

Basic MS Excel

Version 0.5

Authors: Maciej Swat, Ariel Balter, James A. Glazier

Please send corrections and Revisions to: mawat@indiana.edu

Extra credit may be offered

Last Updated: Friday, February 10, 2006

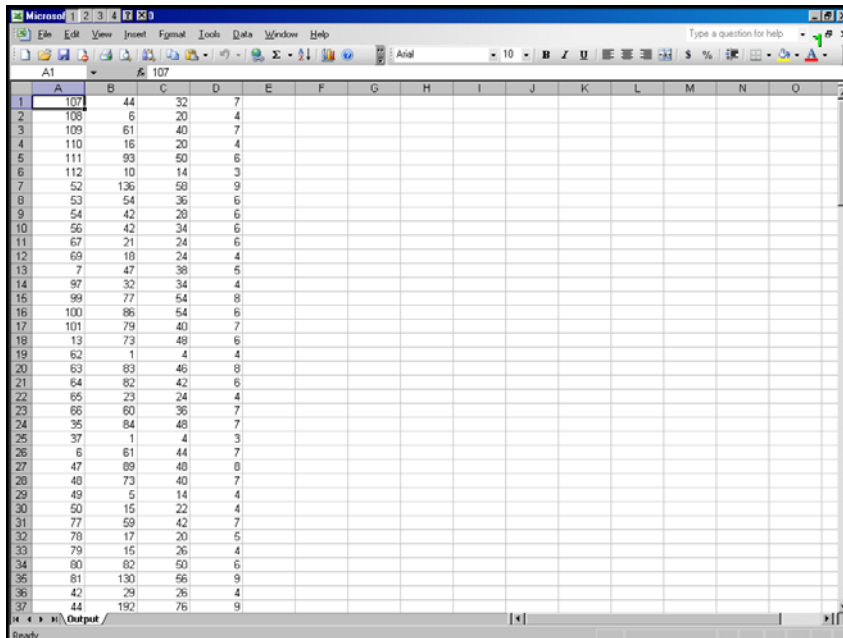
Creating histograms in MS Excel

MS Excel often will serve as a useful analytical tool for doing quantitative analysis of your data. In this write-up I will try to guide you through the steps necessary to create histograms using Excel.

Let's imagine you have the following data to analyze:

73	131	64	8
74	46	32	6
75	57	50	5
76	16	18	5
77	106	58	8
79	6	10	3
80	84	48	7
13	70	46	5
2	128	66	7

Where the first column denotes the cell id, the second the cell volume, the third the cell surface area and the fourth the cell's number of sides. Now, suppose we want to plot a histogram of how often a cell with a given number of sides appears in our data. Of course we could do it manually, by counting all the occurrences, but a faster and better way is to use Excel capabilities. It is interesting that certain features of Excel are underutilized and producing histograms like those described above is a perfect example of such underutilization. So without too much ado, let's begin. I am importing a text file with data analogous to the above example into MS Excel:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	107	44	32	7											
2	108	6	20	4											
3	109	61	40	7											
4	110	16	20	4											
5	111	93	50	6											
6	112	10	14	3											
7	52	136	50	9											
8	53	54	36	6											
9	54	42	20	6											
10	56	42	34	6											
11	67	21	24	6											
12	69	18	24	4											
13	7	47	38	5											
14	97	32	34	4											
15	99	77	54	8											
16	100	86	54	6											
17	101	79	40	7											
18	13	73	48	6											
19	62	1	4	4											
20	63	83	46	8											
21	64	82	42	6											
22	65	23	24	4											
23	66	80	36	7											
24	35	84	48	7											
25	37	1	4	3											
26	6	61	44	7											
27	47	89	48	8											
28	48	73	40	7											
29	49	5	14	4											
30	50	15	22	4											
31	77	59	42	7											
32	78	17	20	5											
33	79	15	26	4											
34	80	82	50	6											
35	81	130	56	9											
36	42	29	26	4											
37	44	192	76	9											

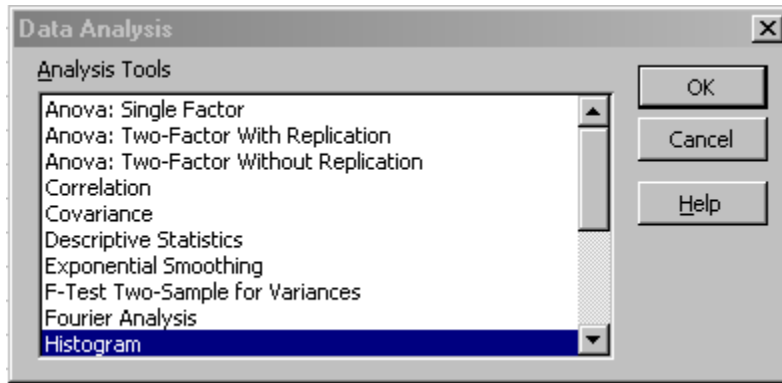
The first step in creating the histogram is to input bin ranges in one column in ascending order. Because we want to plot how often 1-, 2-, 3-, 4,..., sided bubbles appear, our bin ranges are simply consecutive integers. Well, 1 or 2 sided bubbles do not make sense, but I am just illustrating the concept of bin ranges that MS Excel uses. By analogy if you were to count how many people are of age 0-10, 10-20, 20-30 years, your bin ranges would be 0, 10, 20, 30. Here is how your Excel spreadsheet with bin range inputs looks (column F):

	A	B	C	D	E	F	G	H	I	J	K
1	107	44	32	7		0					
2	108	6	20	4		1					
3	109	61	40	7		2					
4	110	16	20	4		3					
5	111	93	50	6		4					
6	112	10	14	3		5					
7	52	136	58	9		6					
8	53	54	36	6		7					
9	54	42	28	6		8					
10	56	42	34	6		9					
11	67	21	24	6		10					
12	69	18	24	4							
13	7	47	38	5							
14	97	32	34	4							
15	99	77	54	8							
16	100	86	54	6							
17	101	79	40	7							
18	13	73	48	6							
19	62	1	4	4							
20	63	83	46	8							
21	64	82	42	6							
22	65	23	24	4							
23	66	60	36	7							
24	35	84	48	7							
25	37	1	4	3							
26	6	61	44	7							
27	47	89	48	8							
28	48	73	40	7							
29	49	5	14	4							

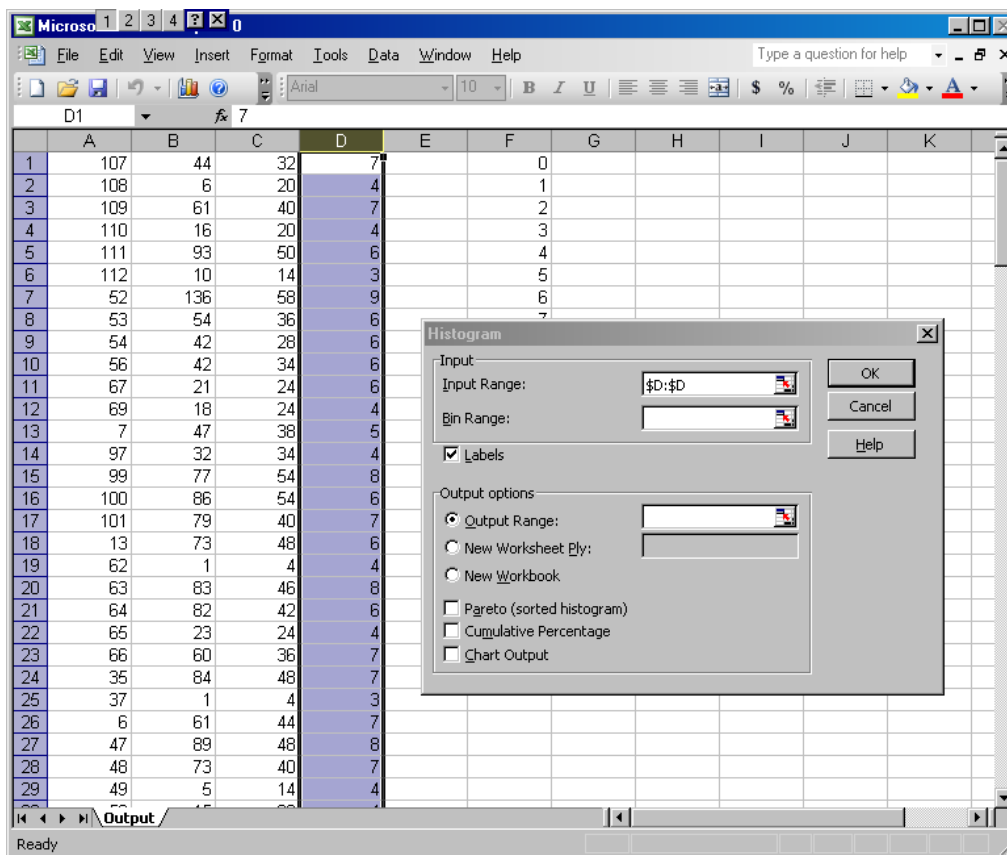
As you may have already noticed, Column D contains information on how many sides a given bubble has.

Let's do the histogram. Before going further, **please unselect** any items in the spreadsheet. Now select column D (the one that we will be analyzing) by clicking on the letter D at the top of the column.

Now, go to [Tools->Data Analysis...](#) .The following dialog will pop-up:



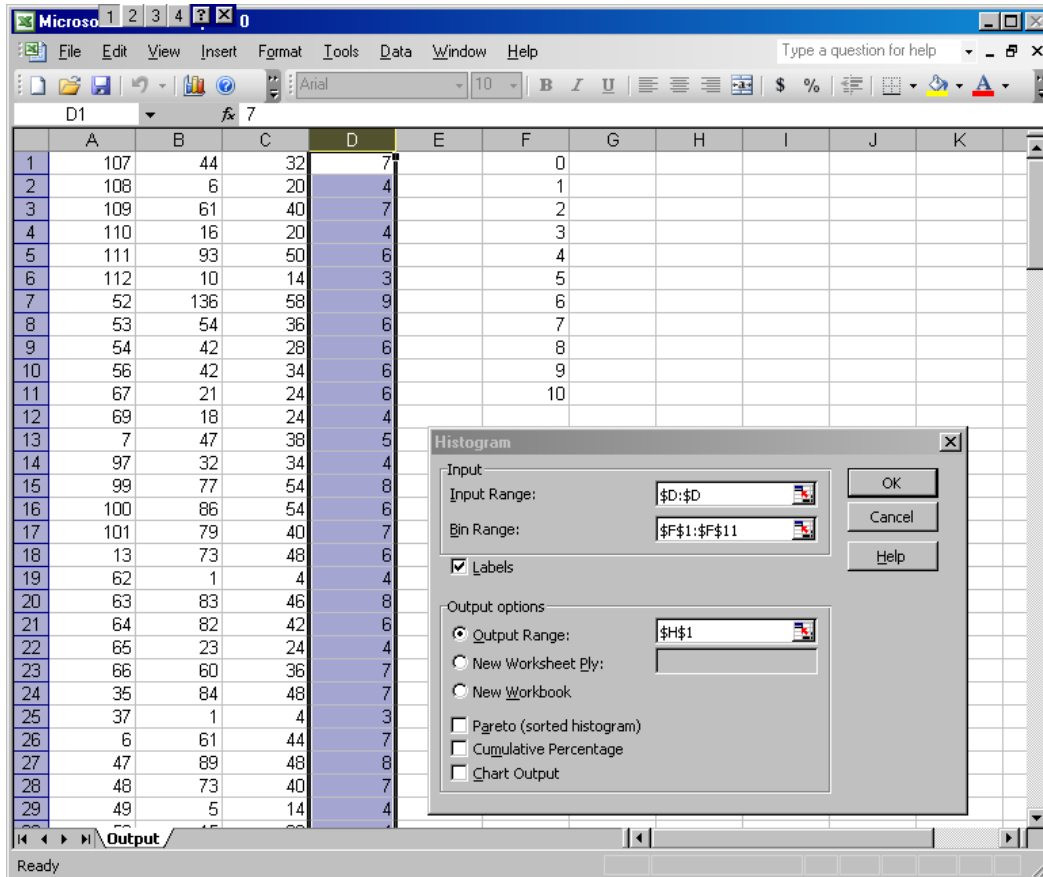
Select Histogram from the list of analysis tools and click OK.



Do not click OK now. You need to do a little bit more work.

In the *Histogram* dialog click on the text edit field called *Bin Range*: once there, go to the spreadsheet and select the cells that contain the bin ranges (remember, they are the numbers 1,2,3,..., in column F). Once you have made your selection, the field *Bin Range*: will be filled with the correct range (in the format understood by Excel). You can also edit this field manually, but at this time, let's stick to the point-and-click method.

Now, select where you want the histogram data output. – this has to be a place in your spreadsheet which has enough space for Excel’s histogram data. Go to the field called **Output Range** Click there, then go to the spreadsheet and click on the cell that will mark the beginning of the region where your histogram data will go. See the screenshot:



As you can see, I have chosen cell H1 to be the beginning of the region where Excel will output the histogram data. Now click **OK**. Hopefully you will obtain the following result:

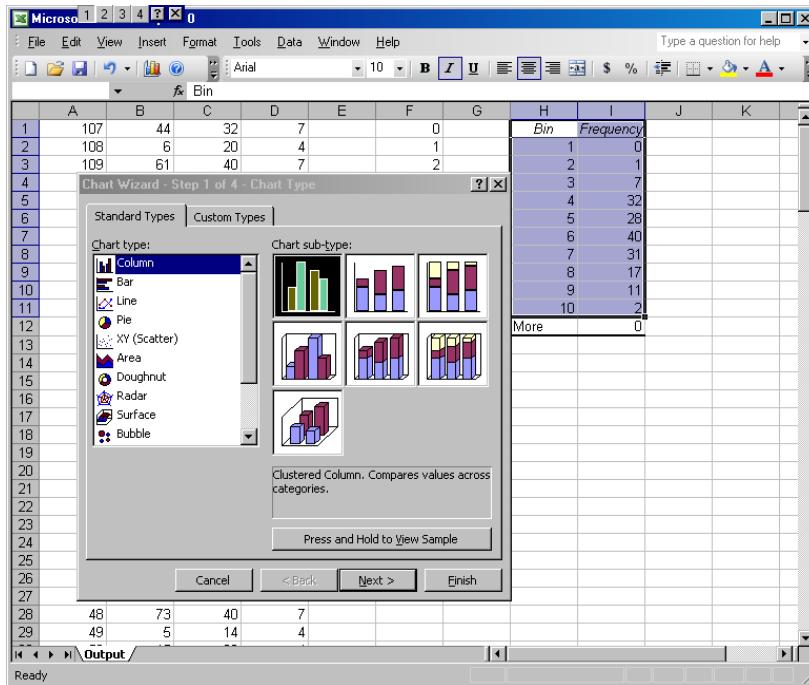
	A	B	C	D	E	F	G	H	I	J	K
1	107	44	32	7		0		Bin	Frequency		
2	108	6	20	4		1		1	0		
3	109	61	40	7		2		2	1		
4	110	16	20	4		3		3	7		
5	111	93	50	6		4		4	32		
6	112	10	14	3		5		5	28		
7	52	136	58	9		6		6	40		
8	53	54	36	6		7		7	31		
9	54	42	28	6		8		8	17		
10	56	42	34	6		9		9	11		
11	67	21	24	6		10		10	2		
12	69	18	24	4				More	0		
13	7	47	38	5							
14	97	32	34	4							
15	99	77	54	8							
16	100	86	54	6							
17	101	79	40	7							
18	13	73	48	6							
19	62	1	4	4							
20	63	83	46	8							
21	64	82	42	6							
22	65	23	24	4							
23	66	60	36	7							
24	35	84	48	7							
25	37	1	4	3							
26	6	61	44	7							
27	47	89	48	8							
28	48	73	40	7							
29	49	5	14	4							

As you can see, we have all the data that tells us how many occurrences of a bubble with a given number of sides were found.

Remark: If you are on Mac every time I mention the **Control** key you will use the **Apple** key.

Charts

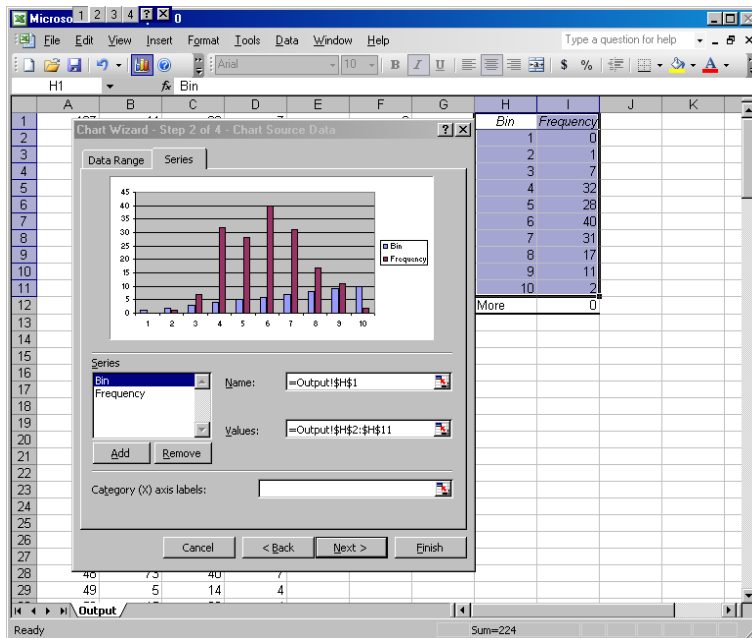
We can now plot the data. Select the columns **Bin** and **Frequency** and go to **Insert->Chart.....** Your screen should look like this:



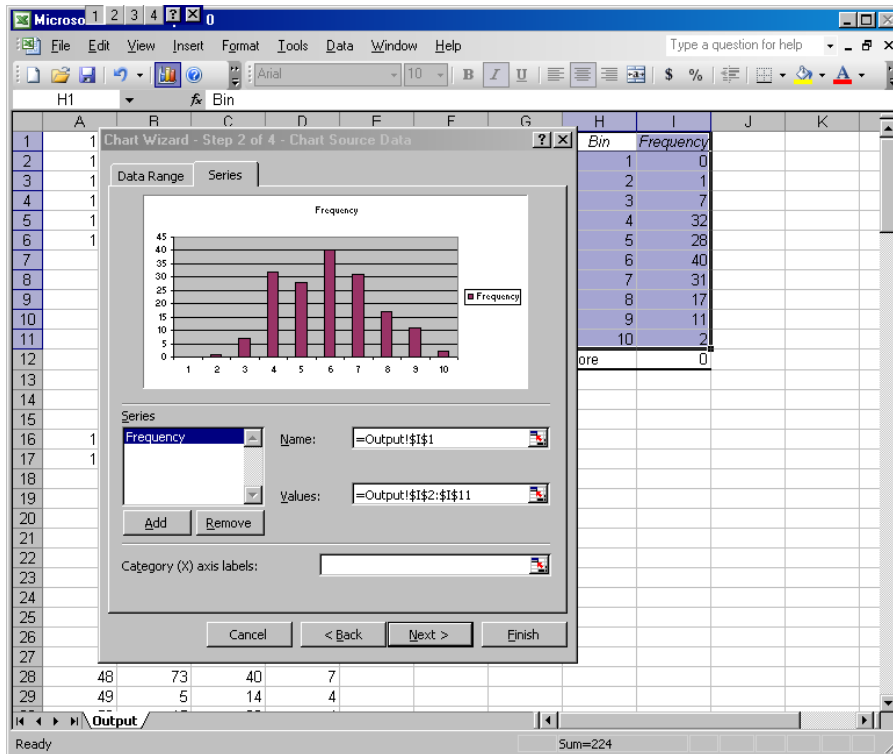
Click **Next>** in the Chart Wizard and in the next screen, click the tab called **Series** and **Remove Series** called **Bin** or any other series that appears first on the list.

????I don't understand the preceding sentence at all???

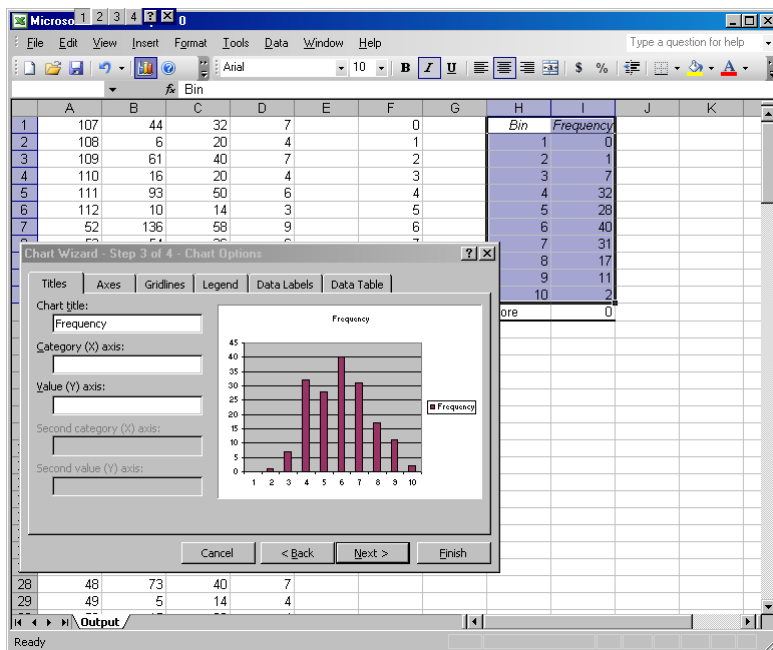
See the screenshot:

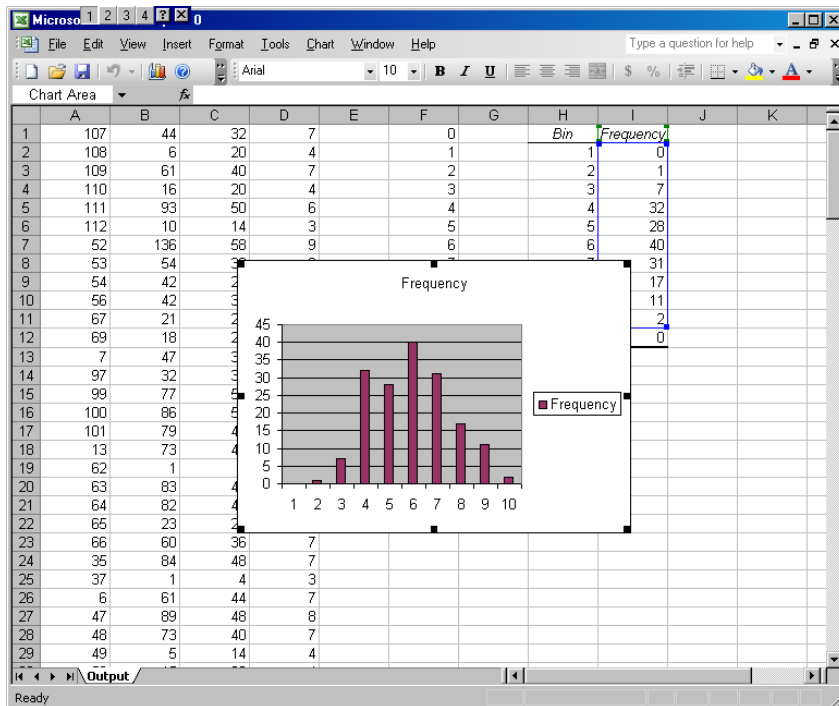


Your screen should look like the one below:



If so, click [Next >](#). On the following screen you may format the graph if you wish (I will not discuss formatting here). Clicking [Next >](#) should insert the graph into your spreadsheet. The next two screenshots illustrate this process:



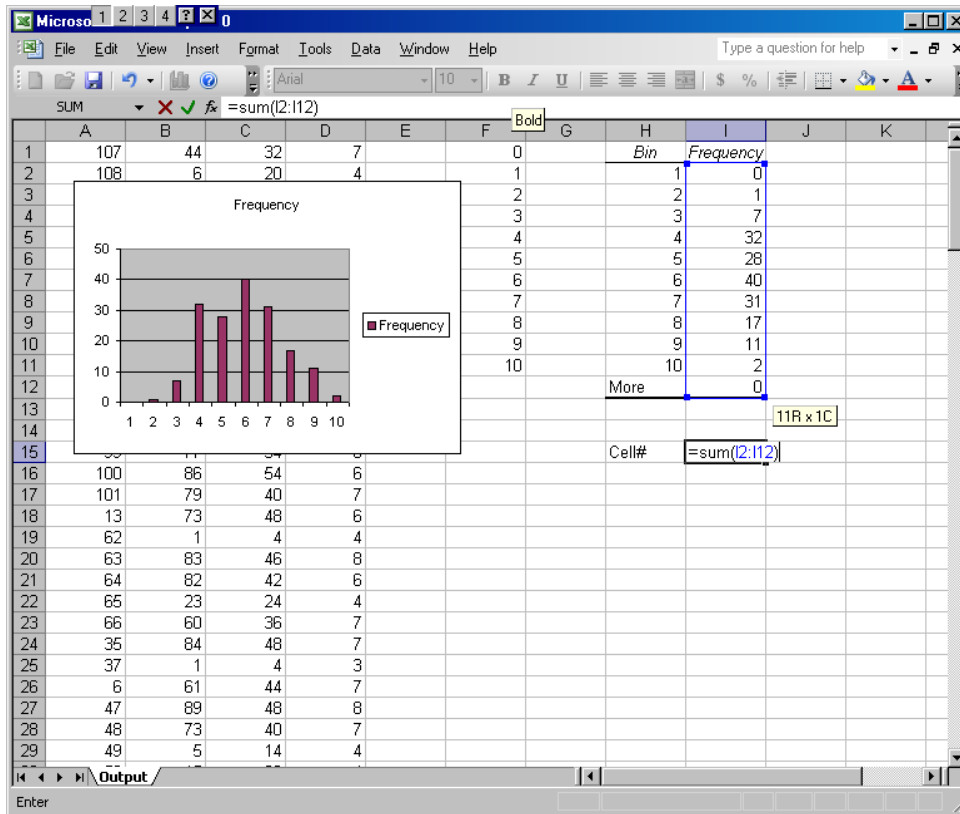


Normalizing Data

Now suppose we want to normalize the data so that instead of plotting absolute values we want to plot percentages. That is, every entry in the frequency column will be divided by the total number of occurrences (total number of cells).

Let's sum the number of cells. In Excel, pick a cell and type "**Cell#**" there. Click the cell next to it and enter the formula there. Click this cell and type "=". This is how you type formulae in Excel. If you ever need to edit a formula you can do it by clicking the cell and editing in the edit field located below the main toolbar. Anyway, let's come back to the formula. Type "=sum(" then select the cells in the column that you want to sum over (those are cells in the frequency column that have numbers **????Another place I don't undertand????**), then close the bracket - - type ")".

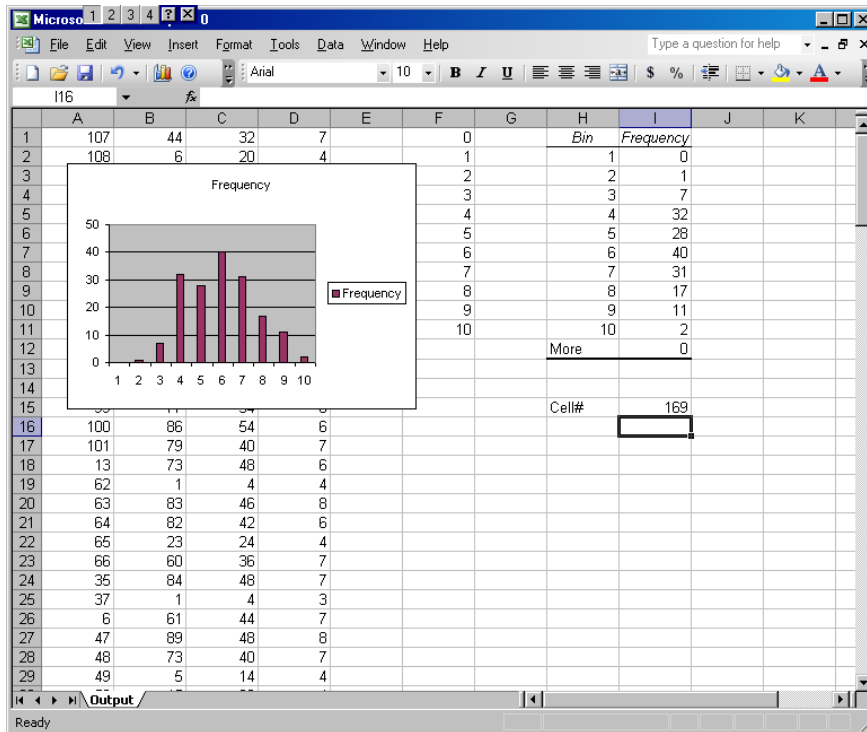
Your screen should look like the one below:



Notice the formula bar below the main toolbar .You can type the whole formula there without using the mouse. In our case the formula would be :

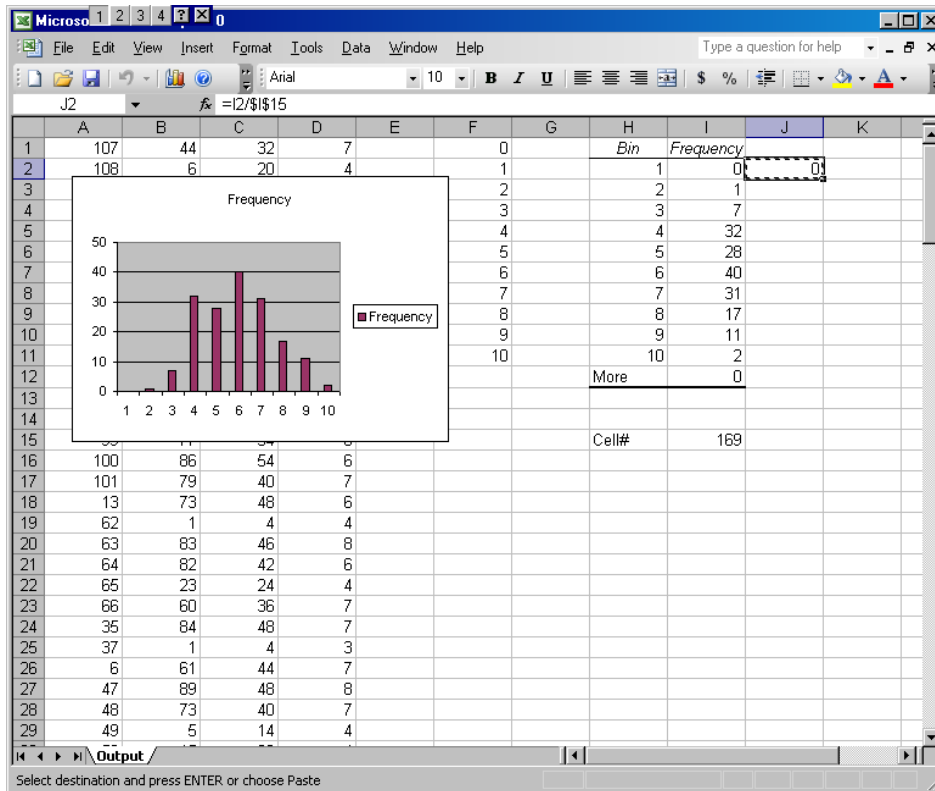
“=sum(I2:I12)”. You can also use the summation tools provided in Excel, but I will not discuss these here.

We are simply summing cells I2 to I12. After you are done with formula editing, hit [Enter](#). You should see the result of your formula displayed:

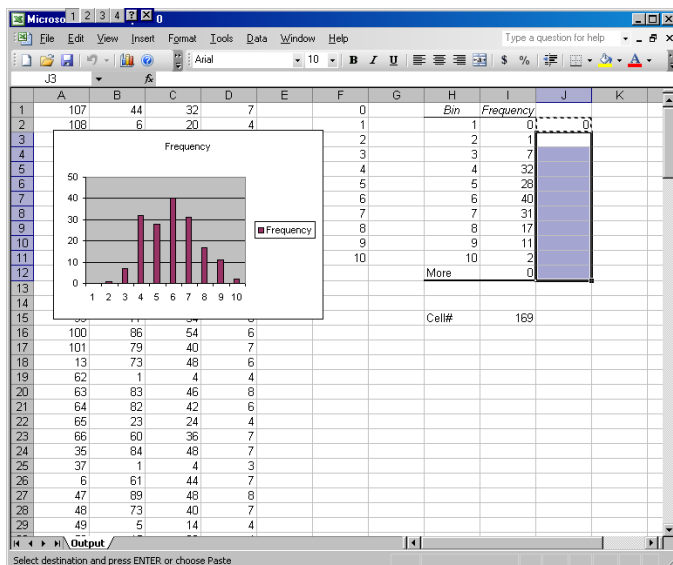


Now we will use this number to normalize **Frequency** to obtain percentages. Again we will use a simple formula. In the cell next to the first row of the histogram data, type the formula: **=I2/\$I\$15**.

What we want to do now is divide every number in the column **Frequency** by the total number of cells. Right now we have managed to do this operation for one entry. The **\$I\$15** is a special way of telling Excel that number appearing in the denominator of the above formula are to be taken **exactly** from cell **I15**. If you forget the **\$s** Excel will treat the reference to cell **I15** as relative and when you copy the cell with the formula you will get errors. More on that in a second. Right now, make sure you have correctly typed the formula.

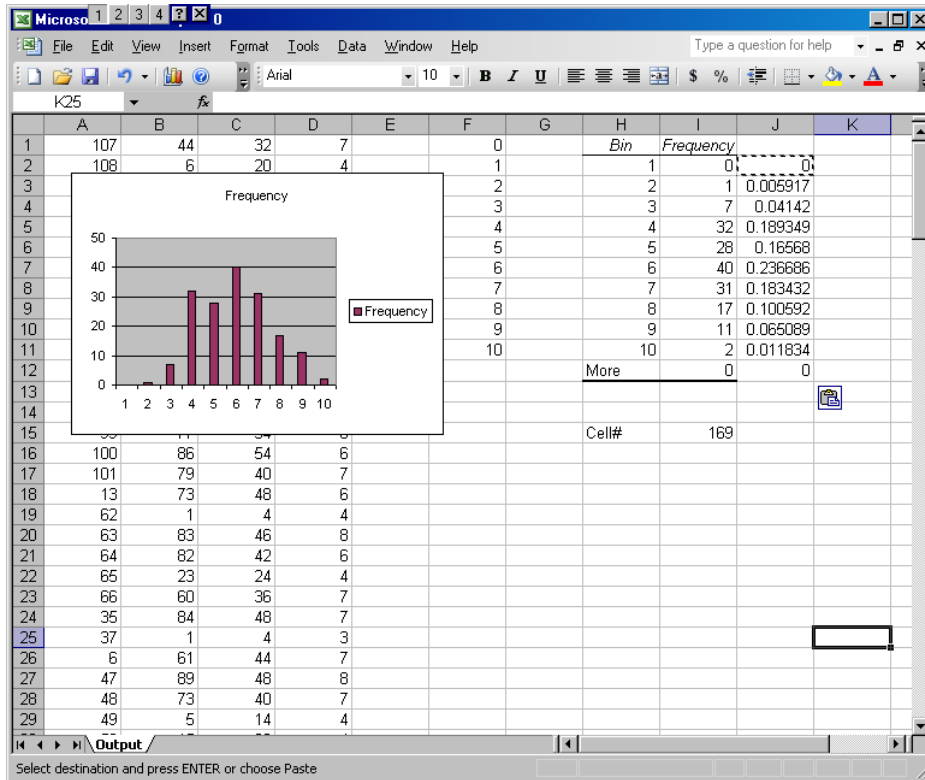


Our goal right now is to quickly apply the above formula to the other cells. To do that, click on the cell into which you entered the last formula and go to [Edit->Copy](#) (or hit [Ctrl-C](#)). Then highlight all other cells to which you want to apply the formula that you used for the last cell (cell [J2](#) in the above screenshot). This is what you should highlight:



As you can see, I have highlighted cells [J3](#) through [J12](#). Now, in order to apply formula that I

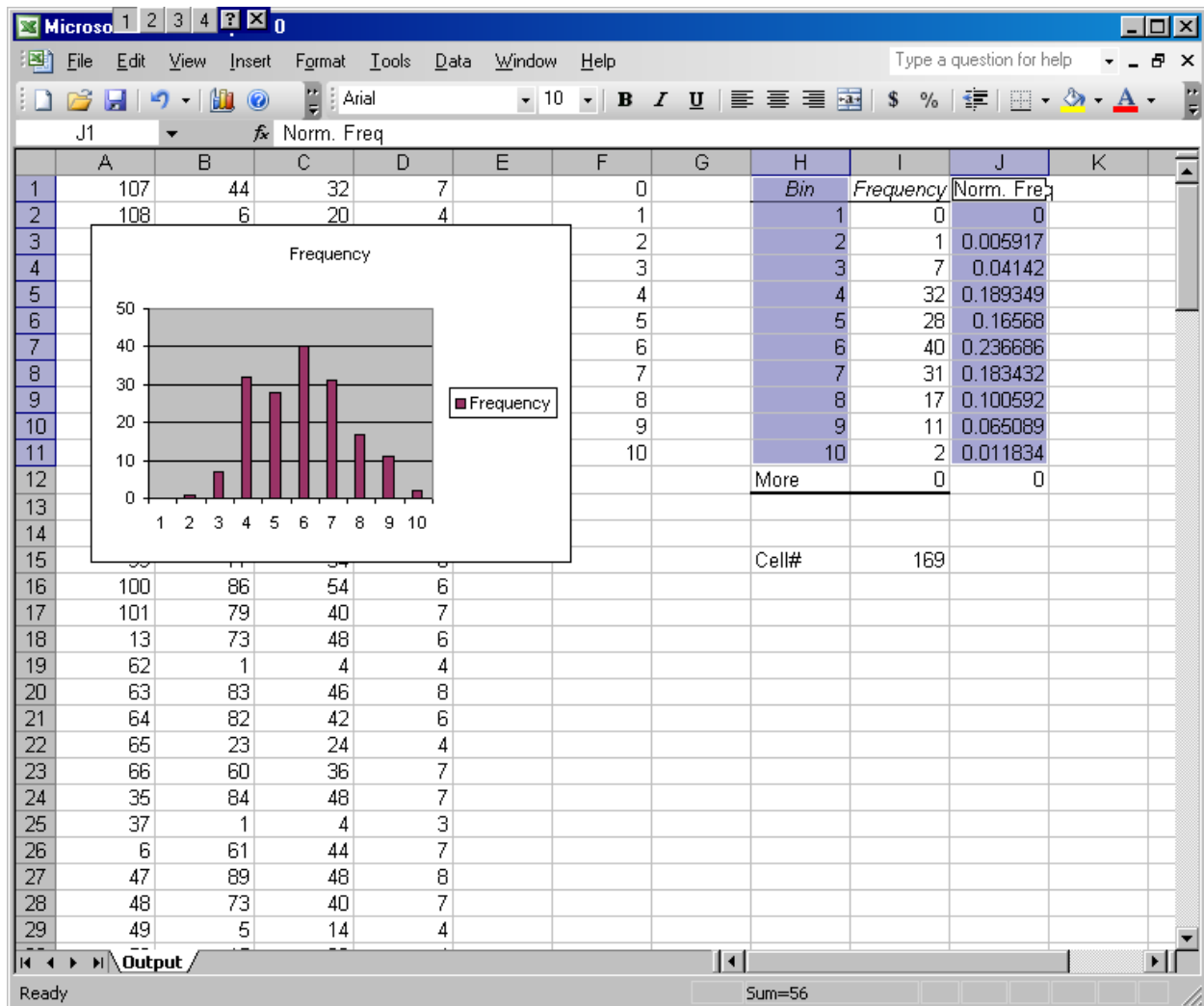
have prepared for cell J2 all I need to do is go to **Edit** and choose **Paste** (or hit **Ctrl-V**). Your screen should look like mine right now:



Notice how much typing we saved ourselves by applying this simple and standard Excel technique for formula copying.

Now we can plot normalized values, *i.e.* we will be plotting column H *vs.* column J. So first let's highlight the two.

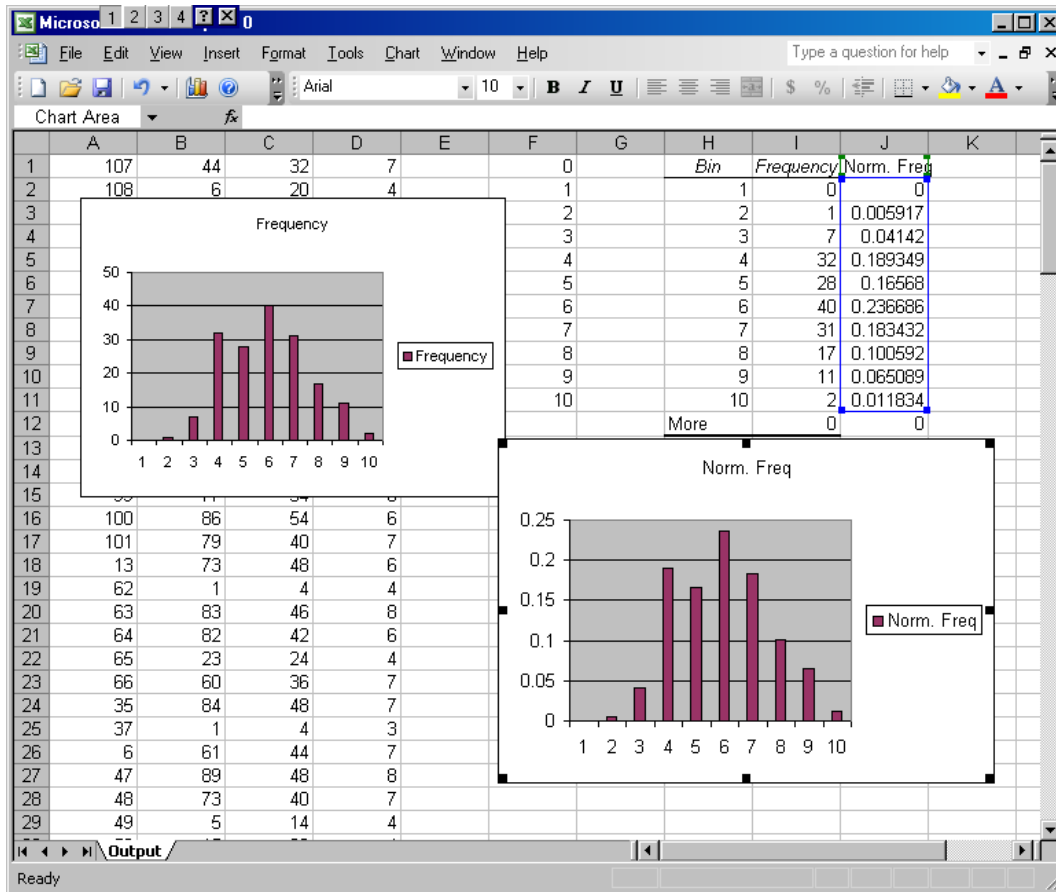
Hint: In Excel to highlight regions that are not adjacent to each other hold down the **Control** key (the **Apple key** if you are on a Mac) and select the two regions. **Before doing so, make sure everything else is unselected. Otherwise your selection may contain unwanted elements.**



Notice that in cell J1 I have added column label so that on the plot I got proper title automatically (you can do it later if you skipped this step).

Now that you have selected the two columns, produce chart. Go to [Insert->Chart...](#) I hope that you will be able to complete this task on your own. If you have trouble go back to the section where I discussed how to create charts in Excel.

In any case, this screenshot shows what I obtained:



Now I would suggest saving your work and trying another exercise.

For more information on Excel you may use IU resources or go to Amazon.com and find a good book on using MS Excel. Overall, I think knowing Excel is a useful skill to have for your career. Especially if you might want to work for an investment bank...