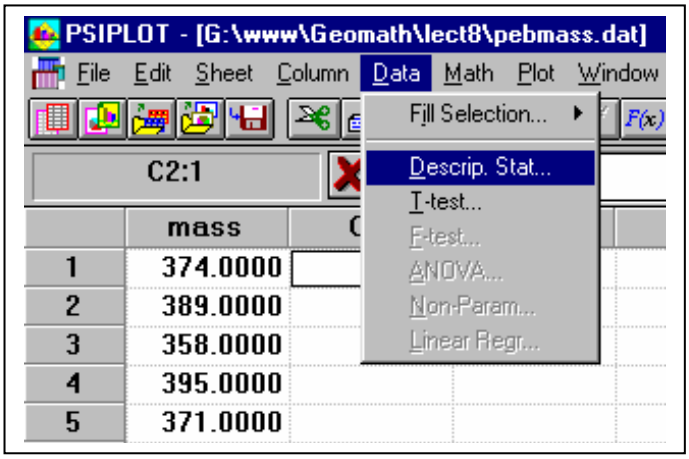
Geology 351 Mathematics for Geologists
In-Class Exercise on Statistics
Generating histograms and Computing Probabilities with PSIPlot

The following computer activities are designed to accompany today's lecture. The exercise uses data presented by Waltham in Chapter 7, Table 7.1. The data consist of the masses of pebbles collected at a beach. This dataset - *pebmass.pdw* – will be provided for you in my shared directory or on the H:\drive. Copy that file to your G:/Drive.

I. Descriptive Statistics:
In our discussion we describe various statistical properties of the sample, such as its mean, variance and standard deviation. PSIPlot can be easily used to generate the descriptive statistics of your data. To do that, first -

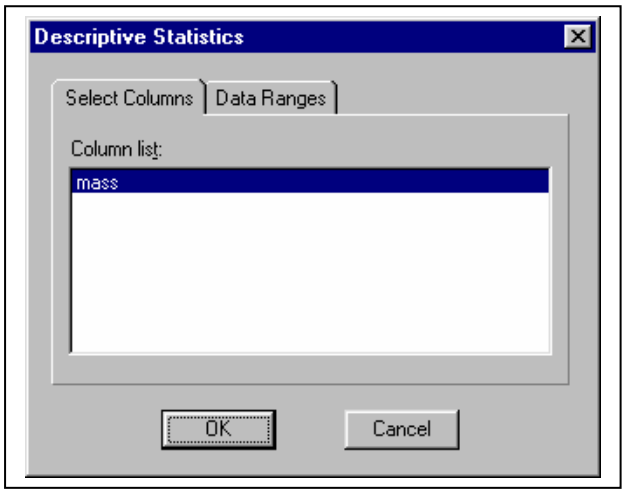
OPEN PSIPlot and then from FILE OPEN **select** the file *pebmass.pdw*.

Then click on **Data, Descriptive Statistics** (see below).



The following window should appear

You will have only one variable in your column list so just **click OK**.



The following window will appear:

Note that the mean, standard deviation, variance and other statistical properties are summarized in this table for the list of pebble masses.

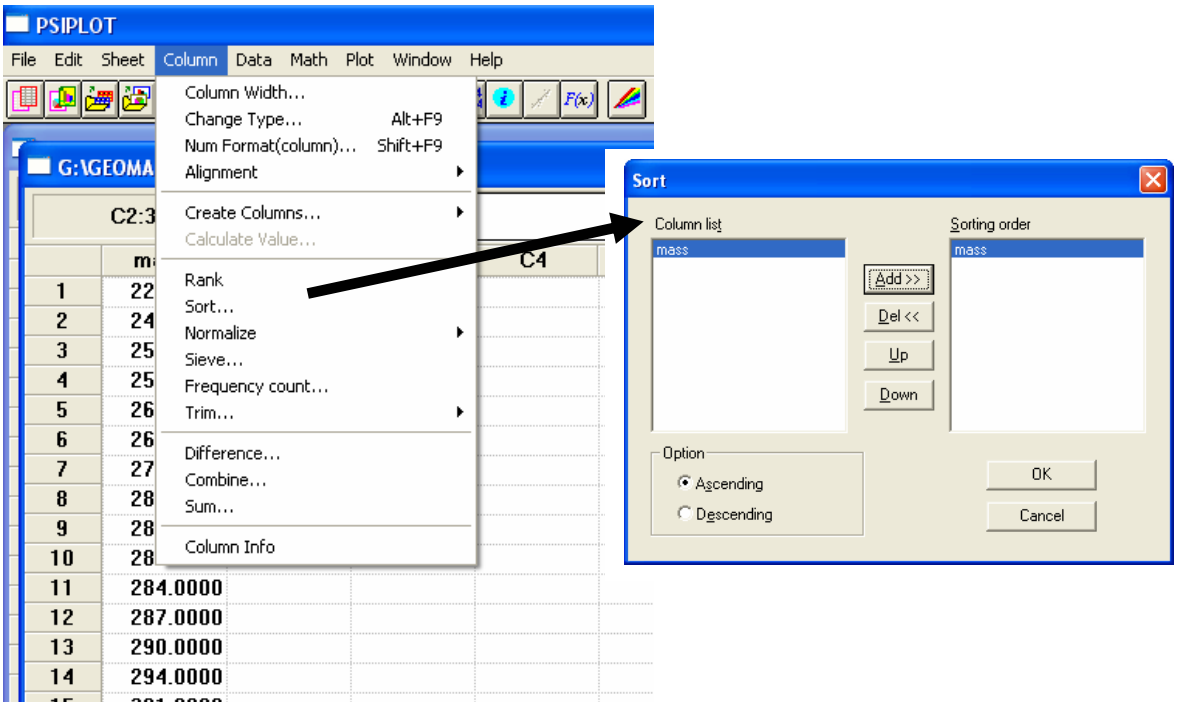
COLUMN NAME:	MASS
Number of rows:	100
Number of valid points:	100
Number of missing points:	0
Number of negative values:	0
Number of positive values:	100
Number of zero values:	0
Minimum value:	224.00000000
Maximum value:	454.00000000
Inter range value:	230.00000000
Median:	352.50000000
Sum of row value:	35018.0000000
Sum of absolute value:	35018.0000000
Arithmetric mean:	350.18000000
Geometric mean:	346.83461326
Quadratic mean:	353.37806949
Harmonic mean:	343.33964895
Absolute mean:	350.18000000
Sum of squares:	12487606.0000
Variance:	2272.75515152
Standard deviation:	47.67342186
Absolute deviation:	38.37640000
Standard error:	4.76734219

II. Examining the Distribution of Pebble Mass using the Histogram:

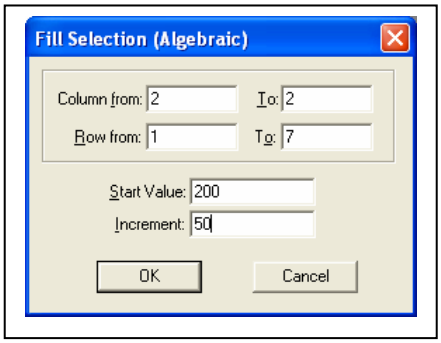
The instructions below will take you through the generation and plotting of a histogram using PSIPlot. The histogram provides a graphical display of the distribution of your data.

In order to generate a histogram you have to subdivide your data. These subdivisions or **bins** usually correspond to intervals of data having the same size. First, let's sort the data into ascending order. Click on the **Column button** (see below) on the top menu bar, click on **sort**, and then **select ascending** in the **sort window.** Click Add to transfer the column over into to the sorting order window. Then, click OK. Mass will now be sorted in increasing order, from the smallest to largest mass. What are the limits (maximum and minimum values) of the pebble mass data?

Minimum Mass = _____; the Maximum Mass = _____

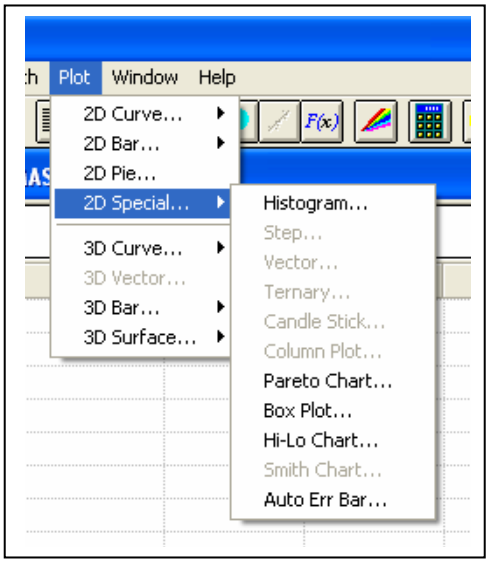


Based on the range, we will adopt the 50gram subdivisions or "bins." That Waltham uses to subdivide the pebble mass data. In **column3** type in 200 (in cell 1), 250,300, … etc to 500. Label that column **BIN** (or use the fill selection option – see right).



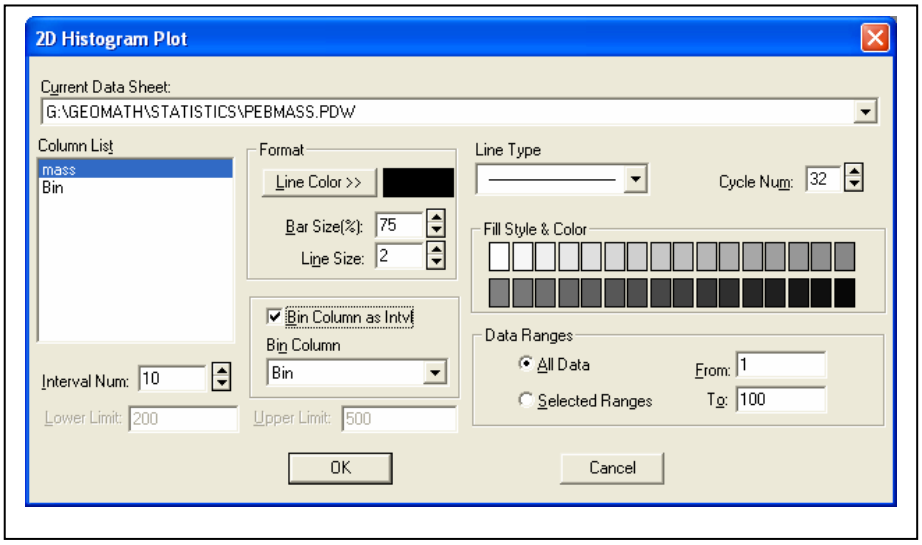
The histogram is just a bar-plot of the number of data points that fall in each subdivision or bin. In the "old-days," one just sat down and counted the number of data points falling into each subdivision. PSIPlot will do this for you. Create the histogram data by going to **Plot - 2D Special**

Click on Plot - 2D Special - Histogram

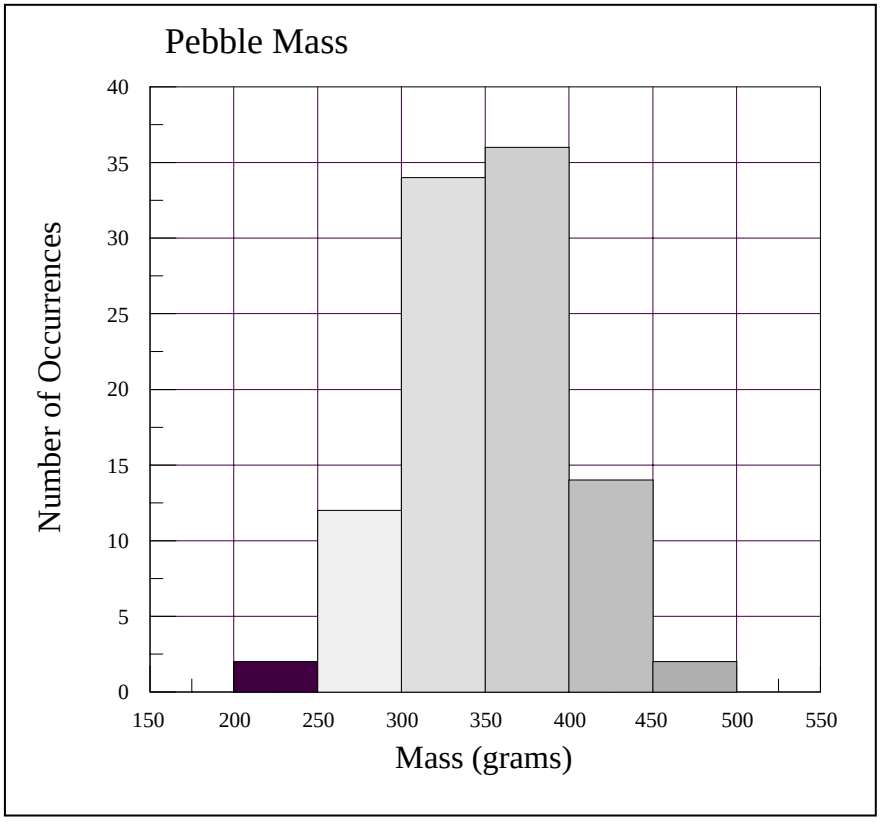


The following window opens. **Highlight** the Mass column, Click **Bin Column as Interval** checkbox with 10 intervals as the default.

Note that instead of using the Bin column, we could have specified the lower and upper limits and values and number of intervals.



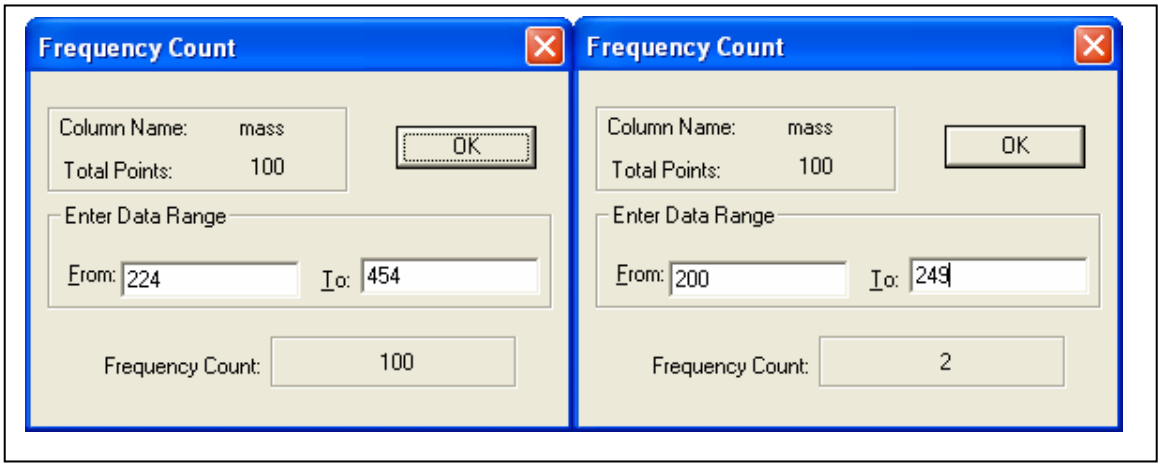
Click OK - You should get the following plot called a histogram.



We did not have to sort the values first, but the relationship of the sorted data to the histogram is a more direct one. As you scroll down through the column of sorted masses you see that there are a much greater number of values with masses in the 340, 350, 360 gm range. The frequency of occurrence is greater over this range than over others.

Frequency Count

PsiPlot does not automatically return the frequency count (number of samples falling in the interval). But there is a relatively easy way to obtain this information. Start by clicking on **Column** and then **Frequency Count.** Note that the following window will appear. You can type in the individual ranges used to construct the histogram and quickly obtain the number of sample values falling in that interval (see blow for interval extending from masses of 200 grams to 250 grams).

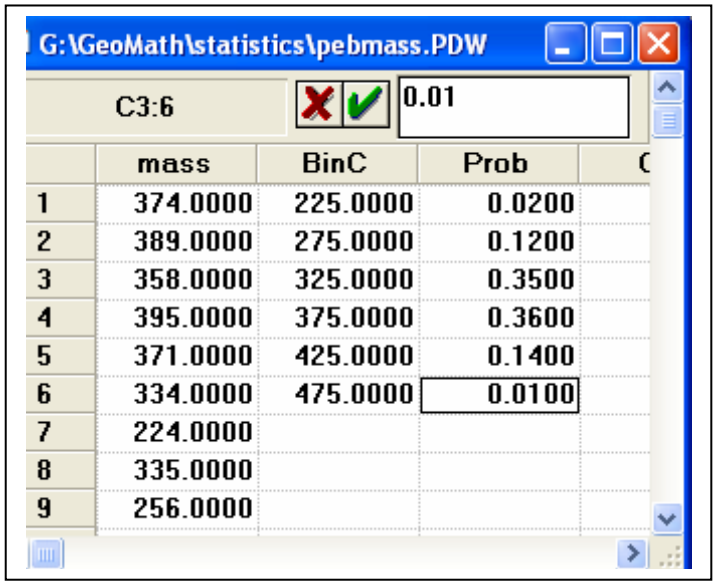


Range (g)	Number
201-250	2
251-300	12
301-350	35
351-400	36
401-450	14
451-500	1

See Table 7.3 of Waltham.

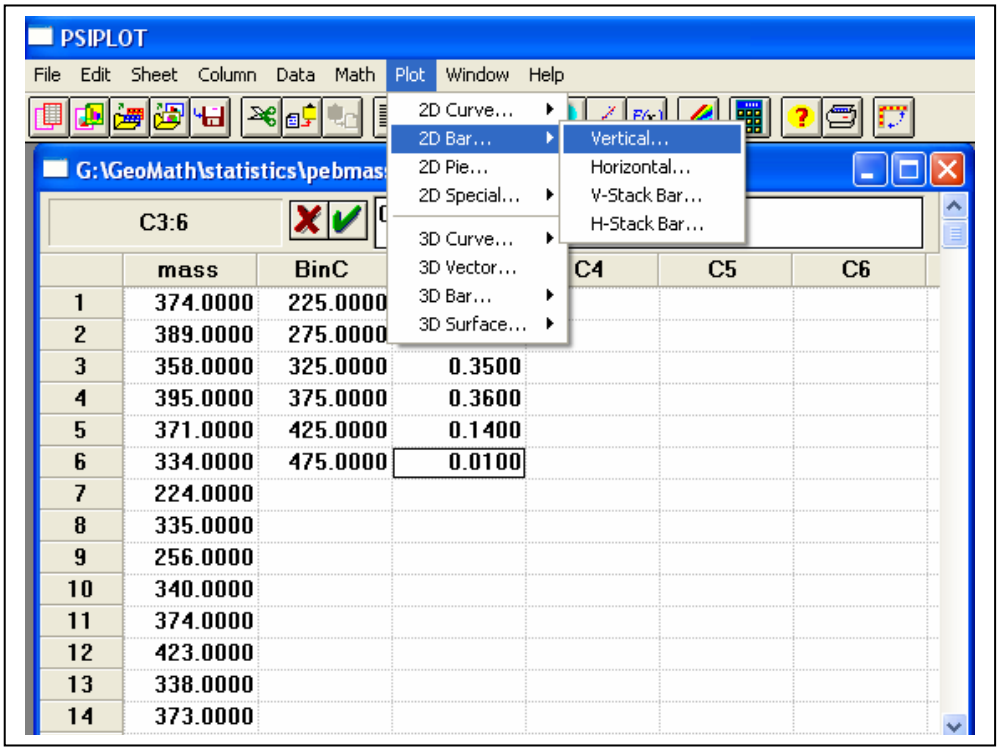
The number of occurrences observed in each of these intervals, divided by the total number of observations in the sample represents the probability of finding a pebble with a mass in that 50 gram interval in that area of the beach. We are assuming that the sample is representative of the distribution of the pebbles in that area.

Construct two additional columns, BinC (bin centers) and PROB (see below), in your worksheet.



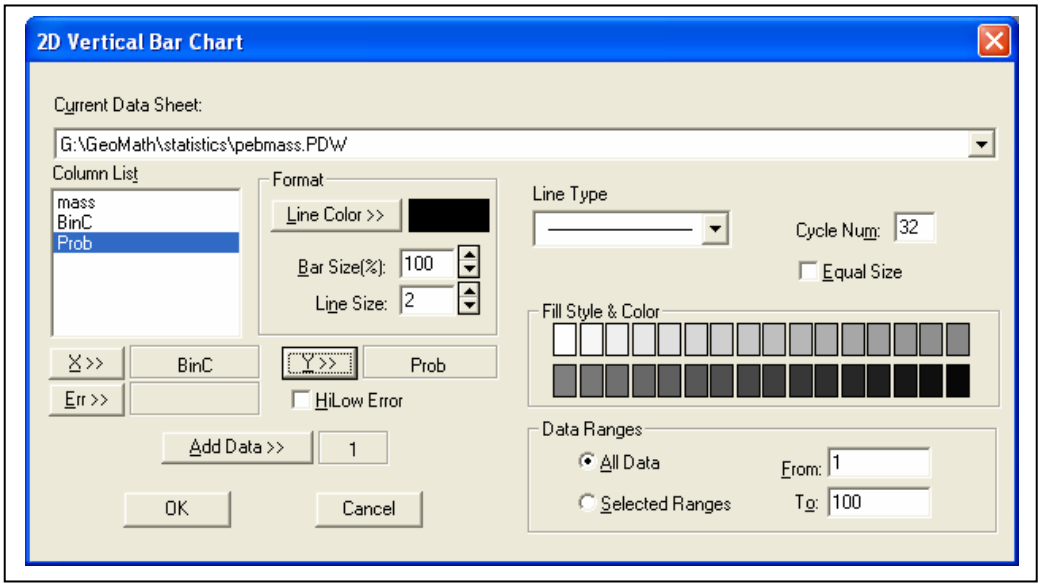
	mass	BinC	Prob
1	374.0000	225.0000	0.0200
2	389.0000	275.0000	0.1200
3	358.0000	325.0000	0.3500
4	395.0000	375.0000	0.3600
5	371.0000	425.0000	0.1400
6	334.0000	475.0000	0.0100
7	224.0000		
8	335.0000		
9	256.0000		

Now construct a histogram, but use the **Plot - 2D Bar - Vertical** plot option (see below).

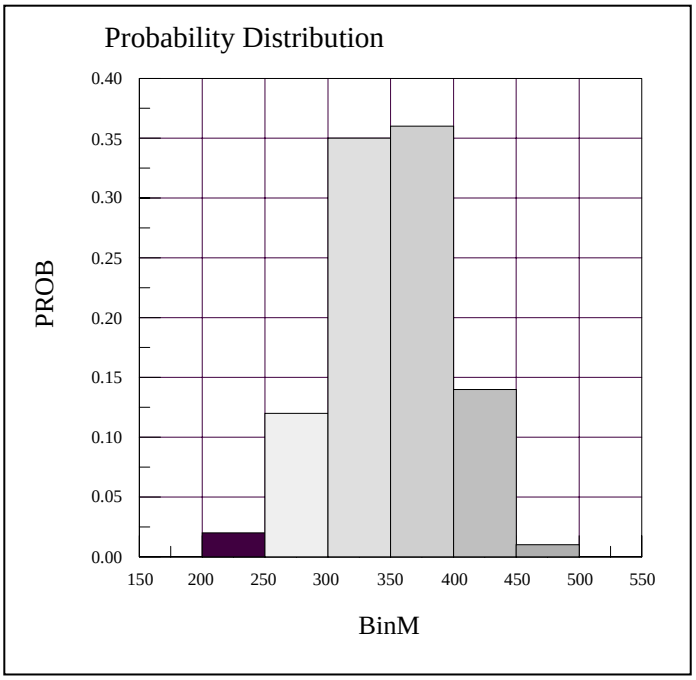


	mass	BinC		C4	C5	C6
1	374.0000	225.0000				
2	389.0000	275.0000				
3	358.0000	325.0000	0.3500			
4	395.0000	375.0000	0.3600			
5	371.0000	425.0000	0.1400			
6	334.0000	475.0000	0.0100			
7	224.0000					
8	335.0000					
9	256.0000					
10	340.0000					
11	374.0000					
12	423.0000					
13	338.0000					
14	373.0000					

Declare **BinM** as your x-variable and **Prob** as your y-variable (see below). Don't worry about the **Err>** or **High Low Error** options. **Add data** and **OK**



You should get the following plot. Note the similarity of the ***probability distribution*** to the frequency histogram.



Thursday - April 8th - 2004
Today's Assignment -

<u>Before you leave class,</u>

1) **generate a histogram of pebble masses**
2) **place your name in a text box on the plot**
3) **Write down the average and standard deviation of the pebble masses on your plot**
4) **print out the histogram of pebble masses**
5) **turn it in before leaving today**