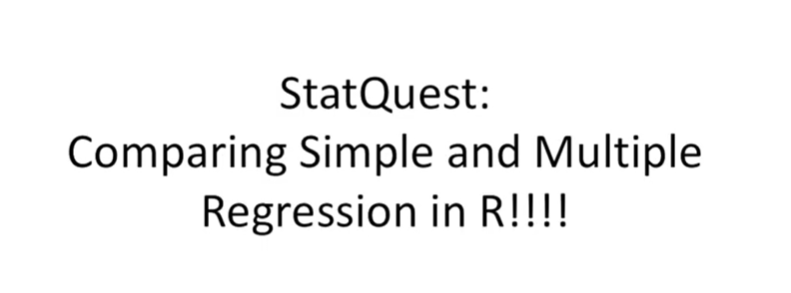
<https://www.youtube.com/watch?v=hokALdIst8k&list=PLblh5JKOoLUIzaEkCLIUxQFjPIlapw8nU&index=7>



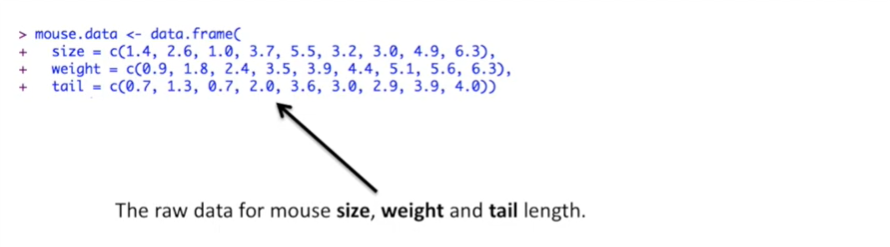
Hello I'm Josh stormer and welcome to stat quest.

Stat quest is brought to you by the friendly folks in the genetics department at the University of North Carolina at Chapel Hill.

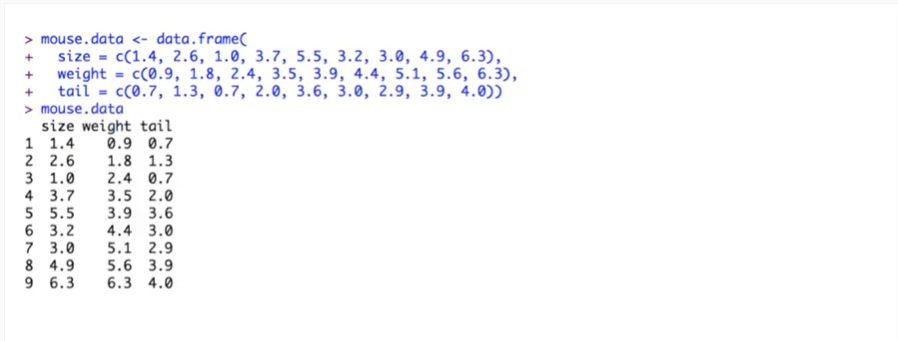
Today we're going to compare simple and multiple regression.

And our and just so you know the R code used in this video is available on the stack quest website.

Stack quest or G this stat quest picks up from where the stat quest on multiple regression left off.



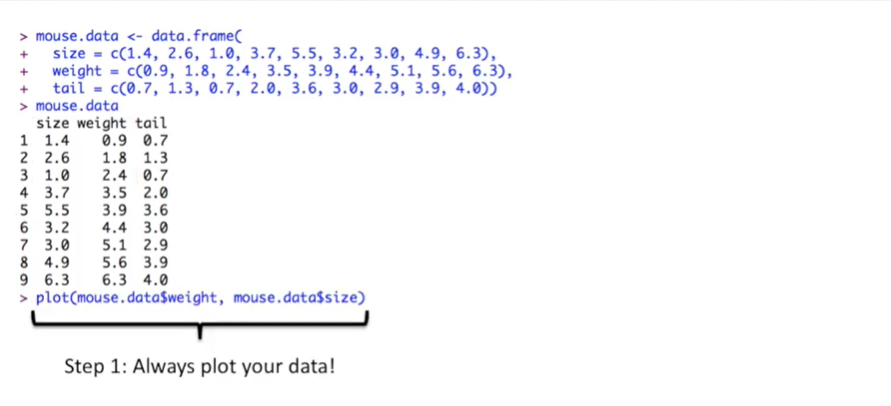
Here's the raw data that we're going to use in our analysis.



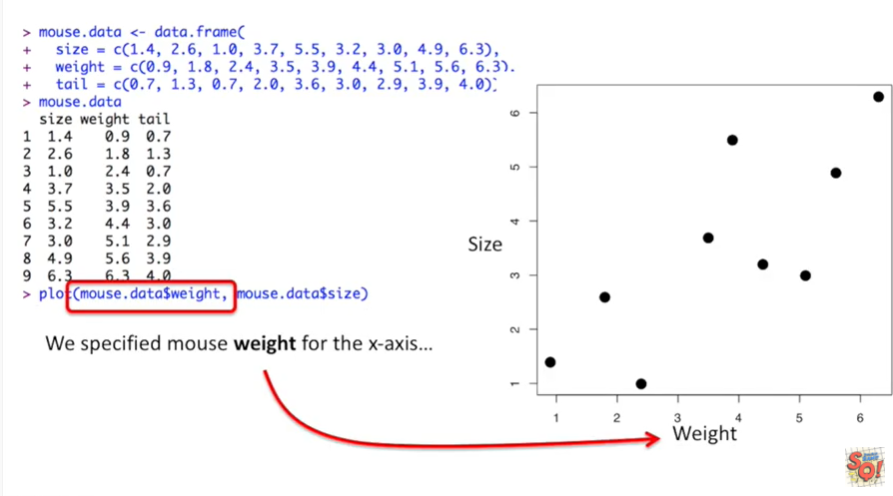
I've created a data frame with measurements for size weight and tail.



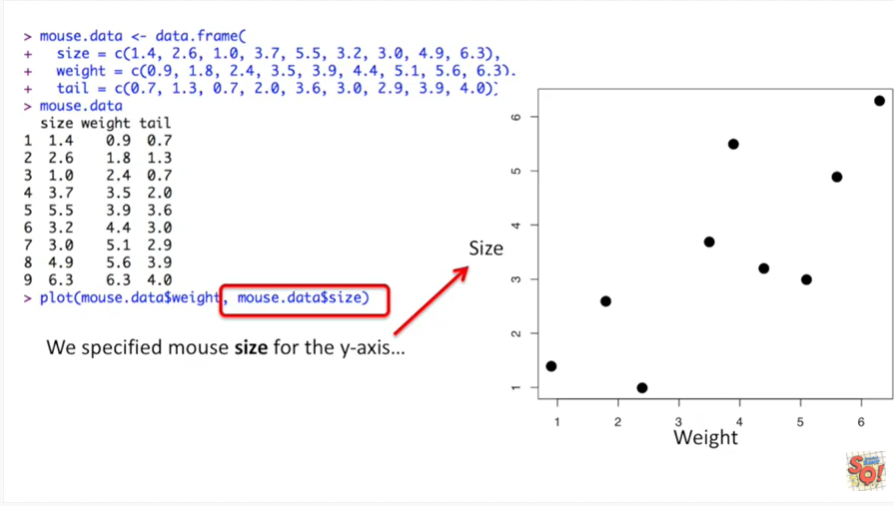
For simple regression we will focus on how well weight predicts size.



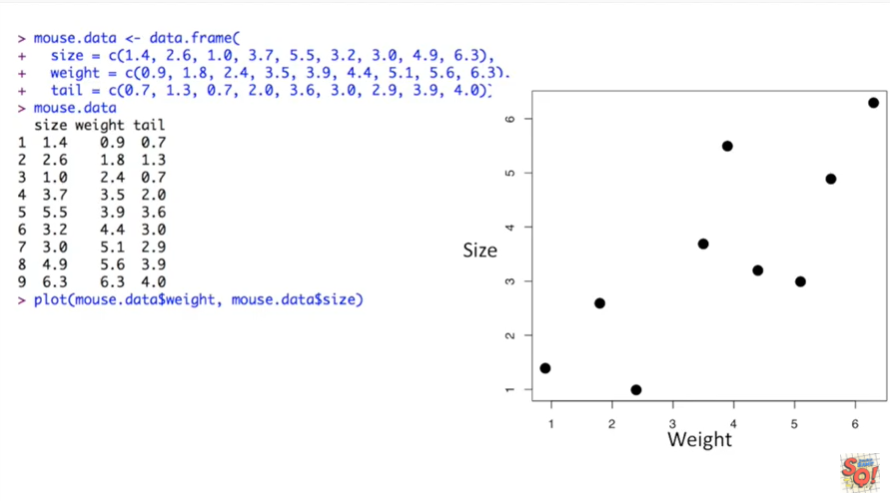
Step 1 always plot your data.



We specified weight for the x-axis.



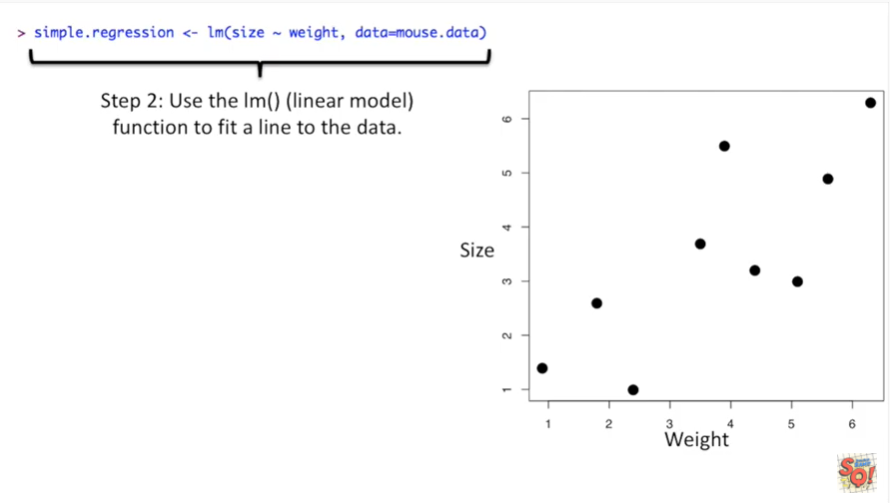
And we specified size for the y-axis.



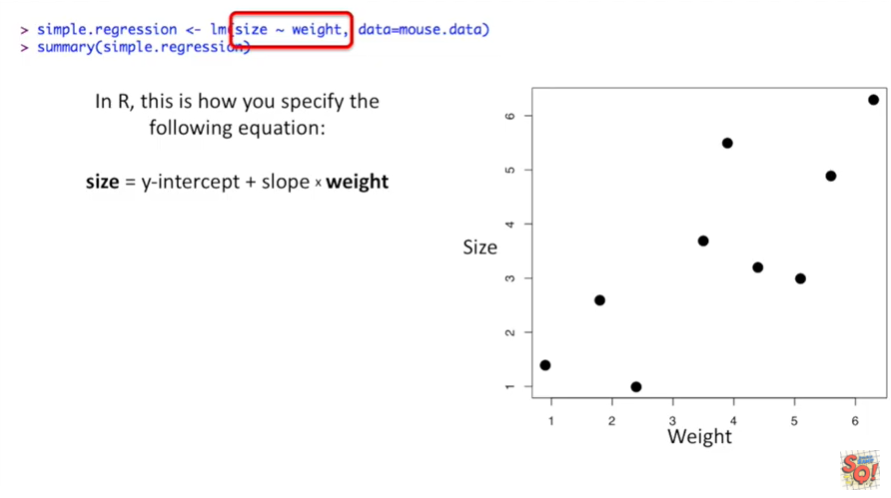
Plotting your data as a first step is super important because it allows us to evaluate whether doing a linear regression to begin with is a good idea.

Can we see a relationship in the data between size and weight ?

In this case we can and that means doing a regression makes sense.



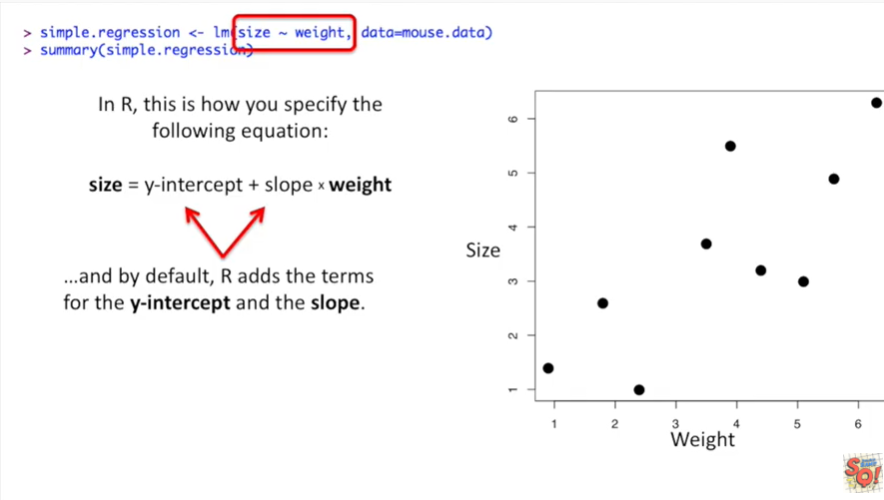
Step 2 use the LM function where L M stands for linear model to fit a line to the data.



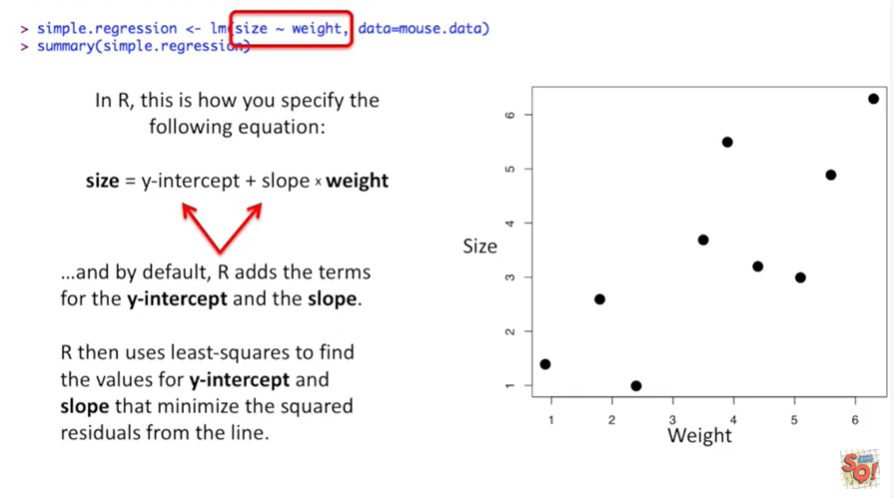
In R this is how you specify the following equation.



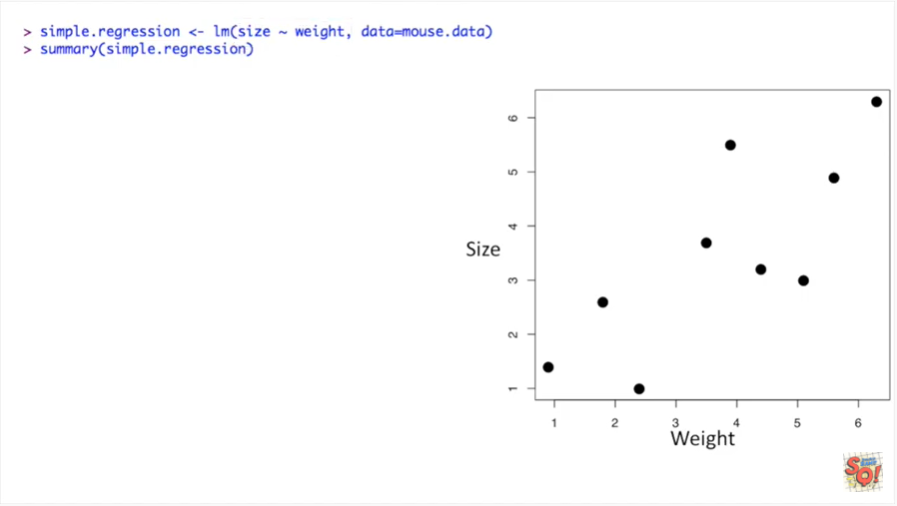
We specify size is predicted by weight by using the tilde character between size and weight.



And by default our adds the terms for the y-intercept and the slope.

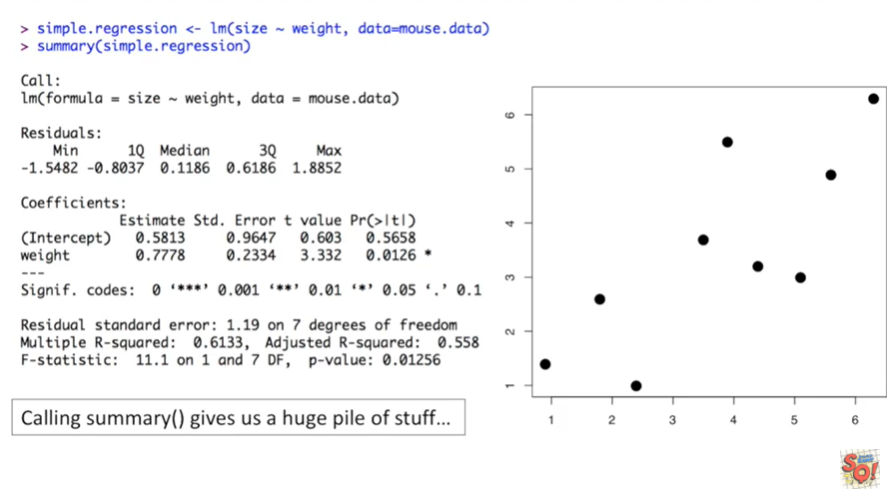


R then uses least-squares to find the values for the y-intercept and the slope that minimize the squared residuals from the line.

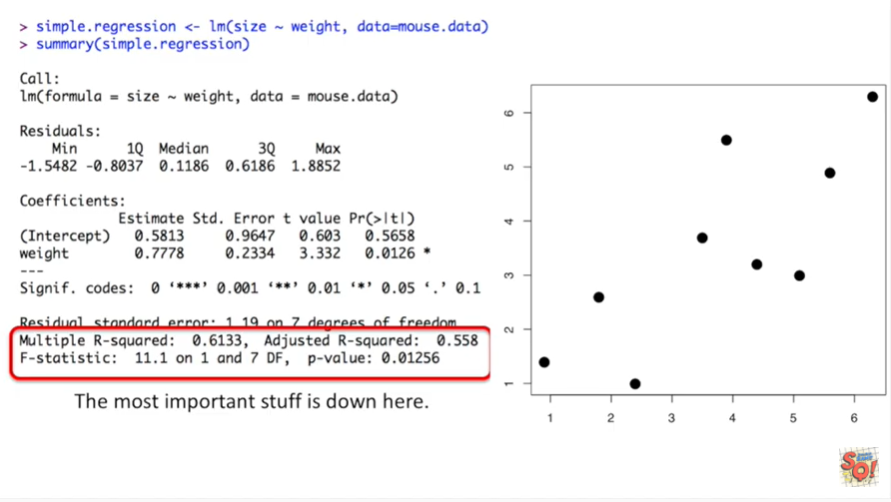


Once we've run the linear models function and save the output in a variable called simple dot regression.

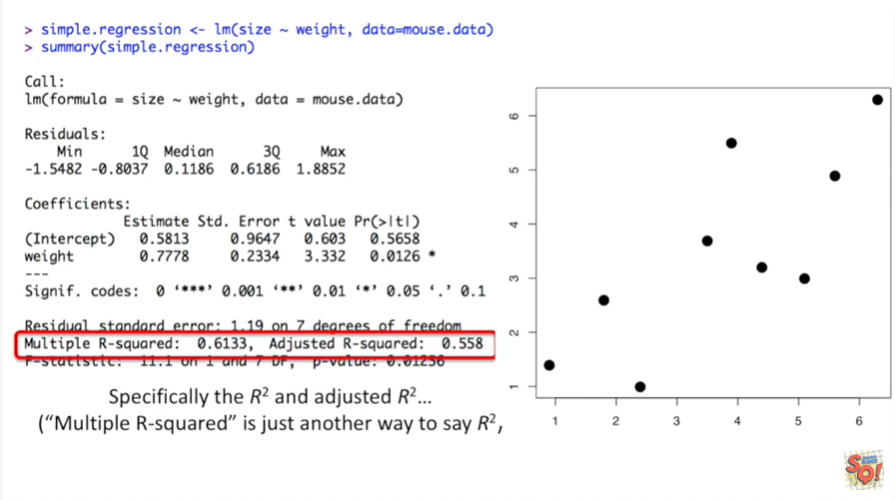
We can get a summary of that regression using the summary function.



Calling the summary function on simple dot regression gives us a huge pile of stuff.



The most important stuff is down here.



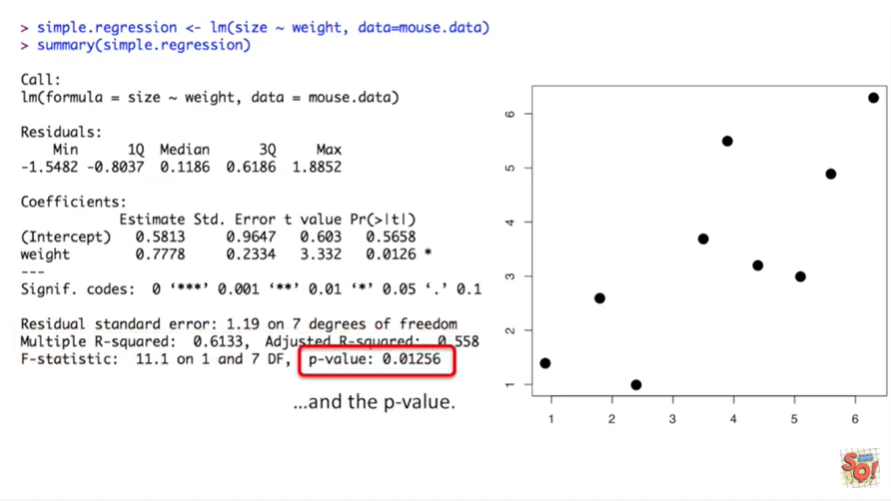
Specifically the r-squared and the adjusted R squared.

Multiple r-squared is just another way to say R squared by the way.

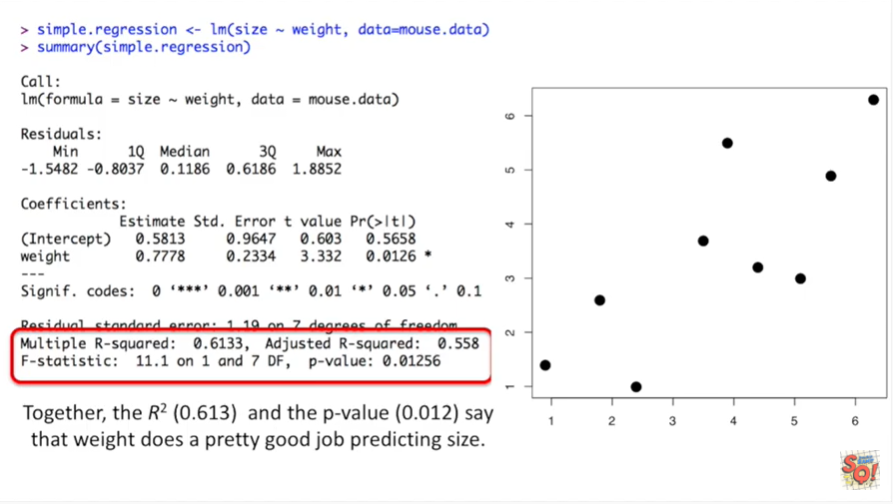
Also for simple regression the multiple r-squared value or just plain old r-squared is one we're interested.

In the adjusted R squared only applies when we have more complicated models.

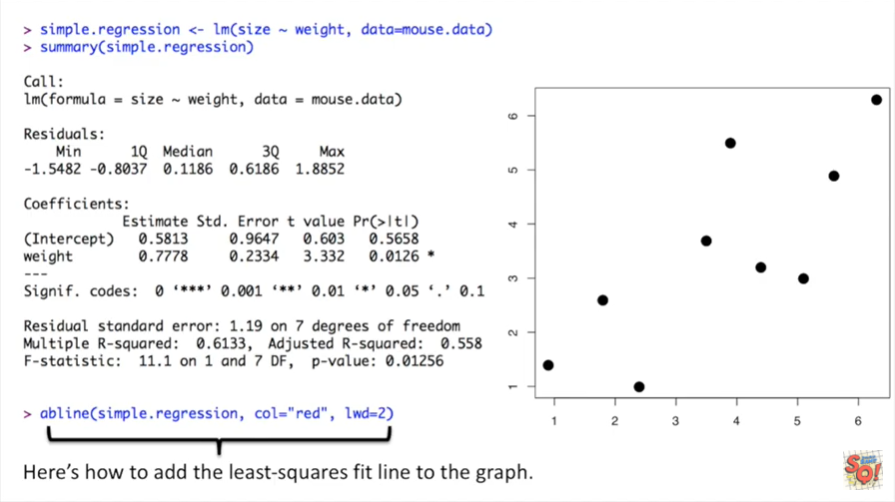
We'll use it later when we do multiple regression.



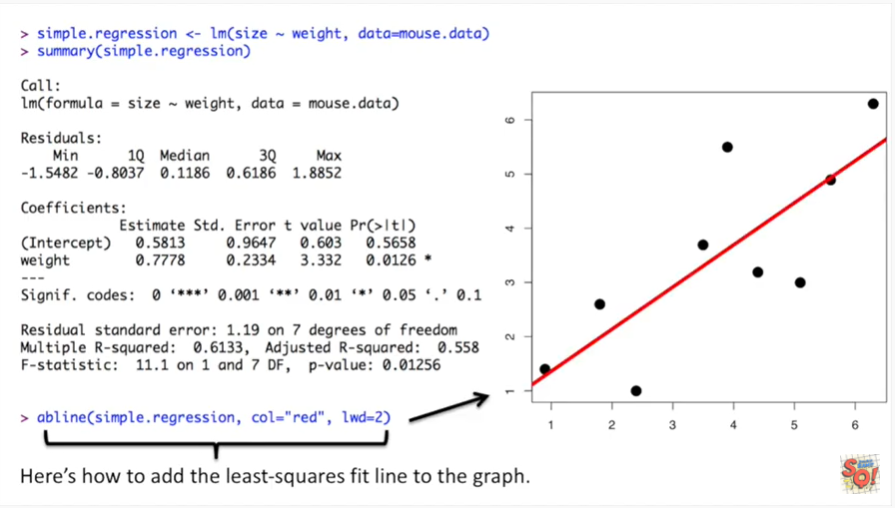
There's also the p-value down here.



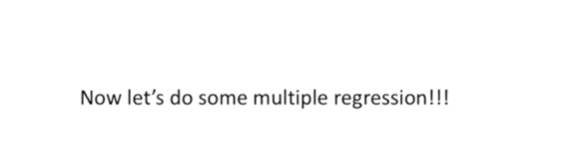
Together the R squared which equals zero point six one three and the p-value which equals zero point zero one to say that weight does a pretty good job predicting size.



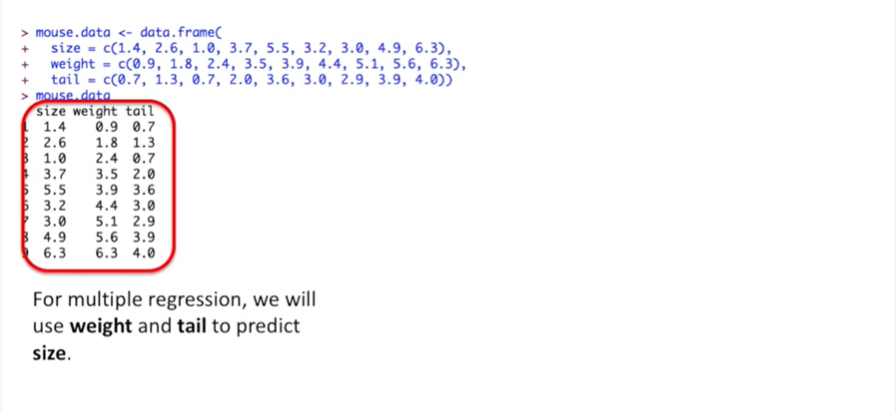
The last thing we want to do for our simple regression is add a line that shows the least squares fit on the graph.



We do this using the a beeline function.



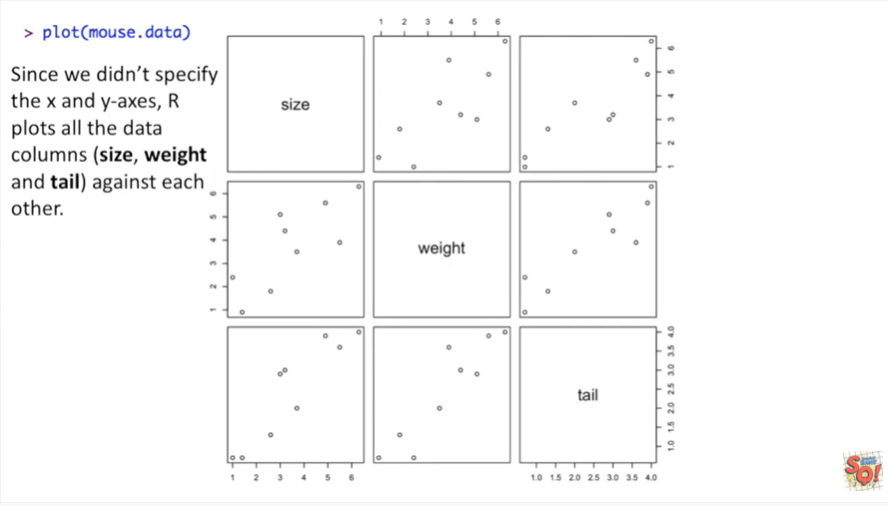
Now let's do some multiple regression.



For multiple regression we will use weight and tail to predict size.

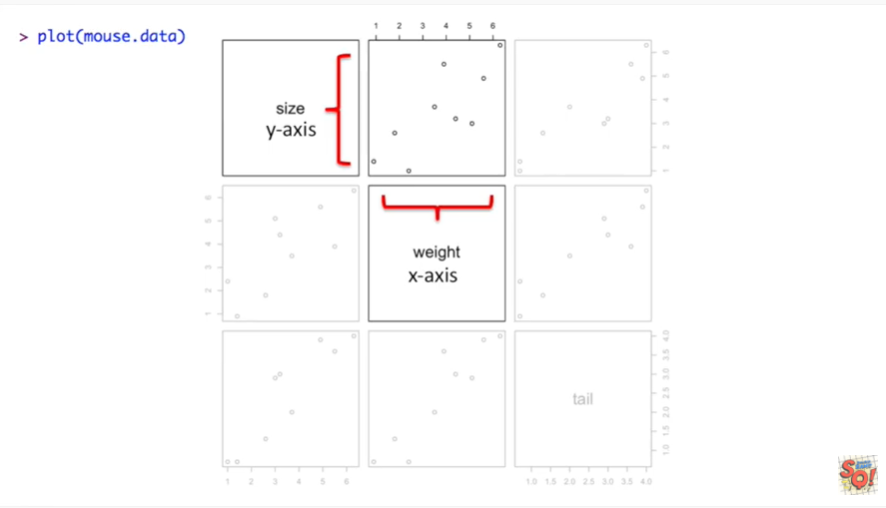


Step one always plot your data.

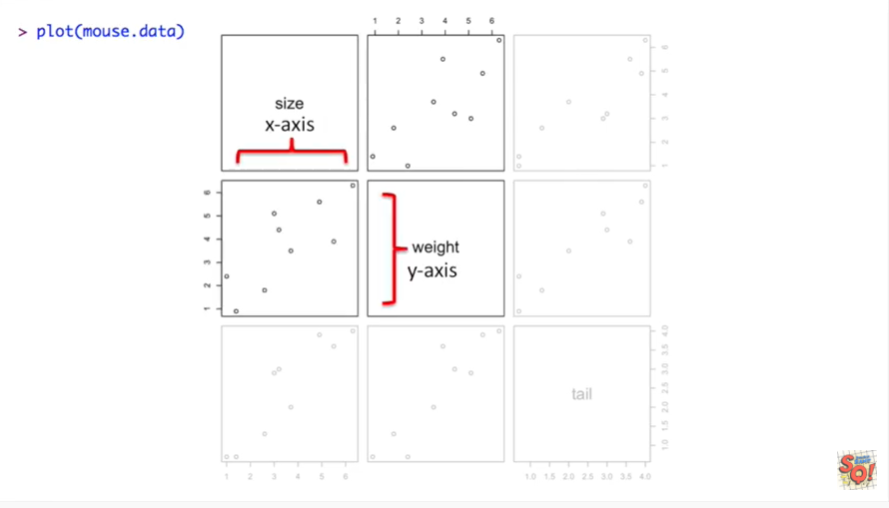


Since we didn't specify the x and y axes our plots all the data columns size weight and tail against each other.

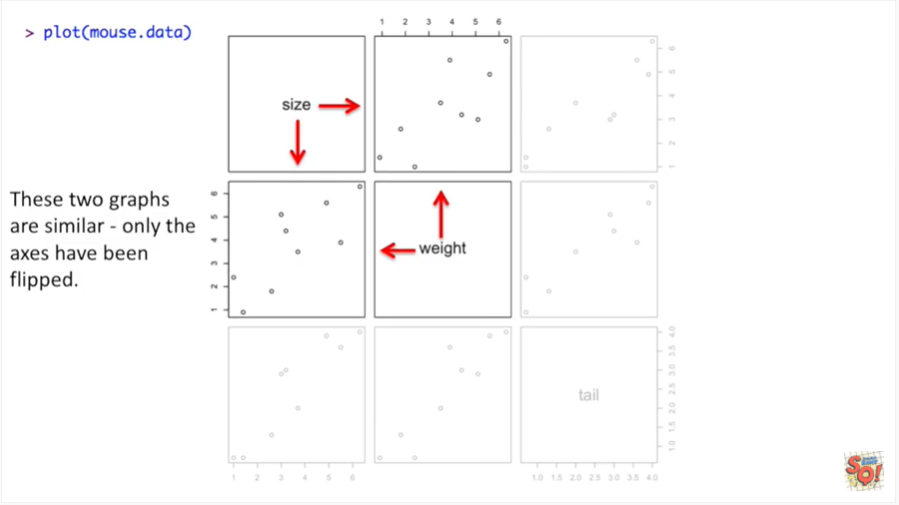
This is super useful because it generates all the plots we need to decide whether doing a multiple regression with this data makes sense or not.



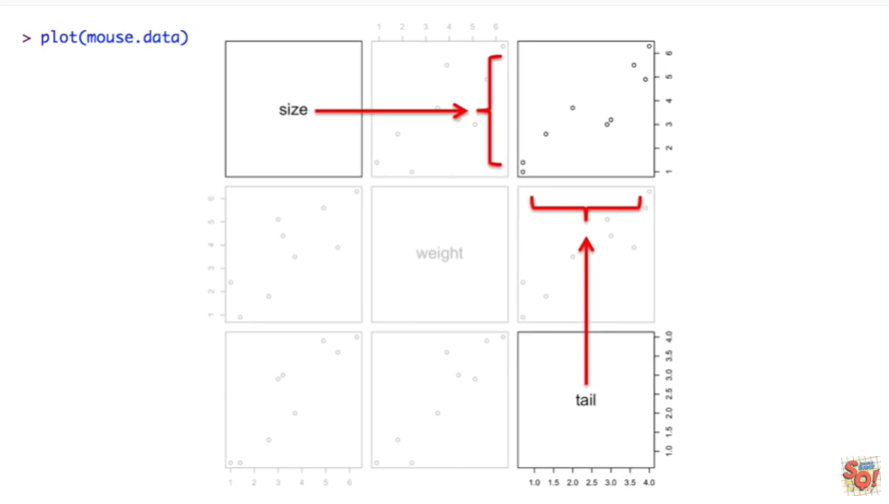
This graph plots size on the y axis and weight on the x axis this is what we used before in the simple regression.



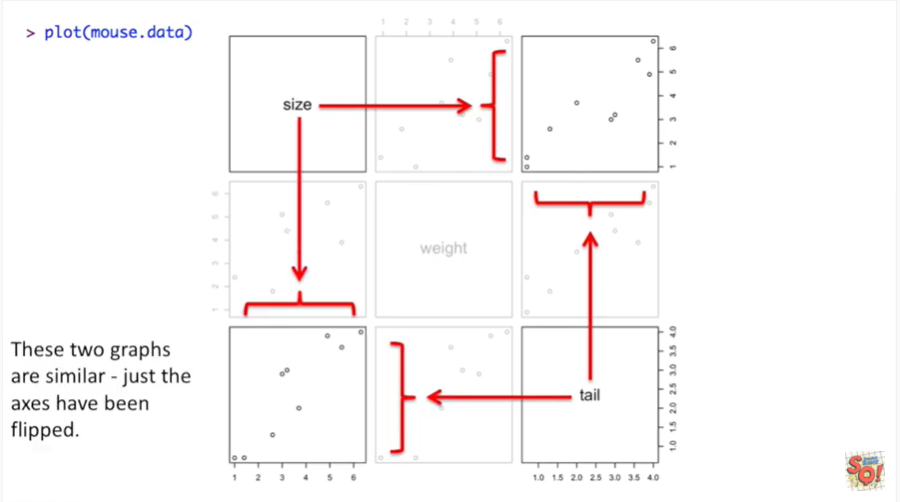
Down here we have the same exact data however this time sizes on the x axis and weight is on the y axis.



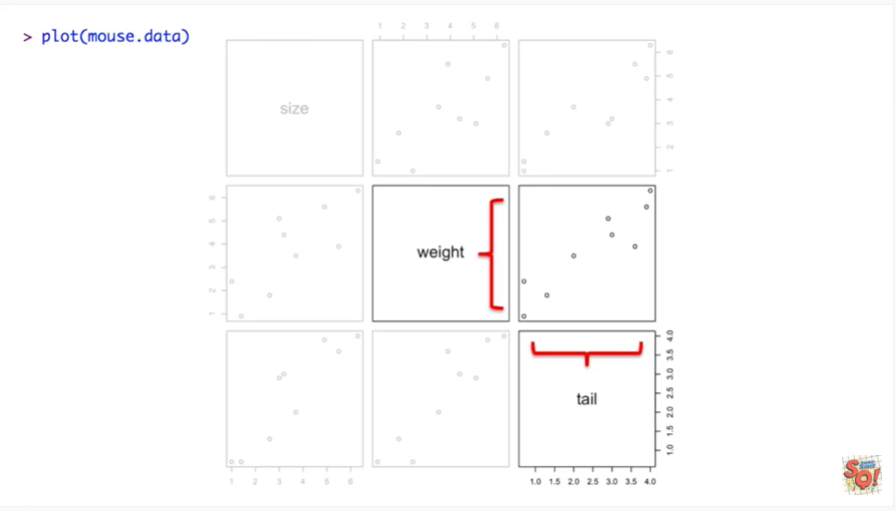
These two graphs are similar only the axes have been flipped.



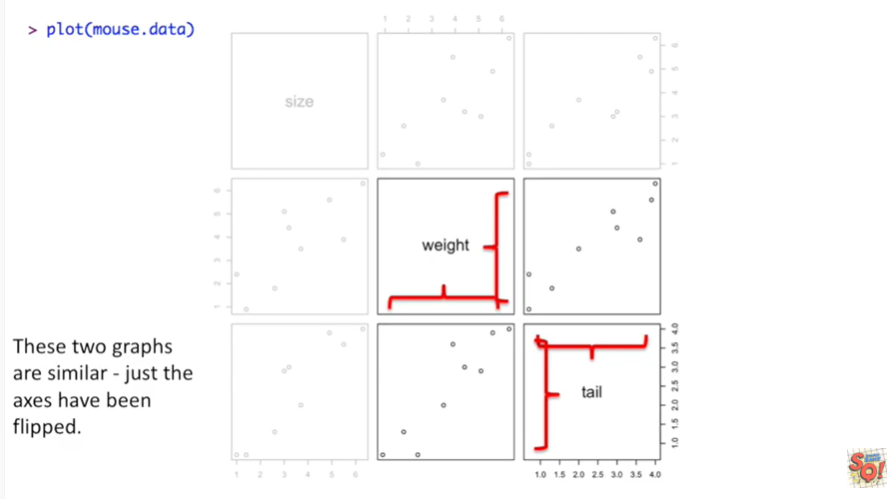
The graph in the upper right hand corner has size on the y axis and tail on the x axis.



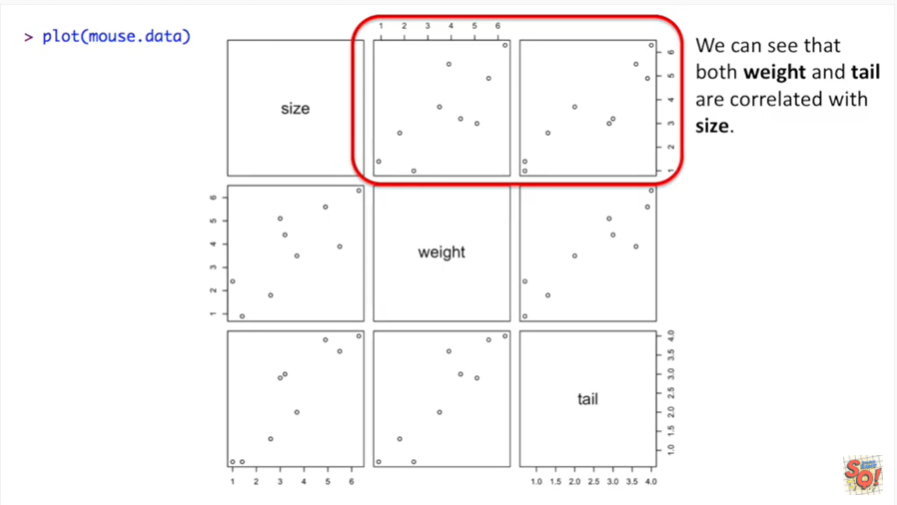
And in the lower left hand corner we have another graph that's very similar just the axes have been flipped.



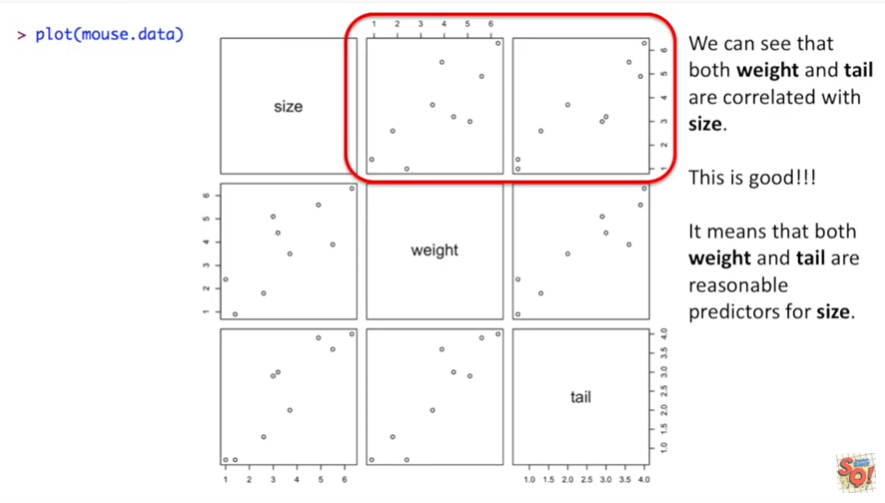
This graph has weight on the y-axis and tail on the x-axis.



And just like for the other graphs these two graphs are similar just the axes have been flipped.

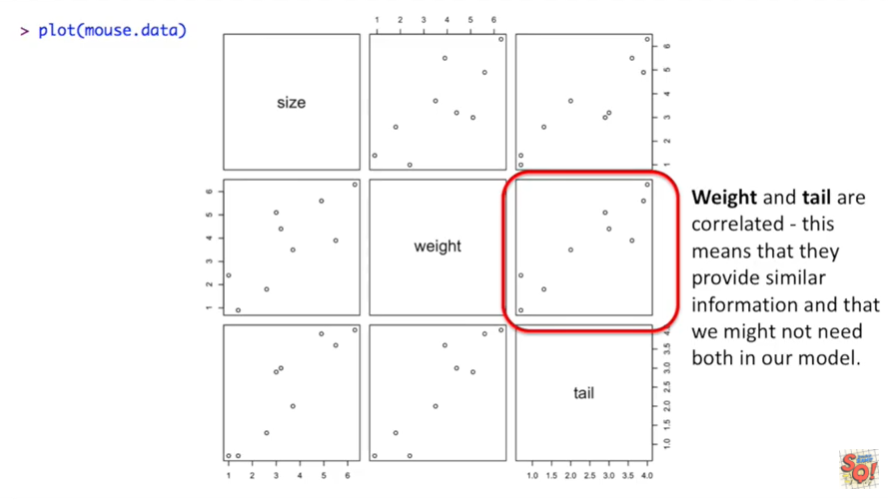


We can see that both weight and tail are correlated with size.



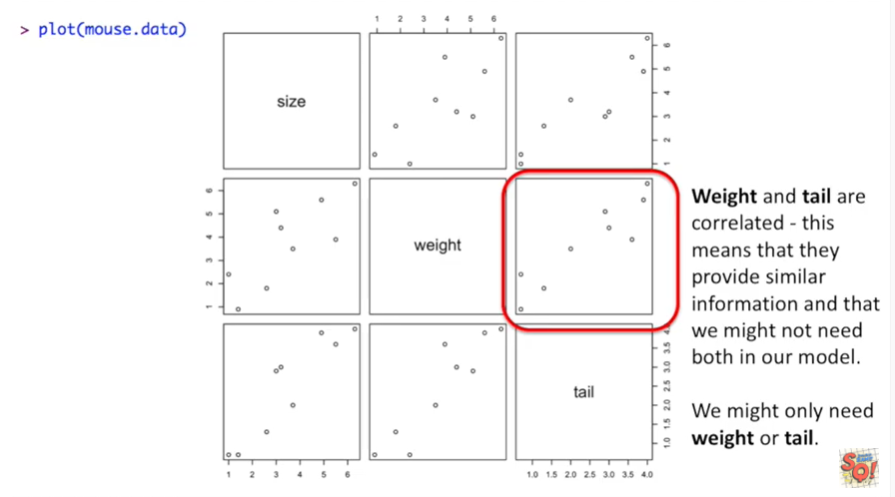
This is good.

It means that both weight and tail are reasonable predictors for size.

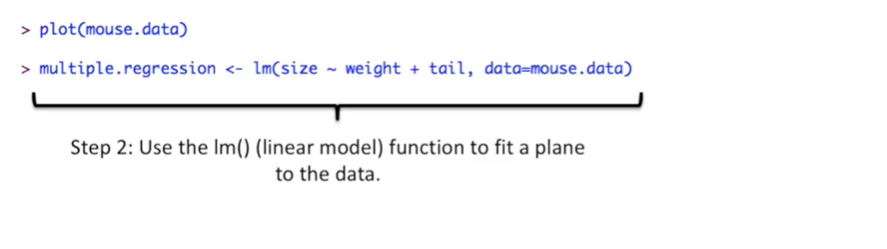


We can also see that weight and tail are correlated.

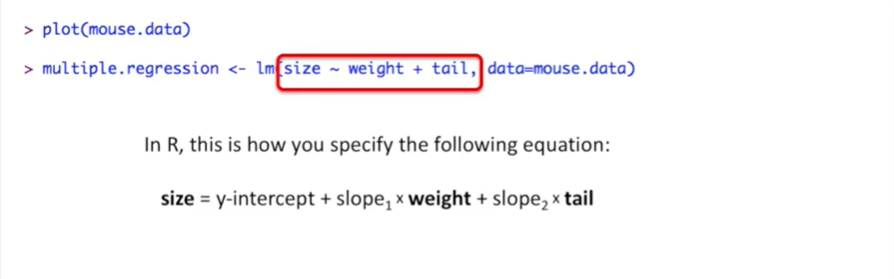
This means they provide similar information and that we might not need both in our model.



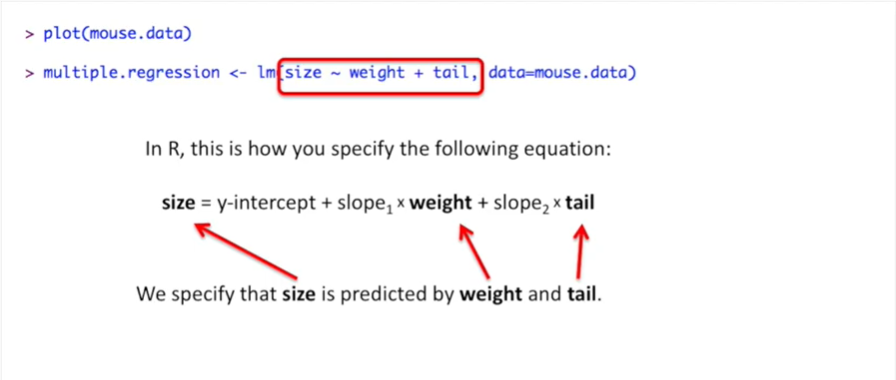
We might only need weight or tail.



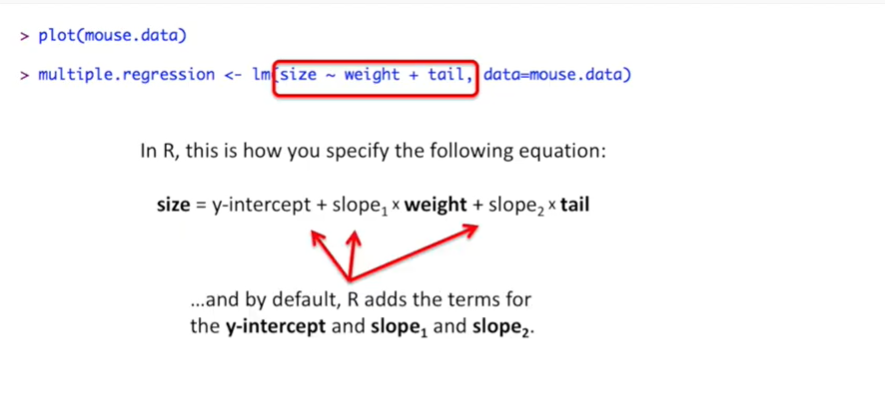
Step 2 use the linear model function to fit a plane to the data.



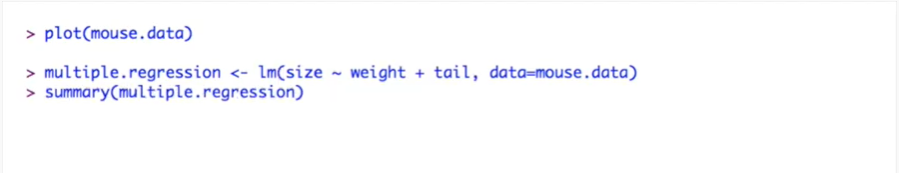
In R this is how you specify the following equation.



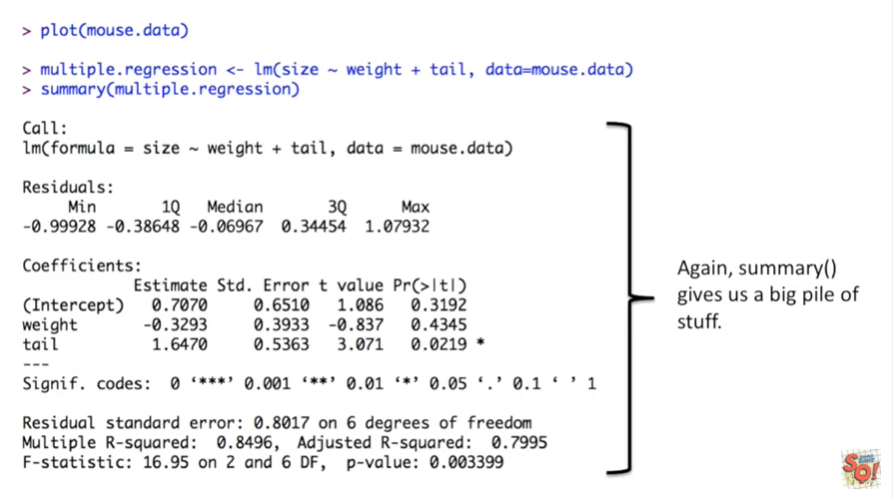
Using the tilde and the plus symbols we specify that size is predicted by weight and tail.



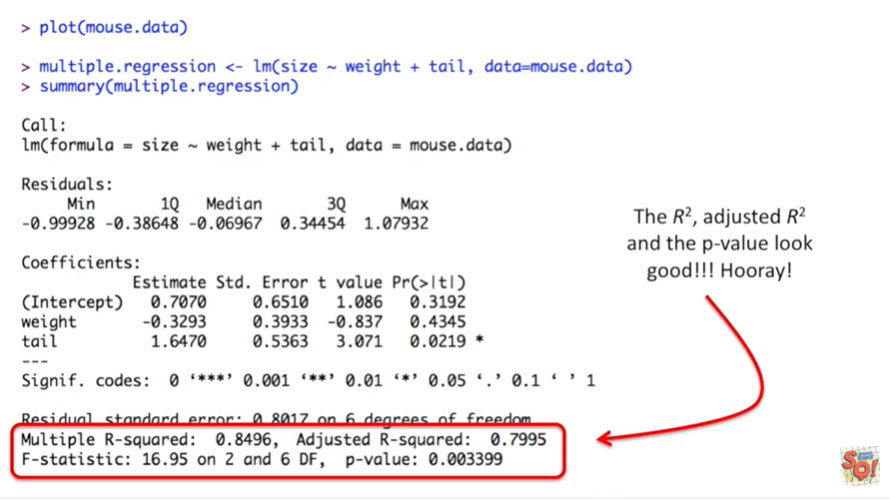
And by default R as the terms for the y-intercept and slope 1 and slope 2.



Once we've run the linear models function we can print out a summary of the results using the summary function.



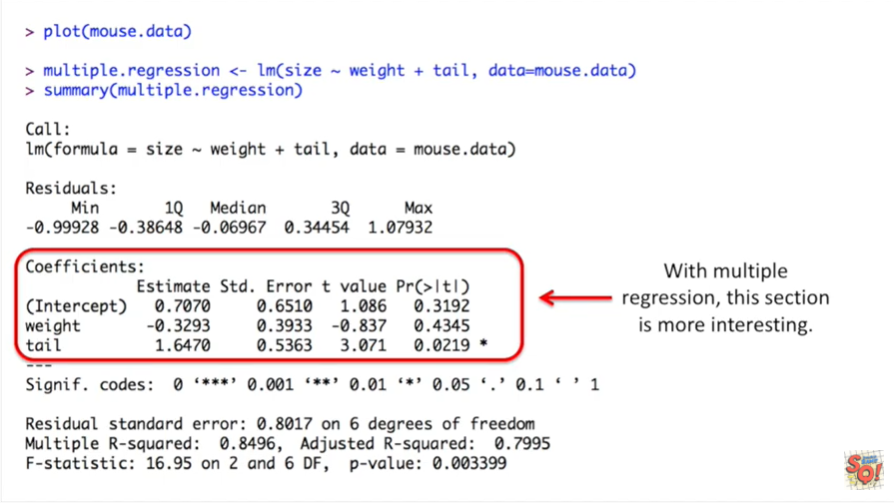
Again summary gives us a big pile of stuff.



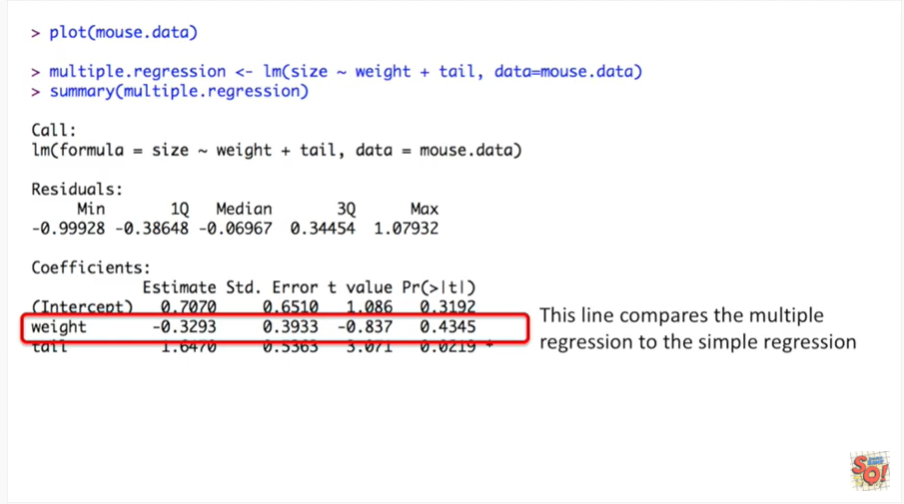
The r-squared adjusted r-squared and the p-value look good.

Hooray.

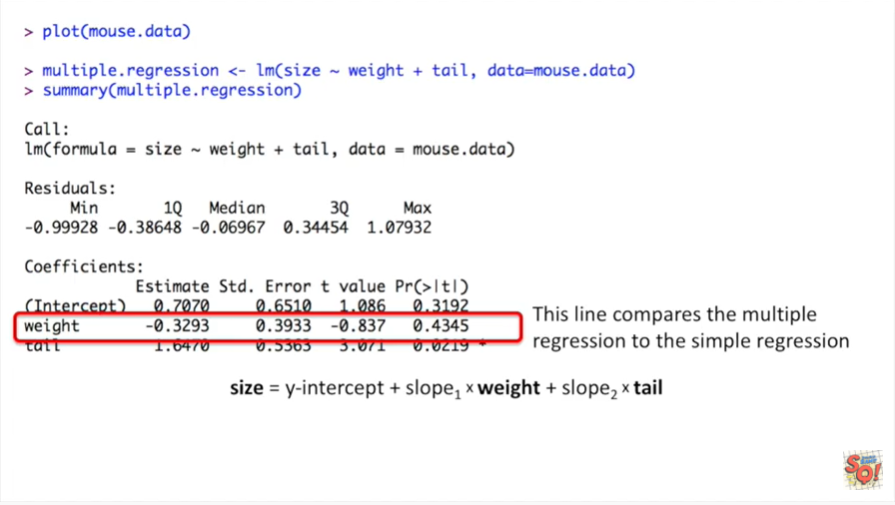
Note : since we're doing multiple regression we're now more interested in the adjusted r-squared value.



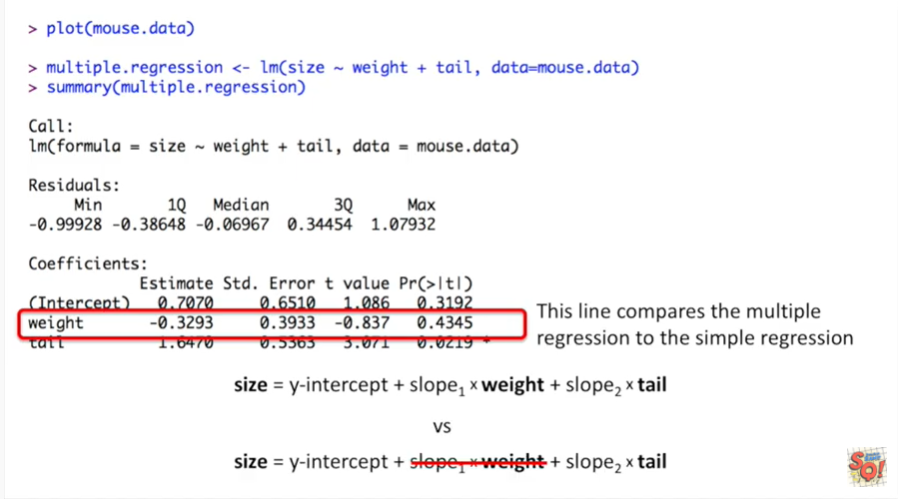
With multiple regression this section is more interesting.



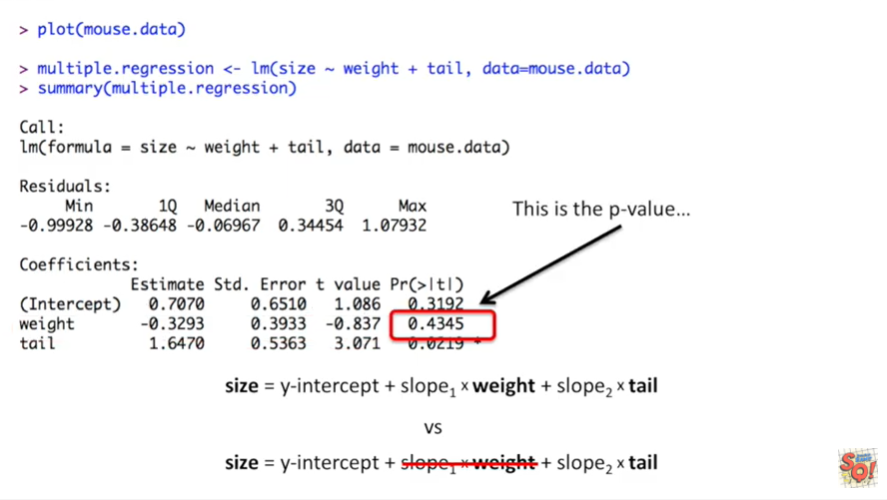
This line compares the multiple regression to the simple regression.



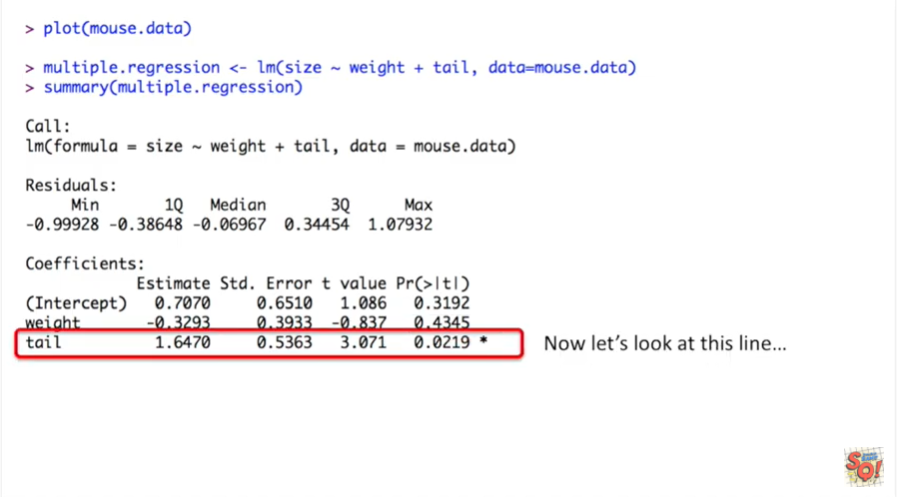
It compares this model which uses both weight and tail to predict size



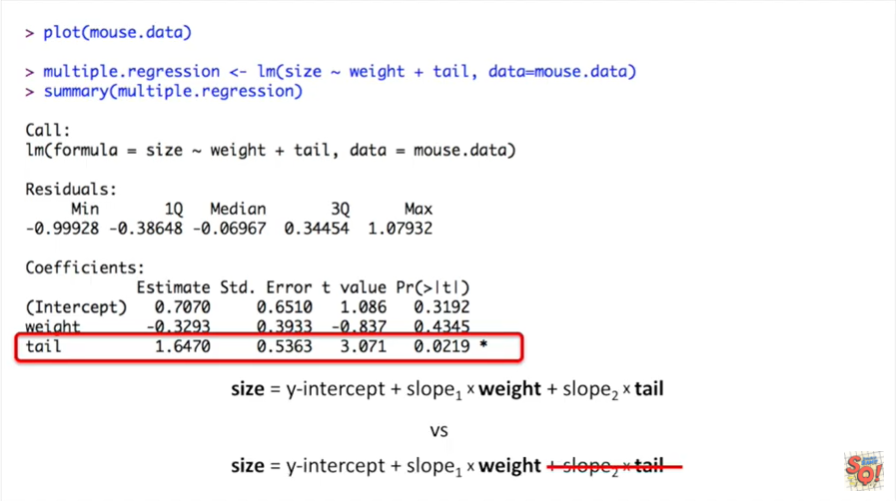
to this simpler model where we're just using tail to predict size.



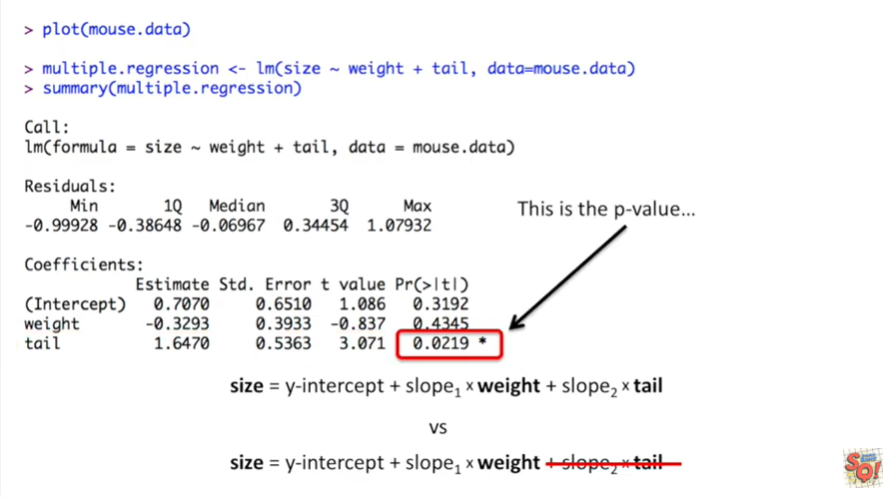
This is the p-value it means that using weight and tail isn't significantly better than using tail alone to predict size.



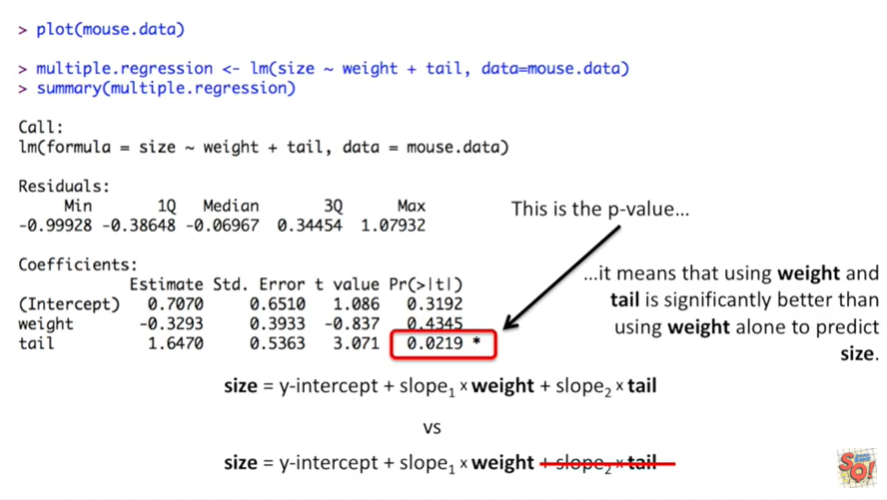
Now let's look at this line.



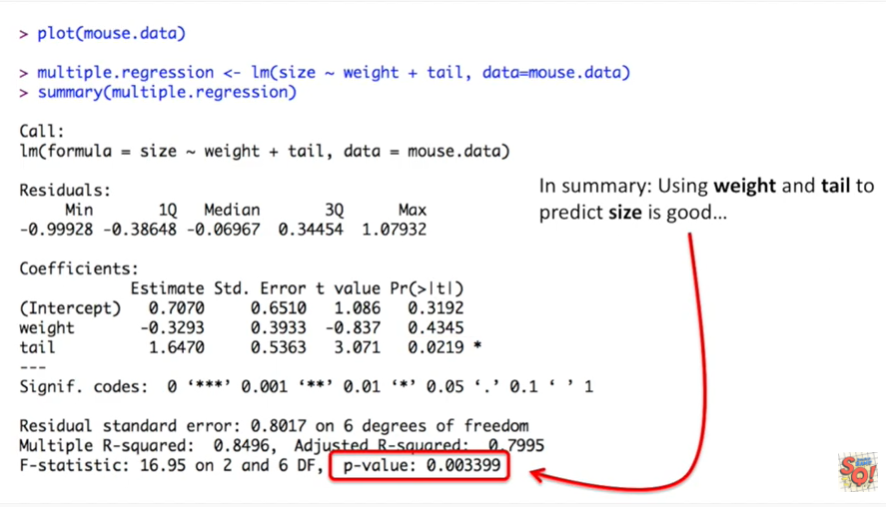
It compares the fancy multiple regression where we use weight and tail to predict size to a simple regression where we just use weight to predict size.



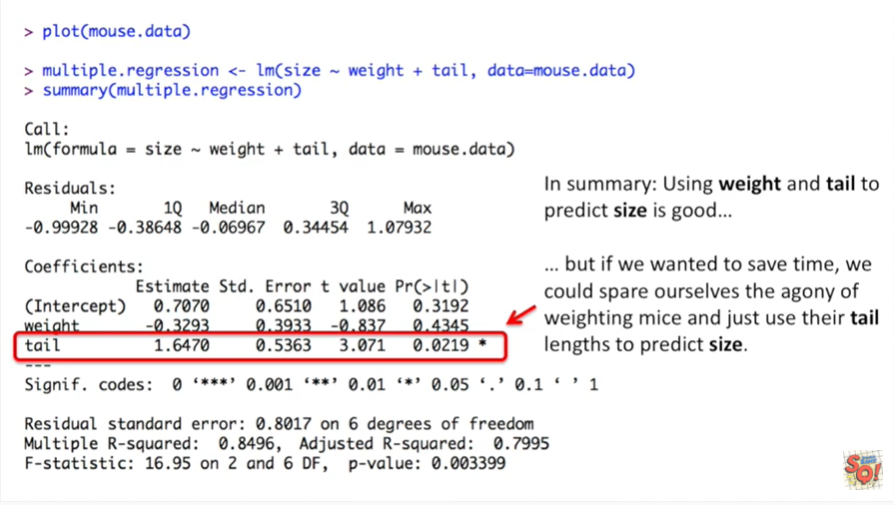
This is the p-value.



It means that using weight and tail is significantly better than using weight alone to predict size.



In summary using weight and tail to predict size is good.



But if we wanted to save time we could spare ourselves the agony of weighing mice and just use their tail lengths to predict size.