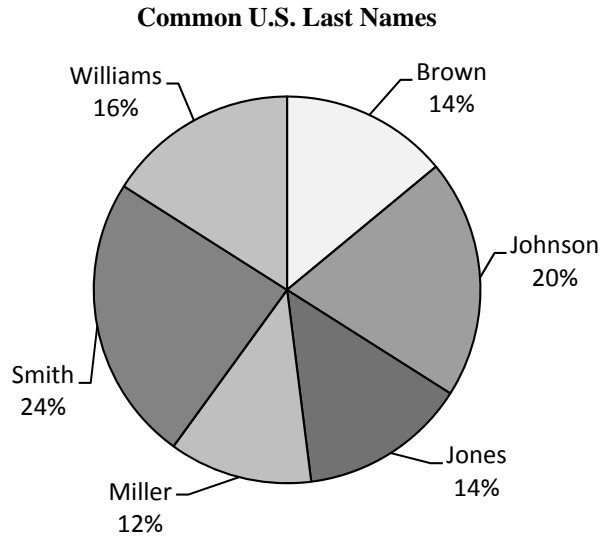
Name	Frequency	Relative Frequency	Percent
Brown	7	0.14	14%
Johnson	10	0.20	20%
Jones	7	0.14	14%
Miller	6	0.12	12%
Smith	12	0.24	24%
Williams	8	0.16	16%
Total:	50	1	100%

b.

Common U.S. Last Names



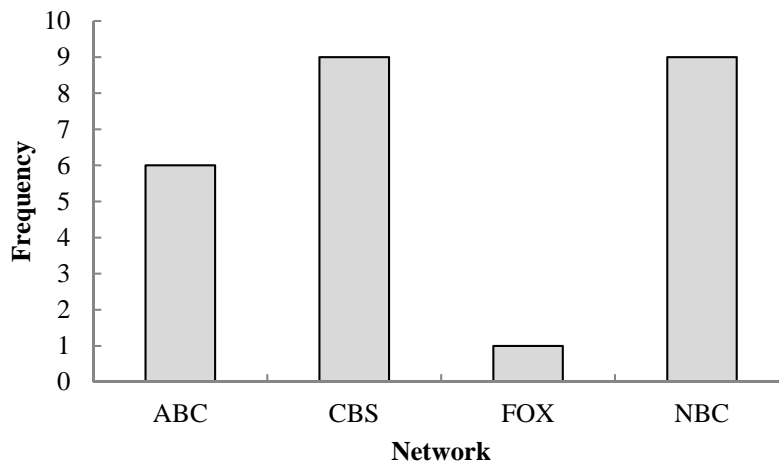
c.



d. The three most common last names are Smith (24%), Johnson (20%), Williams (16%)

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	<u>2</u>	<u>4</u>
	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	<u>7</u>	<u>0.127</u>
	55	1.000

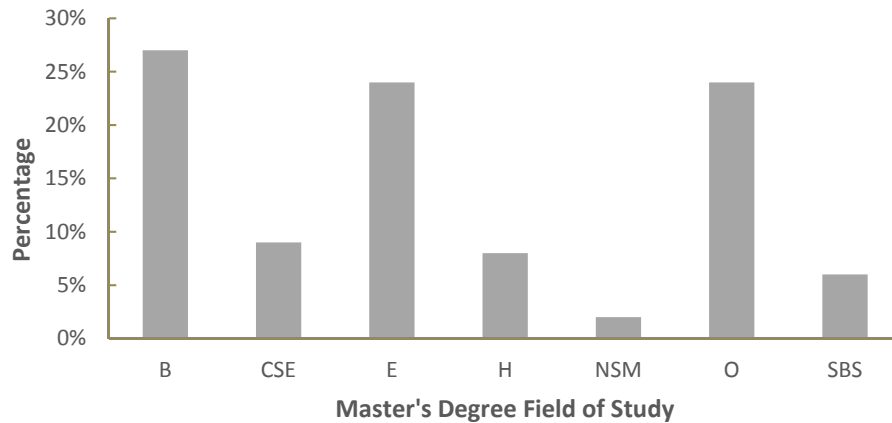
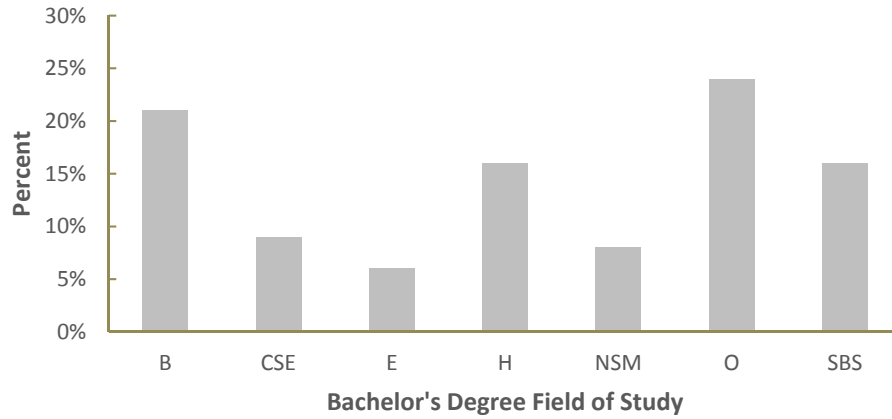
- b. Pitchers (Almost 31%)
- c. 3rd Base (3 – 4%)
- d. Right Field (Almost 13%)

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

9. a.

	Bachelor's	Master's
B	21%	27%
CSE	9%	9%
E	6%	24%
H	16%	8%
NSM	8%	2%
SBS	16%	6%
O	24%	24%
Total	100%	100%

b.



c. The lowest percentage for a Bachelor's is Education (6%) and for Master's Natural Sciences and Mathematics (2%).

d. The highest percentage for a Bachelor's is Other (24%) and for a Master's is Business (27%).

e.

	Bachelor's	Master's	Difference
B	21%	27%	6%
CSE	9%	9%	0%
E	6%	24%	18%
H	16%	8%	-8%
NSM	8%	2%	-6%
SBS	16%	6%	-10%
O	24%	24%	- 0%

Education has the largest increase in percent: 18%

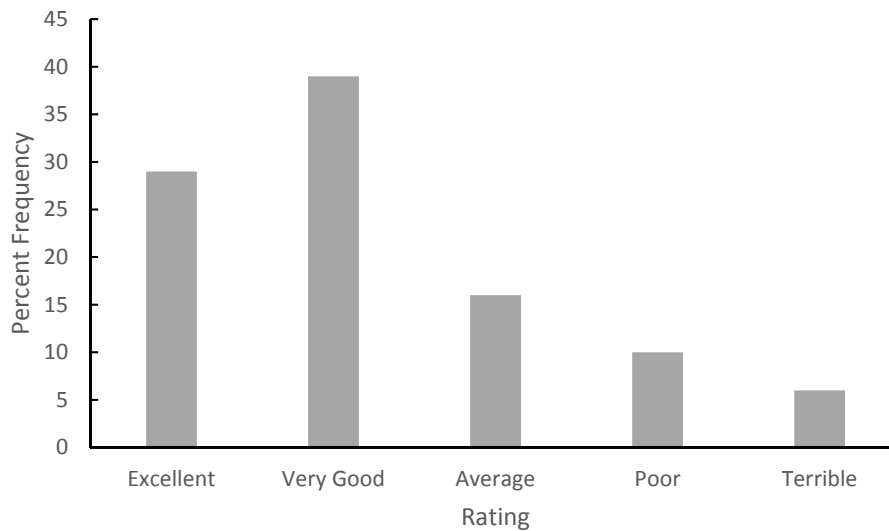
10. a.

Rating	Frequency
Excellent	187
Very Good	252
Average	107
Poor	62
Terrible	41
Total	649

b.

Rating	Percent Frequency
Excellent	29
Very Good	39
Average	16
Poor	10
Terrible	6
Total	100

c.



- d. $29\% + 39\% = 68\%$ of the guests at the Sheraton Anaheim Hotel rated the hotel as Excellent or Very Good. But, $10\% + 6\% = 16\%$ of the guests rated the hotel as poor or terrible.
- e. The percent frequency distribution for Disney's Grand Californian follows:

Rating	Percent Frequency
Excellent	48
Very Good	31
Average	12
Poor	6
Terrible	3
Total	100

$48\% + 31\% = 79\%$ of the guests at the Sheraton Anaheim Hotel rated the hotel as Excellent or Very Good. And, $6\% + 3\% = 9\%$ of the guests rated the hotel as poor or terrible.

Compared to ratings of other hotels in the same region, both of these hotels received very favorable ratings. But, in comparing the two hotels, guests at Disney's Grand Californian provided somewhat better ratings than guests at the Sheraton Anaheim Hotel.

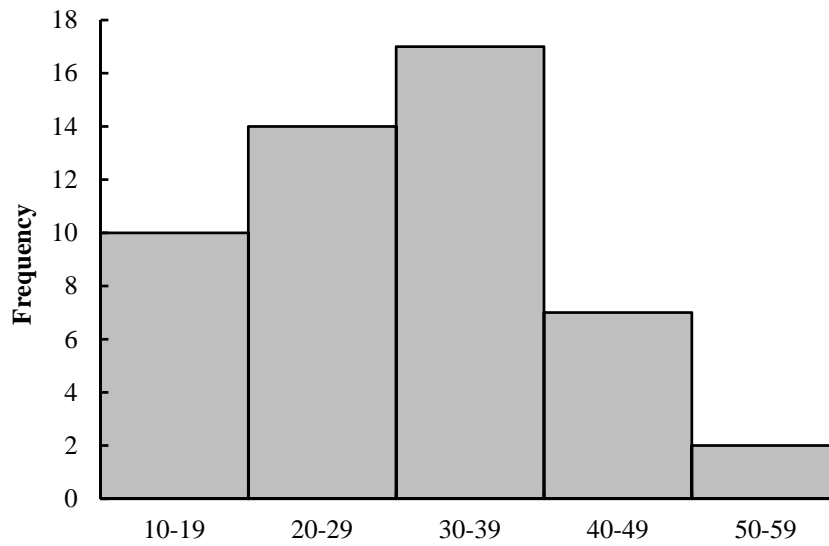
11.

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	<u>9</u>	<u>0.225</u>	<u>22.5</u>
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	<u>3</u>	<u>15</u>
	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	<u>1</u>	<u>0.05</u>
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.

PPG	Frequency
10-12	1
12-14	3
14-16	7
16-18	19
18-20	9
20-22	4
22-24	2
24-26	0
26-28	3
28-30	<u>2</u>
Total	50

Chapter 2

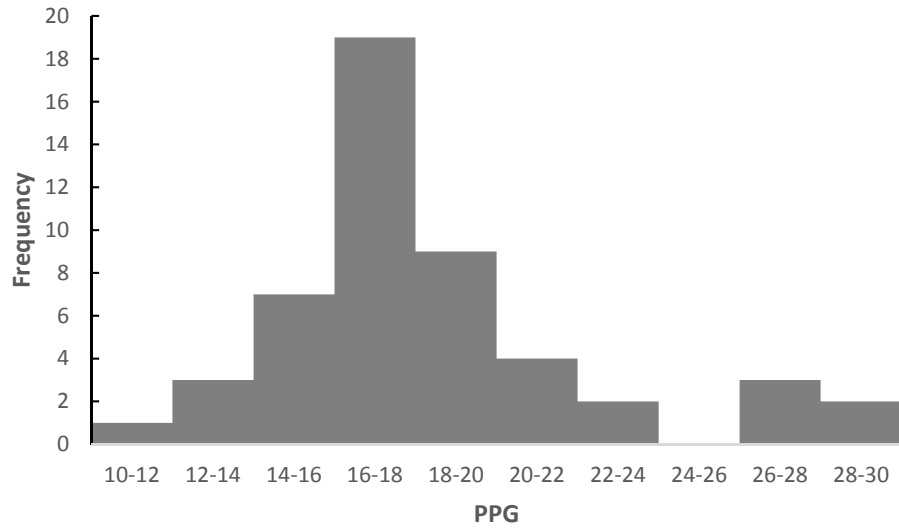
b.

PPG	Relative Frequency
10-12	0.02
12-14	0.06
14-16	0.14
16-18	0.38
18-20	0.18
20-22	0.08
22-24	0.04
24-26	0.00
26-28	0.06
28-30	0.04
Total	1.00

c.

PPG	Cumulative Percent Frequency
less than 12	2
less than 14	8
less than 16	22
less than 18	60
less than 20	78
less than 22	86
less than 24	90
less than 26	90
less than 28	96
less than 30	100

d.



e. There is skewness to the right.

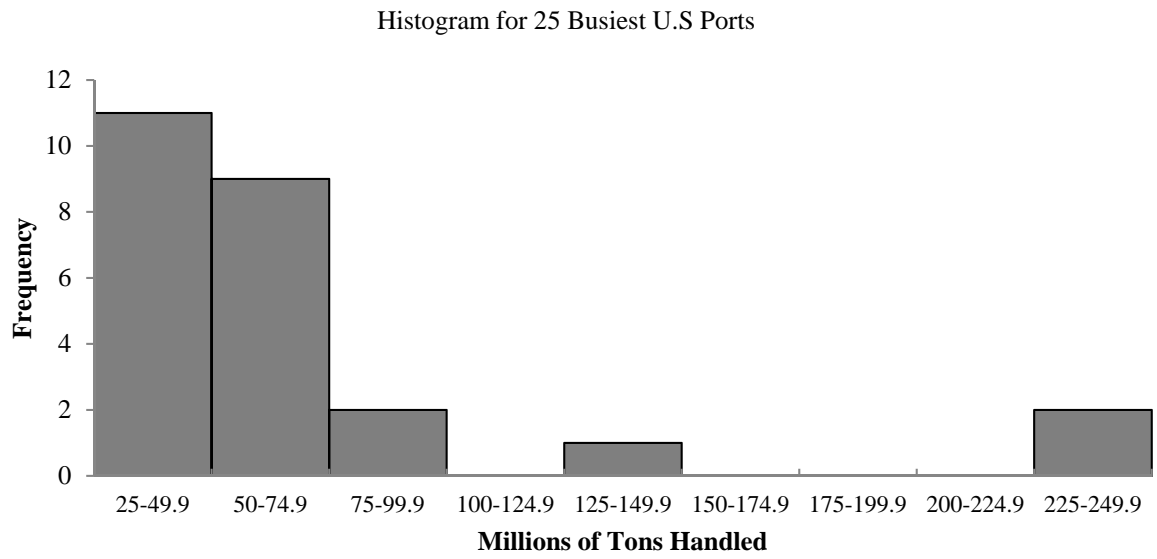
f. $(11/50)(100) = 22\%$

19. a. The largest number of tons is 236.3 million (South Louisiana). The smallest number of tons is 30.2 million (Port Arthur).

b.

<u>Millions Of Tons</u>	<u>Frequency</u>
25-50	11
50-75	9
75-100	2
100-125	0
125-150	1
150-175	0
175-200	0
200-225	0
225-250	2

c.



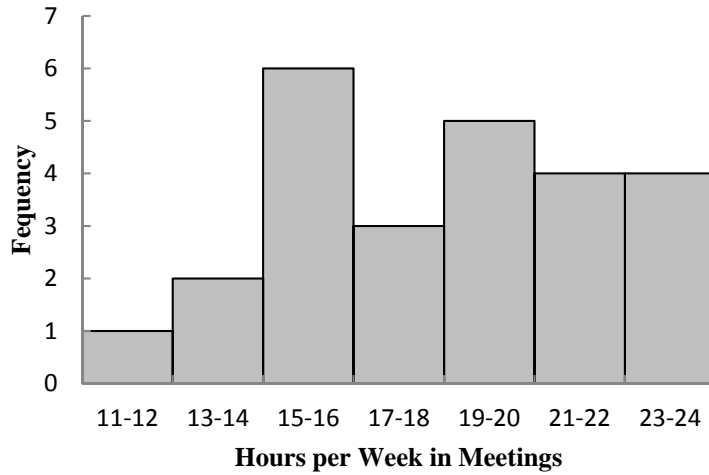
Most of the top 25 ports handle less than 75 million tons. Only five of the 25 ports handle above 75 million tons.

20. a. Lowest = 12, Highest = 23

b.

Hours in Meetings per Week	Percent	
Week	Frequency	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.



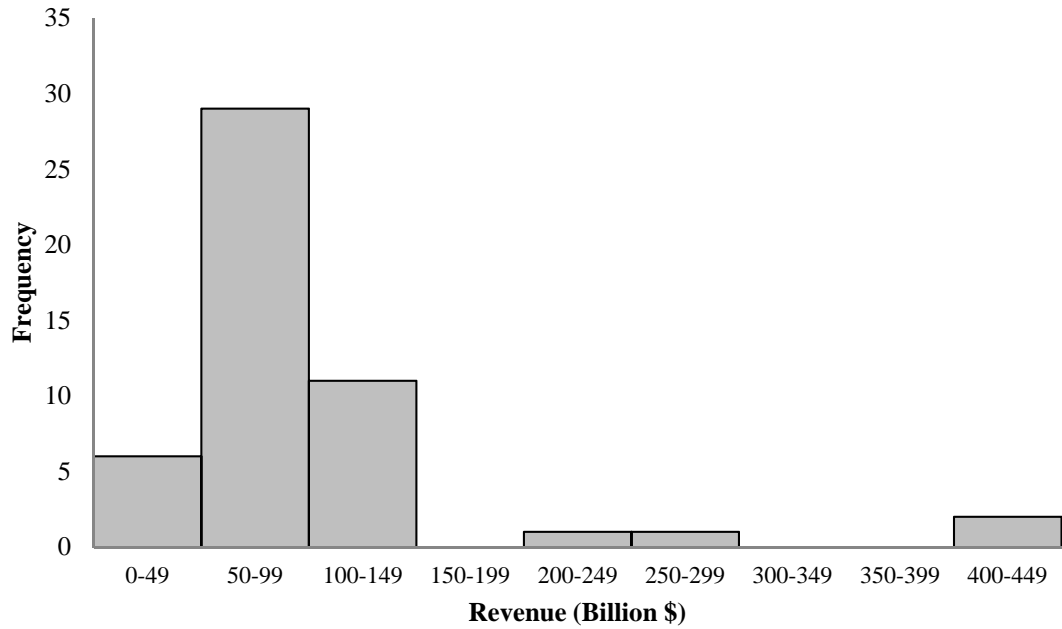
The distribution is slightly skewed to the left.

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.



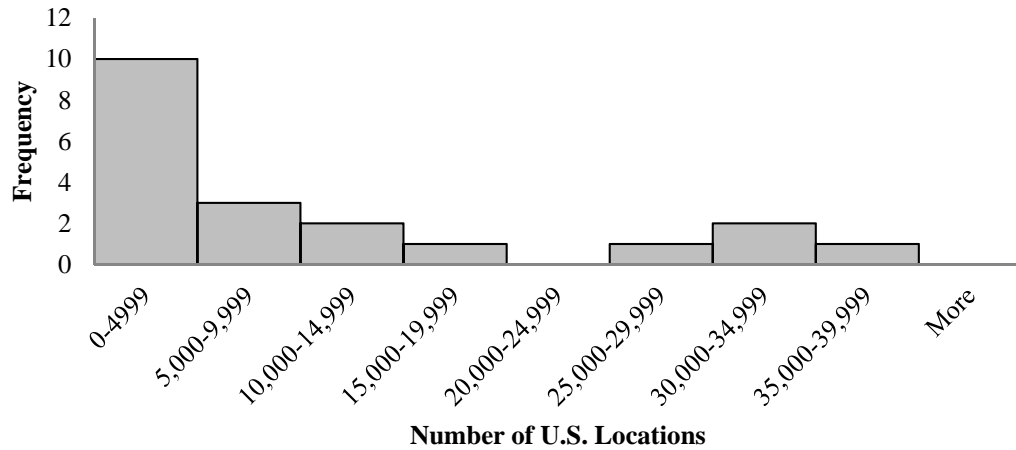
The histogram shows the distribution is skewed to the right with four corporations in the \$200 to \$449 billion range.

- 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.



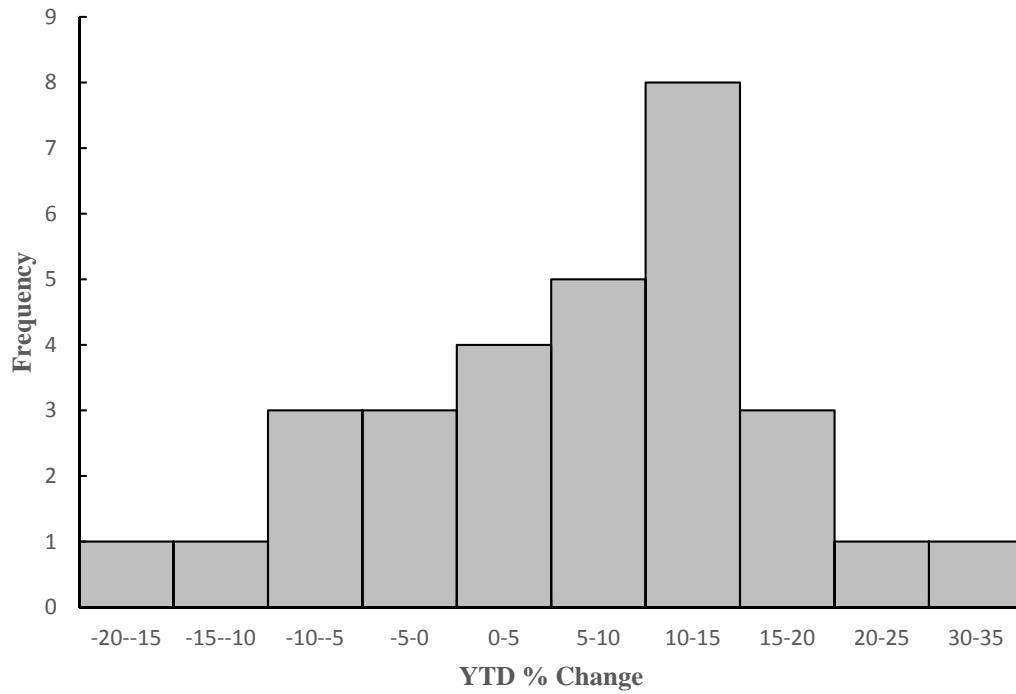
c. The distribution is skewed to the right. The majority of the franchises in this list have fewer than 20,000 locations ($50\% + 15\% + 15\% = 80\%$). McDonald's, Subway and 7-Eleven have the highest number of locations.

23. a. The highest positive YTD % Change for Japan's Nikkei index with a YTD % Change of 31.4%.

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

<u>YTD % Change</u>	<u>Frequency</u>
-20--15	1
-15--10	1
-10--5	3
-5-0	3
0-5	4
5-10	5
10-15	8
15-20	3
20-25	1
30-35	1

c.



The general shape of the distribution is skewed to the left. Twenty two of the 30 indexes have a positive YTD % Change and 13 have a YTD % Change of 10% or more. Eight of the indexes had a negative YTD % Change.

d. A variety of comparisons are possible depending upon when the study is done.

24.

Starting Median Salary

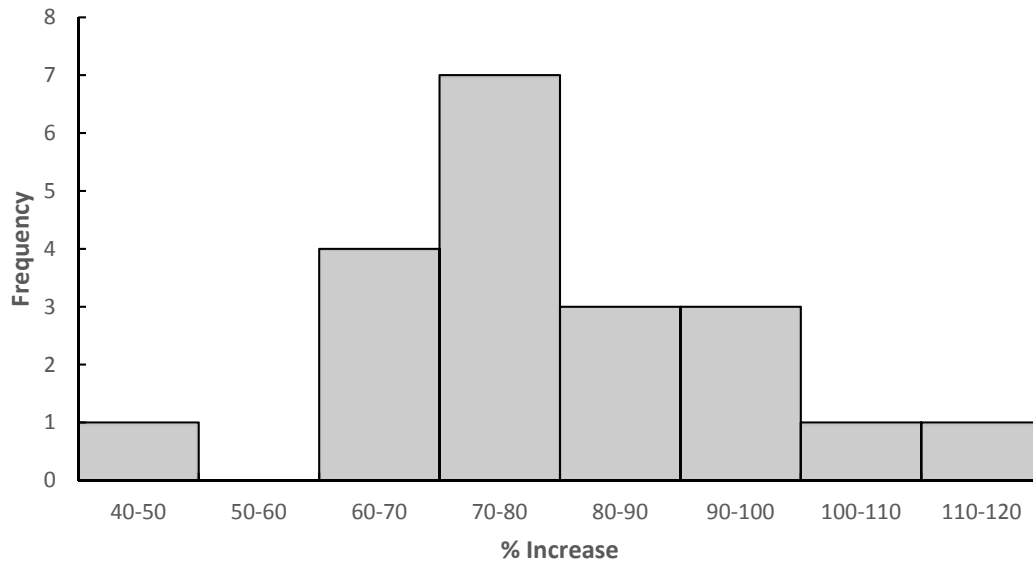
4	6	8						
5	1	2	3	3	5	6	8	8
6	0	1	1	1	2	2		
7	1	2	5					

Mid-Career Median Salary

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

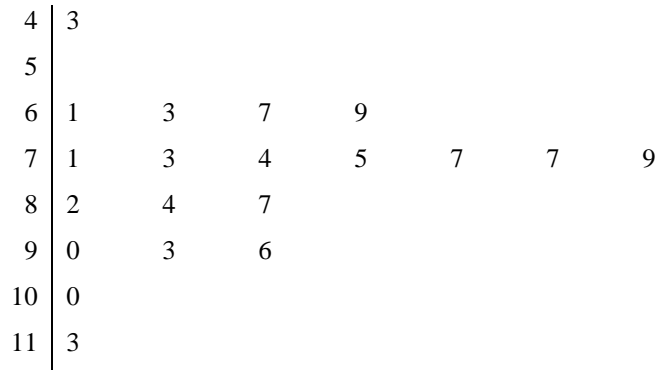
There is a wider spread in the mid-career median salaries than in the starting median salaries. Also, as expected, the mid-career median salaries are higher than the starting median salaries. The mid-career median salaries were mostly in the \$93,000 to \$114,000 range while the starting median salaries were mostly in the \$51,000 to \$62,000 range.

25. a.



b. The histogram is skewed to the right.

c.



d. Rotating the stem-and-leaf display counterclockwise onto its side provides a picture of the data that is similar to the histogram as shown in part (a). Although the stem-and-leaf display may appear to offer the same information as a histogram, it has two primary advantages: the stem-and-leaf display is easier to construct by hand; and the stem-and-leaf display provides more information than the histogram because the stem-and-leaf shows the actual data.

Chapter 2

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
x	A	27.8	0.0
	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				Grand Total
		20-39	40-59	60-79	80-100	
x	10-29			1	4	5
	30-49	2		4		6
	50-69	1	3	1		5
	70-90	4				4
	Grand Total	7	3	6	4	20

b.

		y				Grand Total
		20-39	40-59	60-79	80-100	
x	10-29			20.0	80.0	100
	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
x	10-29	0.0	0.0	16.7	100.0
	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

Chapter 2

29. a.

Make	Average Miles per Hour					Total
	130-139.9	140-149.9	150-159.9	160-169.9	170-179.9	
Buick	100.00	0.00	0.00	0.00	0.00	100.00
Chevrolet	18.75	31.25	25.00	18.75	6.25	100.00
Dodge	0.00	100.00	0.00	0.00	0.00	100.00
Ford	33.33	16.67	33.33	16.67	0.00	100.00

b. $25.00 + 18.75 + 6.25 = 50$ percent

c.

Make	Average Miles per Hour				
	130-139.9	140-149.9	150-159.9	160-169.9	170-179.9
Buick	16.67	0.00	0.00	0.00	0.00
Chevrolet	50.00	62.50	66.67	75.00	100.00
Dodge	0.00	25.00	0.00	0.00	0.00
Ford	33.33	12.50	33.33	25.00	0.00
Total	100.00	100.00	100.00	100.00	100.00

d. 75%

30. a.

Average Speed	Year					Total
	1988-1992	1993-1997	1998-2002	2003-2007	2008-2012	
130-139.9	16.7	0.0	0.0	33.3	50.0	100
140-149.9	25.0	25.0	12.5	25.0	12.5	100
150-159.9	0.0	50.0	16.7	16.7	16.7	100
160-169.9	50.0	0.0	50.0	0.0	0.0	100
170-179.9	0.0	0.0	100.0	0.0	0.0	100

b. It appears that most of the faster average winning times occur before 2003. This could be due to new regulations that take into account driver safety, fan safety, the environmental impact, and fuel consumption during races.

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

Gender	Green Condition		Total
	Too Fast	Fine	
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. Row percentages are shown below.

Region	Under \$15,000	\$15,000 to \$24,999	\$25,000 to \$34,999	\$35,000 to \$49,999	\$50,000 to \$74,999	\$75,000 to \$99,999	\$100,000 and over	Total
Northeast	12.72	10.45	10.54	13.07	17.22	11.57	24.42	100.00
Midwest	12.40	12.60	11.58	14.27	19.11	12.06	17.97	100.00
South	14.30	12.97	11.55	14.85	17.73	11.04	17.57	100.00
West	11.84	10.73	10.15	13.65	18.44	11.77	23.43	100.00

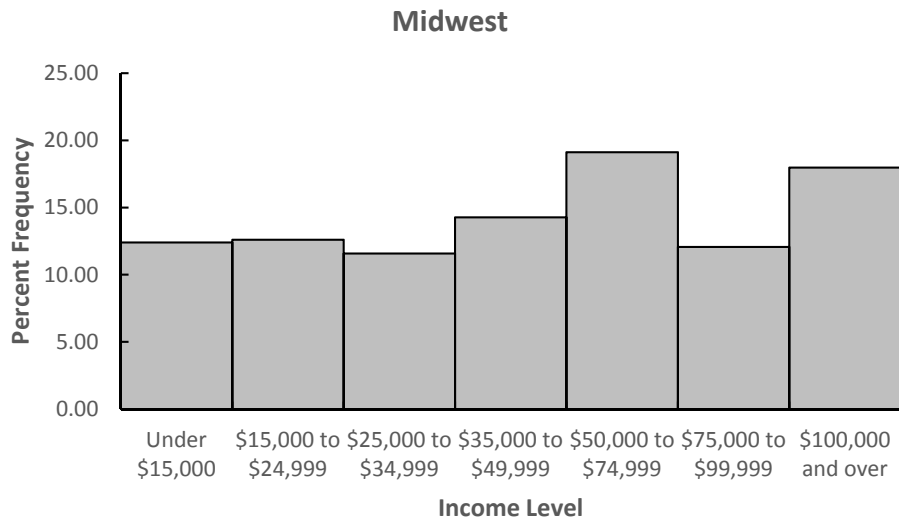
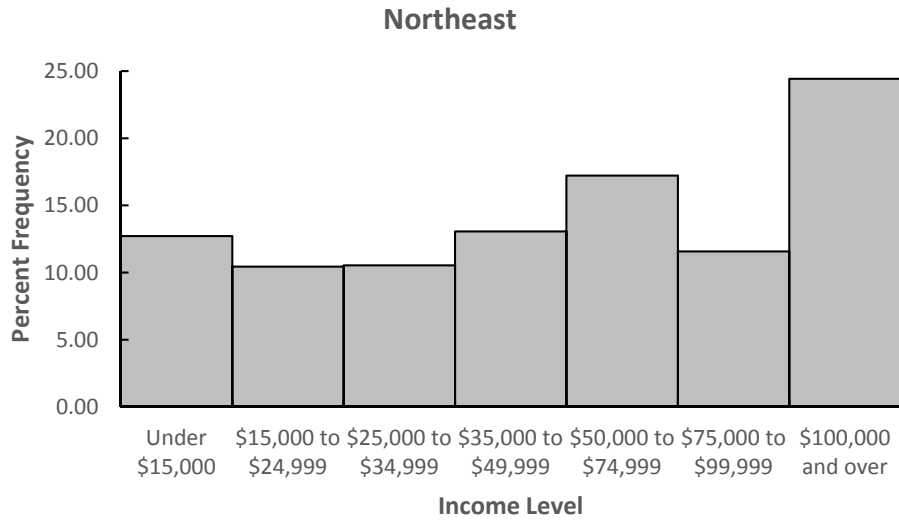
The percent frequency distributions for each region now appear in each row of the table. For example, the percent frequency distribution of the West region is as follows:

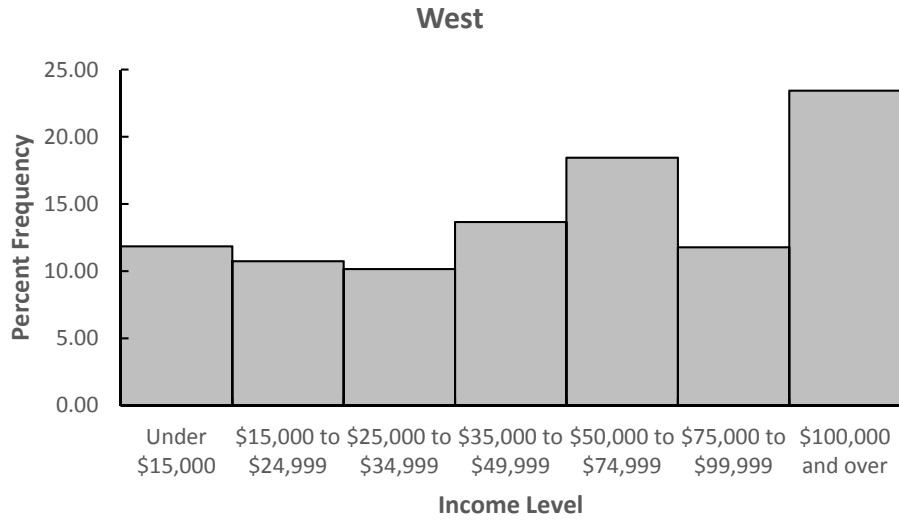
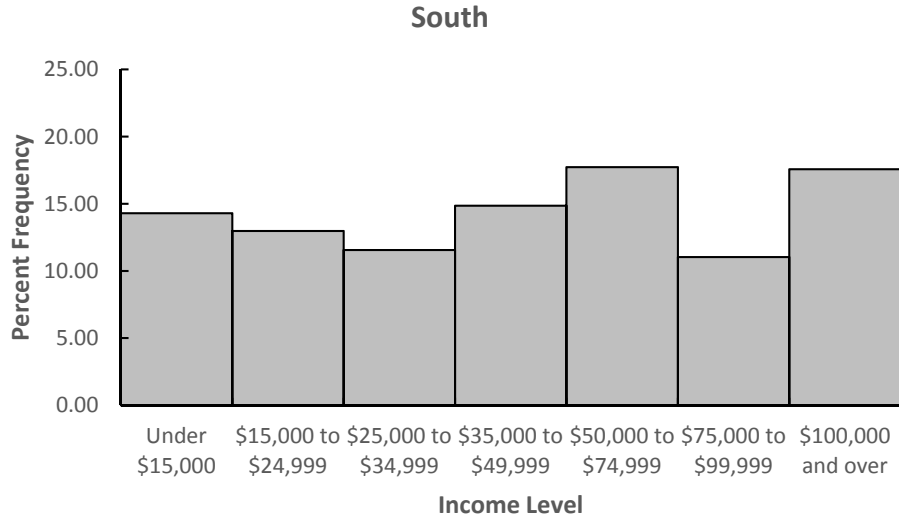
Income Level	Percent Frequency
Under \$15,000	11.84
\$15,000 to \$24,999	10.73
\$25,000 to \$34,999	10.15
\$35,000 to \$49,999	13.65
\$50,000 to \$74,999	18.44
\$75,000 to \$99,999	11.77
\$100,000 and over	23.43
Total	100.00

- b. West: $18.44 + 11.77 + 23.43 = 53.64\%$

South: $17.73 + 11.04 + 17.57 = 46.34\%$

c.





The largest difference appears to be a higher percentage of household incomes of \$100,000 and over for the Northeast and West regions.

d. Column percentages are shown below.

Region	Under \$15,000	\$15,000 to \$24,999	\$25,000 to \$34,999	\$35,000 to \$49,999	\$50,000 to \$74,999	\$75,000 to \$99,999	\$100,000 and over
Northeast	17.83	16.00	17.41	16.90	17.38	18.35	22.09
Midwest	21.35	23.72	23.50	22.68	23.71	23.49	19.96
South	40.68	40.34	38.75	39.00	36.33	35.53	32.25
West	20.13	19.94	20.34	21.42	22.58	22.63	25.70
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00