

Technical Intro to Creating Histograms in Excel

There is more than one way to do this! We will present two ways:

- Option #1: Using the Histogram Tool in the Data Analysis Package, and
- Option #2: Using the Frequency Function.

Choose whichever works best for you, given the settings available in your version of Excel.

Option #1: Using the Histogram Tool in the Data Analysis Package

Excel has several 'Add-Ins' that can be activated for the 'Quick Access Toolbars.' To see if the Data Analysis Add-In is activated on your version of Excel, click on the 'Data' tab and look for this Add-In in the 'Analysis' portion of the toolbar. (If it is not activated, you can try activating it by using the following steps: Position mouse over the 'Data' tab and right-click; select 'Customize Quick Access Toolbar...'; click on 'Add-Ins'; Select Manage 'Excel Add-Ins' and click 'Go'; check the Analysis ToolPack and click 'OK; agree to activate the Add-In. Now you should see the Data Analysis Add-In in the Analysis portion of the Data tab.)

Step 0: Make a plan! Always have a plan in mind before you ask Excel to do anything! Looking at the full range of the data you wish to histogram, decide on a reasonable number of equal-sized bins and figure out what the boundaries of these bins should be. Type these boundaries into a column in Excel. (For example, if we histogram test grades, we might select bins ranging from 0 to 100 points, with 10 bins, each 10 points large. Here the bins are 0-10, 11-20, 21-30, 31-40, 41-50, ... 81-90, 91-100. These bin 'boundaries' are 0, 10, 20, 30, 40, 50, ... , 90, and 100.)

Step 1: Activate the Histogram Analysis Tool. Find the Data Analysis Add-In and click on it. Choose the 'Histogram' analysis tool and click 'OK.' Select the data you wish to create a histogram of and put this into the 'Input Range' box. If you have decided what the boundaries of your bins will be, select the boundaries and put this into the 'Bin Range' box. (If you don't enter anything into the Bin Range box, Excel will calculate these on its own—but, as we know, Excel does not always make good choices when left alone!) Click the 'Output Range' bubble and enter a cell in your spreadsheet (this is where the results from the Histogram tool will be displayed). Then click 'OK' and the analysis will be run.

Step 2: Understand the Histogram Output. The Histogram analysis tool will output both Bin and Frequency data. An Example of such output is given here. The Frequency indicates the number of items from your data that fell between the previous and the current bin; i.e., a Frequency of 16 for Bin 10 says that the data you analyzed fell between the values of 5 and 10 a total of 16 times. The Bin 'More' represents the number of times the data you analyzed has a value more than your maximum bin boundary; i.e., a Frequency of 0 for the More Bin indicates that none of the data you analyzed had a higher value than 25.

<i>Bin</i>	<i>Frequency</i>
0	1
5	3
10	16
15	36
20	33
25	30
More	0

Step 3: Create the Histogram Plot. Now you just need to make a bar graph of this data. Highlight the frequencies and click on the 'Insert' tab, then click 'Column' and select '2-D Column.' To adjust the horizontal axis labels, click on the plot, right click, and choose 'Select Data,' then choose to edit the Horizontal Axis Labels. Title and label the plot appropriately. Tada!

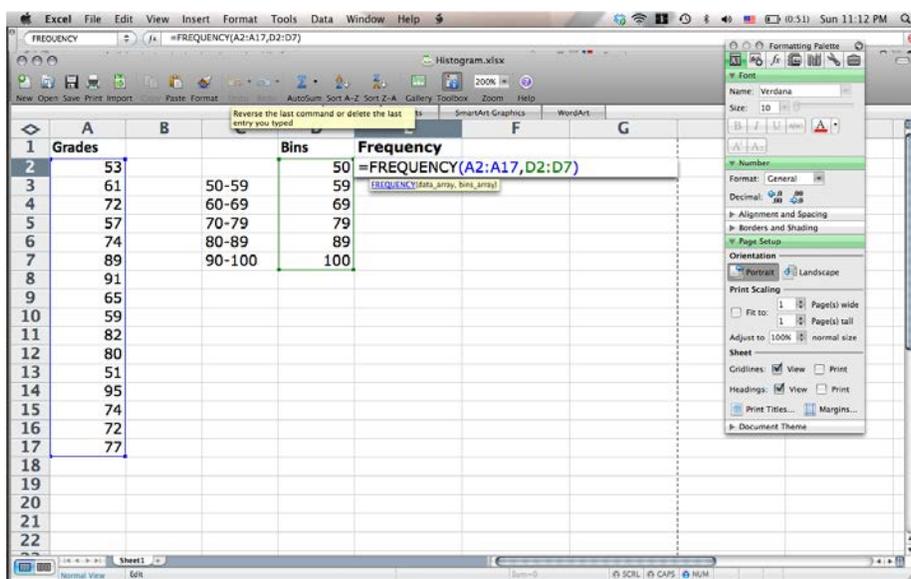
Option #2: Using the Frequency Function

Suppose you want to produce a histogram showing the distribution of student grades on a recent exam. The following guide outlines the procedure you would need to follow.

Step 1: Scan your data to get a sense for the overall range of values. For our example, the grades fall between 50 and 100, so this is the range that our bins must span. The next decision is how fine you want the increment of your bins to be—the finer the increment, the more bins, and thus the more bars in our histogram. For our sample data set, a bin increment of 10 seems appropriate. Create a column next to the raw exam score data that shows the bin ranges, and a column to the right of that which shows the **maximum values of your bins**.

Step 2: Now use the Excel function **FREQUENCY** to determine how many values fall within each of the bins that you have defined. The FREQUENCY function is an array function, returning values to a range of cells. Follow the following steps to enter the FREQUENCY function:

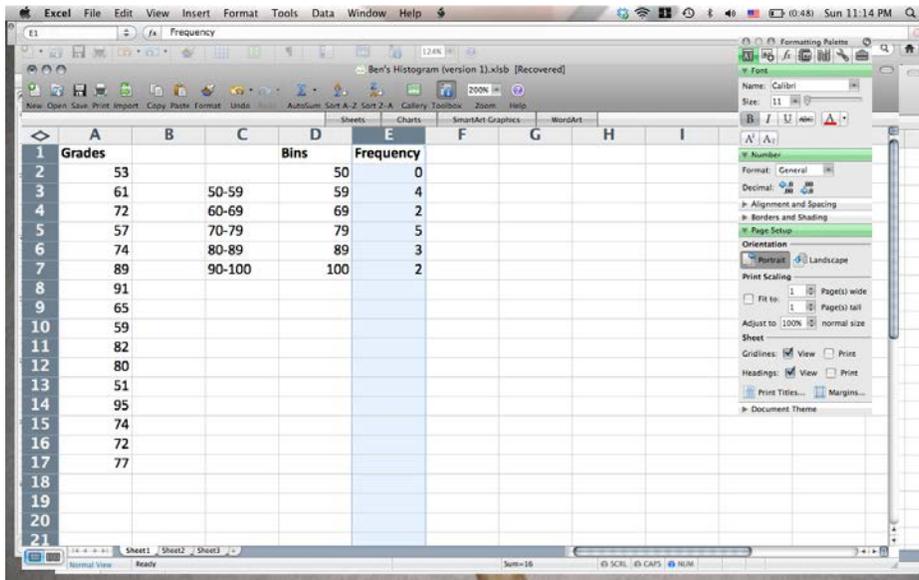
- **Highlight** the **range of cells** which will hold the frequency counts (**E2:E7**). These will be all of the Frequency Count cells next to the max bin values.
- Choose **Insert>Function...**, pick the **Statistical** Function category and scroll down in the box on the right and choose **FREQUENCY** as the Function name.
- Use the dialogue box to enter the function. With the **Data_array** box selected, go to the spreadsheet page and **highlight** the data values (**A2:A17**). The dialogue box with "roll up" while you highlight these values and then "roll down" when you are done.
- Repeat this process by selecting the **Bins_array** box and then go out the spreadsheet and **highlight** the bin limits cells (**D2:D7**).
- Click OK. The completed formula is seen in the formula bar and the correct count value is seen in the Bin Limit 50 count cell (E2).



Step 3: Now copy the array function down to the other Frequency Count cells. This is a bit different than typical cell copying:

- With the Frequency Count cells **still highlighted (E2:E7)**, click on the **FREQUENCY** function into the formula bar (i.e., =FREQUENCY(A2:A17,D2:D7)).
- **Propagate** the function by typing **Control-Shift-Enter** on a PC (type **Command-Return** on a Mac).

The frequency values should now fill the cells next to the bin increments. Note that your first bin increment, 50, holds all the grades at 50 and below. The next bin, 59, holds measurements from 50-59, and so on.



Step 4: Create a bar chart plotting the frequency count (Column E) as a function of the student grade increments (Column C).

