This tutorial talks about modified box-and-whisker plots. A modified box-and-whisker plot is one where the whiskers only extend to the highest and lowest nonoutlier value. Outliers represented with dots here. So here we have our lowest nonoutlier value being marked with the whisker with the line here.

And the outlier is represented by this blue dot. Group two also has an outlier. They have an outlier on the upper end. So their outlier is marked with this blue dot. And then, this line here is no longer indicating the true maximum, but rather the highest nonoutlier value.

Now, here, when we define outliers, as a reminder, outliers are data values that are significantly higher or lower than a majority of the values. And again, it's not just an extreme value. You could have values that are very large or very small that become an outlier when there's a gap between that extreme value in the bulk the rest of the values. And the way we decide whether or not we have an outlier is with the 1.5 times IQR rule. And as a quick reminder, IQR since interquartile range.

So if a data point is more than 1.5 times the IQR away from the first or third quartile, then it's an outlier. So this rule here helps us to mathematically determine whether or not a data value is significantly higher or lower than the majority of the values or it's nonsignificant and it's not an outlier.

Let's walk through an example. In this example here, we are looking at the heights in inches of third graders. So in order to determine whether or not I have an outlier, I need to know the third and first quartile, in order determine the interquartile range. And using the 5-number summary kind of helps to keep me organized.

So first, we need to find the median. And I'm going to switch to a darker color. I like to use dots, instead of actually crossing off, so that I can still see the numbers. And my numbers are already in order, which is good. So I can just start right away by crossing off my values to find my middle number.

And as we can see, in this example here, we end up with a median of 39. And just to double check, I have 1, 2, 3, 4, 5 values below and 1, 2, 3, 4, 5 values above. I don't actually need to know that. I just needed to find it so that I can find my first and third quartile.

So for my first quartile, I'm going to cross off the values from the minimum up to the maximum. And I find out that my first quartile is 37. Now, I'm going to do the same thing in order to find my third quartile. I'm going to cross off from the median up to the maximum to find the middle value here. And in this case, we find out that it's 41. So 41 is our third quartile.

Now, once I have the first and third quartile, I'm going to be calculating the interquartile range. So my interquartile range, my IQR is the same as Q3 minus Q1. So in this case, it's the same as 41 minus 37. And 41 minus 37 is 4. So the IQR for this example is 4.

Now, our rule for finding outliers is I take 1.5 and multiply it by the IQR. And that gets me a value of 6. So if a number is 6 lower than the first quartile or 6 higher than the third quartile, then it's an outlier. So I'm going to take the 37 and subtract 6. And I get 31. And we're going to take the 41 and add 6. And we get 47.

So now, I'm looking through my data values to determine whether or not I have anything less than 31 or more than 47. So I look and my lowest number is 34. So I'm OK there. And my highest number is 51. Now, 51 is above 47. So this is going to be an outlier.

Now, 42 is my next value. So that is underneath 47. So that's OK. You can definitely have more than one outlier in a particular data set. Once we know what our outliers are, we can make a modified box-and-whisker plot. So we're going to do that now. And we already found out the 5-number summary. Our minimum is 34, our first quartile is 37, the median's 39, the third quartile is 41.

Now, the maximum was 51, but that's an outlier. So now, we're going to draw the line where the next highest nonoutlier value is. So that's 42. So while it's not the maximum for the whole data set, that's still the 51. It is the highest nonoutlier value. So we're going to make the same marks that we do in a traditional box plot.

And then, here, we're going to have a bar at 42 to indicate the highest non outlier value. And we're going to put a dot at 51, to indicate our outlier. We're going to connect first and third quartile to make the box. The minimum connects to the first quartile with a line.

The third quartile and maximum connect with a line. So this is our modified box plot. We're showing that we have an outlier out here at 51. And then, the remainder of our data here is indicated in the box. So this has been the tutorial on creating a modified box plot.