Math 100
Distributions

Histograms display the distributions of quantitative data. For instance, 150 iris plants were allowed to grow and, after a fixed period of time, their sepal lengths were measured. Here is the distribution of lengths.

The mean and median lengths are 5.843 cm and 5.8 cm respectively. The distribution is pretty much symmetric, so it is not surprising these two numbers are close.

When we break down this data by species (there are 3 represented in the sample) we see that there is some difference in sepal lengths when viewed by species, at least, in our sample.
Statistics is used to quantify how likely it is that this difference is due to variability from one species to the next instead of other random factors.

The mean and median are less useful in describing a *bimodal distribution* as appears next. This plot represents the frequencies of various lengths of eruptions of the geyser Old Faithful in Yellowstone National Park.

Here it is probably more useful to describe the *mode* or modes of the distribution.