Homework Unit 1

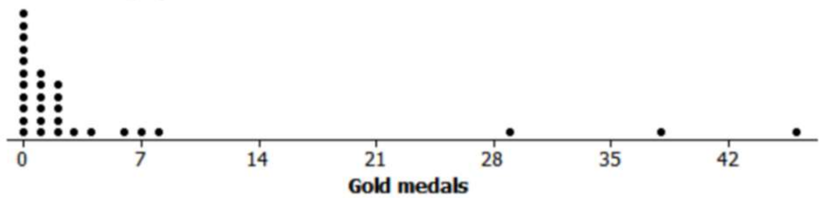
Assignment 1

1.2 Gender, race, and smoker status are categorical. Age, systolic blood pressure and level of calcium are quantitative.

1.4 (a) The individuals are roller coasters that were opened in a recent year.
(b) The categorical variables are type and design. The quantitative variables are height, speed, and duration.
(c) The highlighted roller coaster is the Prowler, a wood, sit-down type coaster. Its height is 102.3 feet, its speed is 51.2 mph, and the duration of the ride is 150 seconds.

1.6 Student answers will vary. Here is one possible answer: Two categorical variables are whether or not a student watches reality shows and how the student receives the television programming (antenna, cable, satellite, Internet). Two quantitative variables are the amount of time spent watching TV per week and the number of different channels viewed during a week.

1.38 (a) The graph is shown below.



The distribution is strongly skewed to the right with a peak at 0, which indicates that many countries did not win any gold medals. The midpoint is 1 gold medal and the number of gold medals varies from 0 to 46. There are outliers at 29 (Great Britain), 38 (China), and 46 (United States).
(b) No, this does not seem to be a representative sample since 19 out of the 30 countries in the sample (or about 63%) won gold medals. Overall, only about 54/205 = 26% won gold medals.

1.48 (a) and (b) The stemplots are shown below. The stemplot with the split stems shows the skewness, gaps, and outliers more clearly.
(c) The distribution of the amount of money spent by shoppers at this supermarket is skewed to the right, with a center around \$28, and values that vary from \$3 to \$93. There are a few gaps (from \$62 to \$69 and \$71 to \$82) and some outliers on the high end (\$83, \$86, and \$93).

(a)

```
0 | 399
1 | 1345677889
2 | 000123455668888
3 | 25699
4 | 1345579
5 | 0359
6 | 1
7 | 0
8 | 366
9 | 3
```

Key: 9 | 3 = \$93

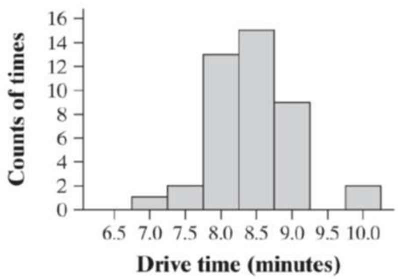
(b)

```
0 | 3
0 | 99
1 | 134
1 | 5677889
2 | 0001234
2 | 55668888
3 | 2
3 | 5699
4 | 134
4 | 5579
5 | 03
5 | 59
6 | 1
6 |
7 | 0
7 |
8 | 3
8 | 66
9 | 3
```

Key: 9 | 3 = \$93

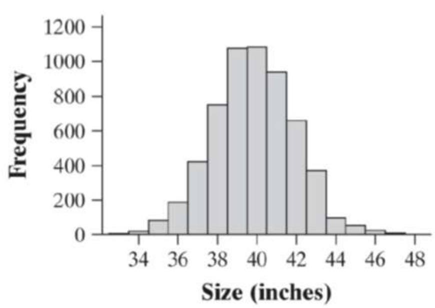
1.52 The distribution of lengths of words in Shakespeare's plays is skewed to the right and unimodal. The center is around 4 letters. The lengths of the words vary from 1 to 12 letters.

1.56 The graph is given below:



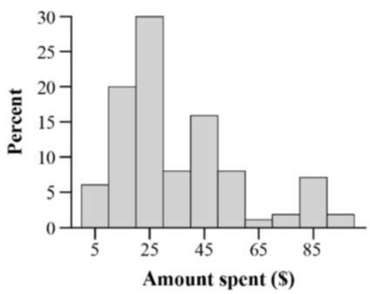
This distribution of drive times is roughly symmetric with a center around 8.5 minutes and values that vary from 6.75 to 10.17 minutes. There are no clear outliers, although times around 10 minutes might be considered outliers.

1.58 (a) The histogram is given below:



(b) The distribution of chest sizes is roughly symmetric with center around 40 inches and values that vary from 33 inches to 48 inches. This information is important so that the military can order the correct distribution of uniform sizes.

2.9 (a) First find the quartiles. The first quartile is the 25th percentile. Find 25 on the y-axis, read over to the line and then down to the x-axis to get about Q_1 = \$19. The 3rd quartile is the 75th percentile. Find 75 on the *y*-axis, read over to the line and then down to the *x*-axis to get about Q_3 = \$46. So the interquartile range is IQR = \$46 – \$19 = \$27.
(b) The person who spent \$19.50 is just above what we have called the 25th percentile. It appears that \$19.50 is at about the 26th percentile.
(c) The graph is below:



R1.1 (a) The individuals are movies.
(b) The variables are Year (quantitative), Rating (categorical), Time (quantitative), Genre (categorical), and Box office sales (quantitative). *Note*: Year might be considered categorical if we want to know how many of these movies were made each year rather than the average year.
(c) This movie is Avatar, released in 2009. It was rated PG-13, runs 162 minutes, is an action film, and had box office sales of \$2,781,505,847.

R1.6 (a) A stemplot is shown below.

48	8
49	
50	7
51	0
52	6799
53	04469
54	2467
55	03578
56	12358
57	59
58	5

Key: 48 | 8 = 4.88

(c) Since the distribution is roughly symmetric, we can use the mean to estimate the Earth's density to be about 5.45 times the density of water.

(b) The distribution is roughly symmetric with one possible outlier at 4.88. The center of the distribution is between 5.4 and 5.5. The densities vary from 4.88 to 5.85.

R2.2 (a) Reading up from 7 hours on the x-axis to the graphed line and then across to the y-axis, we see that 7 hours corresponds to about the 58th percentile.
(b) To find Q_1, start at 25 on the y-axis, move across to the line and down to the x-axis. Q_1 is approximately 2.5 hours. To find Q_3, start at 75 on the y-axis, move across to the line and down to the x-axis. Q_3 is approximately 11 hours. Thus, $IQR = 11 - 2.5 = 8.5$ hours per week.

Assignment 2

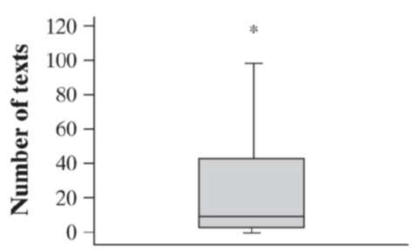
1.84 The mean house price is \$263,200 and the median is \$224,200. The distribution of house prices is likely to be quite skewed to the right because of a few very expensive homes. When a distribution is skewed to the right, the mean is bigger than the median.

1.86 The mean salary is $\dfrac{5(22,000)+2(50,000)+270,000}{8} = \$60,000$. Seven of the eight employees (everyone but the owner) earned less than the mean. The median is \$22,000. An unethical recruiter would report the mean salary as the "typical" or "average" salary. However, the median is a more accurate depiction of a "typical" employee's earnings, because it is not influenced by the outlier of \$270,000.

1.88 (a) Estimate the frequencies of the bars (from left to right): 15, 11, 15, 11, 8, 5, 3, 3, 3 (although answers may vary slightly, the frequencies must sum to 74). We estimate the median by finding the average of the 37th and 38th values. The median is 2. The first quartile is the median of the 37 observations below the median, which is the 19th observation. Thus, $Q_1 = 1$. The third quartile is the median of the 37 observations above the median, which is the 56th observation. Thus, $Q_3 = 4$.

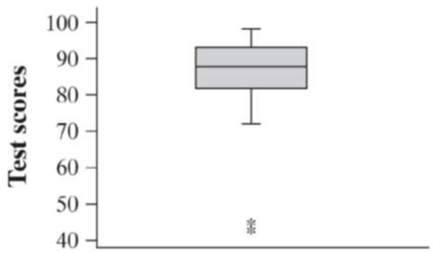
(b) Using these values, we can estimate the mean by adding 0 fifteen times, 1 eleven times, and so on. This gives us a sum of 194. The mean is then calculated by dividing by the number of responses: $\bar{x} = \frac{194}{74} = 2.62$. Alternatively, we can estimate the mean by finding the balance point of the distribution , which appears to be a little higher than 2.5.

1.91 (a) Putting the data in order: 0, 0, 0, 1, 1, 3, 3, 5, 5, 7, 8, 8, 9, 14, 25, 25, 26, 29, 42, 44, 52, 72, 92, 98, 118. The median is 9, the first quartile is 3, and the third quartile is 43. The *IQR* is 43 – 3 = 40. Outliers are anything below $3 - 1.5(40) = -57$ or above $43 + 1.5(40) = 103$. This means that the value of 118 is an outlier. The boxplot is shown below.



(b) The article claims that teens send 1742 texts a month, which works out to be about 58 texts a day (assuming a 30 day month). Nearly all of the members of the class (21 of 25) sent fewer than 58 texts per day, which seems to contradict the claim in the article.

1.92 (a) Putting the data in order: 43, 45, 72, 73, 78, 80, 81, 82, 82.5, 85, 85.5, 86, 86, 87.5, 87.5, 88, 88, 89, 91, 91, 92, 93, 93, 93.5, 93.5, 94.5, 94.5, 95, 96, 98. The median is 87.75, the first quartile is 82, and the third quartile is 93. The IQR is 93 – 82 = 11. Any observation above $Q_3 + 1.5IQR = 93 + 1.5(11) = 109.5$ or below $Q_1 - 1.5IQR = 82 - 1.5(11) = 65.5$ is considered an outlier. Thus, the scores 43 and 45 are outliers. The boxplot is shown below.



(b) Most students did quite well. In fact, more than 75% of the class scored higher than 80. Only two of the students did very poorly.

Assignment 3

1.97 (a) The mean phosphate level is $\bar{x} = \frac{32.4}{6} = 5.4$ mg/dl.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
5.6	0.2	0.04
5.2	−0.2	0.04
4.6	−0.8	0.64
4.9	−0.5	0.25
5.7	0.3	0.09
6.4	1.0	1.00
32.4	0	2.06

The variance is $s_x^2 = \frac{2.06}{5} = 0.412$ and the standard deviation is $s_x = \sqrt{\frac{2.06}{5}} = 0.6419$ mg/dl.

(b) The phosphate level typically varies from the mean by about 0.6419 mg/dl.

1.98 (a) Mean = $\bar{x} = \dfrac{7+7+9+9}{4} = 8$ hours.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
7	−1	1
7	−1	1
9	1	1
9	1	1
32	0	4

The variance is $s_x^2 = \dfrac{4}{3} = 1.33$ and the standard deviation is $s_x = \sqrt{\dfrac{4}{3}} = 1.15$ hours.

(b) The hours of sleep typically varies from the mean by about 1.15 hours.
(c) No, it would not be safe to make this generalization. The first 4 students to arrive in the classroom are not likely to be representative of the entire class in terms of the amount of sleep they got last night.

1.103 (a) One possible answer is 1, 1, 1, 1.
(b) 0, 0, 10, 10.
(c) For part (a), any set of four identical numbers will have $s_x = 0$. For part (b), however, there is only 1 possible answer. We want the values to be as far from the mean as possible, so the squared deviations from the mean can be as big as possible. If we choose 0, 10, 10, 10—or 10, 0, 0, 0—we make the first squared deviation 7.5^2, but the other three are only 2.5^2. Our best choice is two values at each extreme, which makes all four squared deviations equal to 5^2.

2.19 (a) The mean and the median both increase by 18 so the mean is 69.188 + 18 = 87.188 inches and the median is 69.5 + 18 = 87.5 inches.
(b) The standard deviation (3.20 inches) and *IQR* (71 – 67.75 = 3.25 inches) do not change because adding a constant to each value in a distribution does not change the spread.

2.20 (a) The mean and median salaries will each increase by \$1000 because adding a constant to each value in a distribution increases measures of center by the same amount.
(b) The extremes and quartiles will also each increase by \$1000 because adding a constant to each value in a distribution increases all the measures of location by the same amount. The standard deviation will not change because adding a constant to each value in a distribution does not change the spread.

2.22 (a) The mean and median will each increase by 5% because each value in the distribution is being multiplied by 1.05. Multiplying each value in a distribution by a constant multiplies measures of center by the same amount.
(b) Yes. The *IQR* and standard deviation will both increase by 5% because each value in the distribution is being multiplied by 1.05. Multiplying each value in a distribution by a constant multiplies measures of spread by the same amount.

R2.2 (a) Reading up from 7 hours on the x-axis to the graphed line and then across to the y-axis, we see that 7 hours corresponds to about the 58^{th} percentile.
(b) To find Q_1, start at 25 on the y-axis, move across to the line and down to the x-axis. Q_1 is approximately 2.5 hours. To find Q_3, start at 75 on the y-axis, move across to the line and down to the x-axis. Q_3 is approximately 11 hours. Thus, $IQR = 11 - 2.5 = 8.5$ hours per week.