# Chapter 4 Solutions

**4.1. (a)** Time spent studying is explanatory; the grade is the response variable. **(b)** Explore the relationship; there is no reason to view one or the other as explanatory. **(c)** Time spent on extracurricular activities is explanatory, GPA is the response variable. **(d)** Explore the relationship.

**4.2.** Water temperature is explanatory, and weight change (growth) is the response variable. Both are quantitative.

**4.3.** For example: Weight, gender, 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.
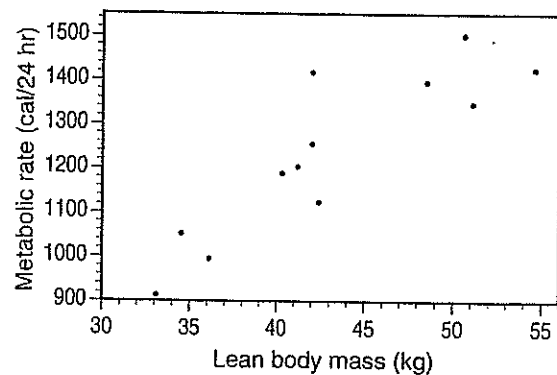


**4.5.** Outsourcing is viewed as explanatory, so it belongs on the horizontal axis. Delays are generally higher for heavily-outsourced airlines.
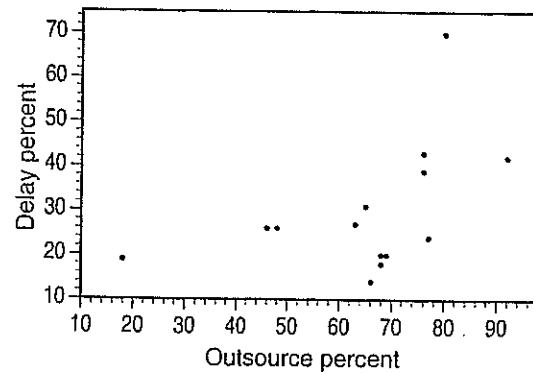


**4.6.** The scatterplot shows a positive, linear, moderately strong association.

   **Note:** *The question asks for direction, form, and strength; the answers are "positive, linear, moderately strong" (respectively).*

**4.7.** The scatterplot shows a positive association. The outlier, in the upper right corner, is Hawaiian Airlines. Aside from that point, the relationship is roughly linear, but only moderately strong.

**4.8.** (a) At right; 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.
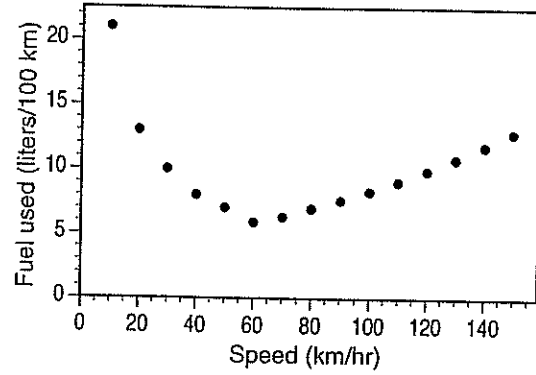


**4.9.** (a) Women are marked with filled circles, men with open circles. (b) For both men and women, the association is linear and positive. The women's points show a stronger association. As a group, males typically have larger values for both variables.
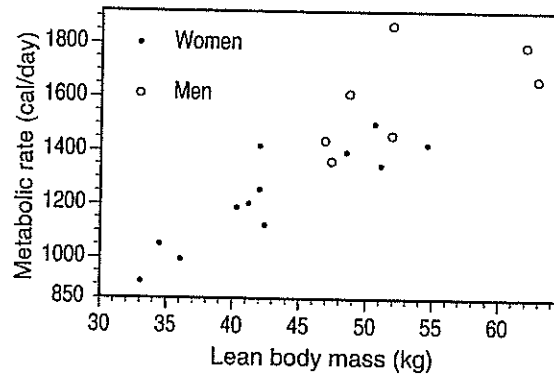


**4.10.** (a) Distance is explanatory (and so is on the horizontal axis). (b) $\bar{x} = 3.5$ and $s_x \doteq 1.3784$ home ranges; $\bar{y} = 31.\overline{3}$ and $s_y \doteq 16.1328$ days. The standardized values are below; the correlation is $r = \frac{4.8117}{5} \doteq 0.9623$. (c) Obviously, the software value should be the same.

| $z_x$ | $z_y$ | $z_x z_y$ |
|---|---|---|
| −1.8137 | −1.6943 | 3.0729 |
| −0.3627 | −0.6405 | 0.2323 |
| 0.3627 | 0.1033 | 0.0375 |
| 0.3627 | 0.5992 | 0.2174 |
| 0.3627 | 0.7232 | 0.2623 |
| 1.0882 | 0.9091 | 0.9893 |
| | | 4.8117 |
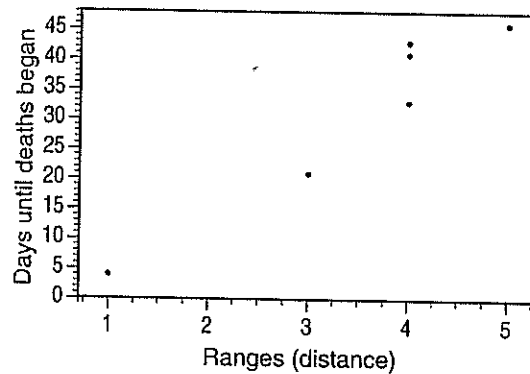


**4.11.** $r$ would not change; units do not affect correlation.

**4.12. (a)** $r \doteq 0.8765$. **(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.
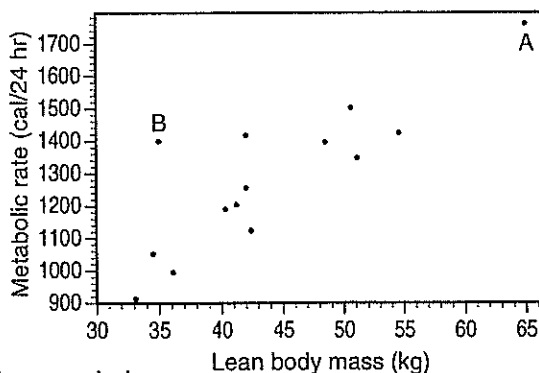


**4.13.** In computing the correlation, note $\bar{x} = 40$ and $s_x \doteq 15.8114$ mph, while $\bar{y} = 26.8$ and $s_y \doteq 2.6833$ mpg. The details of the computation are below; as an alternative to finding standard scores, note that the denominators are all the same ($s_x \, s_y$), and the sum of the numerators of the terms is

$$\sum(\bar{x} - x_i)(\bar{y} - y_i) = (-20)(-2.8) + (-10)(1.2) + (0)(3.2) + (10)(1.2) + (20)(-2.8)$$
$$= 56 - 12 + 0 + 12 - 56 = 0.$$

The correlation is 0 because these variables do not have a straight-line relationship; the association is neither positive nor negative.

| $z_x$ | $z_y$ | $z_x \, z_y$ |
|---|---|---|
| $-1.2649$ | $-1.0435$ | $1.3199$ |
| $-0.6325$ | $0.4472$ | $-0.2828$ |
| $0$ | $1.1926$ | $0$ |
| $0.6325$ | $0.4472$ | $0.2828$ |
| $1.2649$ | $-1.0435$ | $-1.3199$ |
| | | $0$ |
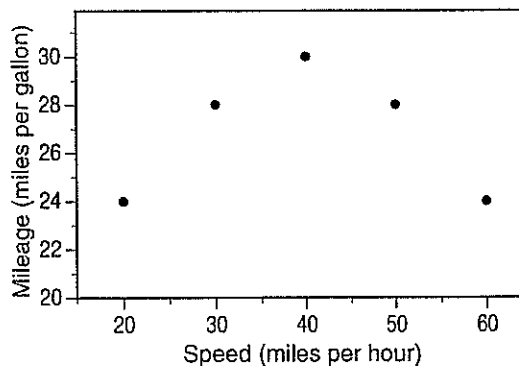


**4.14. (a)** We would expect that the price of a barrel of oil has an effect on the price of gasoline, rather than the reverse.

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

**4.16. (c)** The low point is for an IQ of about 124 and a reading score around 10.

**4.17. (a)** There is a positive association, but not so strong that the correlation would be 0.95.

**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 an example.)

**4.20. (b)** The correlation is negative because the slope of the line is negative.

**4.21. (a)** This would be a perfect linear relationship: if the husband's age is $y$ and the wife's age is $x$, the linear relationship $y = x + 2$ would hold, and hence $r = 1$ (because the slope is positive).

**4.22. (b)** Correlation is unaffected by units.

**4.23. (a)** Computation with calculator or software gives $r \doteq 0.8770$.

**4.24. (a)** Two golfers shot a 69 in the first round. In the second round, they shot a 73 and a 75. **(b)** Villegas shot 85 in the second round, after shooting 80 in the first round. **(c)** The correlation is closest to 0.2; there is some evidence of a positive association, but it is not strong enough for making firm predictions. (In Exercise 5.5, the correlation is given as 0.192.)

**4.25. (a)** The association is (weakly) positive. **(b)** The estimate is 4, which is an overestimate; that child had the lowest score on the test.
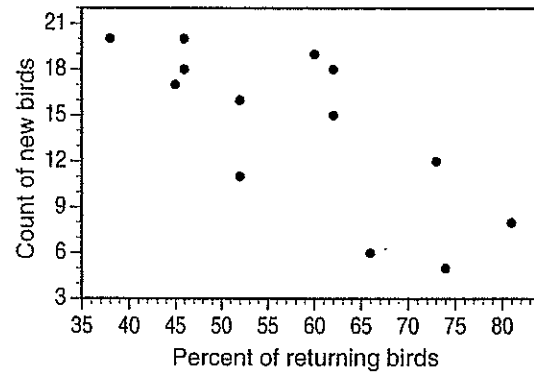
**4.26. (a)** The scatterplot shows a moderate positive association, so $r$ should be positive, but not close to 1. **(b)** The correlation is $r = 0.5653$. There is some evidence that taller women tend to date taller men (and shorter women date shorter men), but it is hardly overwhelming—and the small sample size makes any conclusion suspect.
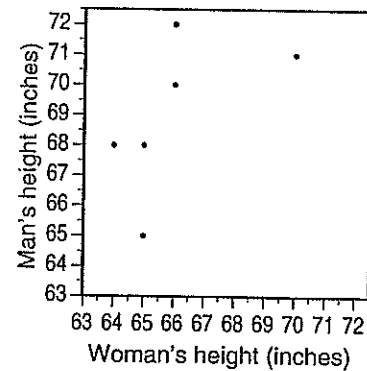
**4.27. (a)** Price is explanatory (and so is on the horizontal axis). The plot shows a positive linear association. **(b)** The correlation is $r \doteq 0.9552$. This supports a connection between coffee price and loss of forest. **(c)** $r$ would not change; units do not affect correlation.
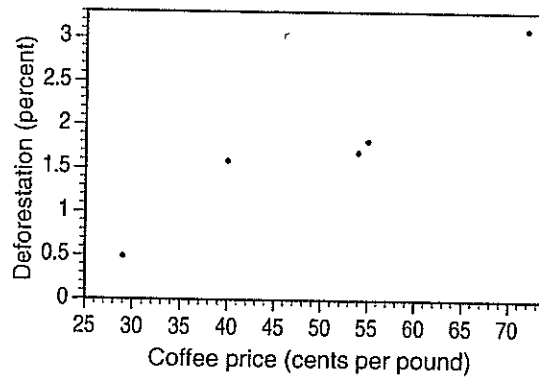
**4.28.** (a) The scatterplot shows a linear negative relationship. Because the relationship is linear, correlation is an appropriate measure of strength: $r \doteq -0.7485$. (b) Because this association is negative, we conclude that the sparrowhawk is a long-lived territorial species.
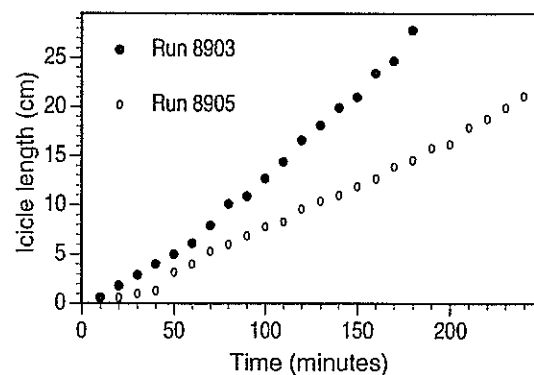
**4.29.** (a) The scatterplot is shown on the right; note that neural activity is explanatory (and so should be on the horizontal axis). (b) The association is moderately strong, positive, and linear. The outlier is in the upper right corner. (c) For all points, $r \doteq 0.8486$. Without the outlier, $r \doteq 0.7015$. 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.
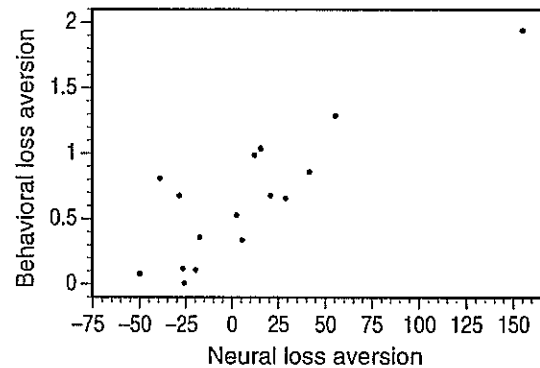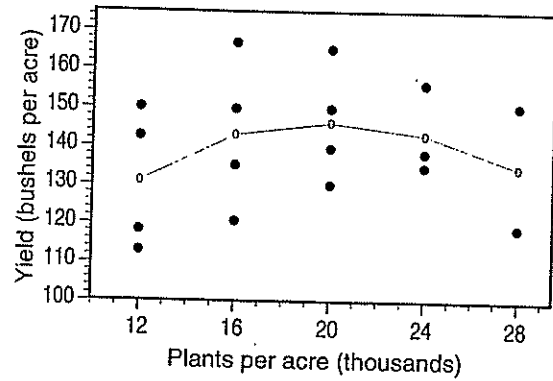
**4.30.** (a) SRD is the explanatory variable, so it should be on the horizontal axis. (b) The scatterplot shows a positive linear association. The correlation coefficient is $r \doteq 0.9685$, which is consistent with the strength of the association visible in the scatterplot.
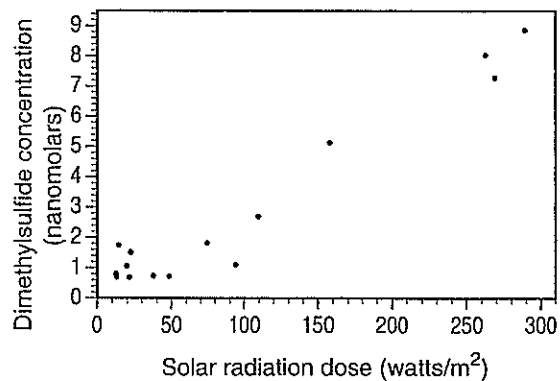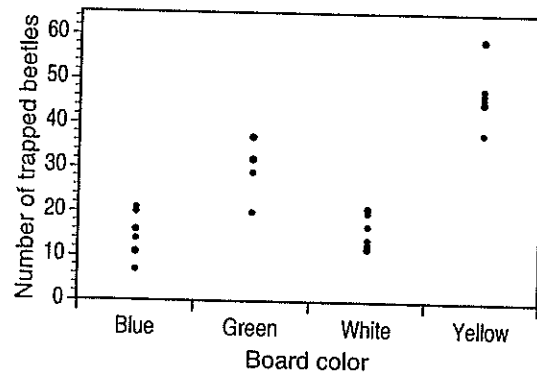
**4.31.** (a) In the scatterplot on the right, the open circles represent Run 8905, the higher flow rate. (b) Icicles seem to grow faster when the water runs more slowly. (Note that there is no guarantee that the pattern we observe with these two flow rates applies to rates much faster than 29.6 mg/s, or slower than 11.9 mg/s.)

**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 who 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—for example, knowing that a professor is a good researcher gives little information about whether she is a good or bad teacher.

**4.39.** (a) Because gender has a nominal scale, we cannot compute the correlation between sex and anything. (There is a strong *association* between gender and income. Some writers and speakers use "correlation" as a synonym for "association." It is much better to retain the more specific meaning.) (b) A correlation $r = 1.09$ is impossible because $-1 \le r \le 1$ always. (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 below, left. (b) Answers will vary, but the correlation will decrease, and can be made negative by dragging the point down far enough (see below, right).
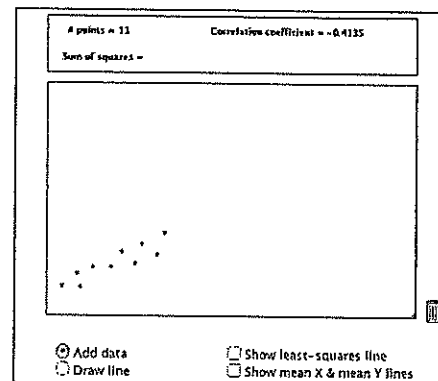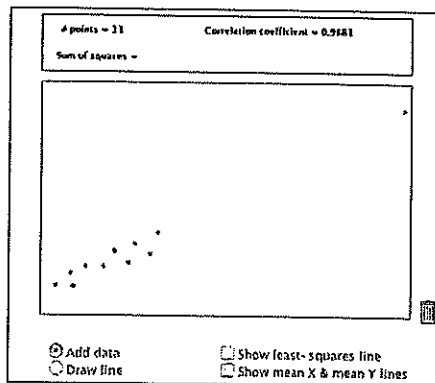
**4.41.** (a) Because two points determine a line, the correlation is always 1. (b) Sketches will vary; an example is shown on the right. Note that the scatterplot must be positively sloped, but $r$ is affected only by the scatter about the line, not by the steepness of the slope. (c) The first nine 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 (based on empirical evidence—that is, from a reasonable amount of playing around with the applet). One possibility is shown below, left. (d) To have $r \doteq 0.7$, the curve must be higher at the right than at the left. One possibility is shown below, right.
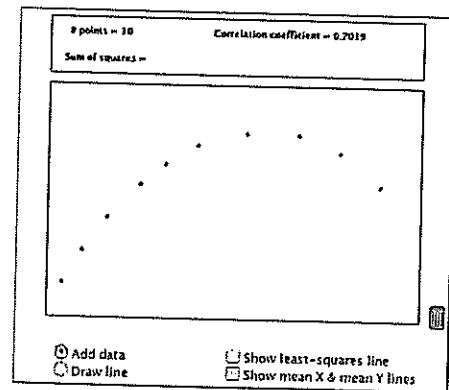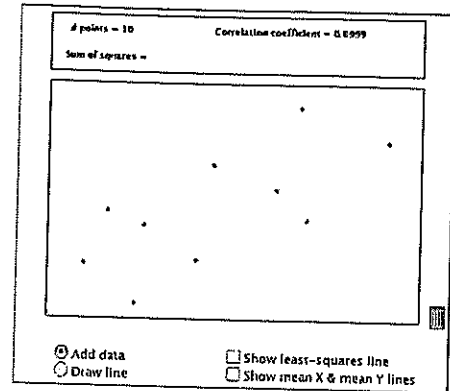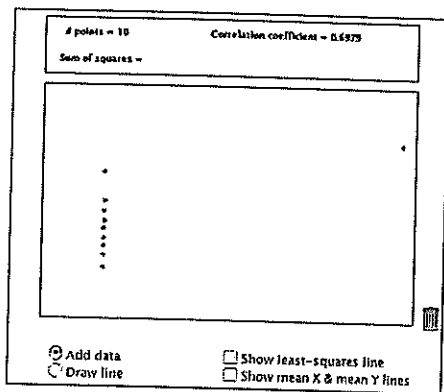
**4.42.** PLAN: To describe the change in solar radiation over time, we begin with a scatterplot (with time as the explanatory variable). If appropriate for the relationship, we compute the correlation coefficient to measure the strength of the association.

SOLVE: The plot suggests that sunlight has brightened overall, but the increase has not been steady—from 1992 through 1999, there seems to be a curved relationship. Although it might not be the best summary of this (not entirely linear) relationship, the correlation is $r \doteq 0.773$.

CONCLUDE: Over time, sunlight has gotten brighter, but with quite a bit of fluctuation.
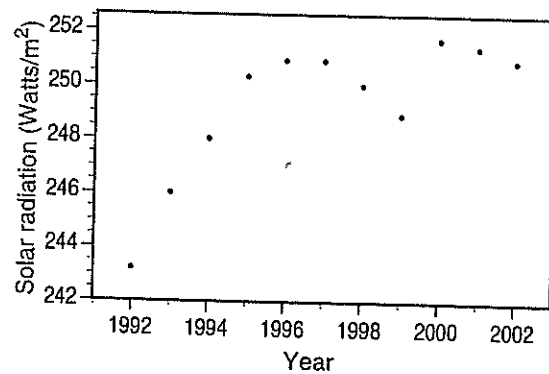
**4.43.** PLAN: To see the effect of the solar panels, we begin with a scatterplot, with degree-days as the explanatory variable. SOLVE: The new data in the scatterplot are plotted with open circles. Both sets of points show a strong positive linear association between degree-days and gas usage. The new points are generally slightly lower than the pre-solar-panel points. CONCLUDE: There is some visual evidence that gas usage dropped slightly after the solar panels were installed.



**Note:** *The before and after correlation coefficients are 0.9953 and 0.9912, but those numbers are not particularly useful for this comparison. We are not interested in how the strength of the relationship changed; we want to judge if there was a shift in the nature of the relationship. Specifically, we note that for any given value of degree-days per day, gas usage appears to be slightly lower with solar panels. Using linear regression, we can make more specific statements about this change; see the solution to Exercise 5.55.*
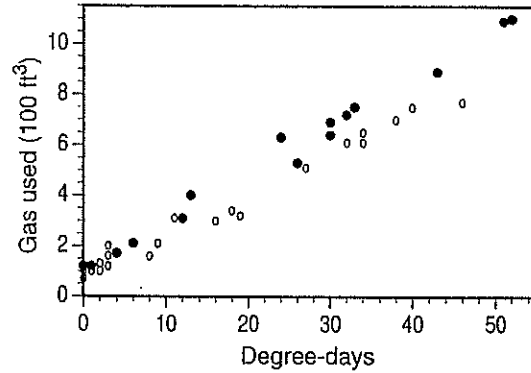
**4.44.** PLAN: We wish to observe the effect of the number of breeding pairs on the survival rates (the percent of males returning). We begin with a scatterplot, and compute the correlation if appropriate. SOLVE: A scatterplot of the percent returning against the number of breeding pairs shows the expected negative association. Though slightly curved, it is reasonable to compute $r \doteq -0.7943$ as a measure of the strength of the association.



CONCLUDE: This supports the theory: a smaller percent of birds survive following a successful breeding season.

**Note:** *We examine the percent returning, rather than the number returning, because with the latter response variable, we might not see the relationship: areas with many breeding pairs would correspondingly have more males that might potentially return. (In the given numbers, the number of breeding pairs varies only from 28 to 38, but considering hypothetical data with 10 and 100 breeding pairs makes more apparent the reason for using percents rather than counts.)*
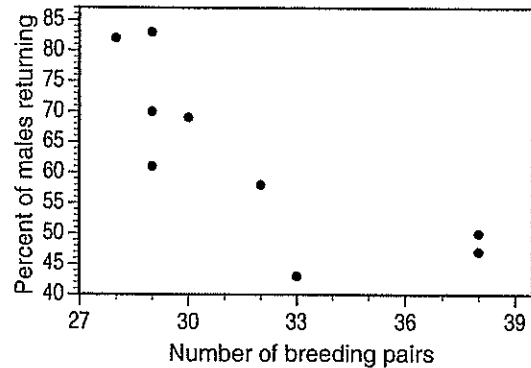
**4.45.** PLAN: We wish to explore the relationship between social distress and brain activity. We begin with a scatterplot, and compute the correlation if appropriate.
SOLVE: A scatterplot shows a fairly strong, positive, linear association. There are no particular outliers; each variable has low and high values, but those points do not deviate from the pattern of the rest. The relationship seems to be reasonably linear, so we compute $r \doteq 0.8782$.
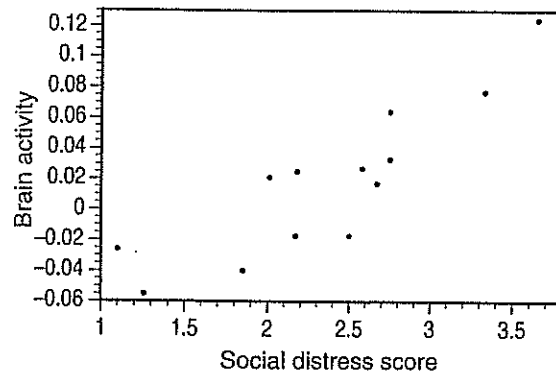
CONCLUDE: Social exclusion does appear to trigger a pain response: higher social distress measurements are associated with increased activity in the pain-sensing area of the brain.

**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 ($r \doteq 0.8042$) is an appropriate measure of the strength of the association.
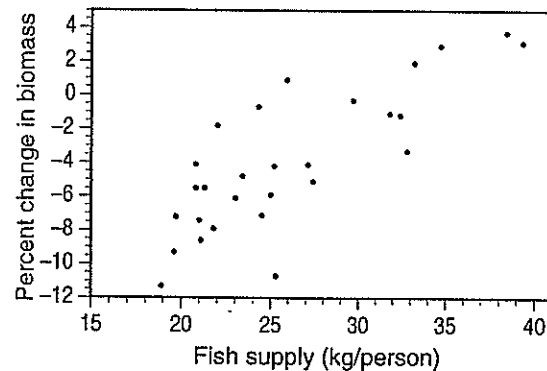
CONCLUDE: The positive association supports the idea that animal populations decline when the fish supply is low. The four years with the greatest fish supply were four of the five years in which biomass increased.