**Transcript: Boxplots in R Commander**

Next, we're going to go over how to create a boxplot in R Commander. Boxplots are great visualizations for when you have outliers or other skewed data sets, and you want to be able to see that in the visualization.

To create a boxplot, we will navigate to “Graphs” and “Boxplot”. First, we need to select which variable we want to use in our boxplot, which in this case is “wages”.

If we go to options, we can select x-axis, y-axis labels and graph titles to be more informative. In this case, our y-axis label is “Wages ($ per Hour)”. We actually don't need to fill in an x-axis label because we're only doing one boxplot. This might be useful if you were, say, separating things by group. But we're going to leave it blank here, and for our title, we'll put in something more informative such as “Distribution of Wages in Ontario”. One other option we have here is whether we want to identify outliers. We're going to select “No”. If you click “Identify Outliers” “Automatically” or “With mouse”, it's going to put the research ID for the values that are shown as outliers and that can really just clog up the plot. It can be useful though if you don't have that many outliers and it's easier to see. Once we do that, we can click “OK”.

And we can look at our boxplot here. You can see here the median in this dark black line. This grey box is the interquartile range. The bottom bar is the minimum and the top bar is the maximum, excluding outliers. All of these dots here are outliers, so they're more than 1.5 times the interquartile range, which is kind of a typical definition for an outlier. You can see that there's a lot of outliers in this data set, but also you know that the majority of individuals are landing in this $10 to $20 an hour range.

So this is a great example of when boxplots can be really useful in terms of understanding how many outliers there are and whether they're high or low, as well as the more typical range for the variable.