Everybody, welcome. This is a very brief demo of total correlations.

As always, I encourage you to follow along in

R. The only package we need for this is the GGally package,

just for the plot, and this is a simile of which we will call a simulation study.

We would demonstrate on a random data set.

We have Variable 1,

which is standard normal random variable x2,

which is basically x1 plus again standard normal.

Then we have variable y,

so that's the dependent variable,

and it's going to be x1 times 2 plus x2 plus some additional noise.

Now, the idea here is that x1 and x2 are correlated

with each other and also they're all correlated with y.

Here's what they look like together.

Positive correlation throughout, and again,

by varying the coefficients and the signs we can make them positive,

negative, or zero if we wanted to.

Now, let's compute the total correlation of y,

given the rest with the two x's.

Well, one simple one as you're probably familiar with if you've had regression,

it's we take y,

regress it on x1 and x2,

and then get the summary of it,

and right here we have our multiple R squared,

0.9055, 91 percent of the squared variation y is explained by the two x's together.

We can also re-express it differently.

We could say that, well,

let's instead make a prediction from this fit,

so these are white hats,

and compute the correlation between the prediction and the original values.

Now we get a correlation here,

so to get the R squared,

we would square that and we once again end up with the same value.

Down to rounding error.

We could also use that shortcut from the correlation matrix.

This is a trick where we take the correlation between y,

then we join the correlation between y, x1 and x2.

Invert it. Take the first element,

that is the one in the upper left-hand corner,

and then take one minus one over that.

Then we get the same value as we would expect.