# Chapter 2
Organizing and Graphing Data

Section 2.1

2.1   Data in their original form are often too large and unmanageable. It is easier to make sense of grouped data than ungrouped data and easier to make decisions and draw conclusions using grouped data.

2.2   The relative frequency for a category is obtained by dividing the frequency of that category by the sum of the frequencies of all categories. The percentage for a category is obtained by multiplying the relative frequency of that category by 100. Example 2–2 in the text is an example which shows how relative frequencies and percentages are calculated.

2.3   **a.** and **b.**

Category	Frequency	Relative Frequency	Percentage
A	8	8/30 = .267	26.7
B	8	8/30 = .267	26.7
C	14	14/30 = .467	46.7

   c.   26.7 % of the elements in this sample belong to category B.

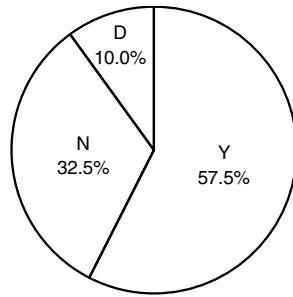   **d.**   22/30 = 73.3% of the elements in this sample belong to category A or C.

   e.
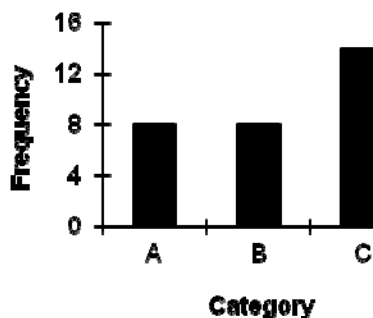
2.4   **a.** and **b.**

Category	Frequency	Relative Frequency	Percentage
Y	23	23/40 = .575	57.5
N	13	13/40 = .325	32.5
D	4	4/40 = .100	10.0

   c.   57.5% of the elements belong to category Y.

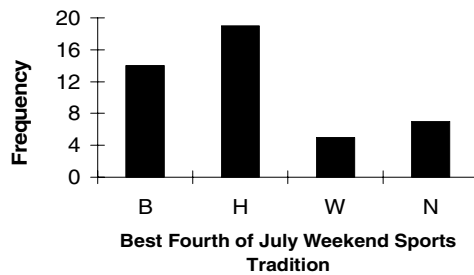   **d.**   17/40 = 42.5% of the elements belong to categories N or D.

e.

2.5   **a.** and **b.**

Category	Frequency	Relative Frequency	Percentage
B	14	14/45 = .311	31.1
H	19	19/45 = .422	42.2
W	5	5/45 = .111	11.1
N	7	7/45 = .156	15.6

c.   B + W = 14 + 5 = 19; 19/45 ≈ .422 = 42.2%
About 42.2% of the respondents mentioned Major League Baseball or Breakfast at Wimbledon.

d.

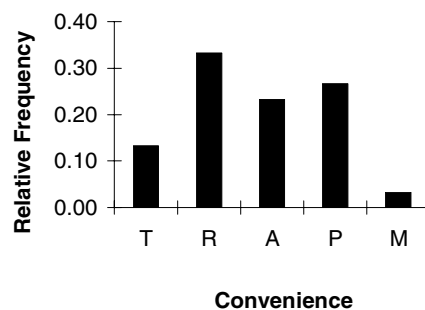**2.6**   **a.** and **b.**

Category	Frequency	Relative Frequency	Percentage
T	4	4/30 = .133	13.3
R	10	10/30 = .333	33.3
A	7	7/30 = .233	23.3
P	8	8/30 = .267	26.7
M	1	1/30 = .033	3.3

c.   (10 + 7)/30 = 56.7% of the adults ranked refrigerators or air conditioning as the convenience that they would find most difficult to do without.
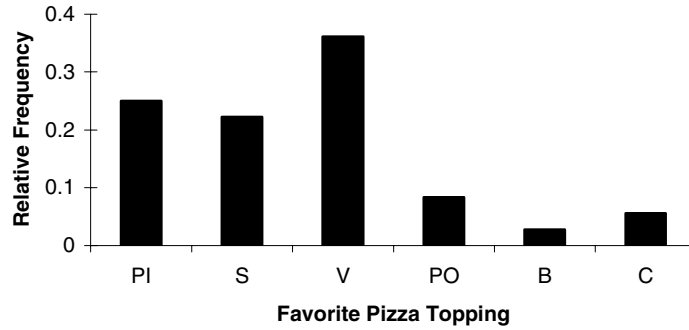
d.

2.7   **a.** and **b.**

Category	Frequency	Relative Frequency	Percentage
PI	9	9/36 = .25	25
S	8	8/36 = .222	22.2
V	13	13/36 = .361	36.1
PO	3	3/36 = .083	8.3
B	1	1/36 = .028	2.8
C	2	2/36 = .056	5.6

c. V + PO + C = 13 + 3 + 2 = 18; 18/36 = .5 = 50%
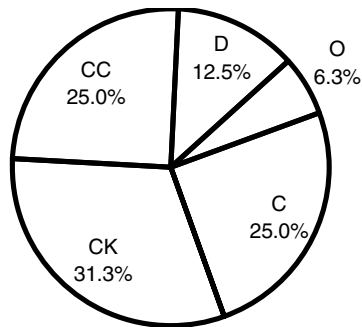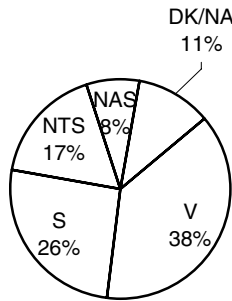50% of the respondents mentioned vegetables and fruits, poultry, or cheese.

d.

2.8   **a.** and **b.**

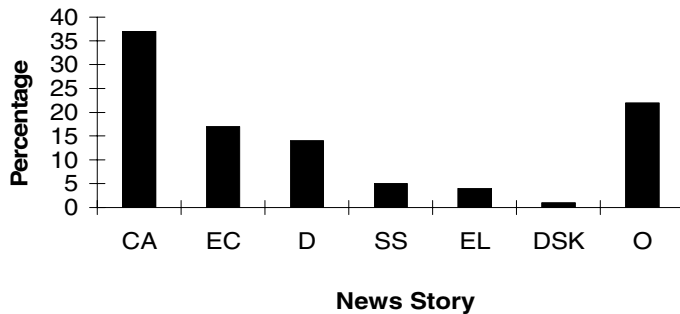| Category	Frequency	Relative Frequency	Percentage
C	4	4/16 = .250	25.0
CK	5	5/16 = .313	31.3
CC	4	4/16 = .250	25.0
D	2	2/16 = .125	12.5
O	1	1/16 = .063	6.3

c.

2.9   Let the four categories listed in the table be denoted by V, S, NTS, and NAS respectively, and let DK/NA represent "did not know or had no opinion."

2.10   Let the seven categories listed in the table be denoted by CA, EC, D, SS, EL, DSK and O respectively.

Section 2.2

2.11   The three decisions that have to be made to group a data set in the form of a frequency distribution table are
   1. The number of classes to be used to group the given data.
   2. The width of each class.
   3. The lower limit of the first class.

2.12   The relative frequency for a class is obtained by dividing the frequency of that class by the sum of frequencies of all classes. The percentage for a class is obtained by multiplying the relative frequency of that class by 100. Example 2-4 is an example that illustrates the calculation of relative frequencies and percentages.

2.13   A data set that does not contain fractional values is usually grouped by using classes with limits. Example 2–4 is an example of the writing classes using limits method. A data set that contains fractional values is grouped by using the *less than* method. Example 2–5 is an example of the *less than* method. Single-valued classes are used to group a data set that contains only a few distinct (integer) values. Example 2–6 is an example of the single-valued classes method.

2.14   **a.** $31 + 78 + 49 + 81 + 117 + 13 = 369$ customers were served.

b. Each class has a width of 4.

Gallons of Gas (Class Limits)	Class Width	Class Midpoint
0 to less than 4	4	$\frac{0+4}{2} = 2$
4 to less than 8	4	$\frac{4+8}{2} = 6$
8 to less than 12	4	$\frac{8+12}{2} = 10$
12 to less than 16	4	$\frac{12+16}{2} = 14$
16 to less than 20	4	$\frac{16+20}{2} = 18$
20 to less than 24	4	$\frac{20+24}{2} = 22$

c.

Gallons of Gas	Number of Customers	Relative Frequency	Percentage
0 to less than 4	31	$31/369 \approx .084$	8.4
4 to less than 8	78	$78/369 \approx .211$	21.1
8 to less than 12	49	$49/369 \approx .133$	13.3
12 to less than 16	81	$81/369 \approx .220$	22.0
16 to less than 20	117	$117/369 \approx .317$	31.7
20 to less than 24	13	$13/369 \approx .035$	3.5

d. $22.0 + 31.7 + 3.5 = 57.2\%$ of the customers purchased 12 gallons or more.

e. The number of customers who purchased 10 gallons or less cannot be determined exactly because 10 is not a boundary value.

2.15   **a.** $32 + 67 + 44 + 20 + 11 = 174$ containers of yogurt were inspected.

b. Each class has a width of 6.

Number of Days (Class Limits)	Class Boundary	Class Width	Class Midpoint
0 to 5	−0.5 to less than 5.5	$5.5 - (-0.5) = 6$	$\frac{0+5}{2} = 2.5$
6 to 11	5.5 to less than 11.5	$11.5 - 5.5 = 6$	$\frac{6+11}{2} = 8.5$
12 to 17	11.5 to less than 17.5	$17.5 - 11.5 = 6$	$\frac{12+17}{2} = 14.5$
18 to 23	17.5 to less than 23.5	$23.5 - 17.5 = 6$	$\frac{18+23}{2} = 20.5$
24 to 29	23.5 to less than 29.5	$29.5 - 23.5 = 6$	$\frac{24+29}{2} = 26.5$

c.

Number of Days (Class Limits)	Frequency	Relative Frequency	Percentage
0 to 5	32	$32/174 \approx .184$	18.4
6 to 11	67	$67/174 \approx .385$	38.5
12 to 17	44	$44/174 \approx .253$	25.3
18 to 23	20	$20/174 \approx .115$	11.5
24 to 29	11	$11/174 \approx .063$	6.3

d. $25.3 + 38.5 + 18.4 = 82.2\%$ of the containers would expire in less than 18 days.

e. The exact number of containers that have already expired cannot be determined because 0 is included in the class 0 to 5.

f. The largest number of containers that could already have expired is 32.

2.16   **a.** and **b.**

Class Limits	Class Boundaries	Class Midpoints
1 to 200	.5 to less than 200.5	100.5
201 to 400	200.5 to less than 400.5	300.5
401 to 600	400.5 to less than 600.5	500.5
601 to 800	600.5 to less than 800.5	700.5
801 to 1000	800.5 to less than 1000.5	900.5
1001 to 1200	1000.5 to less than 1200.5	1100.5

2.17   **a., b.,** and **c.**

Class Limits	Class Boundaries	Class Width	Class Midpoint
1 to 25	.5 to less than 25.5	25	13
26 to 50	25.5 to less than 50.5	25	38
51 to 75	50.5 to less than 75.5	25	63
76 to 100	75.5 to less than 100.5	25	88
101 to 125	100.5 to less than 125.5	25	113
126 to 150	125.5 to less than 150.5	25	138

2.18   **a.** and **b.**

Median Household Income	Frequency	Relative Frequency	Percentage
37,000 to 41,999	8	.157	15.7
42,000 to 46,999	12	.235	23.5
47,000 to 51,999	13	.255	25.5
52,000 to 56,999	9	.176	17.6
57,000 to 61,999	5	.098	9.8
62,000 to 66,999	4	.078	7.8

c. The data are skewed slightly to the right.

d. $(9 + 5 + 4)/51 = 35.3\%$ of these states had a median household income of $52,000 or more.

2.19   **a.** and **b.**

Number of Births per 1000 People	Frequency	Relative Frequency	Percentage
2 to less than 5	3	.054	5.4
5 to less than 8	8	.143	14.3
8 to less than 11	23	.411	41.1
11 to less than 14	7	.125	12.5
14 to less than 17	9	.161	16.1
17 to less than 20	3	.054	5.4
20 to less than 23	3	.054	5.4

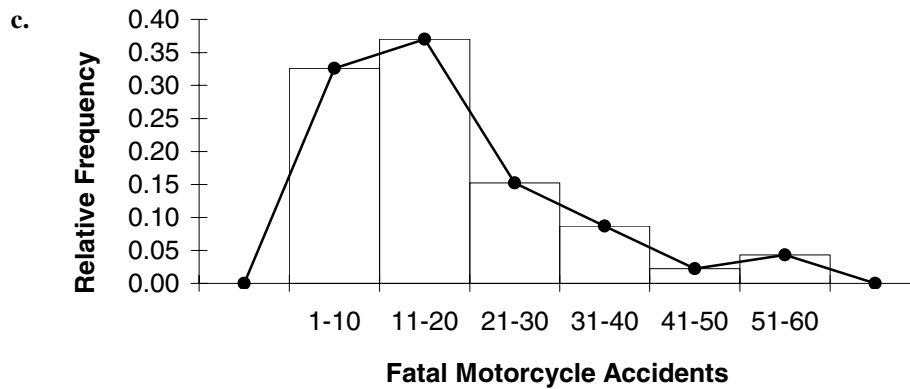**c.**

Birth Rate per 1000 People
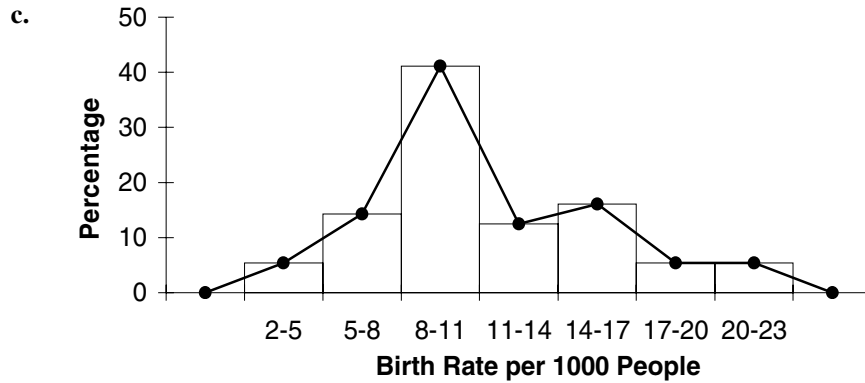
d.   (8 + 3 + 23)/56 = 60.7% of the counties had a birth rate of less than 11 births per 1000 people.

2.20   **a.** and **b.**

Fatal Motorcycle Accidents	Frequency	Relative Frequency	Percentage
1 to 10	15	.326	32.6
11 to 20	17	.370	37.0
21 to 30	7	.152	15.2
31 to 40	4	.087	8.7
41 to 50	1	.022	2.2
51 to 60	2	.043	4.3

c.

**Fatal Motorcycle Accidents**

d.   (7 + 4)/46 = 23.9% of the counties had between 21 and 40 fatal motorcycle accidents during 2009.

2.21   **a.** and **b.**

Charitable Contributions (millions of dollars)	Frequency	Relative Frequency	Percentage
25 to less than 65	25	.625	62.5
65 to less than 105	8	.200	20.0
105 to less than 145	4	.100	10.0
145 to less than 185	1	.025	2.5
185 to less than 225	0	.000	0.0
225 to less than 265	0	.000	0.0
265 to less than 305	1	.025	2.5
305 to less than 345	1	.025	2.5

c.

Charitable Contributions (millions of dollars)
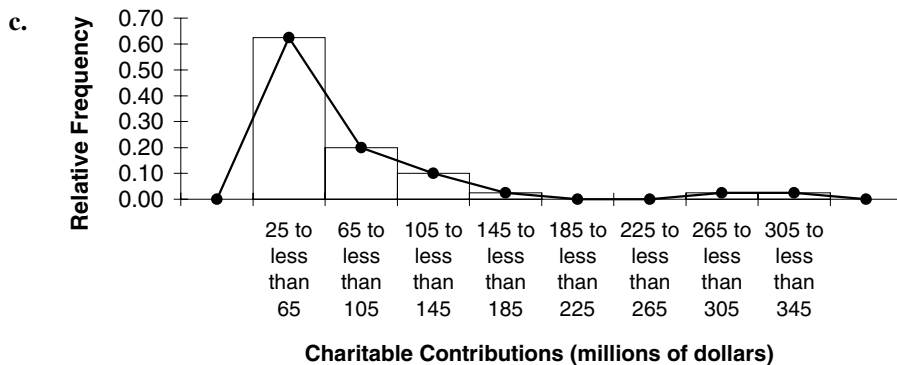
d. The donation amounts 332.0, 279.2, and 162.5 stand out because they are much larger than the rest of the donation amounts.
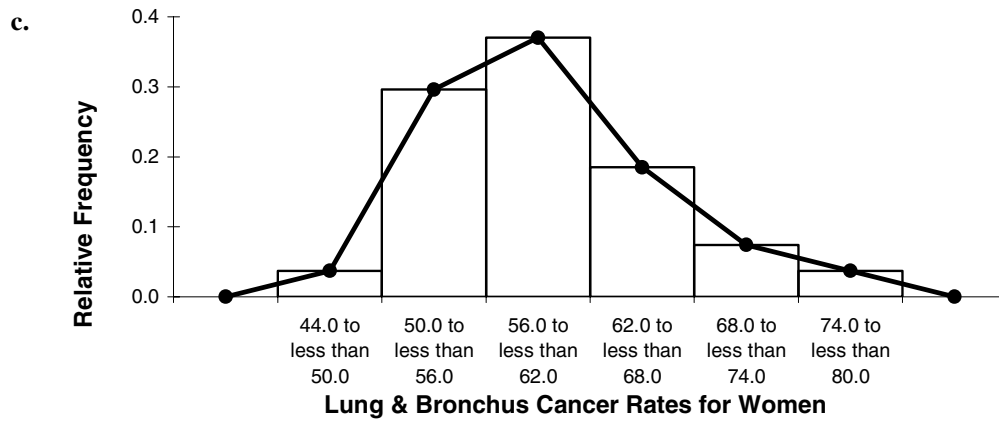
2.22 **a.** and **b.** The minimum colon and rectum cancer rate for women is 40.4, and the maximum rate is 48.9. The following table groups these data into six classes of equal width (1.5) with a starting point of 39.5.

Colon & Rectum Cancer Rates (Females)	Frequency	Relative Frequency	Percentage
39.5 to less than 41	3	.111	11.1
41 to less than 43.5	9	.222	33.3
43.5 to less than 45	5	.185	18.5
45 to less than 46.5	4	.148	14.8
46.5 to less than 48	4	.148	14.8
48 to less than 49.5	2	.074	7.4

2.23 **a.** and **b.** The minimum colon and rectum cancer rate for men is 49.4, and the maximum rate is 68. The following table groups these data into six classes of equal width (4) with a starting point of 46.0.

Colon & Rectum Cancer Rates (Males)	Frequency	Relative Frequency	Percentage
46.0 to less than 50.0	1	.037	3.7
50.0 to less than 54.0	1	.037	3.7
54.0 to less than 58.0	8	.296	29.6
58.0 to less than 62.0	11	.407	40.7
62.0 to less than 66.0	4	.148	14.8
66.0 to less than 70.0	2	.074	7.4

2.24 **a.** and **b.** The minimum lung and bronchus cancer rate for women is 46.3, and the maximum rate is 78.2. The following table groups these data into six classes of equal width (6) with a starting point of 44.0.

Lung & Bronchus Cancer Rates (Females)	Frequency	Relative Frequency	Percentage
44.0 to less than 50.0	1	.037	3.7
50.0 to less than 56.0	8	.296	29.6
56.0 to less than 62.0	10	.370	37.0
62.0 to less than 68.0	5	.185	18.5
68.0 to less than 74.0	2	.074	7.4
74.0 to less than 80.0	1	.037	3.7

c.

**Lung & Bronchus Cancer Rates for Women**

2.25   **a.** and **b.**   The minimum lung and bronchus cancer rate for men is 76.8, and the maximum rate is 131.3. The following table groups these data into six classes of equal width (10) with a starting point of 72.0.

Lung & Bronchus Cancer Rates (Males)	Frequency	Relative Frequency	Percentage
72.0 to less than 82.0	6	.222	22.1
82.0 to less than 92.0	8	.296	29.6
92.0 to less than 102.0	8	.296	29.6
102.0 to less than 112.0	2	.074	7.4
112.0 to less than 122.0	2	.074	7.4
122.0 to less than 132.0	1	.037	3.7

c.

Lung & Bronchus Cancer Rates for Men
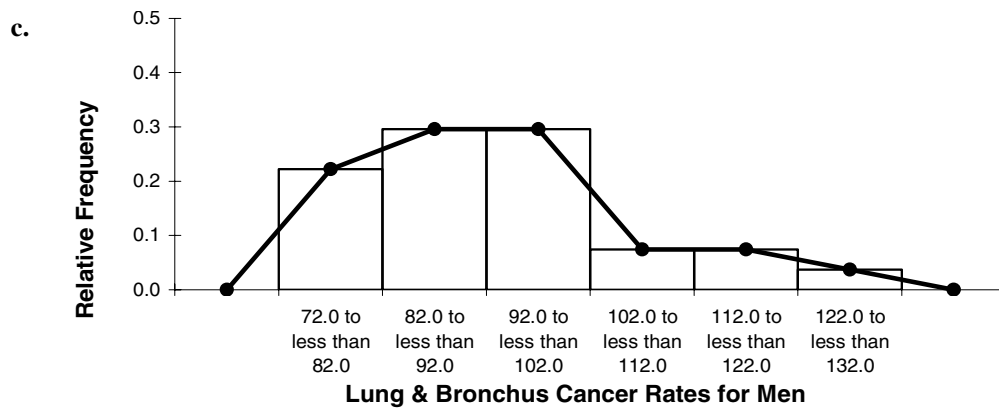
2.26   **a.** and **b.**   The minimum non-Hodgkin lymphoma cancer rate for women is 13.4, and the maximum rate is 19.1. The following table groups these data into four classes of equal width (2) with a starting point of 12.0.

Non-Hodgkin Lymphoma Cancer Rates (Females)	Frequency	Relative Frequency	Percentage
12.0 to less than 14.0	3	.111	11.1
14.0 to less than 16.0	6	.222	22.2
16.0 to less than 18.0	12	.444	44.4
18.0 to less than 20.0	6	.222	22.2

c.
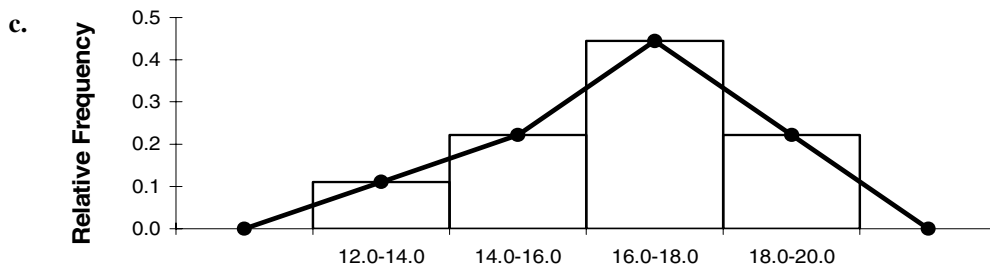
Non-Hodgkin Lymphoma Cancer Rates for Women

2.27 **a.** and **b.**

Strikeouts per Game	Frequency	Relative Frequency	Percentage
5.50 to less than 6.30	4	.133	13.3
6.30 to less than 7.10	12	.400	40.0
7.10 to less than 7.90	11	.367	36.7
7.90 to less than 8.70	2	.067	6.7
8.70 to less than 9.50	1	.033	3.3

2.28 **a.** and **b.**

| Turnovers	Frequency	Relative Frequency	Percentage
1	4	.160	16.0
2	5	.200	20.0
3	3	.120	12.0
4	3	.120	12.0
5	7	.280	28.0
6	2	.080	8.0
7	0	.000	0.0
8	1	.040	4.0

c. 3 + 7 = 10 games had four or five turnovers. The relative frequency is 10/25 = .400.

d.
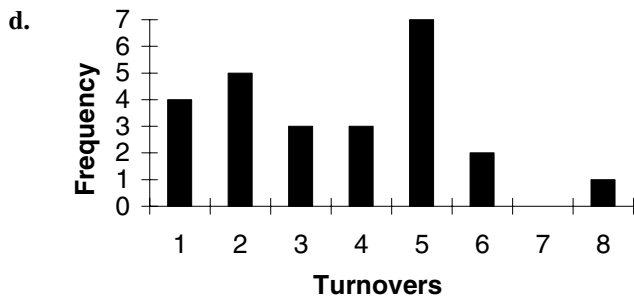
2.29 **a.** and **b.**

Number of Hot Dogs	Frequency	Relative Frequency	Percentage
0	4	0.167	16.7
1	4	0.167	16.7
2	7	0.292	29.2
3	4	0.167	16.7
4	3	0.125	12.5
5	1	0.042	4.2
6	1	0.042	4.2

c. 4 + 4 + 7 + 4 = 19 patrons ate fewer than 4 hot dogs. The relative frequency is 19/24 = .792

d.

2.30
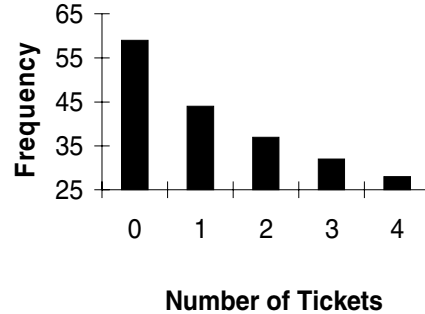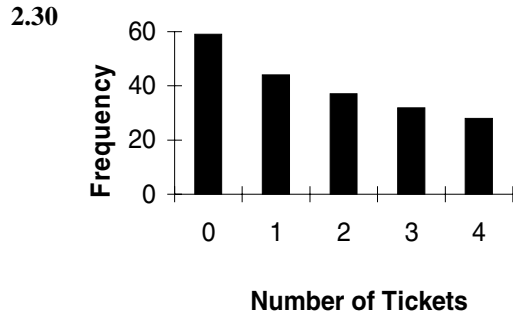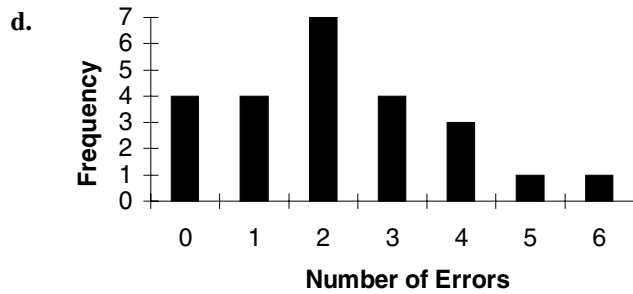
The truncated graph exaggerates the difference in the number of students with different numbers of tickets.

2.31
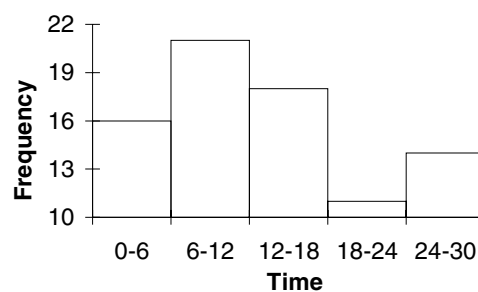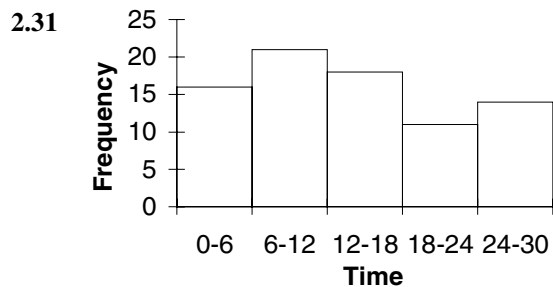
The graph with the truncated frequency axis exaggerates the differences in the frequencies of the various time intervals.

Section 2.3

2.32   The **cumulative frequency distribution** gives the total number of values that fall below the upper boundary of each class. The cumulative relative frequencies are obtained by dividing the cumulative frequencies by the total number of observations in the data set. The cumulative percentages are obtained by multiplying the cumulative relative frequencies by 100.

2.33   An **ogive** is drawn for a cumulative frequency distribution, a cumulative relative frequency distribution, or a cumulative percentage distribution. An ogive can be used to find the approximate cumulative frequency (cumulative relative frequency or cumulative percentage) for any class interval.

2.34   **a.** and **b.**

Gallons of Gasoline	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
0 to less than 4	31	31/369 = .084	8.4
0 to less than 8	31 + 78 = 109	109/369 = .295	29.5
0 to less than 12	31 + 78 + 49 = 158	158/369 = .428	42.8
0 to less than 16	31 + 78 + 49 + 81 = 239	239/369 = .648	64.8
0 to less than 20	31 + 78 + 49 + 81 + 117 = 356	356/369 = .965	96.5
0 to less than 24	31 + 78 + 49 + 81 + 117 + 13 = 369	1.000	100.0

c. 64.8% of the customers purchased less than 16 gallons.

d.

Gallons of Gasoline
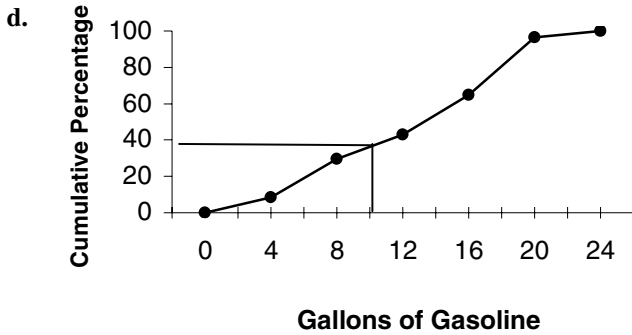
e. Approximately 38% of customers purchased less than 10 gallons of gasoline, as indicated on the ogive in part d.

2.35   **a.** and **b.**

Number of Days	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
0 to 5	32	.184	18.4
0 to 11	32 + 67 = 99	.569	56.9
0 to 17	32 + 67 + 44 = 143	.822	82.2
0 to 23	32 + 67 + 44 + 20 = 163	.937	93.7
0 to 29	32 + 67 + 44 + 20 + 11 = 174	1.000	100.0

c. 100 – 56.9 = 43.1% of the containers will expire in 12 or more days.

d.

**Number of Days to Expiry Date**
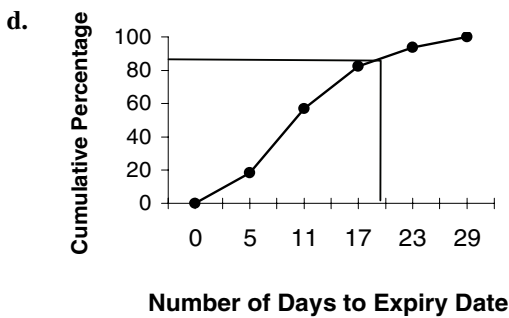
e. Approximately 85% of the containers will expire in less than 20 days, as indicated on the ogive in part d.

2.36

Median Household Income	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
37,000 to 41,999	8	0.157	15.7
37,000 to 46,999	20	0.392	39.2
37,000 to 51,999	33	0.647	64.7
37,000 to 56,999	42	0.824	82.4
37,000 to 61,999	47	0.922	92.2
37,000 to 66,999	51	1.000	100.0

2.37

Number of Births per 1000 People	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
2 to less than 5	3	.054	5.4
2 to less than 8	11	.196	19.6
2 to less than 11	34	.607	60.7
2 to less than 14	41	.732	73.2
2 to less than 17	50	.893	89.3
2 to less than 20	53	.946	94.6
2 to less than 23	56	1.000	100.0

2.38

Fatal Motorcycle Accidents	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
1 to 10	15	.326	32.6
1 to 20	32	.696	69.6
1 to 30	39	.848	84.8
1 to 40	43	.935	93.5
1 to 50	44	.957	95.7
1 to 60	46	1.000	100.0

2.39

Colon & Rectum Cancer Rates (Males)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
46.0 to less than 50.0	1	.037	3.7
46.0 to less than 54.0	2	.074	7.4
46.0 to less than 58.0	10	.370	37.0
46.0 to less than 62.0	21	.778	77.8
46.0 to less than 66.0	25	.926	92.6
46.0 to less than 70.0	27	1.000	100.0

2.40

Lung & Bronchus Cancer Rates (Males)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
72.0 to less than 82.0	6	.222	22.2
72.0 to less than 92.0	14	.519	51.9
72.0 to less than 102.0	22	.815	81.5
72.0 to less than 112.0	24	.889	88.9
72.0 to less than 122.0	26	.963	96.3
72.0 to less than 132.0	27	1.000	100.0

2.41

| Non-Hodgkin Lymphoma Cancer Rates (Females)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
12.0 to less than 14.0	3	.111	11.1
12.0 to less than 16.0	9	.333	33.3
12.0 to less than 18.0	21	.778	77.8
12.0 to less than 20.0	27	1.000	100.0

2.42

Charitable Contributions (millions of dollars)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
25 to less than 65	25	.625	62.5
25 to less than 105	33	.825	82.5
25 to less than 145	37	.925	92.5
25 to less than 185	38	.950	95.0
25 to less than 225	38	.950	95.0
25 to less than 265	38	.950	95.0
25 to less than 305	39	.975	97.5
25 to less than 345	40	1.000	100.0
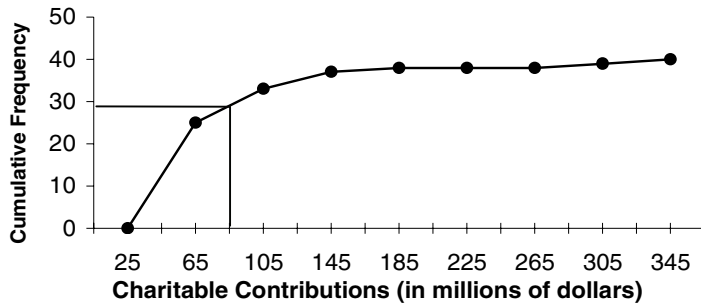
Approximately 29 individuals made charitable contributions of $85 million or less.

2.43

Points Scored	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
5.50 to less than 6.30	4	.133	13.3
5.50 to less than 7.10	16	.533	53.3
5.50 to less than 7.90	27	.900	90.0
5.50 to less than 8.70	29	.967	96.7
5.50 to less than 9.50	30	1.000	100.0
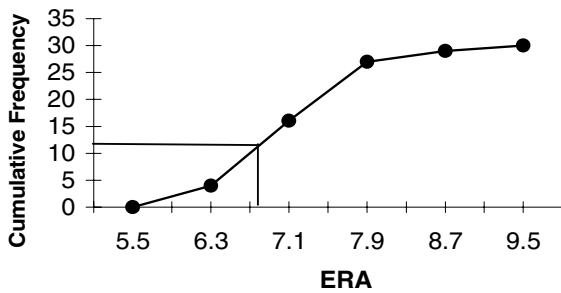
Approximately 12 of the teams had 6.8 or fewer strikouts per game.

Section 2.4

2.44 To prepare a **stem-and-leaf display** for a data set, each value is divided into two parts; the first part is called the *stem* and the second part is called the *leaf*. The stems are written on the left side of a vertical line and the leaves for each stem are written on the right side of the vertical line next to the corresponding stem. Example 2-9 is an example of a stem-and-leaf display.

2.45 The advantage of a stem-and-leaf display over a frequency distribution is that by preparing a stem-and-leaf display we do not lose information on individual observations. From a stem-and-leaf display we can obtain the original data. However, we cannot obtain the original data from a frequency distribution table. Consider the stem-and-leaf display from Example 2–8:

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

The data that were used to make this stem-and-leaf display are: 52, 50, 57, 65, 69, 61, 68, 64, 75, 79, 71, 72, 76, 79, 77, 71, 72, 80, 87, 81, 86, 83, 84, 87, 96, 93, 95, 92, 92, 98

2.46 The data that were used to make this stem-and-leaf display are: 43, 46, 50, 51, 54, 55, 63, 64, 66, 67, 67, 67, 68, 69, 72, 72, 73, 75, 76, 76, 79, 80, 87, 88, 89

2.47 The data that were used to make this stem-and-leaf display are: 218, 245, 256, 329, 367, 383, 397, 404, 427, 433, 471, 523, 537, 551, 563, 581, 592, 622, 636, 647, 655, 678, 689, 810, 841

2.48
```
0 | 3  5  5  6  8
1 | 0  4  5  6  7  7  9
2 | 1  2  3  5
3 | 0  1  1  4
```

2.49
```
 7 | 45  75
 8 | 00  48  57
 9 | 21  33  67  95
10 | 09  24
11 | 33  45
12 | 75
```

2.50
```
0 | 3  3  3  3  3  3  5  5  5  6  6  7  7  8
1 | 0  1  1  2  2  2  2  3  4  4  7  7  7  7  8  8
2 | 0  0  1  3  3  8  9  9
3 | 0  4  5  8
4 | 0  4
5 | 1
6 | 0
```

2.51 **a.**
```
0 | 2  3  5  5  6  6  7  7  7  8  8  8  8  9  9  9  9
  | 9  9
1 | 0  0  0  0  0  0  0  0  0  0  1  1  1  1  1  1  2
  | 2  2  2  4  4  4  4  4  5  5  5  5  5  6  8  9
2 | 0  0  2  3
```

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

c. The stem-and-leaf of part (b) is better because for 56 values, five stems are easier to read than three stems.

2.52   **a.**   0 | 27 28 29 30 30 30 30 30 32 32 33 35 39 40 41 42 45 49 50 50 50
            | 50 53 59 62 67 80 84 88
          1 | 00 00 01 01 10 17 18 20 63
          2 | 79
          3 | 32

   b.   0 | 27 28 29 30 30 30 30 30 32 32 33 35 39 40 41 42 45 49
          0 | 50 50 50 50 53 59 62 67 80 84 88
          1 | 00 00 01 01 10 17 18 20
          1 | 63
          2 |
          2 | 79
          3 | 32

   c.   For 40 values, the stem-and-leaf of part (b) is easier to read than that of part (a).

2.53   0 | 5 7
          1 | 0 1 5 7 9
          2 | 1 2 3 6 6 9
          3 | 2 3 9
          4 | 3 8
          5 | 0
          6 | 5

2.54   **a.**   0 | 6 5 8 9
          1 | 0 2 5 3 8
          2 | 0 5 4 0 4 6 2 8
          3 | 7 0 3 6 6 2 8
          4 | 0 5
          5 |
          6 |
          7 | 2 0 2
          8 | 0 4 6
          9 | 6 2 0

   b.   0-2 | 6 5 8 9 * 0 2 5 3 8 * 0 5 4 0 4 6 2 8
          3-5 | 7 0 3 6 6 2 8 * 0 5 *
          6-9 | * 2 0 2 * 0 4 6 * 6 2 0

2.55   **a.**   2 | 58
          3 | 20 45
          4 | 30 38 60 90
          5 | 05 30 38 50 60 65 70 75
          6 | 10 17 20 35 38
          7 | 02 05 06 20 21 28 65 87
          8 | 40 45 68 70 90
          9 | 57 68

   b.   2-4 | 58 * 20 45 * 30 38 60 90
          5-6 | 05 30 38 50 60 65 70 75 * 10 17 20 35 38
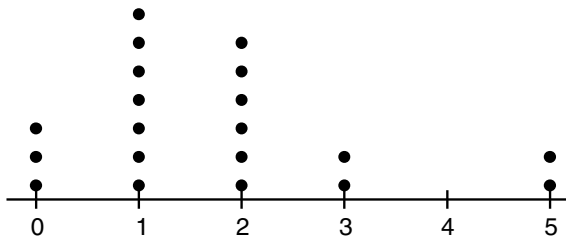   7-9 | 02 05 06 20 21 28 65 87 * 40 45 68 70 90 * 57 68
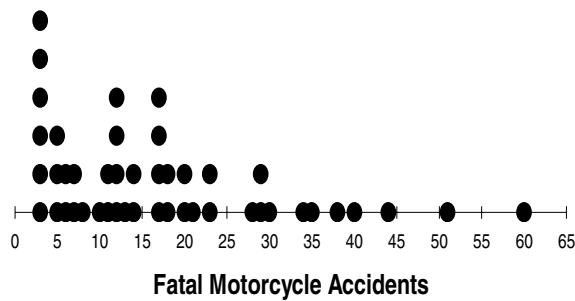
Section 2.5

2.56   In order to prepare a **dotplot**, first we draw a horizontal line with numbers that cover the given data set. Then we place a dot above the value on the number line that represents each measurement in the data set. Example 2-12 illustrates this procedure.

2.57   A **stacked dotplot** is used to compare two or more data sets by creating a dotplot for each data set with numbers lines for all data sets on the same scale. The data sets are placed on top of each other.
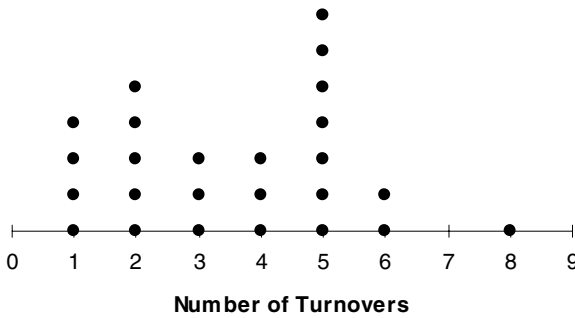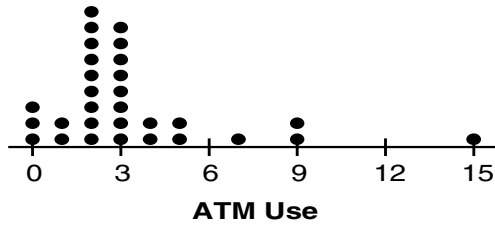
2.58

2.59

Fatal Motorcycle Accidents

2.60

Number of Turnovers
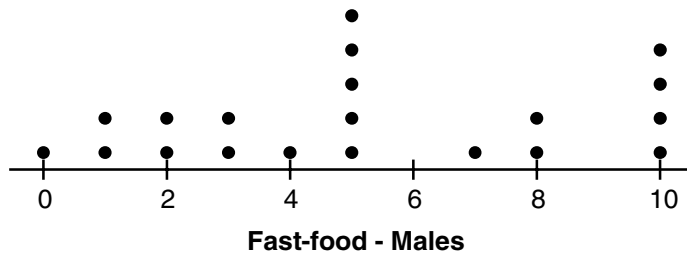
2.61

**Number of Hot Dogs**

2.62

ATM Use
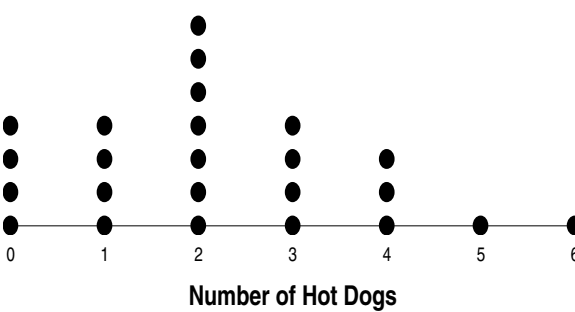
There are two clusters in the data; most of the values lie in the cluster between zero and five, with only three data points between seven and nine. The value 15 appears to be an outlier.

2.63

Fast-food - Males

The data for males is clustered in two groups with the first group having values from zero to five, and the second having values from seven to 10.

2.64

Fast-food - Females
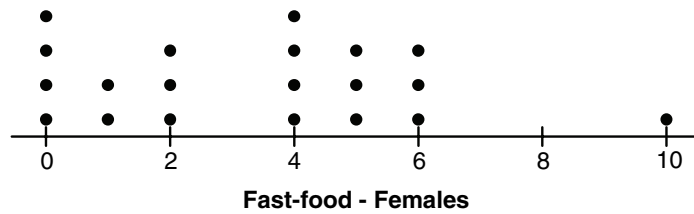
The data for females is also clustered in two groups with the first group having values from zero to two, and the second having values from four to six; 10 appears to be an outlier. With these clusters in different areas, it appears that the female students ate at fast-food restaurants less often than did males during a seven-day period.
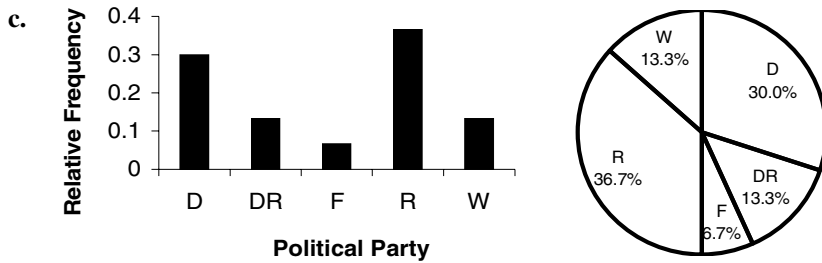
2.65

Number of Double-Doubles
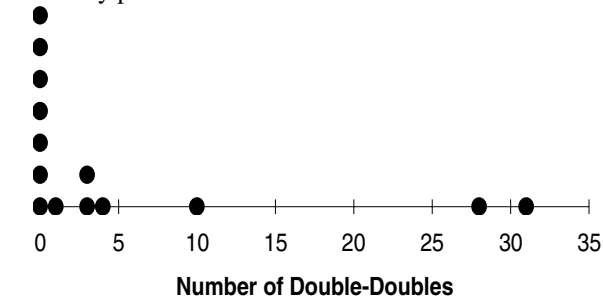
The data set contains a cluster from zero to four. The values 28 and 31 are outliers for number.

Supplementary Exercises
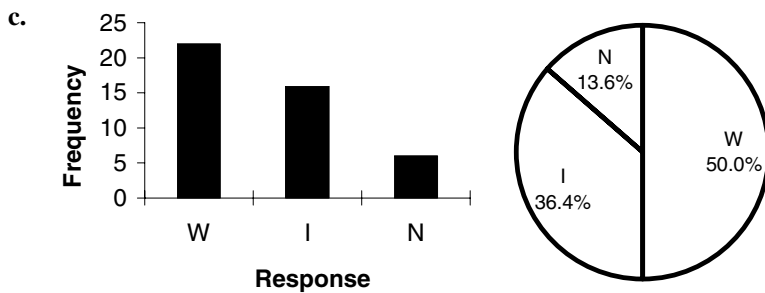
2.66   **a.** and **b.**

Political Party	Frequency	Relative Frequency	Percentage
D	9	.300	30.0
DR	4	.133	13.3
F	2	.067	6.7
R	11	.367	36.7
W	4	.133	13.3

c.

d.   13.3% of these presidents were Whigs.

2.67   **a.** and **b.**

Response	Frequency	Relative Frequency	Percentage
W	22	.500	50.0
I	16	.364	36.4
N	6	.136	13.6

c.

d.   50.0% of these respondents said "wrong priorities".

2.68   **a.** and **b.**

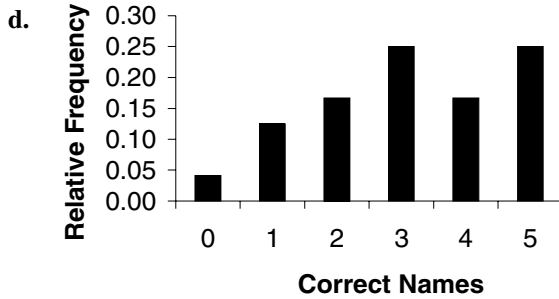| TV sets owned	Frequency	Relative Frequency	Percentage
0	1	.025	2.5
1	14	.350	35.0
2	14	.350	35.0
3	8	.200	20.0
4	3	.075	7.5

c.
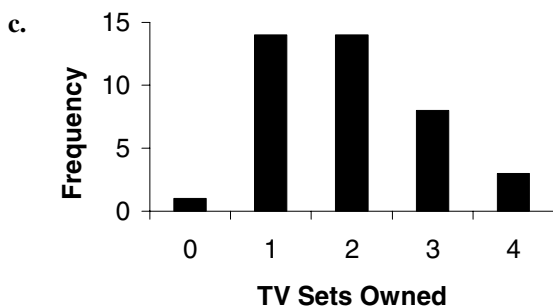
d.   (14 + 8 + 3)/40 = 62.5% of the households own two or more television sets.

2.69 **a.** and **b.**

Correct Names	Frequency	Relative Frequency	Percentage
0	1	.042	4.2
1	3	.125	12.5
2	4	.167	16.7
3	6	.250	25.0
4	4	.167	16.7
5	6	.250	25.0

c. $(1 + 3)/24 = 16.7\%$ of the students named fewer than two of the representatives correctly.

d.

2.70 **a.** and **b.**

Number of Text Messages	Frequency	Relative Frequency	Percentage
32–37	10	.250	25.0
38–43	9	.225	22.5
44–49	13	.325	32.5
50–55	6	.150	15.0
56–61	2	.050	5.0

c.
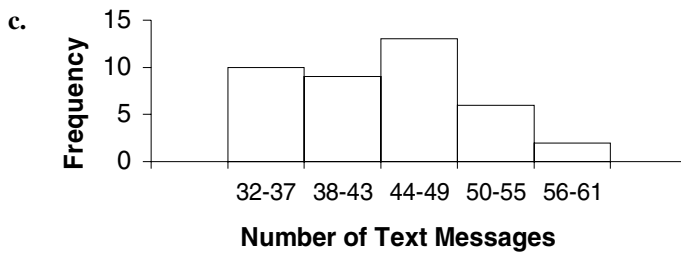
d. On $(13 + 6 + 2)/40 = 52.5\%$ of the 40 days, the student sent more than 44 text messages.

2.71 **a.** and **b.**

Number of Orders	Frequency	Relative Frequency	Percentage
23 – 29	4	.133	13.3
30 – 36	9	.300	30.0
37 – 43	6	.200	20.0
44 – 50	8	.267	26.7
51 – 57	3	.100	10.0

**c.** For $(6 + 8 + 3)/30 = 56.7\%$ of the hours in this sample, the number of orders was more than 36.

2.72 **a.** and **b.**

Concession (dollars)	Frequency	Relative Frequency	Percentage
0 to less than 6	9	.300	30.0
6 to less than 12	10	.333	33.3
12 to less than 18	5	.167	16.7
18 to less than 24	4	.133	13.3
24 to less than 30	2	.067	6.7

c.
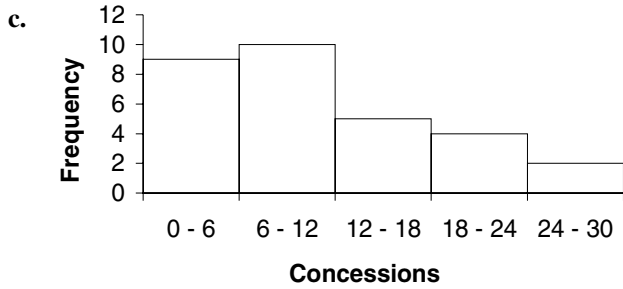
2.73 **a. and b.**

Car Repair Costs (dollars)	Frequency	Relative Frequency	Percentage
1 – 1400	11	.367	36.7
1401 – 2800	10	.333	33.3
2801 – 4200	3	.100	10.0
4201 – 5600	2	.067	6.7
5601 – 7000	4	.133	13.3

c.
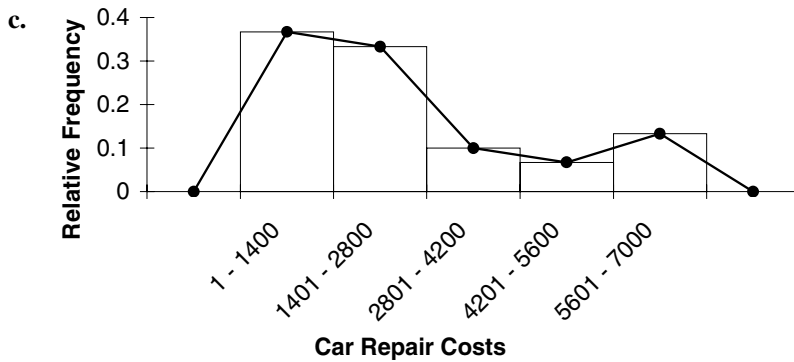
d. The class boundaries of the fourth class are $4200.50 and $5600.50. The width of this class is $1400.

2.74 **a. & b.**

Number of Text Messages	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
32–37	10	.250	25.0
32–43	19	.475	47.5
32–49	32	.800	80.0
32–55	38	.950	95.0
32–61	40	1.000	100.0

2.75

Number of Orders	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
23 – 29	4	.133	13.3
23 – 36	13	.433	43.3
23 – 43	19	.633	63.3
23 – 50	27	.900	90.0
23 – 57	30	1.000	100.0

2.76

Concession (dollars)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
0 to less than 6	9	.300	30.0
0 to less than 12	19	.633	63.3
0 to less than 18	24	.800	80.0
0 to less than 24	28	.933	93.3
0 to less than 30	30	1.000	100.0

2.77

Car Repair Costs (dollars)	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
1 – 1400	11	.367	36.7
1 – 2800	21	.700	70.0
1 – 4200	24	.800	80.0
1 – 5600	26	.867	86.7
1 – 7000	30	1.000	100.0

2.78

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

2.79

```
2 | 8  4  7  7
3 | 4  1  8  5  2  9  3  7  0  8  4  6  0
4 | 4  1  7  6  1  9  5  6  7
5 | 2  3  7  0
```
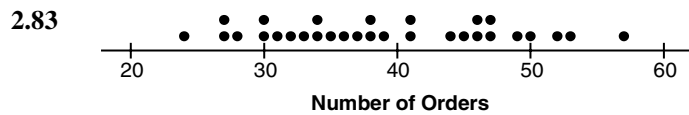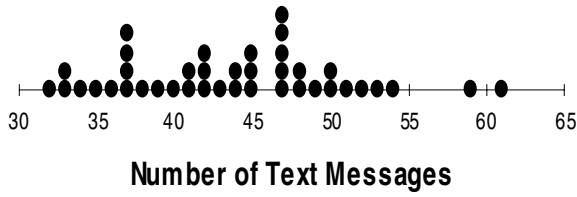
2.80
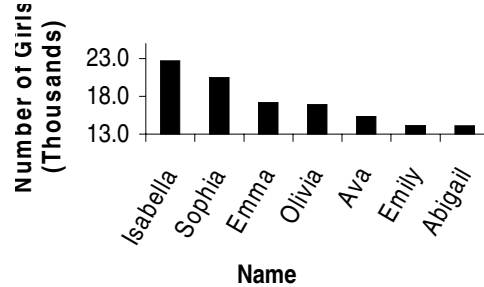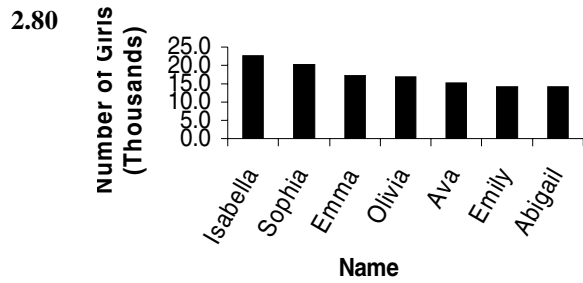
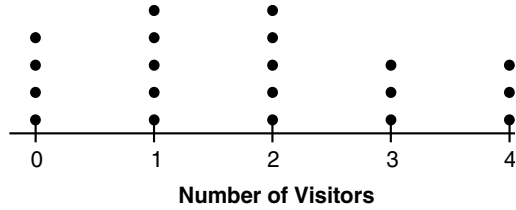The truncated graph exaggerates the differences in the number of girls with the given names.

2.81
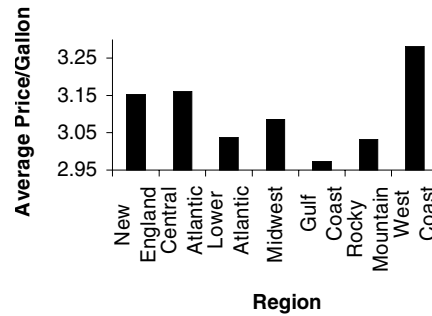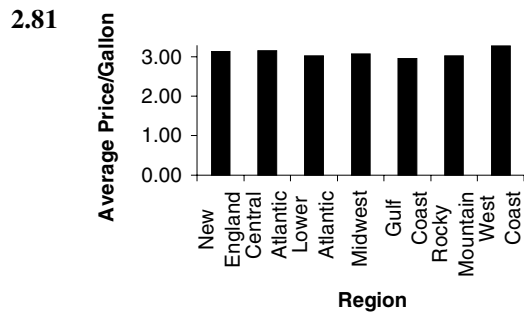
The truncated graph exaggerates the differences in average price per gallon for the period.

2.82

2.83

2.84

Number of Text Messages

2.85

Number of Visitors

2.86   **a.**

Age	Frequency	Relative Frequency
18 to less than 20	7	.060
20 to less than 25	12	.103
25 to less than 30	18	.154
30 to less than 40	14	.120
40 to less than 50	15	.128
50 to less than 60	16	.137
60 and over	35	.299
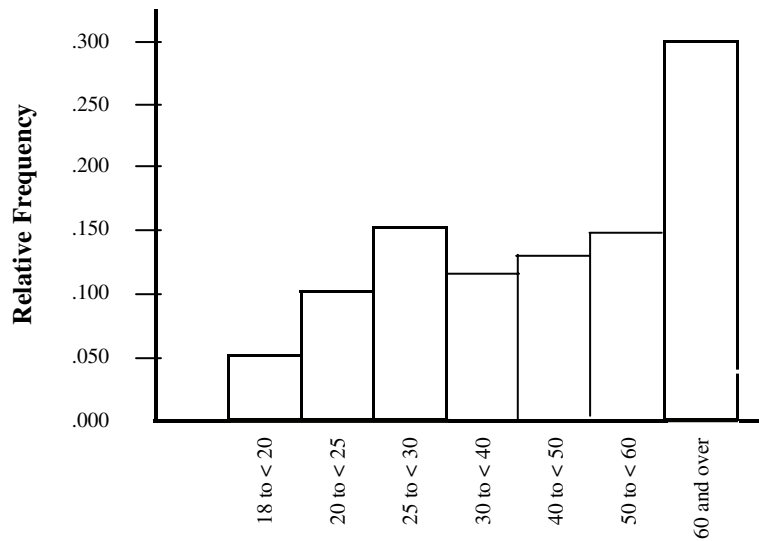
b. and **c.** This histogram is misleading because the class widths differ. If you were to change the frequency
distribution to reflect equal class widths, the resulting histogram would give a clearer picture.

2.87 The greater relative frequency of accidents in the older age group does not imply that they are more accident-prone than the younger group. For instance, the older group may drive more miles during a week than the younger group.

2.88 **a.** Using Sturge's formula:
$$c = 1 + 3.3\log n = 1 + 3.3\log 135$$
$$= 1 + 3.3(2.13033377)$$
$$= 1 + 7.03 = 8.03 \approx 8$$

    b. Approximate class width
$$= \frac{\text{Largest value } - \text{ smallest value}}{\text{Number of classes}}$$
$$= \frac{53 - 20}{8} = 4.125$$
    Use a class width of 5.

2.89 **a.** The top money winners on the men's tour tend to make more money per tournament than those on the women's tour. Earnings on the men's tour begin at $2300, and more of the data points are toward the higher end of the scale. Earnings on the women's tour begin at $800, and more of the data points are toward the lower end of the scale.

    b. Typical earnings per tournament played for the women's tour would be around $2500; typical earnings per tournament played for the men's tour would be around $3650.

    c. The data do not appear to have similar spreads for the two tours. Earnings on the men's tour begin at $2300, the largest grouping is between $2300 and $4800, and go up to $9500. Earnings on the women's tour begin at $800, the largest grouping is between $1100 and $2600, and only go up to $7500.

    d. On the women's tour, the $7500 earnings level appears to be an outlier; on the men's tour, both the $8700 and $9500 earnings levels appear to be outliers.

2.90 **a.** Answers will vary.

    b. i.

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

      ii. The display shows a bimodal distribution, due to the presence of both females and males in the sample. The males tend to be heavier, so their weights are concentrated in the larger values, while the females' weights are found primarily in the smaller values.

    c.

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

2.91 **a.** Top Histogram – Endpoints: – 0.5, 0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5; width = 1
Bottom Histogram – Endpoints: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10; width = 1

b. There is one observation between the left endpoint of the interval and 8. This can be seen by overlaying the histograms and determining the counts for each interval of .5 on the *x* axis starting at the far left. The following table displays these frequencies:

Interval	Frequency
0.0 to less than 0.5	2
0.5 to less than 1.0	2
1.0 to less than 1.5	5
1.5 to less than 2.0	6
2.0 to less than 2.5	6
2.5 to less than 3.0	3
3.0 to less than 3.5	4
3.5 to less than 4.0	5
4.0 to less than 4.5	1
4.5 to less than 5.0	4
5.0 to less than 5.5	3
5.5 to less than 6.0	1
6.0 to less than 6.5	0
6.5 to less than 7.0	1
7.0 to less than 7.5	2
7.5 to less than 8.0	1
8.0 to less than 8.5	2
8.5 to less than 9.0	0
9.0 to less than 9.5	0
9.5 to less than 10.0	2
10.0 to les than 10.5	0

c. The leftmost bar in the first histogram is misleading because it makes it appear as though there are values in the data set less than zero.

2.92
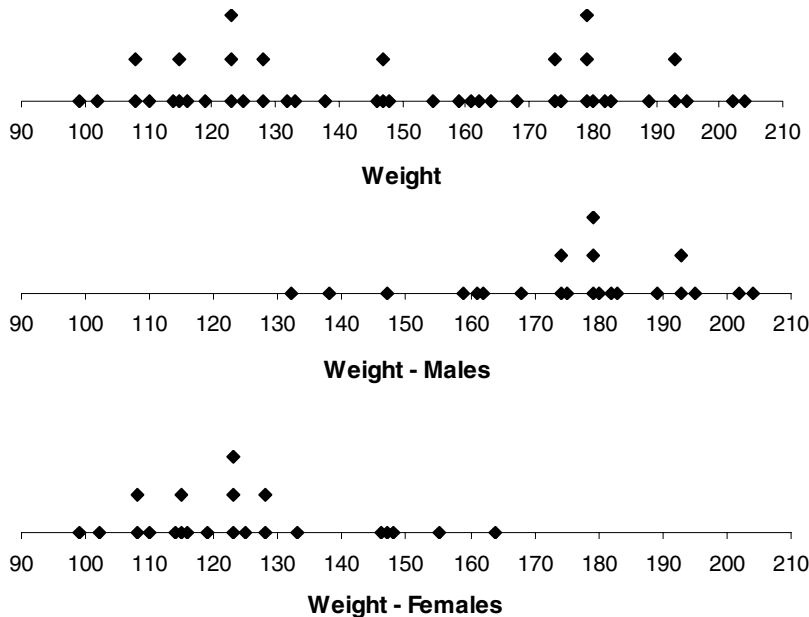
The distribution of all weights is bimodal. The distribution of weights for males is skewed to the left while the distribution for females is skewed to the right. You cannot distinguish between the lightest males and heaviest females in the dotplot of all weights as the distributions overlap in the area between 130 and 170 pounds.

2.93 **a.** Fewer than 50% of the patients are in their 50s since the angle for that classification is slightly less than 180°.

b. More than 75% of the patients are in their 50s and 60s since the angle for the total of the two classifications is slightly more than 270°.

c. The mean and standard deviation of the patients' ages as well as the mean and standard deviation of the ages of the population of men would be helpful. Stacked dotplots comparing the patients' ages to ages in the population would assist in making comparisons. It is likely that there are more men in their 50s and 60s than in their 70s and 80s, and men in these age groups may be more likely to seek medical care than the younger or older groups.

2.94 **a.**

```
    Flying Dog              Sierra Nevada
      Brewery                  Brewery
      7  7  8 | 4 | 4
   1  5  5  6 | 5 | 0  0  6  6  8  9  9
            0 | 6 | 7  8  9
      1  4  8 | 7 | 0
         3  3 | 8 |
         2  9 | 9 | 6
            2 | 10 |
            5 | 11 |
```

b. From the stem-and-leaf display, it appears that the typical alcohol content of the beer from the Flying Dog Brewery is about 7.1. It appears that the typical alcohol content of the beer from the Sierra Nevada Brewery is about 5.9.

c. It appears that the beer from the Flying Dog Brewery has a higher alcohol content than the beer from the Sierra Nevada Brewery. From the stem-and-leaf display, we see that the Sierra Nevada Brewery has only one beer with alcohol content in the 8% to 11% range, while the Flying Dog Brewery has six beers in that range.

d. The beers from the Sierra Nevada Brewery vary from 4.4% to 7.0% with an outlier at 9.6%, while the beers from the Flying Dog Brewery vary from 4.7% to 11.5%. Therefore, the alcohol content distributions do not have the same level of variability.

2.95 **a.** Figure 2.27(a) is the empirical CDF for the men's tour and Figure 2.27(b) is for the women's tour for the following reasons.
1) On Figure 2.27(a), the percentage of earning between $800 and $2300 is 0.
2) On Figure 2.27(b), 100% is reached at $7500.
3) On Figure 2.27(a), the graph takes a large number of vertical steps between $3000 and $5000.

b. The long steps at the top of the graph indicate bigger gaps between observations indicating a few observations that pull the tail of the distribution to the right.

c. Approximate values for $3000 – 30% for the men's tour and 62% for the women's tour
Approximate values for $4000 – 57% for the men's tour and 76% for the women's tour
Approximate percentage between $3000 and $4000 – 27% for the men's tour and 14% for the women's tour.

2.96 **a.** Answers may include 10.0 because it is in the center of the data.

b. There is one outlier in the data. The value 4.8 is an outlier as it is in the tail of the distribution with a large gap preceding it.

c. The distribution is skewed right as the majority of the values are between 7.3 and 12.3 with more values further to the right than to the left.

d. We cannot conclude that Oklahoma had the highest obesity rate nor Alaska the lowest as these data represent the *change* in the obesity rates, not the actual rates.

2.97   **a.** The West has the least variability as the data are clustered together. The South has the most variability as the data are the most widely spread.

      b. The South tends to have the highest obesity rates as a large number of the data points are above 27.0. The West and Northeast tend to have the lowest obesity rates with most of the data points below 25 and only one each above 27.

      c. The West appears to have an outlier at approximately 21.2. The South also appears to have outliers at approximately 22.5 and 34. The Northwest has one outlier at about 28.5. The Midwest has an outlier at about 25.

2.98   **a.** The ACC received more than 25% of the vote. The section of the pie chart representing the ACC is more than one-quarter of the whole pie.

      b. Southeastern and Big East

      c. a = Conference USA, b = Pac 10, c = Others, d = Southeastern, e = Big East, f = Big 12, g = Big Ten, h = ACC.

Self-Review Test

1.   An **ungrouped data set** contains information on each member of a sample or population individually. The first part of Example 2-1 in the text, listing the responses of each of the 30 employees, is an example of ungrouped data. Data presented in the form of a frequency table are called **grouped data**. Table 2.4 in the solution of Example 2-1 is an example of grouped data.

2. **a.** 5   **b.** 7   **c.** 17   **d.** 6.5   **e.** 13.5   **f.** 90   **g.** .30

3.   A histogram that is identical on both sides of its central point is called a **symmetric histogram**. A histogram that is **skewed to the right** has a longer tail on the right side, and a histogram that is **skewed to the left** has a longer tail on the left side. The following three histograms present these three cases. Figure 2.8 in the text provides graphs of symmetric histograms, Figure 2.9a displays a histogram skewed to the right, and Figure 2.9b displays a histogram that is skewed to the left.

4.   **a.** and **b.**

Category	Frequency	Relative Frequency	Percentage
B	8	.40	40
F	4	.20	20
M	7	.35	35
S	1	.05	5

      c. 35% of the children live with their mothers only.

      d.
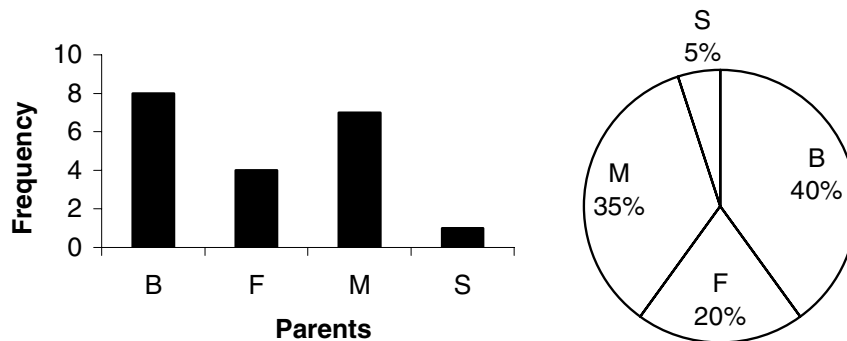
5. a. and **b.**

Number of False Alarms	Frequency	Relative Frequency	Percentage
1 – 3	5	.208	20.8
4 – 6	6	.250	25.0
7 – 9	6	.250	25.0
10 – 12	4	.167	16.7
13 – 15	3	.125	12.5

c. (5 + 6 + 6)/24 = 70.8% of the weeks had 9 or fewer false alarms.

d.
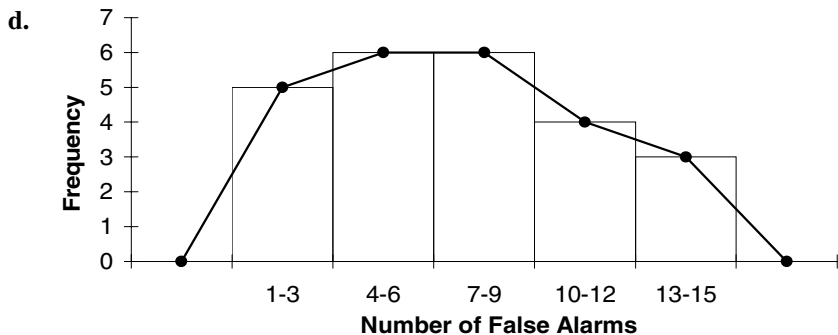
6.

Number of False Alarms	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percentage
1 – 3	5	.208	20.8
1 – 6	11	.458	45.8
1 – 9	17	.708	70.8
1 – 12	21	.875	87.5
1 – 15	24	1.000	100.0
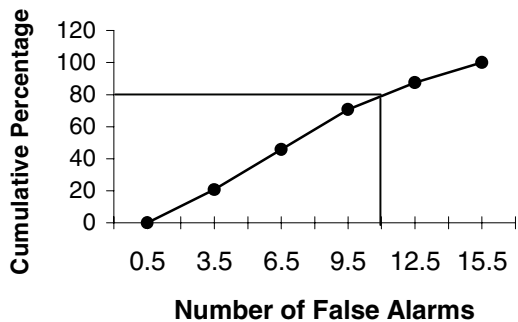
7.

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

8.   30   33   37   42   44   46   47   49   51   53   53   56   60   67   67   71   79

9.