

Drawing Proper Histograms Using Excel

Written by Gregg Harbaugh

Center for Social Science Computation & Research
611 Condon Hall
University of Washington
Seattle, WA 98195 U.S.A
(206) 543-8110

Oct 2007

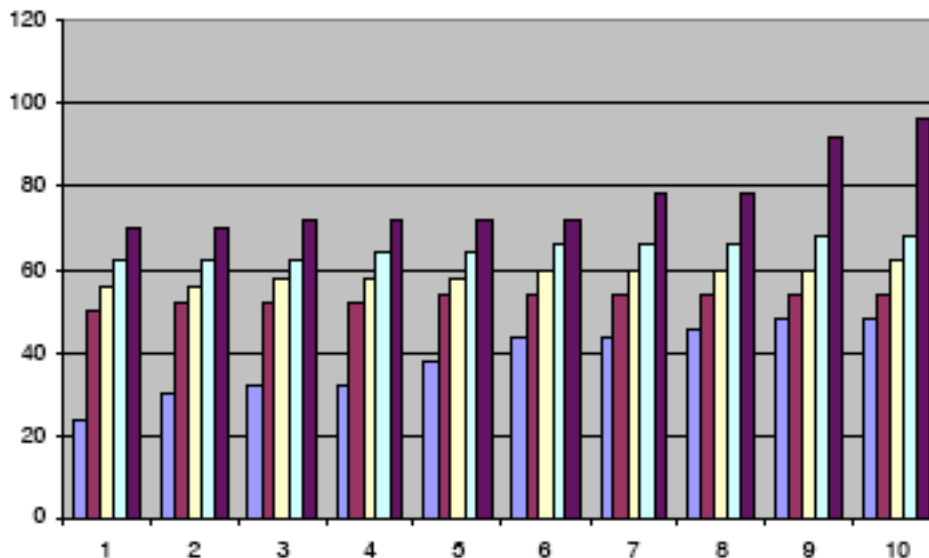
Drawing Proper Histograms Using Excel

The objective of this handout is to help you avoid the common pitfalls encountered when using Excel to draw histograms. We will work with the data set shown on the right.

The most common error is the belief that Excel is perfectly automated. In fact, it is not. It does have an Add-In feature that will help “draw” a histogram, but even this feature is flawed.

The first two graphs demonstrate this misconception of automation. Many students believe that if you simply select the numbers, then Excel can generate a bar graph. Students often select the (matrix) block of numbers (as shown above), then they hit the chart wizard button or select **Insert > Chart** from the menu. This will produce a graph similar to the one shown below. (Note: For convenience, the legend has been disabled on all graphs.)

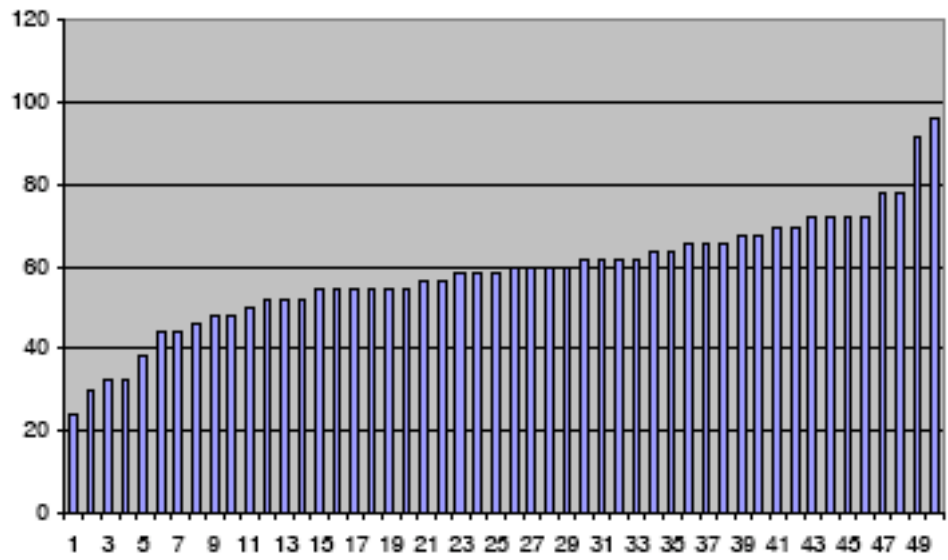
24	50	56	62	70
30	52	56	62	70
32	52	58	62	72
32	52	58	64	72
38	54	58	64	72
44	54	60	66	72
44	54	60	66	78
46	54	60	66	78
48	54	60	68	92
48	54	62	68	96



This clearly is not a histogram. It has not broken the data into disjoint classes. And, it does not report the frequency of data points that fall into those respective classes. In this case, Excel interprets each column as a different data point to be graphed.

The companion graph to this is obtained when students first make a long table of the values. All of the values are entered into one column in Excel, and then this column is sorted (often using the A→Z button).

When data is placed in a column and sorted (as partially shown to the right), Excel will no longer see multiple data sets (or series, using the Microsoft vocabulary). However, Excel still does not have the ability to automatically create a histogram. So, if you select the sorted data list and hit the chart wizard button, Excel produces the following graph:

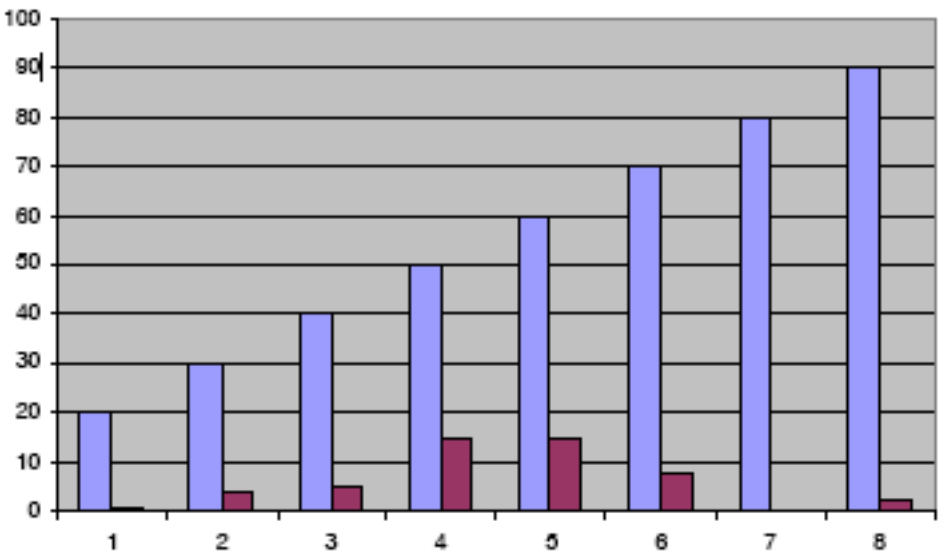


24
30
32
32
38
44
44
46
48
48
50
52
52
52
54
54
54
54
54
...
92
96

Again, this does not produce a histogram. The vertical axis does not report frequencies, and the horizontal axis is not broken into classes. In fact, the horizontal axis simply is reporting the ranked position of each data point in the data set; the height of the bar represents the data value.

Unfortunately, you must do a fair amount of the work to get Excel to draw a histogram. In particular, you need to count the number of instances that a data point falls into a given class. For this data set, we would start at 20, and create classes for every 10 points. By doing this, we would find the following frequencies (see the table 50 15 to the right). But, even this table will produce a flawed histogram.

20	1
30	4
40	5
50	15
60	15
70	8
80	0
90	2

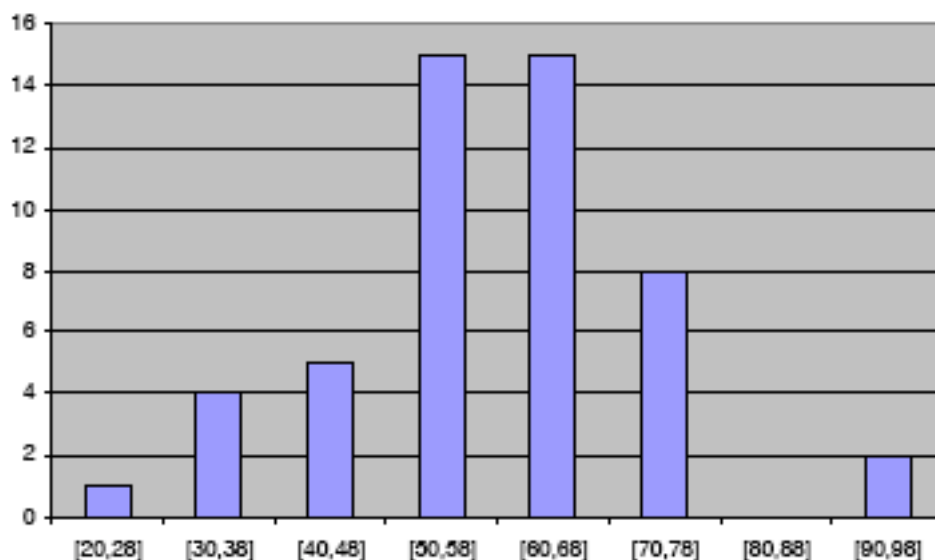


Unfortunately, Excel is not smart enough to realize that the first column represents the classes. The person generating this table understood that the first column of 20, 30, 40, ... represents the classes from 20-29, 30-39, 40-49, etc. However, Excel sees the first column as a set of numbers to be graphed. This is why the blue bars each increase by 10. However, if you can mentally erase the blue bars from the picture, you will see the desired histogram in the red bars at the bottom of the graph.

We are almost there.

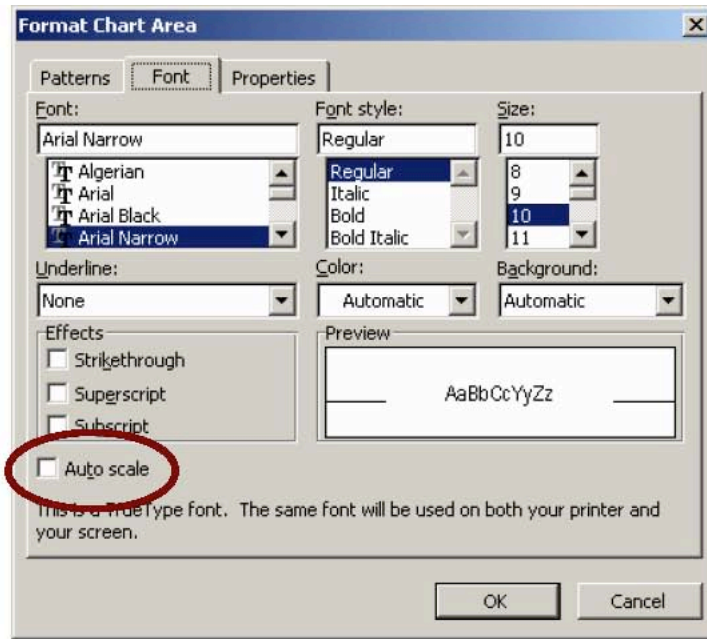
So, we must be very specific for Excel. The first column must be text values in order for Excel to recognize that column as the labels for the classes (as opposed to an additional set of frequencies to be graphed). In this case, the classes were entered using interval notation. (Note, the entire data set is composed of even 15 numbers, so the intervals include only even numbers.) Now, if you were to select this table and click the chart wizard button, then Excel can generate the desired graph, as shown below.

[20,28]	1
[30,38]	4
[40,48]	5
[50,58]	15
[60,68]	15
[70,78]	8
[80,88]	0
[90,98]	2

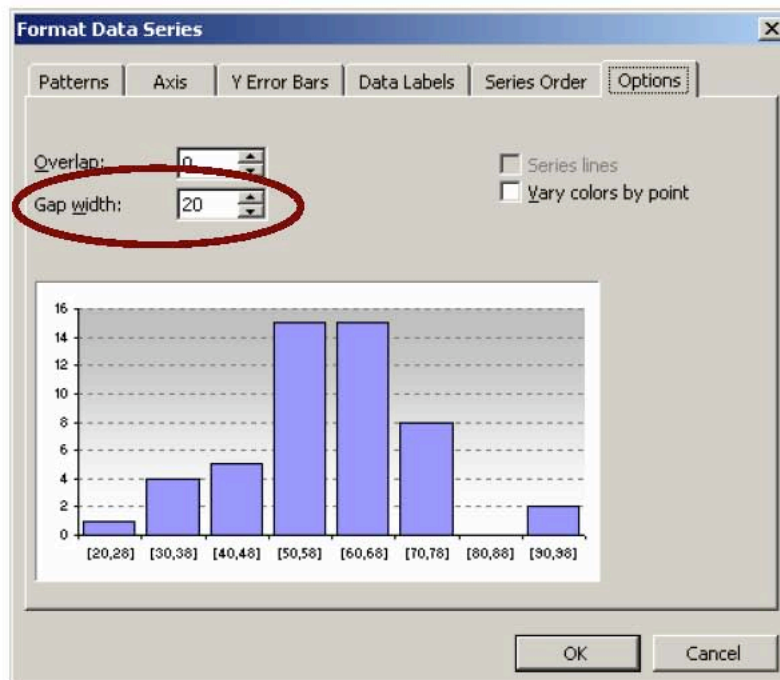


Personally, I am not content with default settings for Excel graphs. I do not like the variable font sizes, the excessive gap between the bars, nor the bland background. Here are a few tricks you can use to obtain a slightly prettier graph.

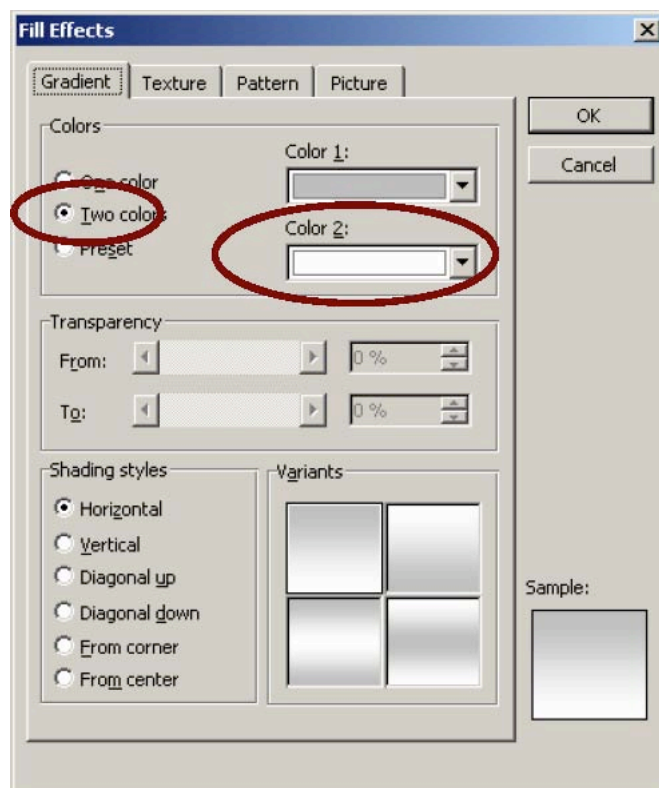
First, let's change the font. A good font and font size for graphs is Arial Narrow 10-point. Select the chart (you should see eight black squares on the border of the graph). From the menu, select **Format > Selected Chart Area...** Move to the **Font** tab at the top of the dialogue box. Change the **Font** and **Size** appropriately (Arial Narrow and 10, respectively). Before clicking the OK button, be sure to uncheck the **Auto scale** check box near the bottom of the dialogue box. (Dialogue box shown at top of next page.)



To change the gap size between the bars, touch a bar in the graph. This should select all of the bars. If this has happened, you will see little gray squares at the center of each and every bar. From the menu, select **Format > Selected Data Series...** Move to the **Options** tab at the top of the dialogue box. Change the **Gap Width** value to 20; then click the OK button.



Finally, click anywhere on the gray background of the graph. From the menu select **Format > Selected Plot Area...** Select the **None** radio-bullet under the Border section of the dialogue box. Under the Area section of the dialogue box, click the **Fill Effects...** button. This will bring up the dialog box to the right. Under the Colors section of the dialogue box, select the **Two colors** radio-bullet. Set **Color 2** to white. Then click the **OK** button. This generates a gradient appearance for the background, not just a solid gray background. To accept these changes, click the **OK** button, again.



The final result should look something like this.

