

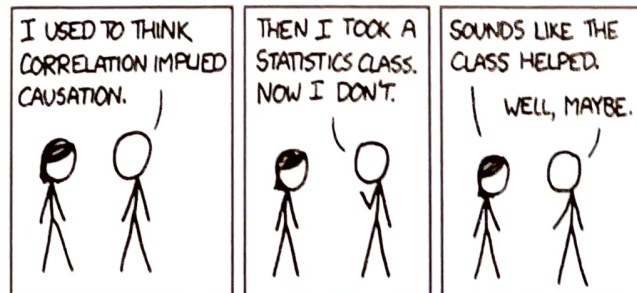
Once again we've learned that doing Statistics right means we have to *Think* about whether our choice of methods is appropriate.

- ◆ The correlation coefficient is appropriate only if the underlying relationship is linear.
- ◆ We'll check the Straight Enough Condition by looking at a scatterplot.
- ◆ And, as always, we'll watch out for outliers!

Finally, we've learned not to make the mistake of assuming that a high correlation or strong association is evidence of a cause-and-effect relationship. Beware of lurking variables!

TERMS

Scatterplots	A scatterplot shows the relationship between two quantitative variables measured on the same cases. In a timeplot the horizontal axis is time. (p. 148)
Association	<ul style="list-style-type: none"> ◆ Direction: A positive direction or association means that, in general, as one variable increases, so does the other. When increases in one variable generally correspond to decreases in the other, the association is negative. ◆ Form: The form we care about most is linear (straight), but you should certainly describe other patterns you see in scatterplots. ◆ Strength: A scatterplot is said to show a strong association if there is little scatter around the underlying relationship. (p. 149)
Outlier	A point that does not fit the overall pattern seen in the scatterplot. (p. 149)
Response variable, Explanatory variable, x-variable, y-variable	In a scatterplot, you must choose a role for each variable. Assign to the y-axis the response variable that you hope to predict or explain. Assign to the x-axis the explanatory or predictor variable that accounts for, explains, predicts, or is otherwise associated with the y-variable. (p. 150)
Correlation coefficient	<p>The correlation coefficient is a numerical measure of the direction and strength of a linear association. (p. 153)</p> $r = \frac{\sum z_x z_y}{n - 1}$
Lurking variable	A variable other than x and y that simultaneously affects both variables, accounting for the association between the two. (p. 160)



© 2013 Randall Munroe. Reprinted with permission. All rights reserved.

ON THE COMPUTER

Scatterplots and Correlation

Statistics packages generally make it easy to look at a scatterplot to check whether the correlation is appropriate. Some packages make this easier than others.

Many packages allow you to modify or enhance a scatterplot, altering the axis labels, the axis numbering, the plot

symbols, or the colors used. Some options, such as color and symbol choice, can be used to display additional information on the scatterplot.