2.14.  (a) Below; speed is explanatory, so it belongs on the \( x \)-axis. (b) The relationship is curved—low in the middle, higher at the extremes. Because low “mileage” is actually \textit{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. (Milage decreases until speed exceeds 60 km/h and then increases beyond that. 60 km/h is probably where the car shifts into high gear, which is more efficient. As speeds increase, wind resistance becomes more of a factor, causing fuel consumption to increase.) (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, and it is very useful for prediction. It would appear that two lines (one for below 60 km/h and one above) would fit the data quite well.

2.32.  See the solution to Exercise 2.14 for the scatterplot. \( r = -0.172 \)—it is close to zero, because the relationship is a curve rather than a line; correlation measures the degree of \textit{linear} association.

2.33.  (a) The Insight seems to fit the line suggested by the other points. (b) Without the Insight, \( r = 0.9757 \); with it, \( r^* = 0.9934 \). The Insight increases the strength of the association (the line is the same, but the scatter about that line is \textit{relatively} less when the Insight is included).
2.29.  (a) The scatterplot shows a moderate positive association, so $r$ should be positive, but not close to 1.  (b) The correlation is $r = 0.5653$.  (c) $r$ would not change if all the men were six inches shorter. A positive correlation does not tell us that the men were generally taller than the women; instead it indicates that women who are taller (shorter) than the average woman tend to date men who are also taller (shorter) than the average man.  (d) $r$ would not change, because it is unaffected by units.  (e) $r$ would be 1, as the points of the scatterplot would fall on a positively-sloped line.
2.46. (a) The least-squares line is \( \hat{y} = 0.7267x + 4.9433 \). This is less steep than the line \( y = x \), reflecting the observation that field measurements tend to be lower for greater depths. (b) The line \( y = x \) has slope 1; the regression line has slope 0.7267. A slope of 1 would mean that for every additional unit of depth as measured in the laboratory, the field measurement would also increase by one unit. The slope of 0.7267 means that on the average, the field measurement increases by only 0.7267 units for every one unit in the lab.

2.55. Women’s heights are the \( x \) values; men’s are the \( y \) values. The slope is \( b = (0.5)(2.7)/2.5 = 0.54 \) and the intercept is \( a = 68.5 - (0.54)(64.5) = 33.67 \).

The regression equation is \( \hat{y} = 33.67 + 0.54x \). Ideally, the scales should be the same on both axes. For a 67-inch tall wife, we predict the husband’s height will be about 69.85 inches.
Calculus Supplement Problems:

2.1 Let's show that (S2.5) and (S2.6) are equal.

(S2.5) \[ b = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} \]

(S2.6) \[ b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \]

Look at the numerator of (S2.6). Write \( \sum (x_i - \bar{x})(y_i - \bar{y}) \) as \( \sum (x_i - \bar{x})y_i - \bar{y} \sum (x_i - \bar{x}) \) 

Since \( \sum (x_i - \bar{x}) = 0 \), the second term is zero. Write the first term as:

\[ \sum x_i y_i - \bar{x} \sum y_i = \sum x_i y_i - \bar{x} n \bar{y} \]

showing that the numerators are equal.

Now look at the denominator of (S2.6), which we write as:

\[ \sum (x_i - \bar{x})(x_i - \bar{x}) = \sum (x_i - \bar{x})x_i - \bar{x} \sum (x_i - \bar{x}) \]

and as above the second term is zero and the first term is

\[ \sum x_i^2 - \bar{x} \sum x_i = \sum x_i^2 - n \bar{x}^2 \]

showing that the denominators are the same.

Now, to check the formula on p. 141: \[ b = r \frac{s_y}{s_x} \]

Re-write the definition of r on p. 127 as:

\[ r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \]

Putting this into the formula for b, we get:

\[ b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x^2} \]

Using the definition of \( s_x \) on p. 48, we see that \( (n-1)s_x^2 = \sum (x_i - \bar{x})^2 \)

Thus, the equation for b on p. 141 is identical to (S2.6).

2.3. (a) It seems plausible to say that something that is on the sun’s surface (r = 0) would have no velocity either towards or away from the sun (i.e. \( v = 0 \)). This does not mean that the data will agree with this supposition.

The minitab output is:

```
Predictor Coef SE Coef T P
Noconstant 423.94 42.15 10.06 0.000
```

The regression equation is

\[ v = 424 r \]
(b) The minitab output fitting the full line with an intercept is:

\[
\text{The regression equation is} \\
v = -40.8 + 454 \; r
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-40.78</td>
<td>83.44</td>
<td>-0.49</td>
<td>0.630</td>
</tr>
<tr>
<td>r</td>
<td>454.16</td>
<td>75.24</td>
<td>6.04</td>
<td>0.000</td>
</tr>
</tbody>
</table>

You can see that the slope is somewhat different, but not a great deal, and that the best fitting line has a negative intercept which may seem illogical, but just gives a better description of the data.

Superimposed on the scatterplot, the two lines do not look very different:

![Scatterplot of v vs r](image)

The value of \( r^2 \) for the full model is 62.4%. An equivalent measure for the restricted line is 61.9%.