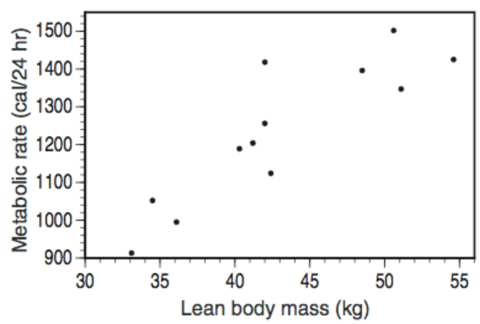
Chapter 4 Solutions

4.1: (a) Explanatory: time spent studying; response: grade. (b) Explore the relationship; there is no reason to view one or the other as explanatory. (c) Time spent online using Facebook is explanatory, GPA is the response variable. (d) Explore the relationship.

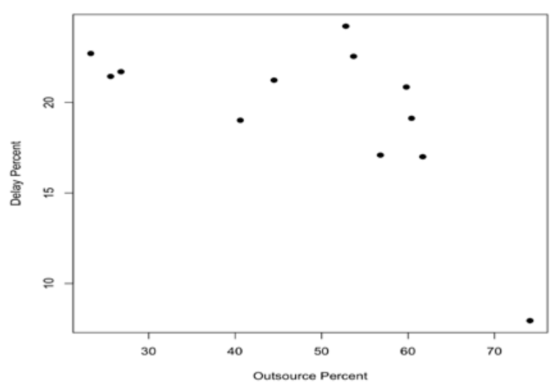
4.2: Sea-surface temperature is the explanatory variable; coral growth is the response variable. Both variables are quantitative.

4.3: For example: weight, sex, other food eaten by the students, type of beer (light, imported, . . .).

4.4: The researchers suspect that lean body mass is explanatory, so it should be on the horizontal axis.



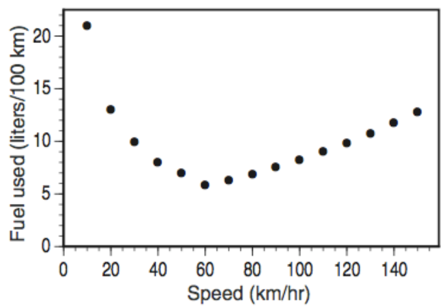
4.5: Outsource percent is the explanatory variable and should be on the horizontal axis. Delay percent is the response and should be on the vertical axis.



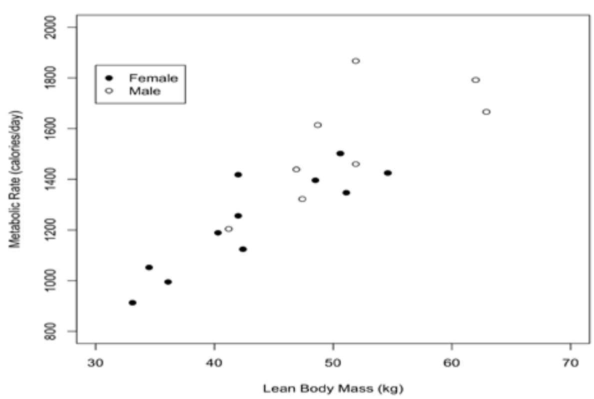
4.6: The scatterplot shows a positive direction, linear form, and moderately strong association.

4.7: There is an outlier (Hawaiian Airlines). Removing it, we would see no association between these variables. Without removing it, there is a very weak, negative association between the variables (which contradicts the suspicions described in Exercise 4.5).

4.8: (a) Below; speed is explanatory. (b) The relationship is curved—low in the middle, higher at the extremes. Because low "mileage" is actually *good* (it means that we use less fuel to travel 100 km), this makes sense: Moderate speeds yield the best performance. Note that 60 km/hr is about 37 mph. (c) Above-average (that is, bad) values of "fuel used" are found with both low and high values of "speed." (d) The relationship is very strong—there is little scatter around the curve, so the curve is very useful for prediction.

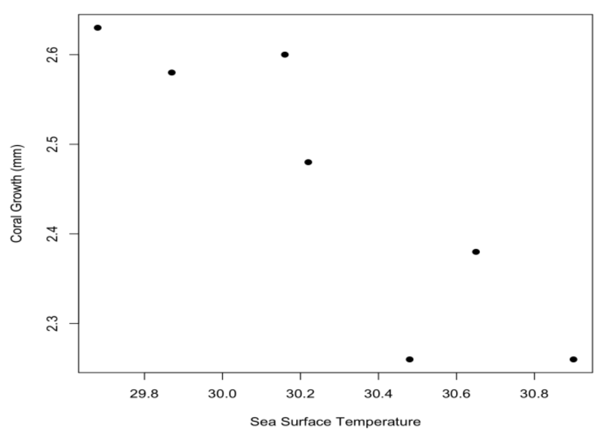


4.9: (a) Women are marked with filled circles, men with open circles. (b) For both men and women, the association is linear and positive. The women's points show a stronger association. As a group, males typically have larger values for both variables (they tend to have more mass, and tend to burn more calories per day).



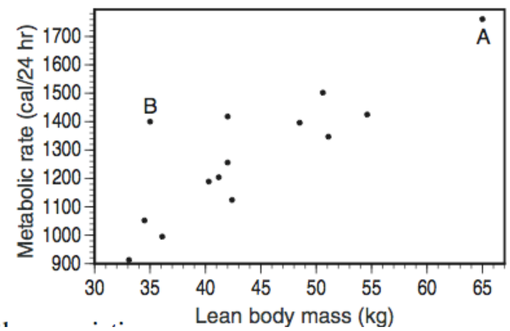
4.10: (a) Temperature is the explanatory variable. A scatterplot is provided below, and shows a fairly strong, *negative* linear association between Temperature and Coral growth. (b) $\bar{x}$ = 30.28 degrees, s_x = 0.43 degrees, $\bar{y}$ = 2.46 mm, s_y = 0.16 mm. See the table below for the standardized scores. The correlation is r = – 5.28/6 = – 0.88. This is consistent with the strong, negative association depicted in the scatterplot. (c) Software will give a value of – 0.8914. The more precision you carry at each step, the closer you'll get to that value. The answer in (c) is erroneous at the hundredths place due to rounding.

z_x	z_y	$z_x z_y$
–1.40	+1.06	–1.48
–0.95	+0.75	–0.71
–0.28	+0.88	–0.25
–0.14	+0.13	–0.02
+0.47	–1.25	–0.59
+0.86	–0.50	–0.43
+ 1.44	–1.25	–1.80
		–5.28



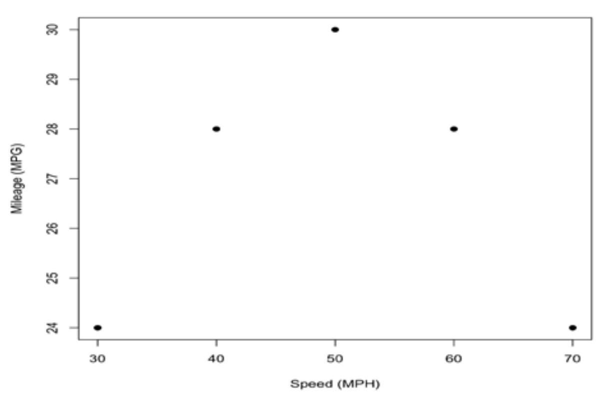
4.11: r would not change; units do not affect correlation.

4.12: (a) r = 0.8765. (b) With Point A included, the correlation increases to 0.9273; with Point B, it drops to 0.7257. (c) Point A fits in with the positive linear association displayed by the other points, and even emphasizes (strengthens) that association because, when A is included, the points of the scatterplot are less spread out (relative to the length of the apparent line suggested by the points). Meanwhile, Point B deviates from the pattern, weakening the association.



4.13: In computing the correlation, note that $\bar{x}$ = 50 mph, s_x = 15.8114 mph, $\bar{y}$ = 26.8 mpg and s_y = 2.6833 mpg. Refer to the table of standardized scores below, then note that r = 0/4 = 0. The correlation is zero because these variables do not have a straight-line relationship; the association is neither positive nor negative. Remember that correlation only measures the strength and direction of a *linear* relationship between two variables.

z_x	z_y	$z_x z_y$
−1.2649	−1.0435	1.3199
−0.6325	0.4472	−0.2828
0	1.1926	0
0.6325	0.4472	0.2828
1.2649	−1.0435	−1.3199
		0



4.14: (a) We would expect that the price of a barrel of oil has an effect on the price of gasoline, rather than the reverse.

4.15: (a) The association should be positive (e.g., if oil prices rise, so do gas prices).

4.16: (b) IQ = 103, GPA = 0.5.

4.17: (a) 0.9. Without the outlier, there is a strong positive linear relationship.

4.18: (c) Correlations range from –1 to 1 inclusive.

4.19: (c) A correlation close to 0 might arise from a scatterplot with no visible pattern, but there could be a nonlinear pattern. See Exercise 4.13, for example.

4.20: (c) Because we are not told how the x and y values vary together, we cannot tell whether the correlation will be –1 or +1.

4.21: (a) 1. There would be a perfect, positive linear association.

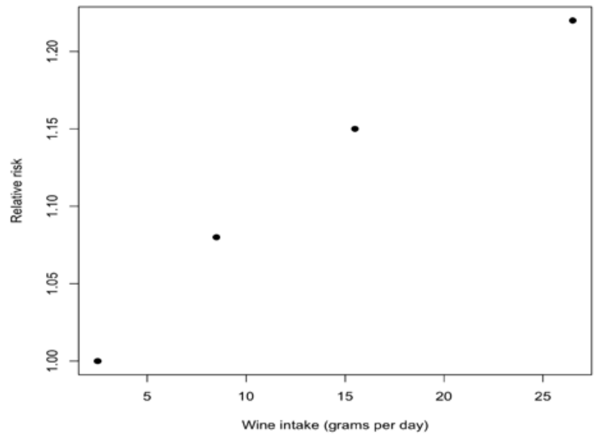
4.22: (b) Correlation is unaffected by units.

4.23: (b) Computation with calculator or software gives $r = 0.8900$.

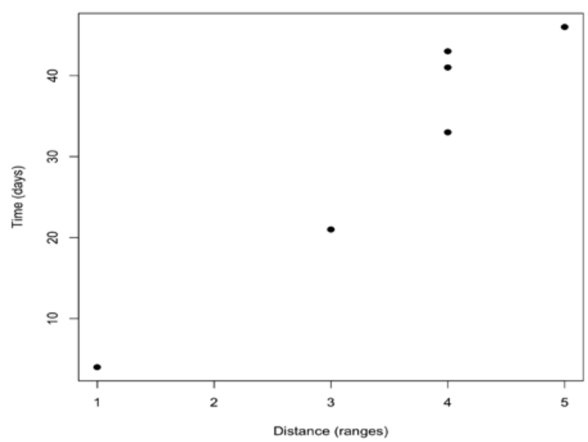
4.24: (a) The lowest first-round score was 66, scored by one golfer. This golfer scored 75 in the second round. (b) Lyle scored 86 in the second round, and 69 in the first round. (c) The correlation is very small, but positive… so closest to 0.1. Knowing a golfer's first-round score would not be useful in predicting a second-round score.

4.25: (a) Overall, there is a slightly negative association between these variables. (b) There is general disagreement — low BRFSS scores correspond to greater happiness, and these are associated with higher-ranked states (the least happy states, according to the objective measure). (c) It is hard to declare any of the data values as "outliers." It does not appear that any of the values are obviously outside of the general pattern. Perhaps one value (Rank = 8, BRFSS = 0.30) is an outlier, but this is hard to say.

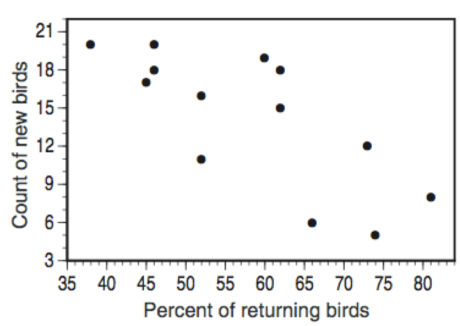
4.26: (a) The scatterplot reveals a very strong, positive linear relationship between wine intake and relative risk for cancer. We expect correlation to be close to +1. (b) Using software, $r = 0.9851$. The data suggest that women who consume more wine tend to have higher risk of breast cancer. However, this is an observational study, and no causal relationship can be determined. The women who drink more wine may differ in many respects from women who drink less wine.



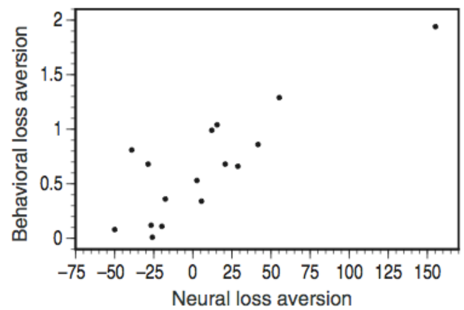
4.27: (a) The scatterplot suggests a strong positive linear association between distance and time with respect to the spread of Ebola. (b) $r = 0.9623$. This is consistent with the pattern described in (a). (c) Correlation would not change, since it does not depend on units.



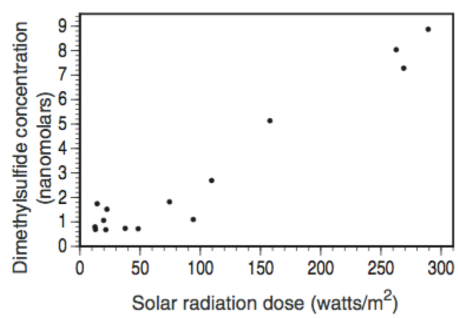
4.28: (a)The scatterplot shows a linear negative relationship. Because the relationship is linear, correlation is an appropriate measure of strength: $r = -0.7485$. (b)Because this association is negative, we conclude that the sparrow hawk is a long-lived territorial species.



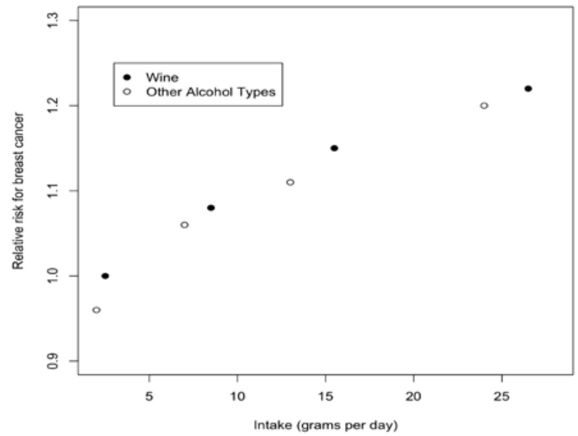
4.29: (a) The scatterplot is shown; note that neural activity is explanatory (and so should be on the horizontal axis). (b) The association is moderately strong, positive, and linear. The outlier is in the upper right corner. (c) For all points, $r = 0.8486$. Without the outlier, $r = 0.7015$. The correlation is greater with the outlier because it fits the pattern of the other points; if one drew the line suggested by the other points, the outlier would extend the length of the line and would therefore decrease the relative scatter of the points about that line.



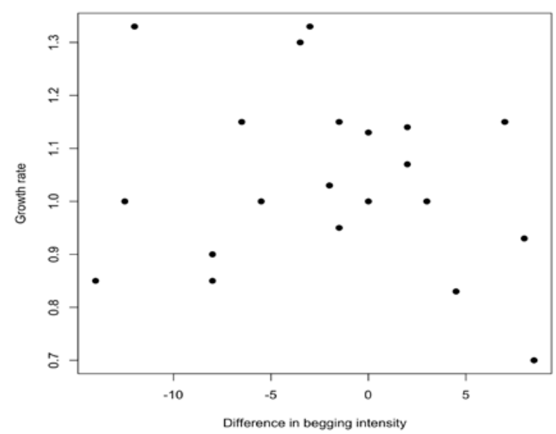
4.30: (a) SRD is the explanatory variable, so it should be on the horizontal axis. (b) The scatterplot shows a positive linear association. The correlation coefficient is $r = 0.9685$, which is consistent with the strength of the association visible in the scatterplot.



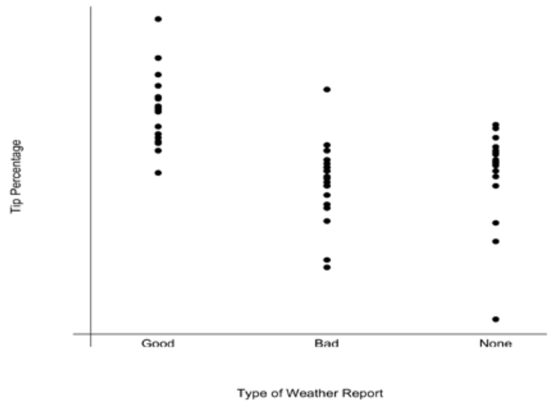
4.31: (a) The scatterplot is provided below. (b) The plot suggests that there is a strong relationship between alcohol intake and relative risk of breast cancer (again, this is an observational study, so no causal relationship is established here). It seems that type of alcohol has nothing to do with the increase since the same pattern and rate of increase is seen for both groups.



4.32: (a) The scatterplot is provided below. (b) The scatterplot suggests that there is not a linear relationship between relative growth rate and difference in begging intensity. Here, $r = -0.1749$. (c) Neither theory is strongly supported, but the latter is more strongly supported. That is, growth rate increases initially as begging intensity increases but then levels off or decreases as parents begin to ignore increases in begging by the foster babies.

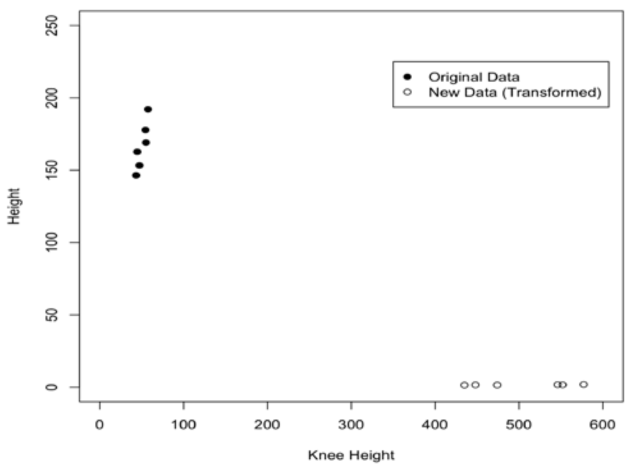


4.33: (a) A plot follows, and suggests that "Good" weather reports tend to yield higher tips. (b) The explanatory variable is categorical, not quantitative, so r cannot be used. Notice that we can arrange the categories any way, and these different arrangements would suggest different associations. Hence, it doesn't make sense to discuss a relationship direction here.



4.34: (a) Correlation would not change, as correlation does not depend on units. (b) Correlation would not change. By subtracting 0.25 from all risks, each point in the scatterplot moves "down" by 0.25, but the strength and direction of the linear relationship between risk and wine intake does not change. (c) There would be a perfect positive linear relationship with $r = +1$.

4.35: (a) The scatterplot is provided below. Changing the units has a dramatic impact on the plot. (b) Nevertheless, units do not impact correlation. For both data sets, $r = 0.8900$.



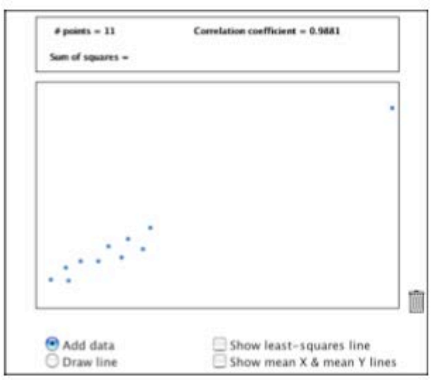
4.36: Explanations and sketches will vary, but should note that correlation measures the strength of the association, not the slope of the line. The hypothetical Funds A and B mentioned in the report, for example, might have a linear relationship having line of slope 2 or ½.

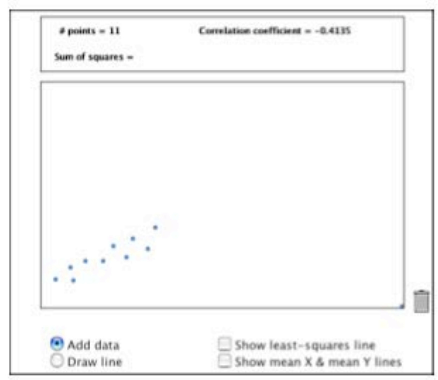
4.37: (a) Small-cap stocks have a lower correlation with municipal bonds, so the relationship is weaker. (b) She should look for a negative correlation (although this would also mean that this investment tends to *decrease* when bond prices rise).

4.38: The person that wrote the article interpreted a correlation close to 0 as if it were a correlation close to –1 (implying a negative association between teaching ability and research productivity). Professor McDaniel's findings mean there is little linear association between research and teaching ability. For example, knowing that a professor is a productive researcher gives little information about whether she is a good or bad teacher. Also, remember that correlation is only meaningful if both variables are quantitative — and here there is no guarantee that this is the case.

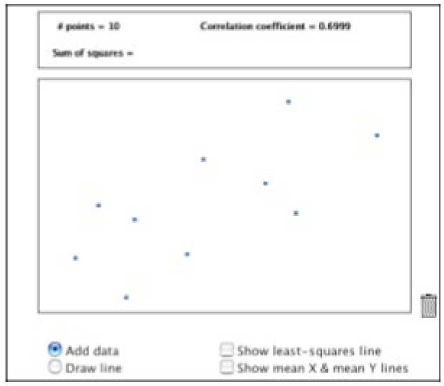
4.39: (a) Because gender has a nominal scale, we cannot compute the correlation between sex and any other variable. There is a strong *association* between sex and income. Some writers and speakers use "correlation" as a synonym for "association," but this is not correct. (b) A correlation of $r = 1.09$ is impossible, because r is restricted to be between –1 and 1. (c) Correlation has no units, so $r = 0.63$ centimeter is incorrect.

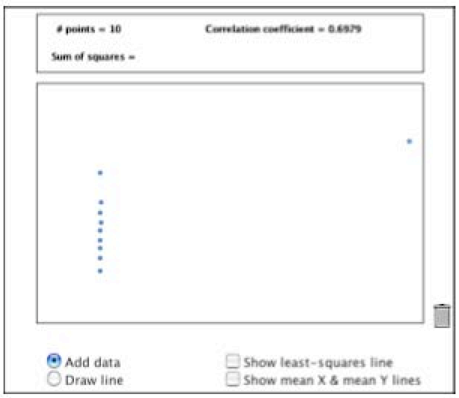
4.40: (a) The correlation will be closer to 1. One possible answer is shown. (b) Answers will vary, but the correlation will decrease, and can be made negative by dragging the point down far enough (see below, right).

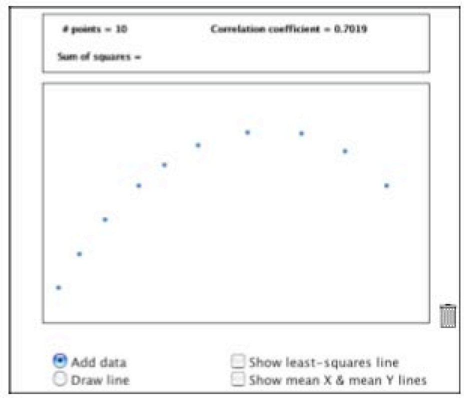




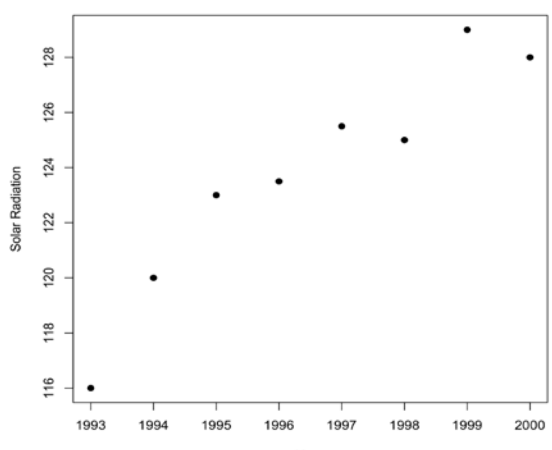
4.41: (a) Because two points determine a line, the correlation is always 1. (b) Sketches will vary; an example is shown. Note that the scatterplot must be positively sloped, but r is affected only by the scatter about the line, not by the steepness of the slope of that line. (c) The first nine points cannot be spread from the top to the bottom of the graph because in such a case the correlation cannot exceed about 0.66 (this is based on experience — lots of playing around with the applet). One possibility is shown. (d) To have $r = 0.7$, the curve must be higher at the right than at the left. One possibility is shown.



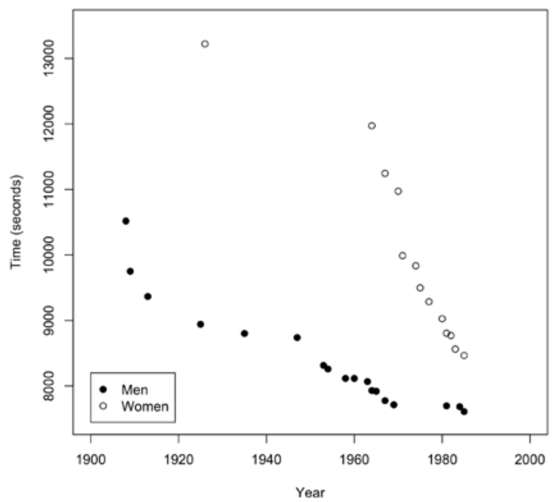




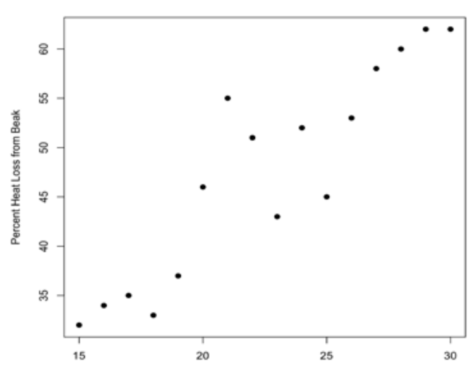
4.42: PLAN: To describe the change in solar radiation over time, we begin with a scatterplot (with year as the explanatory variable). If appropriate for the relationship, we compute the correlation coefficient to measure the strength of the association. SOLVE: The plot suggests that sunlight has brightened overall, and the increase has been relatively steady. Correlation is a useful measure here, and r = 0.9454. CONCLUDE: Over time, sunlight has gotten brighter.



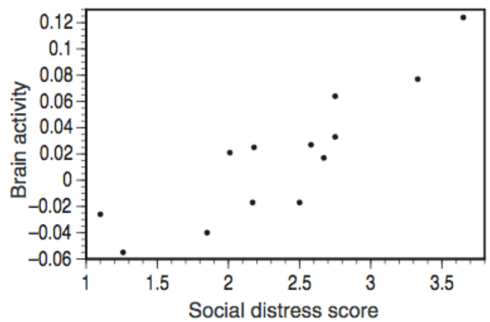
4.43: PLAN: To study the improvements in running times between men and women, we'll plot the data on the same scatterplot. We will not use correlation, but we will examine the plot to see if women are beginning to outrun men. SOLVE: The plot is provided below. By inspection, one might guess that the "lines" that fit these data sets will meet around 1998. This is how the researchers made this leap. CONCLUDE: Men's and women's times have, indeed, grown closer over time. Both sexes have improved their record marathon times over the years, but women's times have improved at a faster rate. In fact, as of 2011, the world record time for men has continued to be faster than the world record time for women. The difference is currently about 686 seconds (under 12 minutes), where in the data plotted, the difference was about 856 seconds.



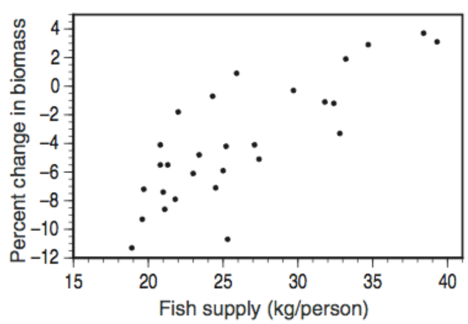
4.44: PLAN: To investigate the relationship between outside temperature and the percent of total heat loss due to beak, we plot heat loss from beak against outside temperature. We'll compute the correlation, if the relationship looks to be reasonably linear. SOLVE: The plot follows. Notice that there is a reasonably strong linear relationship. It seems reasonable to use correlation to describe this relationship's strength and direction. In fact, $r = 0.9143$. CONCLUDE: When the outside temperature increases, a greater percentage of total heat loss is due to beak heat loss. That is, the beak plays a more important role in cooling down the toco toucan as the weather outside becomes hotter.



4.45: PLAN: We wish to explore the relationship between social distress and brain activity. We begin with a scatterplot, and compute the correlation if appropriate. SOLVE: A scatterplot shows a fairly strong, positive, linear association. There are no particular outliers; each variable has low and high values, but those points do not deviate from the pattern of the rest. The relationship seems to be reasonably linear, so we compute $r = 0.8782$. CONCLUDE: Social exclusion does appear to trigger a pain response: higher social distress measurements are associated with increased activity in the pain-sensing area of the brain. However, no cause-and-effect conclusion is possible since this was not a designed experiment.



4.46: PLAN: We wish to explore the relationship between fish supply and animal population. We begin with a scatterplot, and compute the correlation if appropriate. SOLVE: A scatterplot shows a moderately strong, positive, linear association. There are no clear outliers, although a few points fall slightly above (and one slightly below) the cluster. Correlation ($r = 0.8042$) is an appropriate measure of the strength of the association. CONCLUDE: The positive association supports the idea that animal populations decline when the fish supply is low. The four years with the greatest fish supply were four of the five years in which biomass increased.



4.47 and 4.48 are Web-based exercises.