Histograms in Python

Histograms are relatively simple things to make in python. We will use matplotlib.

> import matplotlib.pyplot as plt

The function we will use is plt.hist which has the following syntax

> n,bins,patches = plt.hist(inputarray, bins=100)

This will not only plot the histogram for you, but it will return the number of elements in each bin (n) and the bins themselves (bins). You can name the three variables holding the output anything you want, but you need three. Don’t worry about what the “patches” are.

Keywords for the hist() function

You can denote the color and linestyle in the same way as you do in plt.plot()

- **bins**: number of bins you want to have. can also be a list of bin edges.
- **range**: lower and upper range of the bins
- **normed**: “= True” means you get a probability distribution instead of just raw number counts
- **histtype**: ‘bar’ = traditional style, ‘step’ = a line plot. looks better usually
- **Weights**: this is an array of values that must have the same size as the number of bins you have. This will be a factor by which you will multiply the number count of each bin. In other words, it will make the “number of elements” output be n*weights instead. This is a good way to normalize your histogram outside of just using the normed variable. For example, if you wanted to plot the fraction of objects in each bin, you would set weights equal to an N sized array (N = number of bins you have) where each element of the array is equal to 1/(total # of objects).

Making legends in Python

This is super easy. When you plot anything (even a histogram) just add in the keyword label = “whatever you want to call these data points/lines”. Then, when you have plotted everything you need, type in the command plt.legend() and python will make your legend!

The only really important keyword for legend is loc = “string”

This will set the location of your legend. The strings you can use that will be understood are: