1.72 One possible confounding variable is wealth. People who own a yacht are likely wealthy and can afford a sports car. Other answers are possible. Remember that a confounding variable should be associated with both of the variables of interest.

1.78 We are actively manipulating the explanatory variable (planting trees or not), so this is an experiment.

1.80 Data were collected after the fact from sprinters, marathon runners, and non-athletes. No genes were manipulated. These data came from an observational study.

1.84 A possible confounding variable is amount of snow and ice on the roads. When more snow and ice has fallen, more salt will be needed *and* more people will have accidents. Notice that the confounding variable has an association with *both* the variables of interest.

1.88 Yes, this study provides evidence that louder music causes people to drink more beer, because the explanatory variable (volume of music) was randomly determined by the researchers and an association was found.

1.92 (a) We randomly divide the participants into two groups of 25 each. Half will be given fluoxetine and half will get a placebo.

(b) The placebo pills will look exactly like the fluoxetine pills and will be taken the same way, but they will not have any active ingredients.

(c) The patients won't know who is getting which type of pill (the fluoxetine or the placebo) and the people treating the patients and administering the questionnaire won't know who is in which group.

2.10 A relative frequency table is a table showing the proportion in each category. In this case, the categories we are given are "No piercings", "One or two piercings", and "More than two piercings". The relative frequency with no piercings is $188/361 = 0.521$, the relative frequency for one or two piercings is $82/361 = 0.227$. The total has to add to 361, so there are $361 - 188 - 82 = 91$ students with more than two piercings, and the relative frequency is $91/361 = 0.252$. These are summarized in the relative frequency table below.

Response	Relative Frequency
No piercings	0.521
One or two piercings	0.227
More than two piercings	0.252
Total	1.00

2.13 (a) The sample is the 119 players who were observed. The population is all people who play rock-paper-scissors. The variable records which of the three options each player plays. This is a categorical variable.

(b) A relative frequency table is shown below. We see that rock is selected much more frequently than the others, and then paper, with scissors selected least often.

Option selected	Relative frequency
Rock	0.555
Paper	0.328
Scissors	0.118
Total	1.0

(c) Since rock is selected most often, your best bet is to play paper.

(d) Your opponent is likely to play paper again, so you should play scissors.

2.17 (a) There are two variables, both categorical. One is whether or not the dog selected the cancer sample and the other is whether or not the test was a breath test or a stool test.

(b) We need to include all possible outcomes for each variable when we make a two way table. The result variable has two options (dog is correct or dog is not correct) and the type of test variable has two options (breath or stool). The two-way table below summarizes these data.

	Breath test	Stool test	Total
Dog selects cancer	33	37	70
Dog does not select cancer	3	1	4
Total	36	38	74

(c) The dog got $33/36 = 0.917$ or 91.7% of the breath samples correct and $37/38 = 0.974$ or 97.4% of the stool samples correct.

(d) The dog got 70 tests correct and 37 of those were stool samples, so $37/70 = 0.529$ of the tests the dog got correct were stool samples.

2.18 (a) The table is given.

	HS or less	Some college	College grad	Total
Agree	363	176	196	735
Disagree	557	466	789	1812
Don't know	20	26	32	78
Total	940	668	1017	2625

(b) For the survey participants with a high school degree or less, we see that $363/940 = 0.386$ or 38.6% agree. For those with some college, the proportion is $176/668 = 0.263$, or 26.3% agree, and for those with a college degree, the proportion is $196/1017 = 0.193$, or 19.3% agree. There appears to be an association, and it seems that as education level goes up, the proportion who agree that every person has one true love goes down.

(c) We see that $1017/2625 = 0.387$, or 38.7% of the survey responders have a college degree or higher.

(d) A total of 1812 people disagreed and 557 of those have a high school degree or less, so we have $557/1812 = 0.307$, or 30.7% of the people who disagree have a high school degree or less.

2.19 (a) We compute the percentage of smokers in the female column and in the male column. For females, we see that $16/169 = 0.095$, so 9.5% of the females in the sample classify themselves as smokers. For males, we see that $27/193 = 0.140$, so 14% of the males in the sample classify themselves as smokers. In this sample, a larger percentage of males are smokers.

(b) For the entire sample, the proportion of smokers is $43/362 = 0.119$, or 11.9%.

(c) There are 43 smokers in the sample and 16 of them are female, so the proportion of smokers who are female is $16/43 = 0.372$, or 37.2%.

2.20 (a) This is an observational study since the researchers are observing the results after the fact and are not manipulating the gene directly to force a disruption. There are two variables: whether or not the person has dyslexia and whether or not the person has the DYXC1 break.

(b) Since $109 + 195 = 304$ people participated in the study, there will be 304 rows. Since there are two variables, there will be 2 columns: one for dyslexia or not and one for gene break or not.

(c) A two-way table showing the two groups and gene status is shown.

	Gene break	No break	Total
Dyslexia group	10	99	109
Control group	5	190	195
Total	15	289	304

(d) We look at each row (Dyslexia and Control) individually. For the dyslexia group, the proportion with the gene break is $10/109 = 0.092$. For the control group, the proportion with the gene break is $5/195 = 0.026$.

(e) There is a very substantial difference between the two proportions in part (d), so there appears to be an association between this particular genetic marker and dyslexia for the people in this sample. (As mentioned, we see in Chapter 4 how to determine whether we can generalize this result to the entire population.)

(f) We cannot assume a cause-and-effect relationship because this data comes from an observational study, not an experiment. There may be many confounding variables.

2.30 (a) Table where the vaccine has no effect (10% infected in both groups)

	Vaccine	No vaccine	Total
Malaria	20	30	50
No malaria	180	270	450
Total	200	300	500

(b) Table where the vaccine cuts the infection rate in half (from 10% to 5%).

	Vaccine	No vaccine	Total
Malaria	10	30	40
No malaria	190	270	460
Total	200	300	500

2.55 (a) The distribution is skewed to the left.

(b) The median is the value with half the area to the left and half to the right. The value 5 has way more area on the right so it cannot be correct. If we draw a line at 7, there is more area to the left than the right. The answer must be between 5 and 7 and a line at 6.5 appears to split the area into approximately equal amounts. The median is about 6.5.

(c) Because the data is skewed to the left, the values in the longer tail on the left will pull the mean down. The mean will be smaller than the median.

2.60 (a) We have $\overline{x}_f = 6.40$.

(b) We have $\overline{x}_m = 6.81$.

(c) We see that $\overline{x}_m - \overline{x}_f = 6.81 - 6.40 = 0.41$. In this sample, the males, on average, spent 0.41 more hours per week exercising than the females.

2.63 (a) There are many possible answers. One way to force the outcome is to have a very small outlier, such as
2, 51, 52, 53, 54.
The median of these 5 numbers is 52 while the mean is 42.4.

(b) There are many possible answers. One way to force the outcome is to have a very large outlier, such as
2, 3, 4, 5, 200.
The median of these 5 numbers is 4 while the mean is 42.8.

(c) There are many possible answers. One option is the following:
2, 3, 4, 5, 6.
Both the mean and the median are 4.