In our R Console, let's start by loading our data set.

Don't forget to make sure you're in the directory containing the file wine.csv first.

We'll call our data frame wine, and we'll use the read.csv function to read in the data file wine.csv.

We can look at the structure of our data by using the str function.

We can see that we have a data frame with 25 observations of seven different variables.

Year gives the year the wine was produced, and it's just a unique identifier for each observation.

Price is the dependent variable we're trying to predict.

And WinterRain, AGST, HarvestRain, Age, and FrancePop are the independent variables we'll use to predict Price.

We can also look at the statistical summary of our data using the summary function.

This gives us information about the range of values for each variable in our data set.

Let's now create a one-variable linear regression equation using AGST to predict Price.

We'll call our regression model model1, and we'll use the lm function, which stands for linear model.

This is the function we'll use whenever we want to build a linear regression model.

Then inside parentheses, type Price, our dependent variable, and then a tilde symbol, and then AGST, the independent variable we'll use in this model.

Then after a comma, we need to add data = wine to tell the lm function what data set to use to build the model.

We're saving the output of the lm function to the variable named model1.

So when we hit Enter, we didn't see any output because it's been saved to the variable model1.

Let's take a look at the summary of model1.

The first thing we see is a description of the function we used to build the model.

Then we see a summary of the residuals or error terms.

Following that is a description of the coefficients of our model.

The first row corresponds to the intercept term, and the second row corresponds to our independent variable, AGST.

The Estimate column gives estimates of the beta values for our model.

So here beta 0, or the coefficient for the intercept term, is estimated to be -3.4.

And beta 1, or the coefficient for our independent variable, is estimated to be 0.635.

There's additional information in this table that we'll discuss in the next video.

Towards the bottom of the output, you can see Multiple R-squared, 0.435, which is the R-squared value that we discussed in the previous video.

Beside it is a number labeled Adjusted R-squared.

In this case, it's 0.41.

This number adjusts the R-squared value to account for the number of independent variables used relative to the number of data points.

Multiple R-squared will always increase if you add more independent variables.

But Adjusted R-squared will decrease if you add an independent variable that doesn't help the model.

This is a good way to determine if an additional variable should even be included in the model.

We'll also discuss other ways to select important independent variables in the next video.

Let's also compute the sum of squared errors, or SSE, for our model.

Our residuals, or error terms, are stored in the vector model1$residuals.

By hitting Enter, we can see the values of all of our residuals.

We can compute the Sum of Squared Errors, or SSE, by taking the sum(model1$residuals^2).

If we type SSE and hit Enter, we can see that our sum of squared errors is 5.73.

Now let's add another variable to our regression model, HarvestRain.

We'll call our new model model2.

And again, we'll use the lm function to predict Price, but this time using AGST and HarvestRain.

When you want to use more than one independent variable, you can just separate them with a plus sign like we did here.

Then we again need to indicate that the data that should be used is wine.

Let's take a look at the summary of our new model using the summary function.

We have a third row in our Coefficients table now corresponding to HarvestRain.

The coefficient estimate for this new independent variable is negative 0.00457.

And if you look at the R-squared near the bottom of the output, you can see that this variable really helped our model.

Our Multiple R-squared and Adjusted R-squared both increased significantly compared to the previous model.

This indicates that this new model is probably better than the previous model.

Let's now also compute the sum of squared errors for this new model.

So SSE equals, and then sum(model2$residuals^2).

If we type SSE, we can see that the sum of squared errors for model2 is 2.97, which is much better than the sum of squared errors for model1.

Now let's build a third model, this time with all of our independent variables.

We'll call this one model3.

And again, use the lm function to predict Price using AGST and HarvestRain and WinterRain and Age and FrancePop.

Again, we need to tell the lm function to use the data set wine.

Let's take a look at the summary of model3.

Now the Coefficients table has six rows, one for the intercept and one for each of the five independent variables.

If we look at the bottom of the output, we can again see that the Multiple R-squared and Adjusted R-squared have both increased.

Let's now compute the sum of squared errors for this new model.

SSE equals the sum(model3$residuals^2).

And if we type SSE, we can see that the sum of squared errors for model3 is 1.7, even better than before.

In the next video, we'll determine if we should keep all of these variables in our final model.