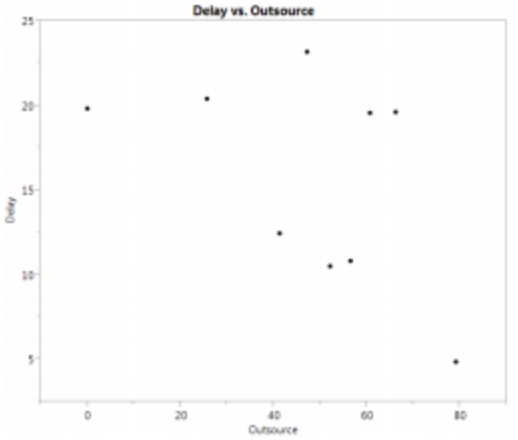
Chapter 4 – Scatterplots and Correlation

4.1 (a) Explanatory: number of lectures attended; Response: grade on final exam. **(b)** Explanatory: time exercising; Response: calories burned. **(c)** Explanatory: time spent online using Facebook; Response: GPA (assuming that more time on Facebook means less time studying). **(d)** Explore the relationship.

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

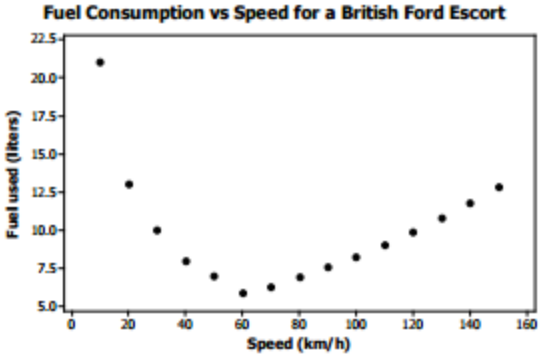
4.5 Outsource percent is the explanatory variable and should be on the horizontal axis. Delay percent is the response and should be on the vertical axis. These data do not support concerns of the critics.



4.7 One could consider there to be two outliers: Frontier has an unusually low outsourcing percent and a high delay percent; Hawaiian has a very high outsourcing percent and a very low delay percent. Without Frontier, there would be a decreasing relationship that is approximately linear and moderately strong. Without Hawaiian, there is really no relationship, but two sets of points: five airlines with high delays,

which don't seem to depend on outsourcing, and three airlines with low delay percentages that, again, don't seem to depend on outsourcing.

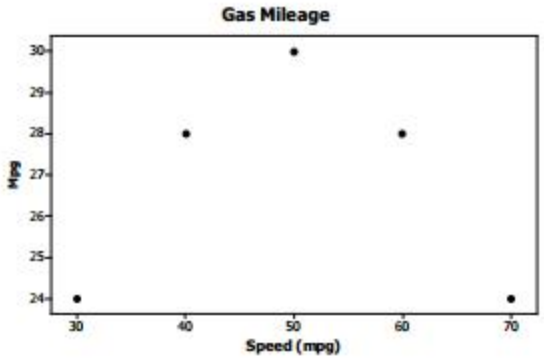
4.8 (a) Speed is explanatory. **(b)** The relationship is curved—low in the middle, higher at the extremes. Because low "mileage" is actually *good* (it means that we use less fuel to travel 100 km), this makes sense: Moderate speeds yield the best performance. Note that 60 km/h is about 37 mph. **(c)** 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—little scatter occurs around the curve, so the curve is very useful for prediction.



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

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	1.19269	0
0.6325	0.4472	0.2828
1.2649	−1.0435	−1.3199
		0



4.14 (a) The researchers wanted to know if reaction time could predict time to death.

4.15 (c) The association should be negative (e.g., if slower reaction times mean less time to death).

4.16 (b) IQ = 103, GPA = 0.5

4.17 (a) 0.9; without the outlier, a strong positive linear relationship exists.

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.

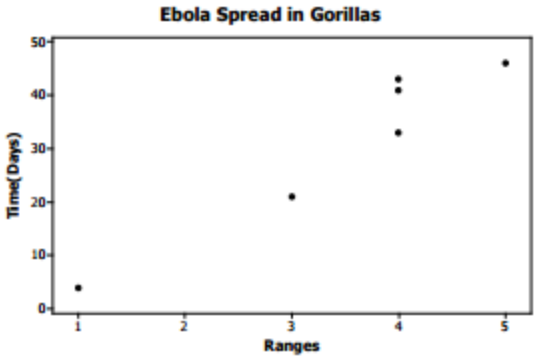
4.20 (c) Because we are not told how the *x* and *y* values vary together, we cannot tell whether the correlation will be **–1** or +1.

4.21 (a) 1. There would be a perfect, positive linear association. The line would be Exam2 = Exam1 – 10.

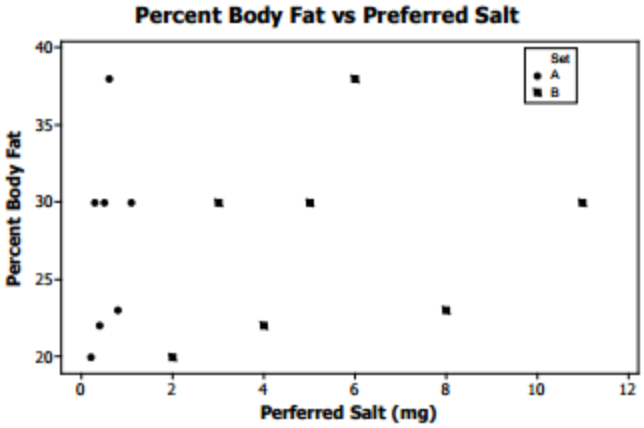
4.22 (b) Correlation is unaffected by units.

4.23 (b) Computation with a calculator or software gives $r = 0.298$.

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, because it does not depend on units.



4.35 (a) The scatterplot is provided at right. Set B (the mad scientist's set) has stretched out the x values, but the pattern is still the same.
(b) Units do not impact correlation. For both data sets, $r = 0.298$.



4.39 (a) Because sex 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 $r = 0.63$ cm is incorrect.