

How to create histograms in Excel

An excerpt from:

Middleton, M.R. 2004. Data Analysis using Microsoft Excel (3rd ed).
Brooks/Cole/Thompson, Belmont CA. 280pp.

4.2 ANALYSIS TOOL: HISTOGRAM

The Histogram analysis tool determines a frequency distribution table for your data and prepares a histogram chart. In addition to individual frequencies there is an option to include cumulative frequencies in the results.

You should determine the intervals of the distribution *before* using this tool. Otherwise, Excel will use a number of intervals approximately equal to the square root of the number of values in your data set, with equal-width intervals starting and ending at the minimum and maximum values of your data set. If you specify the intervals yourself, you can use numbers that are multiples of two, five, or ten—which are much easier to analyze.

To determine intervals, first use the Descriptive Statistics analysis tool to determine the minimum and maximum values of the data set. Alternatively, enter the MIN and MAX functions on your worksheet. Use these extreme values to help determine the limits for your histogram's intervals. Usually 5 to 15 intervals are used for a histogram.

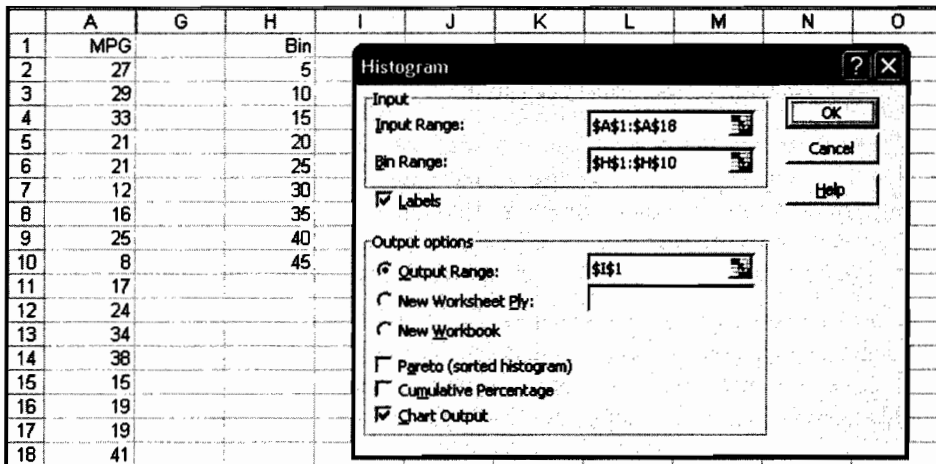
For the gas mileage data, the minimum is 8 and the maximum is 41. A compact histogram could start the first interval at 5, use an interval width of 5, and finish the last interval at 45, requiring 8 intervals. The approach used here adds an empty interval at each end; at the low end is an interval "5 or less," and at the high end is an interval "more than 45."

Excel refers to the maximum value for each interval as a *bin*. Here, the first bin is 5, and the interval will contain all values that are 5 or less. The Histogram tool automatically adds an interval labeled "More" to the bins you specify. Here, the last bin specified is 45, and the last interval (More) will contain all values greater than 45.

Refer to Figure 4.5 and follow these steps to obtain the frequency distribution and histogram.

1. Hide columns B through F. (Select columns B through F by clicking on B and dragging to F. Right-click and select Hide from the shortcut menu. To unhide the columns, select the two adjacent columns, A and G, right-click, and select Unhide. If column A is hidden, click the Select All button in the top-left corner at the intersection of the row and column headings, right-click a column heading, and select Unhide.)
2. Enter **Bin** as a label in cell H1, enter **5** in cell H2, and enter **10** in cell H3. Select H2:H3. Drag the AutoFill square in the lower-right corner of the selected range down to cell H10.
3. From the Tools menu, choose the Data Analysis command and choose Histogram from the Analysis Tools list box.

Figure 4.5 Bins and Histogram Dialog Box



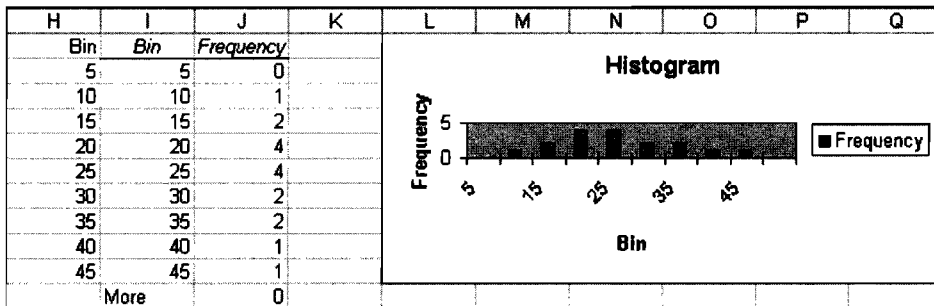
4. **Input Range:** Enter the reference for the range of cells containing the data (A1:A18), including the label.
5. **Bin Range:** Enter the reference for the range of cells containing the values that separate the intervals (H1:H10), including the label. These interval break points, or bins, must be in ascending order.
6. **Labels:** Check this box to indicate that labels have been included in the references for the input range and bin range.
7. **Output Range:** Enter the reference for the upper-left cell of the range where you want the output table to appear (I1). The combined table and chart output requires approximately ten columns.
8. **Pareto:** To obtain a standard frequency distribution and chart, clear the Pareto checkbox. If this box is checked, the intervals are sorted according to frequencies before preparing the chart. (In this example the box has been cleared.)
9. **Cumulative Percentage:** Check this box for cumulative frequencies in addition to the individual frequencies for each interval. (In this example the box has been cleared.)
10. **Chart Output:** Check this box to obtain a histogram chart in addition to the frequency distribution table on the worksheet. (In this example the box has been checked.)

- After you provide inputs to the dialog box, click OK. (If you receive the error message "Cannot add chart to a shared workbook," click the OK button. Then click New Workbook under Output in the Histogram dialog box. Use the Edit | Move or Copy Sheet command to copy the results to the original workbook.)

Excel puts the frequency distribution and histogram on the worksheet. As shown in Figure 4.6, the output table in columns I and J includes the original bins specified. These bins are actually the upper limit for each interval; that is, the bins are actually bin boundaries.

For example, the interval associated with bin value 15 (cell I4) includes mileage values strictly greater than 10 (the previous bin value) and less than or equal to 15. There are two such mileage values in this data set: 12 mpg and 15 mpg. Thus, for bin value 15 the frequency is 2 (cell J4).

Figure 4.6 Histogram Output Table and Chart



Histogram Embellishments

To make the chart more like a traditional histogram and easier to interpret, make the following changes.

- Legend:** Because only one series is shown on the chart, a legend isn't needed. Click on the legend ("Frequency" on the right side of the chart) and press the Delete key.
- Plot area pattern:** The plot area is the rectangular area bounded by the x and y axes. Double-click the plot area (above the bars); in the Format Plot Area dialog box, change Border to None and change Area to None. Click OK.
- Y-axis labels:** If you resize the chart vertically, intermediate values (0.5, 1.5,...) may appear on the y axis, but frequencies must be integer values. Double-click the y-axis (value axis); in the Format Axis dialog box on the Scale tab, set the Major Unit and Minor Unit values to 1. Click OK.

4. **Bar width:** In traditional histograms, the bars are adjacent to each other, not separated. Double-click one of the bars; in the Format Data Series dialog box on the Options tab, change the gap width from 150% to 0%. Click OK.
5. **X-axis labels:** Double-click the x-axis (category axis); in the Format Axis dialog box on the Alignment tab, double-click the Degrees edit box and type 0 (zero). With this setting, the x-axis labels will be horizontal even if the chart is resized. Click OK.
6. **Chart title:** Click on Histogram (chart title). Type **Distribution of Gas Mileage**, hold down Alt and press Enter, type **for 17 cars**, and press Enter. Click the Bold button to change from bold to normal type.
7. **Y-axis title:** Click on Frequency (value axis title). Click the Bold button to change from bold to normal type.
8. **X-axis title:** Click on Bin. Enter **Interval Maximum, in miles per gallon**. Click the Bold button to change from bold to normal type. Excel puts the x-axis values at the center of each interval, not at the marks that separate the intervals. This title makes it clear to the reader that these values are the maximum ones for each interval.
9. **Bar color:** Columns in a dark color may print as black with no gaps, in which case it is difficult to see the boundaries. Click on the center of one of the columns to select the data series. Click the right mouse button, choose Format Data Series, and click the Patterns tab. In the dialog box, leave Border at Automatic and change Area from Automatic to None. Click OK.

To move the chart, click just inside the chart's outer border (chart area) and drag the chart to the desired location. To resize the chart, first click the chart area and then click and drag one of the eight handles.

When you first create a chart, Excel uses automatic scaling for the font sizes of the chart title, the axis titles, and the axis labels. When you resize the chart, the font sizes change and the number of axis labels displayed may change. For example, if the axis labels on the horizontal axis have a large font size and you resize the chart to be narrow, perhaps only every other axis label will be displayed.

One approach to chart and font sizing is to first decide the size of the chart. For this example the chart is 6 columns wide using the standard column width of 8.43 and 14 rows high. The font size of the three titles is Arial 10, and the font size of the two axes is Arial 8 so that all axis labels are displayed. The resulting histogram chart is shown in Figure 4.7.