1.8 Since we expect the amount of fertilizer used to impact the yield (and not the other way around), we think of the amount of fertilizer as the explanatory variable and the yield of the crop as the response variable.

1.12 (a) In addition to the two identification columns, ID *Country* and the abbreviated *Code*, there are 12 variables. We see that *Developed* is a categorical variable, while the other eleven variables are all quantitative.

(b) There are many possible answers, such as "What is the average life expectancy for all countries of the world?" or "What proportion of countries are developed?"

(c) There are many possible answers, such as "Do countries with a greater land area have a larger percent rural?" or "Do countries that spend a relatively large amount on the military spend a relatively small amount on health care?" or "Do developed countries have a longer life expectancy than developing countries?"

1.20 (a) There are 8 cases, corresponding to the 8 rowers. The two variables are number of days to cross the Atlantic and gender. Number of days to cross the Atlantic is quantitative and gender is categorical.

(b) We need two columns, one for each variable. The columns can be in either order. See the table.

Time	Gender
40	Male
87	Male
78	Male
106	Male
67	Male
70	Female
153	Female
81	Female

1.22 (a) The cases are the 41 participants.

(b) There are many variables in this study. The only categorical variable is whether or not the person participated in the meditation program. All other variables are quantitative variables. These variables include (at minimum):

- Brain wave activity before
- Brain wave activity after
- Brain wave activity 4 months later
- Immune response after 1 month
- Immune response after 2 months
- Negative survey before
- Negative survey after
- Positive survey before
- Positive survey after

(c) The explanatory variable is whether or not the person participated in the meditation program.

(d) The data set will have 41 rows (one for each participant) and at least 10 columns (one for each variable).

1.24 We could sample people eligible to vote and ask them each their political party and whether they voted in the last election. The cases would be people eligible to vote that we collect data from. The variables would be political party and whether or not the person voted in the last election. Alternatively, we could ask whether each person plans to vote in an upcoming election.

1.36 The sample is the 1000 households which have databoxes attached to the televisions. The population is all US households with televisions.

1.40 (a) The sample is the girls who are on the selected basketball teams.

(b) The population we are interested in is all female high school students.

(c) A population we can generalize to, given our sample, is female high school students who are on a basketball team.

1.50 From the description, it appears that this method of data collection is not biased.

1.52 Because this was a random sample of parents in Kansas City, the result can be generalized to all parents in Kansas City.

1.56 No. This is a volunteer sample, and there is reason to believe the participants are not representative of the population. For example, some may choose to participate because they LIKE alcohol and/or marijuana, and those in the sample may tend to have more experience with these substances than the overall population. In addition, the advertisements for the study were aired on rock radio stations in Sydney, so only those people who listen to rock radio stations in Sydney would hear about the option to participate.

1.58 (a) The sample is the 300 salons that were contacted.

(b) Yes, the sample was collected in a way that should be representative of all tanning salons.

(c) The salons are more interested in marketing what they offer than in giving facts. The responses are dishonest, and a completely inaccurate portrayal of the facts, because they are trying to get more business.

(d) Yes, the sample is well taken so the study is probably accurate in how salons market to teenage girls.

1.60 (a) Since the NHANES sample is drawn from all people in the US, that is the population we can generalize to.

(b) Since the NHAMCS sample is drawn from patients in emergency rooms in the US, we can generalize the results to all emergency room patients in the US.

(c)
i. NHANES: The question about an association between being overweight and developing diabetes applies to all people in the US, not just those who visit an emergency room.

ii. NHAMCS: This question asks specifically about the type of injury for people who go to an emergency room.

iii. NHAMCS: This question of average waiting time only applies to emergency room patients.

iv. NHANES: This question is asking about all US residents. Note that the proportion would be equal to one for the people sampled in NAMCS since they only get into the sample if they visit an emergency room!