Hi. This tutorial covers what's called the best-fit line, also known as a regression line. So let's start with some data. This data came from-- I believe it's 19 countries. And what it's measuring is the alcohol consumption of-- this is an average for the residents of each country, and it's the number of liters per year.

So if we look at Australia, on average, they would consume 2.5 liters per year. And then there's also the heart disease death rate per 100,000 people for each country. So again, for Australia, their typical heart disease death rate is about 211 people per every 100,000 people. And we can see that there's a pair for each country.

OK, so let's take a look at the graph of this, the scatter plot. And notice, alcohol consumption is on our x-axis is what we're considering as the explanatory variable. And the heart disease rate is the response variable. And that's kind of a surprising association here. It seems to be a negative association so that, as alcohol consumption increases, heart disease rate decreases.

And we can actually see that it's probably a pretty strong association-- so maybe or strong, in terms of our association there. Now, let's define what's called the best-fit line or the regression line. And what it is it's a line that summarizes the tendency of x to explain why-- when x and y have a linear association.

And we could see, in that last example, that did appear pretty linear. This line is also called the trend line. So there's a lot of different lines that describe this-- a lot of different terms that describe the same line. Best-fit line, regression line, trend line-- those are all synonymous. So the best-fit line splits the points on the scatter plot. So let's draw in a best-fit line on the previous scatter plot.

So let's go back to that same scatter plot. And again, I want to draw it in here so that it's kind of splits the data. So there are formal ways of coming up with this type of line. We'll just do it kind of informal. So what I'm going to do is I'm going to place my ruler on here so it seems like it splits the data into two relatively similar groups.

And I would say that's a pretty good estimate there, so I'm going to flatten out my ruler here and draw in my best-fit line. So this is my best-fit line. And then what that does is that this is used to model the data. And it can also be used to help us make some predictions. So we'll come back to the line in a minute. A best-fit line can be used to make y value predictions for given x values. Let's use our best-fit line to predict the heart disease rate for a country that consumes an average of five liters of

alcohol per week.

So in this case, if I know the amount of alcohol consumption for a specific country, in terms of the average person, I can then make a prediction for the y value. So the alcohol consumption is my given x. The y that I want to find then would be the heart disease rate. OK, so essentially what I'm doing is I'm saying x is equal to 5, and then I want to know what y hat is.

y hat represents a predicted y-- so predicted y value. Now, it's different than just a regular y value because a regular y value would be an observed y value. So y hat then is always predicted, and a y hat value will always come from an equation or a graphed line. So let's go back again to the graph. And we want to figure out, about what's the heart disease rate for a country with-- that consumes about 5 liters per year?

So I know I have 4 and 5 here-- or 4 and 6 here. 5 is going to be about halfway-- so should be about here. And then what I'll do is I'll follow this up until it reaches the line, and then I will see about what y value that matches up with. So it seems like it's between 100 and 150-- closer to 150 than 100.

So we'll approximate that to be about 140. So my y hat value here is, we'll say, approximately equal to 140. So for a country that consumes 5 liters of alcohol per year, the heart disease rate will be about 140 people per 10,000. All right, that has been the tutorial on best-fit line/regression line. Thanks for watching.