

---

This tutorial covers histograms. Histograms are a particular display for quantitative data. Now, remember, quantitative data is numerical data. It can be measured. It can be used in arithmetic.

Once you have that collection of quantitative data, you start by placing them into bins. That is intervals of even size. For each interval, you calculate the frequency or the count. Then that count tells you the height of the bin, and this graph can help to display some information about our data collection. When we plot the bins, there there's no space in between, so it's different from a bar graph where there is space.

This term bin I keep using comes for an important process called binning. Binning is how you're deciding to pick and create those bins, those intervals of even size. This is a fairly simple process, but the decisions you make affect the outcomes of how your graph looks. If you change the bin size, you can change the shape and size of your graph significantly. If you have too many bins, your graph's going to appear really flat, pretty straight across, and you're not going to be able to find out much information from it.

I've made a short video showing these effects of bin size. We'll watch that now.

[VIDEO PLAYBACK]

- So this histogram shows the body fat percentage of 252 men. Right now, the interval

size or how wide bins are 10. Now, what this toggle here is change the bin size. So as I get smaller, we're starting to make more and more bins. And the shape of the graph stays pretty consistent. There's a lower part, and a higher part in the middle, and a lower part at the end. You can start to see a lot more variation in the heights of the bins as you make them smaller and smaller.

Now, the same thing can happen on the other end. If I start to make the bin sizes a lot wider, then we're not getting very much data, or we're not able to interpret very much from our histogram because the bins are including so much in a chunk. So here, it appears to be a pretty flat graph because the only two intervals that exist are both pretty similar. And there's this third one down here that's very small.

But when I'm making and changing the size of the bins, I'm manipulating how the data appears in our histogram. So the choice of how wide the bins are has an effect on how our histogram appears.

[END PLAYBACK]

Now that we can see the effects of bin size, I've decided to make bins of 5 feet wide to measure the heights of the cherry trees. Here is my data, and I've set up the frequency chart. I've listed all the categories. Those are my bins. They're all 5 feet apart. It starts at 60, which is my lowest value, and goes up to 90, which is just above my highest value.

Now, for the frequency, I want to count how many times my pieces of data appear. So for the 60 to 65

bin, I have 1, 2, 3, maybe four pieces of information. Now, here's where I have a decision to make. And it doesn't matter where I decide, as long as I stay consistent. If we notice, this category goes from 60 to 65, and this one goes from 65 to 70.

Where would I put a data value of 65? Where does that go? I can either put it in the lower bin with the 60 to 65, or I can put it in the upper bin. It doesn't matter what I choose, as long as I do the same thing throughout my chart. Because I have this 60 here, and that one has to go in this bin, I'm going to say that the 65 goes in the upper bin. And I'm going to make a little note of that in case I come across this situation later. So this goes upper. So for 60 to 65, I have just those three values.

For 65 to 70, I have 1, 2, 3, 3 again. For 70 to 75-- 1, 2, 3, 4, 5, 6, 7, 8. And then we decided that the 75s would go in upper bin, so I have 8. 75 to 80-- 1, 2, 3, 4, 5, 6, 7, 8 9, 10. 80 to 85-- 1, 2, 3, 4, 5. And then 85 to 90, I have 1, 2.

So now that I have my frequency table, I can make the histogram. Here's the histogram that reflects the information in the frequency table. For the first bin, from 60 to 65, there are 3 values. For the second bin, 65 to 70, there are again 3 values. For the 75 to 80 bin, it goes all the way up to 10-- 10 values.

Now, like I mentioned before, there's no spaces in between the bins this time. There's no spaces in between intervals. That's different from a bar graph, where there is spaces. So for the quantitative data for the histogram, there's no spacing. It's because the bins are right next to each other. 65 goes right

next to the next chunk that goes from 65 to 70. This has been your tutorial on histograms.