

## **Video: Comparing Histograms (2:14)**

**(0:00):**

Are final tricky issue is the correct way to use histograms to compare datasets. We are going to compare the data from 751 Daphne island finches and 43 Santa Cruz island finches. We're definitely going to use percentages rather than counts because there are so many more Daphne birds. The vertical scales on the histograms would be totally different. When doing a comparison we want the viewer to be able to use the height of the bars to compare the percentage of values. Now the heights of the bars are equivalent in terms of percentages, but we're not done because the two graphs have different vertical scales. We will need to make them the same for comparison. Let's set the scales from 0-40.

**(0:46):**

Here is the result of setting the vertical scales of the two graphs to be the same. We're closer but not done yet because the horizontal scales are not the same so we cannot directly use the heights of the bars to compare the data. The final step is to use a common set of bins to histogram the data. Here is the final result. We see that on the whole Santa Cruz finches have larger beaks than Daphne finches. The beak sizes of Santa Cruz finches are centered around a value close to 11 millimeters while daphne finches beak sizes are centered around 10 millimeters. There are no Santa Cruz beak sizes less than 8 millimeters while about 10% of the daphne finches are less than 8. A greater percentage of Santa Cruz finches have beak sizes larger than 12 millimeters. These are rough estimates and we can easily calculate the exact values but they give us a quick starting point for comparison.

**(1:41):**

Getting a common set of bins is a bit technical, but here's a way to do it. First you have to find a common range. You do this by finding the larger of the two data set maxima and finding the smaller of the two data set minima and then subtract the smallest from the largest. Then you bin both data sets using the same bin positions in the common range. You will have to experiment with then number of bins to make both histograms look reasonable.