

---

In this tutorial, you're going to learn about a graphical display called a boxplot. They're also sometimes called box and whisker plots. But the real name actually is boxplots. So you might wonder what this thing actually is and what it looks like. A boxplot is a way to graphically display the five-number summary for a data set.

So suppose I have the heights of the Chicago Bulls basketball team. The five-number summary consists of the minimum value, which was the shortest individual on the team at 71 inches, the maximum, which is the highest value in the data set. This is the tallest individual on the team. And the three quartiles-- first quartile, the median, and the third quartile.

This is the value at which 25% of the data falls at or below it. Half the data falls at or below 79. And 3/4 of the data falls at or below 81. And obviously, all the data falls at or below 84.

So let's get to actually making this thing. First, we're going to draw an axis. It can be horizontal or vertical. I've made mine horizontal. You should also scale it with equal increments. So I've gone from the lowest number, 71 and a little bit lower, to the tallest number, 84.

First, make some kind of mark at the five numbers in the five-number summary, so 71, 74, 79, 81, and 84 for our example. I've chosen to make vertical lines.

Second, draw a box from the first quartile to the third quartile. This box is going to show you where 50% of the data lie, the middle 50%. And about 25% of the data falls in this whisker out in the left side. And about 25% of the data falls in this whisker out to the right-hand side. This is why it's sometimes called a box and whisker plot.

Different statistical packages might show this a little bit differently. You'll notice, the boxplots from this statistical package don't have the vertical lines out here at the edges, and that's fine. We can use boxplots to compare two distributions. For instance, if we're talking about the height of girls versus boys, we might be able to compare them by saying the spread or the variation with the girls is much less than the variation with boys.

We can see that not only in the width of the boxes but also in the total width from the minimum to the maximum in each of these two data sets. So we can use boxplots to use as sort of a summary distribution for the boys and for the girls.

So to recap, boxplots will allow us to display, visually, the five-number summary. We can interpret it to

see where the data points are close together-- that's where the vertical lines are close together-- and where the data points are further apart. We can analyze skewness, and we can look for it for symmetry as well. And we can use multiple boxplots on the same set of axes that will help us to compare two or more distributions. Good luck, and we'll see you next time.