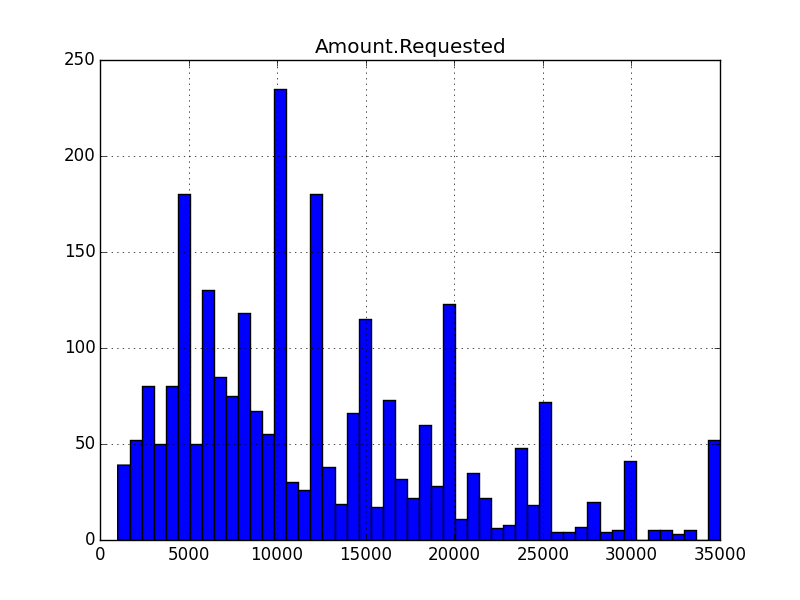
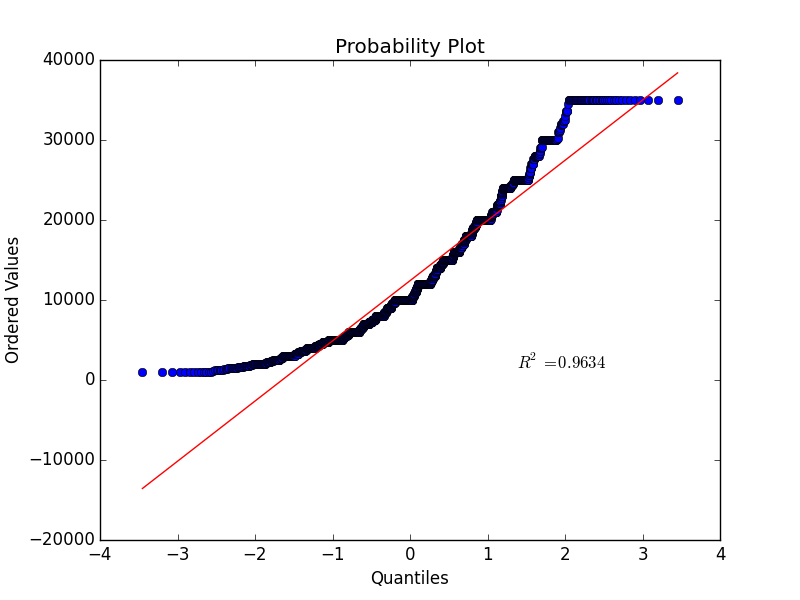
Typical plot from Excel showing the range in values (Y) along the sequence they are occurring (X).

  
This is the frequency (Y) along the values (X). Note that I added binning here to increase the details of the distribution. Histograms are used to show the distribution pattern, but hides the details of the what the exact values might be for both Y and X. For example, along X, the value 10,000 does not actually correspond to this specific value in the data. What it does say though is that for amounts requested, the majority is in the middle to lower quartile. Few amounts actually are at the high end (which we know is over or around 100,000 from looking at the data).

loansData.hist(column='Amount.Requested', bins=50)

plt.savefig("ARhisto.png")

**QQ Plots**



This QQ Plot is comparing one sample of Amount.Requested to the Normal theoretical distribution. If the data did follow the Normal assumed distribution, then the points on the QQ plot will fall approximately on a straight line. You see that they don’t, which backs up what the histogram shows. This is defined as skewed (non straight) with ‘heavy tails’, meaning it has extends similar variates at both ends. It’s amplified on the QQ plot compared to the histogram because it is being compared to the Normal distribution.

This plot is supposed to reveal differences in location, spread and shape more clearly than histograms because it is compares to a distribution. In the Python syntax used, you will see you are comparing a Normal distribution.

The quantile is plotted along X, which is each of any set of values of a variate that divide a frequency distribution into equal groups, each containing the same fraction of the total population. Again, these are not absolute values. The same holds true for Y, which is a representative value compared to the normal.