2.

Observation	Deviation	Squared deviation
67	$67 - 75 = -8$	$(-8)^2 = 64$
72	$72 - 75 = -3$	$(-3)^2 = 9$
76	$76 - 75 = 1$	$1^2 = 1$
76	$76 - 75 = 1$	$1^2 = 1$
84	$84 - 75 = 9$	$9^2 = 81$
Total	**0**	**156**

3. The variance is the sum of the squared deviations (taken from the "Total" row) divided by 4 (the number of observations minus 1). In this case, that means that the variance is $\frac{156}{4} = 39$ inches squared. The standard deviation is the square root of the variance. In this case, that is 6.24 inches. 4. The players' heights vary about 6.24 inches from the mean height of 75 inches on average.

Answers to Odd-Numbered Section 1.3 Exercises

79. The mean of Joey's first 14 quiz scores is $\frac{86 + 84 + \cdots + 93}{14} = \frac{1190}{14} = 85$. If Joey had scored the same number of total points on the first 14 quizzes, but the scores had all been the same, then he would have scored an 85 on each quiz.

81. (a) Putting the scores in order:

74 75 76 78 80 82 84 86 87 90 91 93 96 98

Since there are 14 scores, the median is the mean of the 7th and 8th scores. Therefore, the median is $\frac{84 + 86}{2} = 85$. About half of the scores are smaller than 85 and about half are larger than 85. **(b)** If Joey had a 0 for the 15th quiz, then the sum of his quiz scores would still be 1190, leading to a mean of $\frac{1190}{15} = 79.33$. To find the median, we add the 0 to the beginning of the list in part (a). Since there are now 15 measurements, the median is the 8th measurement, which is 84. Notice that the median did not change much but the mean did. This shows that the mean is not resistant to outliers, but the median is.

83. Mean: \$60,954. Median: \$48,097. The distribution is likely to be quite right-skewed because of a few people who have very large incomes. When a distribution is skewed to the right, the mean is bigger since the tail values pull the mean toward them.

85. \$30 million. No; you would not be able to calculate the team's annual payroll from the median, because you cannot determine the sum of all 25 salaries from the median.

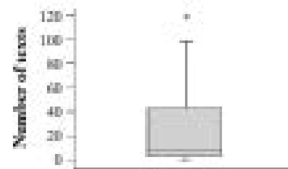
87. (a) Estimate the frequencies of the bars (from left to right): 10, 40, 42, 58, 105, 60, 58, 38, 27, 18, 20, 10, 5, 5, 1, and 3. We estimate the mean by first adding 2 ten times, 3 forty times, . . . , and 17 three times. This gives us a sum of 3504. The mean is then estimated by dividing by the number of responses: $\bar{x} = \frac{3504}{500} = 7.01$. We estimate the median by finding the average of the 250th and 251st values. The median is 6. (b) Since the median is less than the mean, we would use the median to argue that shorter domain names are more popular.

89. (a) Putting the data in order we get:

74 75 76 78 80 82 84 86 87 90 91 93 96 98

There are 14 observations here so the first quartile is the median of the bottom 7 observations. This means that it is the value of the 4th observation. It is 78. The third quartile is the median of the top 7 observations, so it is the value of the 11th observation. It is 91. So $IQR = 13$. The middle 50% of the data have a spread of 13 points. (b) Any outliers are below $Q_1 - 1.5IQR$ or above $Q_3 + 1.5IQR$. These limits are computed to be $78 - 1.5(13) = 58.5$ and $91 + 1.5(13) = 110.5$. There are no points outside these bounds, so there are no outliers.

91. (a) Here is a boxplot.



(b) The article claims that teens send 1742 texts a month. This works out to be about 58 texts a day (assuming a 30-day month). That seems pretty high given this data set. Twenty-one of the 25 students sent fewer than that; in fact, half of the students sent fewer than 10 messages (about 1/6 of the number claimed in the article). The mean number of texts per day was 27.48, less than half of 58.

93. (a) Since the data are recorded as *texts − calls*, positive numbers indicate students who had more text messages than calls. It appears from the boxplot as though the first quartile is 0, which means that approximately 25% made more calls than they texted. But the remaining 75% had more texts than calls. So this does support the article's conclusion. **(b)** No; we cannot make any more general conclusions. The sample was not a random sample, and there may be some commonality among his students that affected their responses to this question.

95. (a) The stock fund varied between about −3.5% and 3%. (b) The median return for the stock fund was slightly positive, about 0.1%, while the median real estate fund return appears to be close to 0%. (c) The stock fund is much more variable. It has higher positive returns, but also higher negative returns.

97. (a) $s_x = 0.6419$ mg/dl. (b) The typical phosphate level is an average of 0.6419 mg/dl different from the mean level.

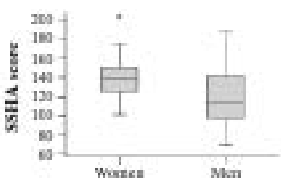
99. (a) It looks like the distribution is skewed to the right because the mean is much larger than the median. (b) The standard deviation is \$21.70. The average distance between the individual amounts spent and the mean amount spent is about \$21.70. (c) The first quartile is 19.27 and the third quartile is 45.4, so the IQR is $45.4 - 19.27 = 26.13$. Any points below $19.27 - 1.5(26.13) = -19.925$ or above $45.4 + 1.5(26.13) = 84.595$ are outliers. Since the maximum point is 93.34, there are outliers.

101. Yes, the IQR is resistant. Answers will vary. Consider the simple data set 1, 2, 3, 4, 5, 6, 7, 8. Median $= 4.5$, $Q_1 = 2.5$, $Q_3 = 6.5$, and $IQR = 4$. Changing any value outside the interval

between Q_1 and Q_3 will have no effect on the IQR. For example, if 8 is changed to 88, the IQR will still be 4.

103. (a) One possible answer is 1, 1, 1, 1. (b) 0, 0, 10, 10. (c) For (a), any set of four identical numbers will have $s_x = 0$. For (b), the answer is unique; here is a rough description of why. We want to maximize the "spread-out-ness" of the numbers (which is what the standard deviation measures), so 0 and 10 seem to be reasonable choices based on that idea. We also want to make each individual squared deviation—$(x_1 - \bar{x})^2$, $(x_2 - \bar{x})^2$, $(x_3 - \bar{x})^2$, and $(x_4 - \bar{x})^2$—as large 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.

105. **State:** Do the data indicate that women have better study habits and attitudes toward learning than men? **Plan:** We will draw side-by-side boxplots for each group. We will compute the five-number summary, the mean, and the standard deviation for the scores of each group. Then we will compare the groups using both graphical and numerical summaries. **Do:** Create boxplots and numerical summaries.



Variable	N	Mean	St. dev.	Min.	Q_1	M	Q_3	Max.
Women	18	141.06	26.44	101.00	126.00	138.50	154.00	200.00
Men	20	121.25	32.85	70.00	98.00	114.50	143.00	187.00

Conclude: It appears from the boxplots and the numerical summaries that the women have higher values for all the components of the five-number summary. They also have a higher mean and a smaller standard deviation. This means that their scores are not only higher but have less variability.

107. d

109. a

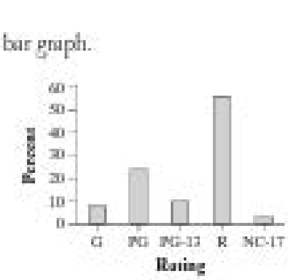
111. (a) Yes; the categories (method of communication) are mutually exclusive (each student chose one method by which they most often communicated with friends) and are parts of a whole. (b) This is a bar graph of categorical data. Skewness describes only quantitative data.

113. Women appear to be more likely to engage in behaviors that are indicative of "habits of mind." They are especially more likely to revise papers to improve their writing (about 55% of females report this, as opposed to about 37% of males). The difference is a little less for seeking feedback on their work. In that case, about 49% of the females did this, as opposed to about 38% of males.

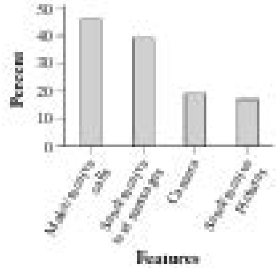
Answers to Chapter 1 Review Exercises

R1.1 (a) Movies. (b) Categorical: name, rating, genre. Quantitative: year (in years), time (in minutes), box office sales (in dollars). (c) *Avatar*, released in 2009, rated PG-13, runs 162 minutes, is an action film, and had box office sales of $2,714,767,458.

R1.2 Here is a bar graph.



R1.3 (a) It is the areas of the phones that should be in proportion, not just the heights. For example, the picture for "send/receive text messages" should be roughly twice the size of the picture for "camera," but it is actually much more than twice as large. (b) No; they do not describe parts of a whole. Students were free to answer in more than one category. (c) Here is a bar graph.



R1.4 (a) 67.6% were Facebook users. This is part of a marginal distribution because it compares the total in one column to the overall total in the table. (b) The percent of younger students who are Facebook users is 95.1%. 52.7% of Facebook users are younger.

R1.5 There does appear to be an association between age and Facebook status. Looking at the conditional distributions of Facebook status for each of the age groups, we get the following table:

	Facebook User?	
Age (years)	**Yes**	**No**
Younger (18–22)	95.1%	4.9%
Middle (23–27)	70.0%	30.0%
Older (28 and up)	31.3%	68.7%

This can also be seen in a bar graph. From both the table and the graph, we can see that the older the student is, the less likely the student is to be a member of Facebook. For younger students, about 95% are members. That drops to 70% for middle students and drops even further to 31.3% for older students.