Welcome to Week 5!
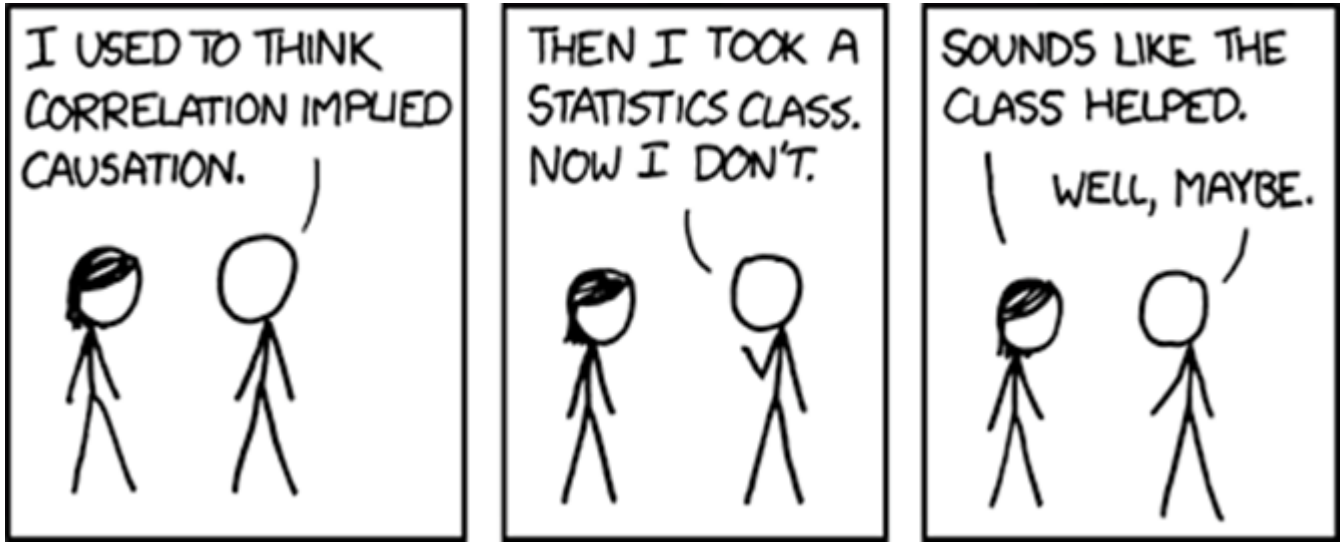
Welcome to Week 5!

One of my favourite comics in statistics is the one below. See if you can work out the message in it:

I USED TO THINK CORRELATION IMPLIED CAUSATION.

THEN I TOOK A STATISTICS CLASS. NOW I DON'T.

SOUNDS LIKE THE CLASS HELPED.

WELL, MAYBE.

Did you figure it out?

Often it is easy to think that correlation implies causation that if two variables have a relationship with each other then one variable must have caused the other. In practice, this is not true, even though there is a correlation between two variables this does not automatically imply that there is causation. To fully understand this comic (and of course, the concept) let us first define correlation and causation individually:

From the Australian Bureau of Statistics:

Correlation is a statistical measure (expressed as a number) that describes the size and direction of a relationship between two or more variables. A correlation between variables, however, does not automatically mean that the change in one variable is the cause of the change in the values of the other variable.

Causation indicates that one event is the result of the occurrence of the other event; i.e. there is a causal relationship between the two events. This is also referred to as cause and effect.

Going back to the comics, we see now that the male character (called Cueball) who took the class has perhaps finally understood the difference between correlation and causation. He used to think that correlation implied causation but after taking the statistics class he doesn't think this anymore. The female character (called Megan) points out the possibility that taking the class has caused Cueball to understand this underlying concept of correlation vs causation however, his final comment "well, maybe" implies that perhaps there was only a correlation. Remember in statistics, we usually talk about likelihoods so nothing is absolute. Welcome to doing research!

So why all this talk about correlation and causation? Well this week we are going to look at two statistical tests: Pearson's correlation and Simple Linear Regression. Pearson's correlation is a test that looks at the relationships between two continuous variables and Simple Linear Regression looks at how a change in one variable is influenced (caused) by a change in the other variable. Let's get to it!