In this tutorial, you're going to learn about the five number summary of a data set. Now the five number summary takes larger data sets and makes a more manageable and easier to understand. By breaking them down from lots of numbers to just five, they can help to summarise the center and variability. Two of the numbers in the five number summary are the smallest and largest, the minimum and the maximum.

So we'll talk about those first real quick here. So suppose you have the example of the Chicago Bulls basketball team. It's easy to see, just by inspection of the list, that the shortest person on the team is here at 71 inches tall. And the tallest person on the team is 84 inches tall. Those are two of the numbers in the five number summary. The three remaining numbers will be based on the median.

So just a little bit of review. The median measures of center of a data set. And it's the middle of an ordered set of data. Currently, this is alphabetical by last name. So we should rearrange it so that it's in height order.

Then we just work our way in until we find the middle number of the ordered data set. That number in the middle is the median, 79. What that leaves us with is two groups-- a low group and a high group.

What we can do is take the median of each of those data sets. So now we have 74 down here, 81 up here, and 79 in the middle. Those three numbers are called quartiles. And they divide the data set into four equal parts.

This is called the first quartile. The median is the second quartile. And then you have the third quartile up here. What you'll notice is that 25% of the data falls at or below the first quartile. Half the data set falls at or below the median. And 75% of the data falls at or below the third quartile.

So the five number summary consists of the minimum, the first quartile, the median, the third quartile, and the maximum. The benefits of this particular summary is that you'll notice is that about 25% of the data falls within each of these bands here. So what you can understand about the data set is where lots of data values lie.

For instance, there are more data values in a narrower range. There's the same amount of data values here between 79 and 81 as there are between 74 and 79. It's the same number of data values, but they fall in a more narrow range. So you can tell the data are more clustered together in this area than they were in this area.

And again, this makes it fairly obvious that 25% of the data falls below the first quartile-- at or below the first quartile. 50% falls at or below the median. 75% falls at or below the third quartile. And obviously, all the data falls at or below the maximum.

So to recap, the five number summary is a nice way to summarize a larger data set. Now it will work for pretty

much any size data set so long as it has five numbers, obviously. Then you can make those your five numbers for your five number summary. But it's ideally for larger data sets to summarise more values.

It consists of the minimum, first quartile, median, third quartile, and the max. And it allows us to understand where clusters of data points might be and where the data might be more spread out.

So we talked about the five number summary, which consists of the min, the max, and the quartiles. And so we had the first quartile, which was called the lower quartile also. The second quartile is the median. Typically we don't use the term second quartile. We use the term median. And then we have the third or upper quartile.

Good luck. And we'll see you next time.