

This tutorial is going to deal with explanatory variables and response variables. Now, we've actually talked about these terms before, but not in the context of a scatter plot. So let's take a look. When examining the relationship between two variables, often we want to see if there's an effect that one has on the other.

So does one variable being high or low help to explain why another variable would be high or low? Why would something being high or low cause one to increase or decrease? And it doesn't have to cause the increase or decrease, it just has to be associated with an increase or decrease in the other.

And so we're going to call the explanatory variable the variable whose increase or decrease we believe helps to explain the increase or decrease in some other variable. The variable that rises or falls in response to a rise or fall in the explanatory variable is called the response variable.

So let's look at an example here. The relationship between the number of firefighters at a fire and the damage caused by that fire in dollars-- so maybe after the insurance claims are filed. So which one helps to explain the other?

By the way, there's a positive association between these two. As one goes up, the other goes up. So which one helps to explain the other? Well, it's the damage that helps to explain the number of firefighters. So a more severe fire will cause more firefighters to arrive on-scene.

Obviously, it's not going to work the other way, where the more firefighters you have, that's going to cause more damage. They are associated, though, with each other. Because the severity of the fire is going to cause more damage, it's also going to cause more firefighters to arrive on-scene.

When we graph it, we're going to put the explanatory variable on the x-axis. And so this is a little mnemonic device-- "explanatory" has an "x" in it. So it's the x-axis, the horizontal axis. The response variable, on the other hand, goes on the y-axis. So how much damage was caused by the fire? And how many firefighters were on-scene to respond to that fire?

Occasionally, there's not a clear explanatory variable. So suppose we have something like kidney cancer rates and lung cancer rates for the 50 states in the United States. So each dot corresponds to a state. And we don't think that one causes the other. We don't really even think that an increase in one corresponds to an increase or decrease in the other. They don't seem to be all that related.

And so when we graph them, it really doesn't matter which one we talk about being the explanatory or response variable. They can be graphed either way. Only if there's some obvious choice for an explanatory or response

variable do we make a huge deal out of which one goes on the x-axis. In situations where there is no clear explanatory variable, more investigation would be required-- for instance, here-- to see what actually does cause kidney cancer.

And so to recap, in a scatter plot, one variable helps to explain an increase or decrease in another. And we call that the explanatory variable. And it's on the x-axis. The variable that appears to increase or decrease due to the increase or decrease in the explanatory variable is called the response variable. And we place it on the y-axis.

In times where it's not clear whether one is associated with an increase or decrease in the other at all, or we don't believe that one causes the other, there's no real association there, it doesn't really matter which one we call the explanatory or response. And so we talked about explanatory variables-- that's the x-axis variable-- and the response variable, the one that we believe increases or decreases in response to the explanatory. Good luck. And we'll see you next time.