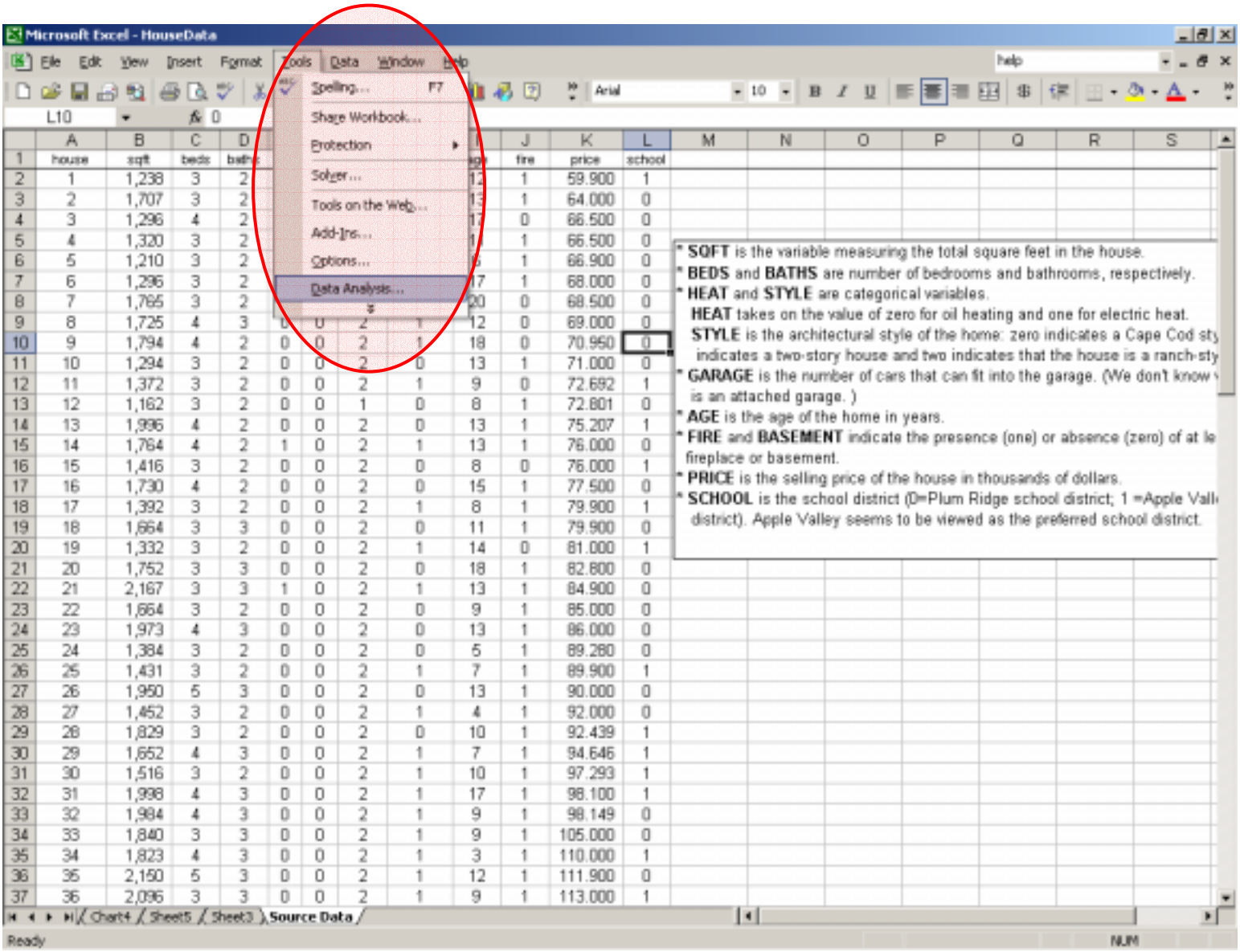
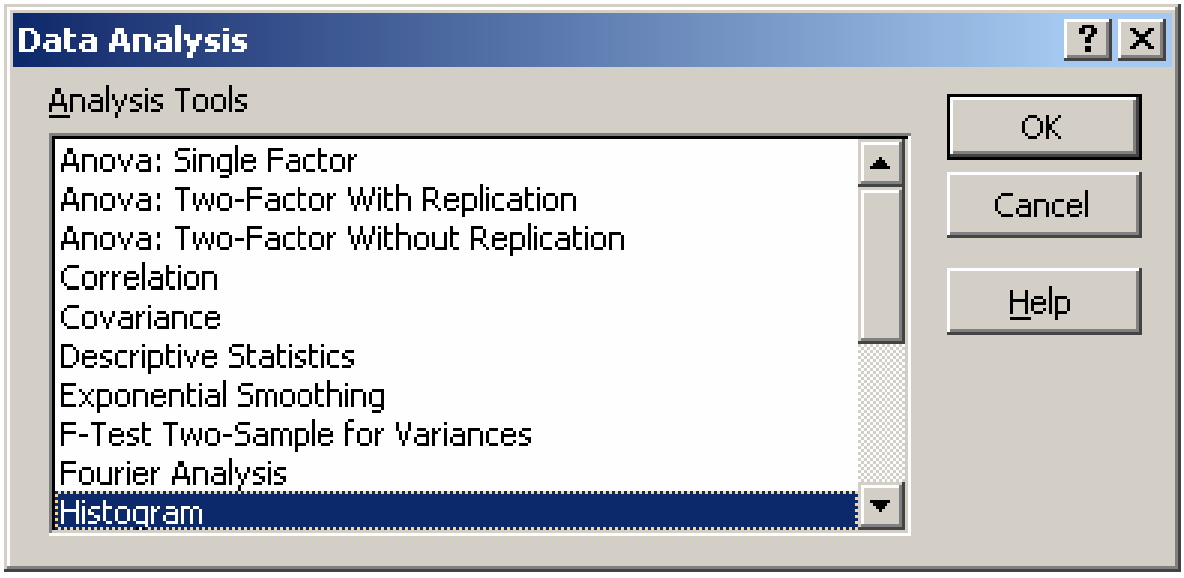
Using Excel's Histogram command

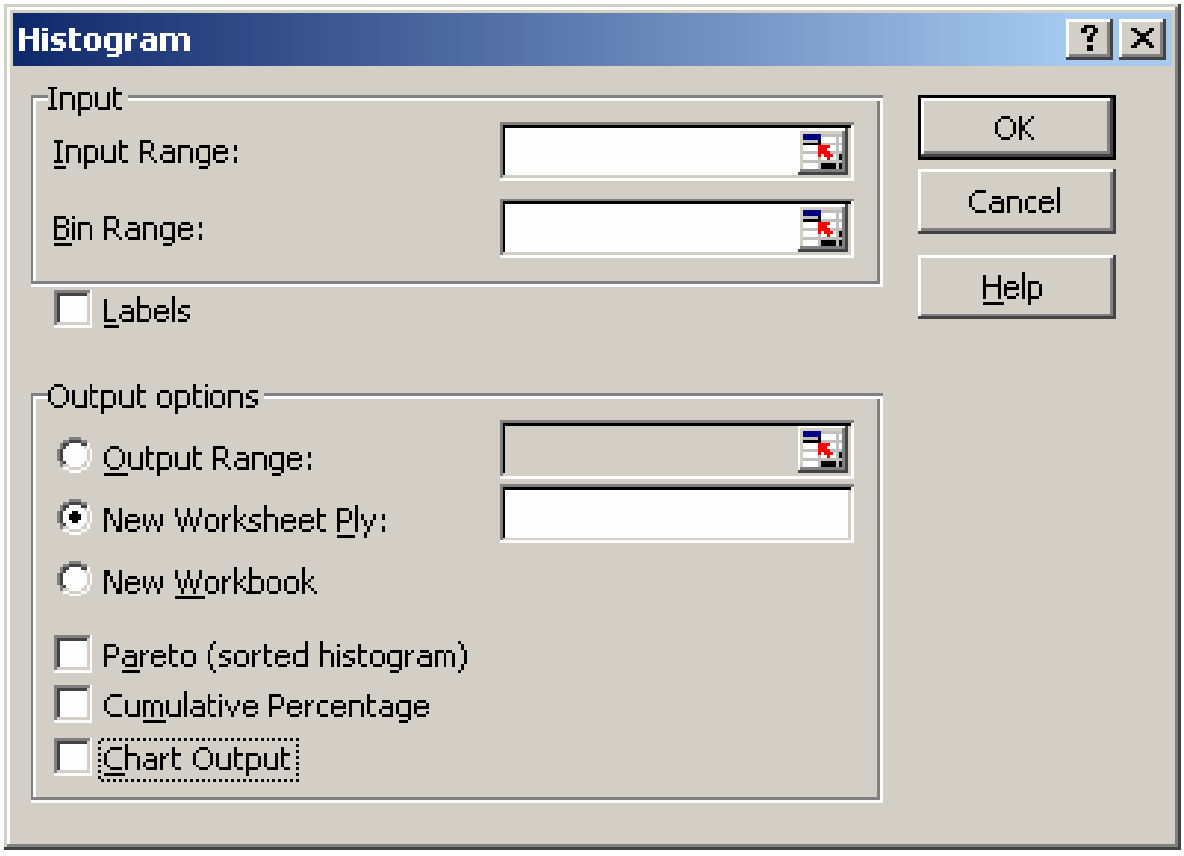
This handout briefly discusses the *Histogram* command in the Data Analysis Add-In of Excel. It assumes that the Data Analysis Add-In has been installed. We will use the Housing Data as the example data set. Our goal is to generate a histogram for the *square footage* of all houses in the data set. From the Command Menu, choose *Tools,* then *Data Analysis* as in the graphic below:



The next dialog box for *Data Analysis* appears, you choose the option *Histogram* from the list and press *OK*.

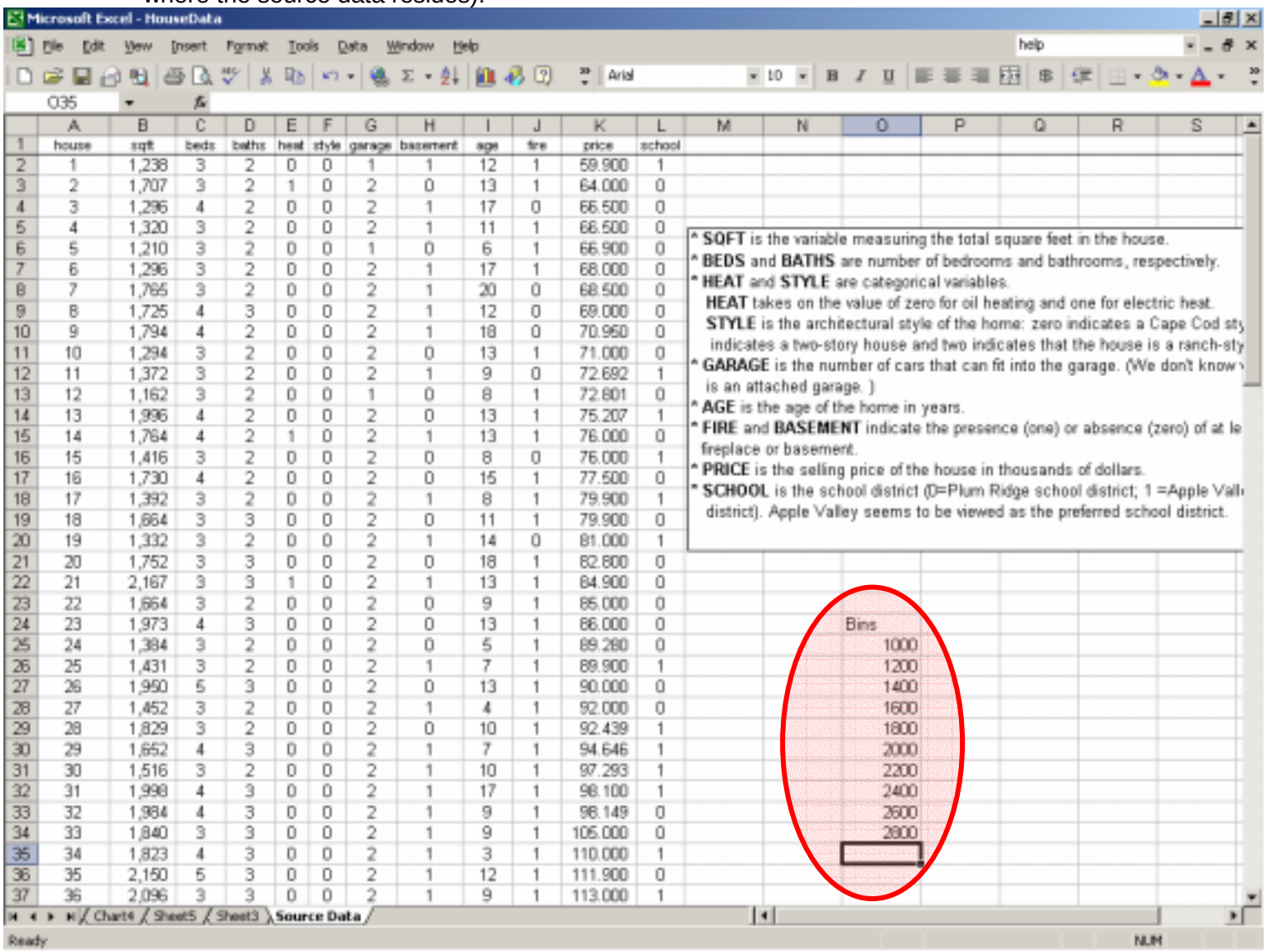


The following dialog box then appears:



Input Range is the range in the Excel spreadsheet that contains the numerical values for which we want to generate a histogram. By pressing the *red arrow button* on the right, Excel responds with:

and we can now freely select the range in the spreadsheet directly. You may also type the starting and ending address of the range manually (e.g.: *B1:B109*). Pressing *Enter* on the keyboard will take you to the *Bin Range* field. *Bins* are classes in Excel language. If you leave this field blank, Excel will automatically compute what it thinks are the "best" classes. But, it will likely end up with class boundaries that you won't like, in this case, e.g., the first class starts at 816 sq.ft. and ends at 1015.3 sq. ft., whereas we would rather like to see class boundaries like 800 through 1000, 1000 through 1200, etc. In this case we need to specify classes for the bins somewhere in the spreadsheet, something like this (the bins could be in another sheet as well, different from the sheet where the source data resides):

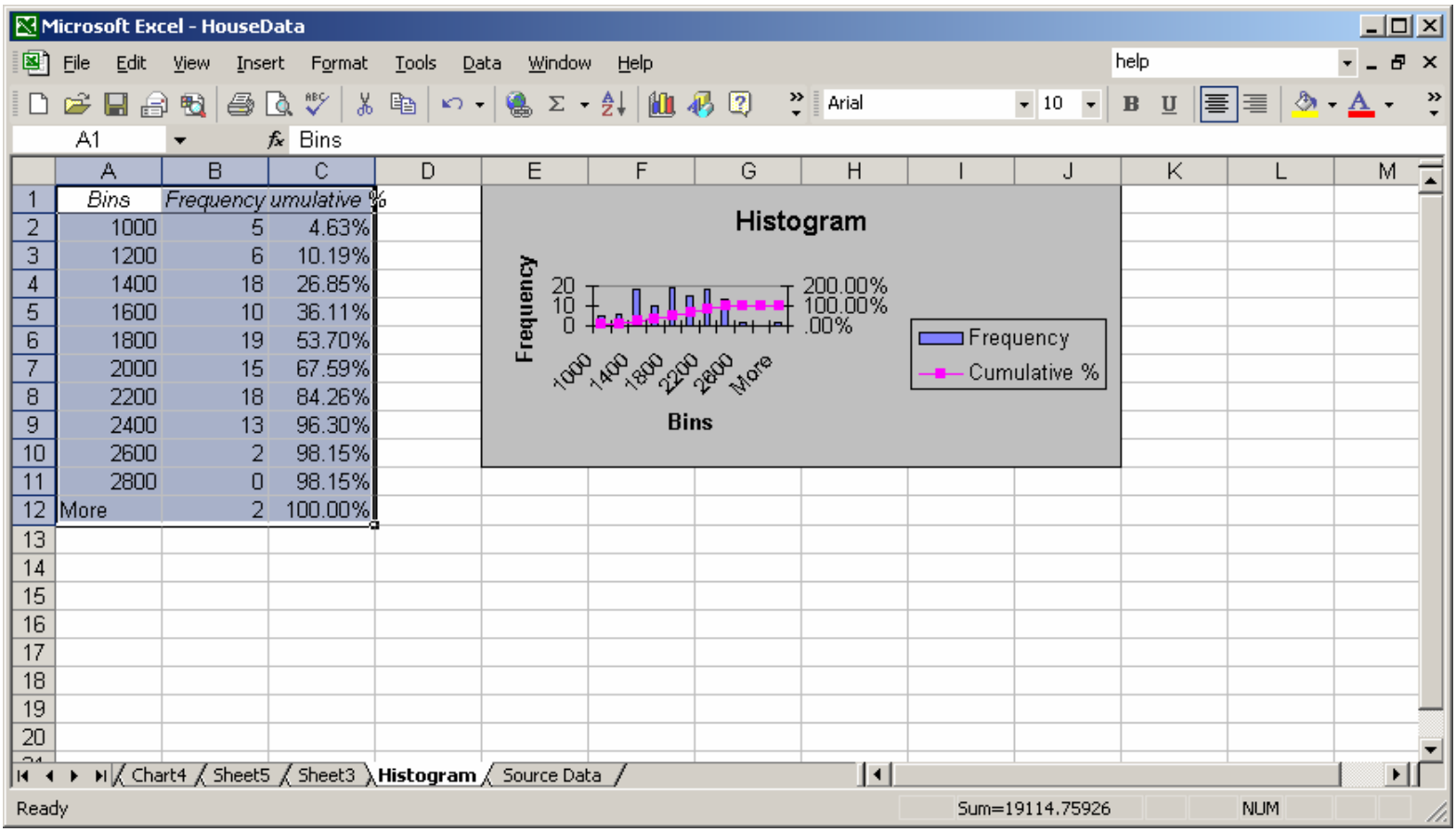


Excel will automatically create a last *bin*, and call it "More" for the values that are higher than 2800. By pressing the red arrow button on the dialog box at the *Bin Range* field,

we can select the cells that contain our classes (or again, type in range manually). Be sure to select the cell above the first one as well in your range if you check the *Labels* box, otherwise Excel thinks that your first value ("1000") is actually a label and not the ending boundary of your first class.

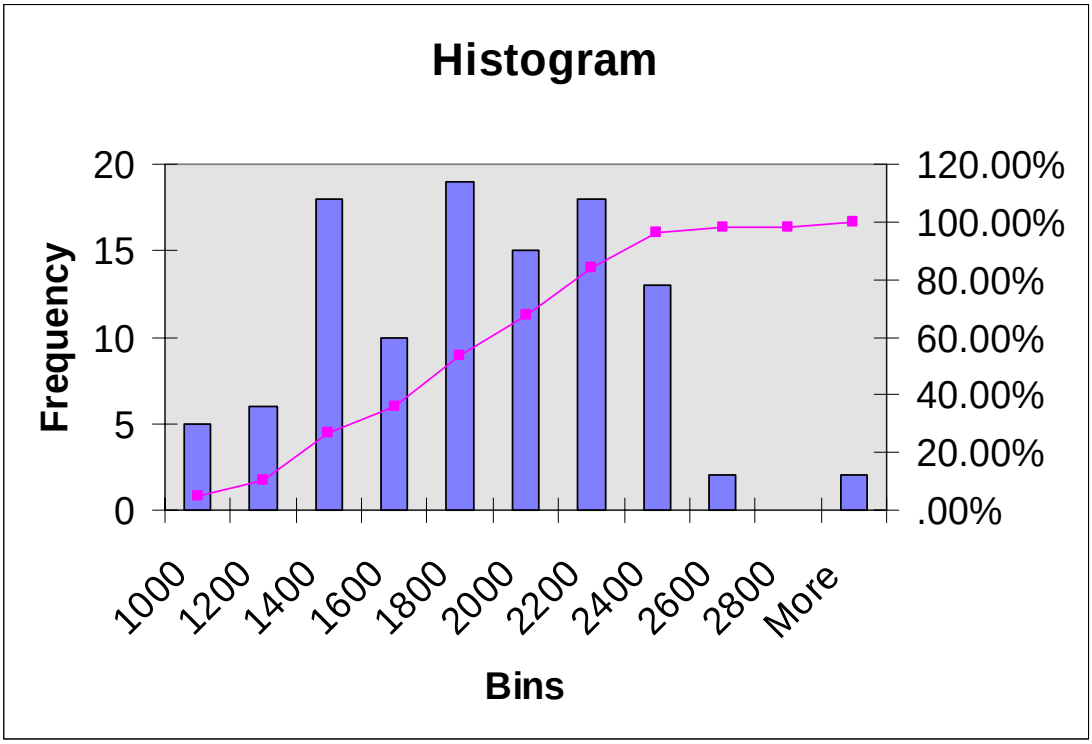
For the output options, we can decide to put the histogram on the same spreadsheet (first option), in which case we need to specify where the results should go (the field "*Output Range"*), or put the results in a new sheet in the same book (we may give it a name such as "histogram" in our example) or we can choose to put the results in a new workbook.

The last two checkboxes allow you to compute a *cumulative frequency distribution* and have the results displayed in a *chart* (graph), which was our goal from the outset. After pressing the *OK* button, the new sheet called "histogram" is created which looks as follows:



	A	B	C
1	*Bins*	*Frequency*	*umulative %*
2	1000	5	4.63%
3	1200	6	10.19%
4	1400	18	26.85%
5	1600	10	36.11%
6	1800	19	53.70%
7	2000	15	67.59%
8	2200	18	84.26%
9	2400	13	96.30%
10	2600	2	98.15%
11	2800	0	98.15%
12	More	2	100.00%

The column *Frequency* (and *Cumulative* %) needs to be interpreted as the number of observations *less than or equal to* the value under *Bins,* i.e., there are 5 observations with a square footage less than or equal to 1000, 6 observations with values between 1000 and 1200, etc.
The charted histogram is probably a little too small, but by clicking on it, you can drag it and stretch it and if you want to do some more basic formatting, you may get it to look like this:



Wow, that's one good looking histogram with very little effort!