**Transcription: Descriptive Statistics in R Commander**

In this video, we're going to go over how to produce descriptive statistics in R Commander. I've already got our data set loaded here. It's the same data set we used in the last lab on data visualization, so it's got data on education spending and education achievement across the 50 U.S. states plus Washington DC.

If we wanted to see some numerical descriptive statistics, what we can do is go to “Statistics”, “Summaries”, and “Numerical summaries...”. This will pull up a window that shows us all of the numeric variables in our data set. We can pick just one of them, or we could pick multiples. Let's go ahead and look at the descriptive statistics for all of the variables in this data set. We can select all of them by holding down shift and then clicking on the variable names to select all six variables.

Next, we can go into “Statistics” and tell R Commander which specific statistics we want to see. In this case, we have pre-selected “Mean”, “Standard Deviation”, “Interquartile Range” as well as some quantiles. If we want to see the mean, standard deviation, and median, we've already got those, but we've got some extra stuff as well. We can deselect “Interquartile Range”, and in this case we don't really care about the 25th or 75th percentile so we can erase those. Of course, if you leave all of this in, it won't be a problem; it will just show you more information and potentially be a little bit confusing in terms of what you're looking for. If you make all of those changes, we can go ahead and click “OK”.

And we will get this table that displays the different statistics we've asked for each of the six variables. We can see here the mean. The mean amount that's spent on public education across these 50 States and Washington, DC is $5175 per student per year. We also have the standard deviation here, which is a measure of variability as was discussed in the lecture. And this 50% right here is the median. If you remember what the definition of the median is, it's the value that's at the 50th percentile or exactly in the middle of the values for the variable if you line them up from smallest to largest. So in this case, our median education spending is $5045 per student per year. We also have our minimum value at 0% and our maximum value at 100% as well as the number of observations. So that's the number of times that there's a value for this variable within this data set.

Now R Commander is not going to produce the range or the variance for you, but there are ways that you can calculate them based on the values that are shown. Remember that variance is just standard deviation squared, so if you wanted to know the variance for education spending, you could plug this standard deviation 1.376166 into your calculator and square it and that would give you the variance.

Similarly, if you wanted to know the range, you could subtract the minimum value from the maximum value. This is just going back to what the definition of range is. So say we wanted to see the range in teacher salary; we would take the maximum value, $43000, minus the minimum value, $22000, and we would get a range of $21,000.

Unfortunately, R Commander doesn't have a way of producing the mode of these data. However, one thing we learned in the last lab was how to make a histogram, and that's an easy way to identify the mode. You can look at which value has the most observations for it, and that's the mode. You do have to be careful of how things are grouped. If you really want to know the mode, it's probably better to use SPSS because that will give you a certain value for what the mode is within that variable. Usually in research, though, we don't report the mode. So, in general, R Commander is going to give you the descriptive statistics that you need.

So there's how you can find the mean, standard deviation, and median in R Commander as well as values that you can use to calculate the variance and the range.