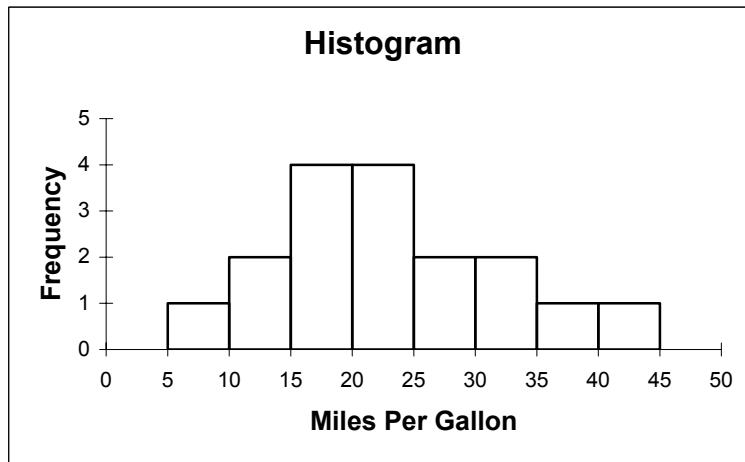


Better Histograms Using Excel

Michael R. Middleton
School of Business and Management, University of San Francisco

A histogram is usually shown in Excel as a Column chart type (vertical bars). The labels of a Column chart are aligned under each bar, and there is no Excel feature for changing the alignment. A better histogram has a horizontal axis with numerical labels aligned under the tick marks between the bars as shown below.

Desired Chart

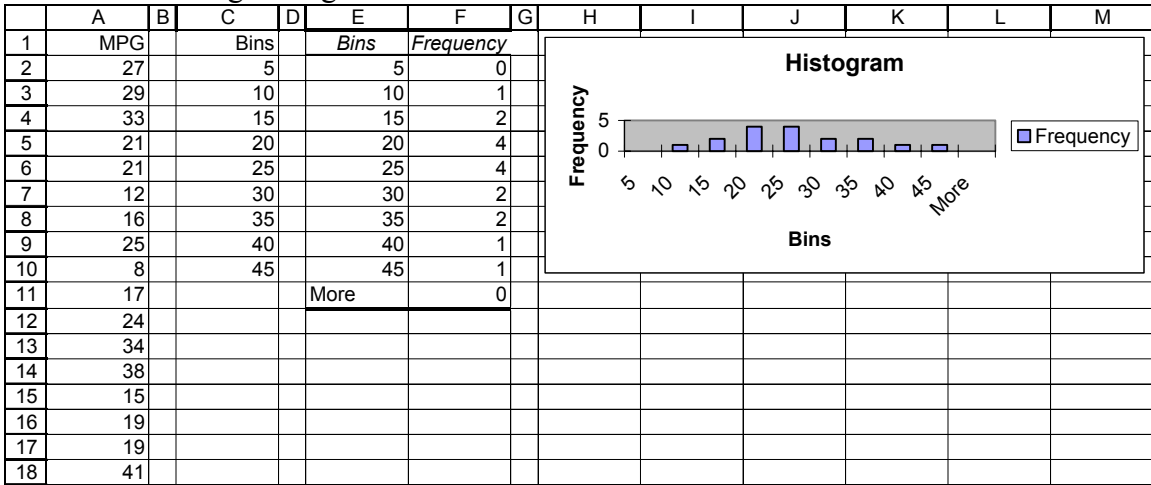


The method described here involves adding a new data series to the Column chart, changing the new series to an XY (Scatter) chart type, using the horizontal axis of the new series, and making the unneeded axes disappear. Since the intermediate steps of this method involve some rather strange looking charts, and a chartist may give up before getting the final result, I have included detailed steps.

The data for this example are gas mileages for 17 cars [1]. The minimum value is 8 mpg, and the maximum is 41 mpg. Histogram intervals bounded by multiples of 2, 4, 5, or 10 are easier to interpret, and I follow the usual guideline of using between 5 and 15 intervals. Since the Column chart's bars are equal width, it is important to choose equal-width intervals. And I like to include an empty interval at each end. Before using Excel's Histogram tool, bin values from 5 to 45 in steps of 5 are entered in a column. These bin values specify the inclusive upper bound for equal-width intervals.

The results (frequency distribution and Column chart) from the Histogram tool are shown below. You could obtain a similar result by creating an Excel Column chart after using (a) the array-entered FREQUENCY worksheet function, (b) multiple COUNTIF worksheet functions, or (c) a Pivot Table.

Initial Chart Using Histogram Tool

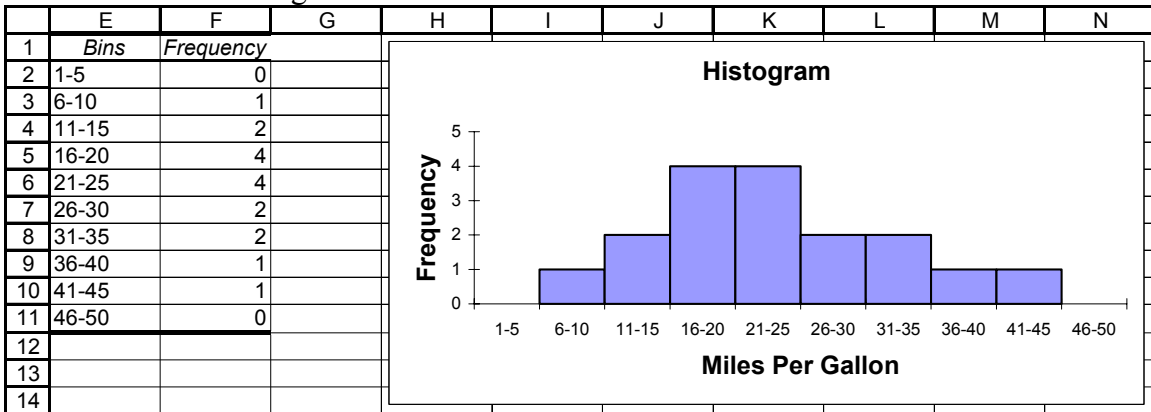


The Histogram tool adds an interval "More" to the original bins. Instead, if you use the FREQUENCY worksheet function to obtain the frequency distribution, you can select a range with one more cell than the bin range before array-entering the function. Excel may display only every other label on the horizontal axis of the Column chart. To see more labels, you can use a smaller font, change the alignment, or make the chart larger. In this Column chart, the labels associated with each interval are 5, 10, ..., 45, More.

Some Less Desirable Solutions

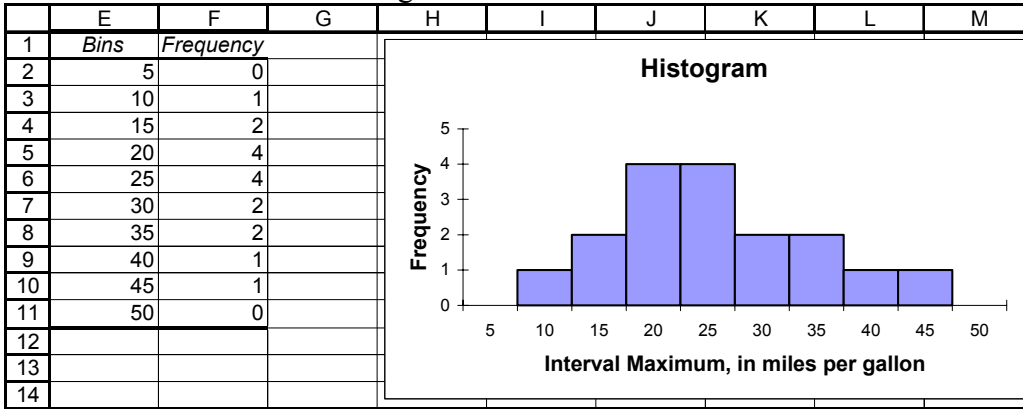
If the data values are discrete, as opposed to continuous, you can change the text labels in column E to specify discrete ranges for the horizontal axis, as shown below. For example, this approach is quite appropriate for integer-valued data. If the text labels have many characters, you can format the horizontal axis to use a smaller font.

Chart Suitable for Integer-Valued Data



If the data values are continuous or many-valued, and if you want the horizontal axis to reflect this characteristic, you can change the axis title to indicate that the axis values are the inclusive upper bound for each interval, as shown below.

Embellished Chart From Histogram Tool

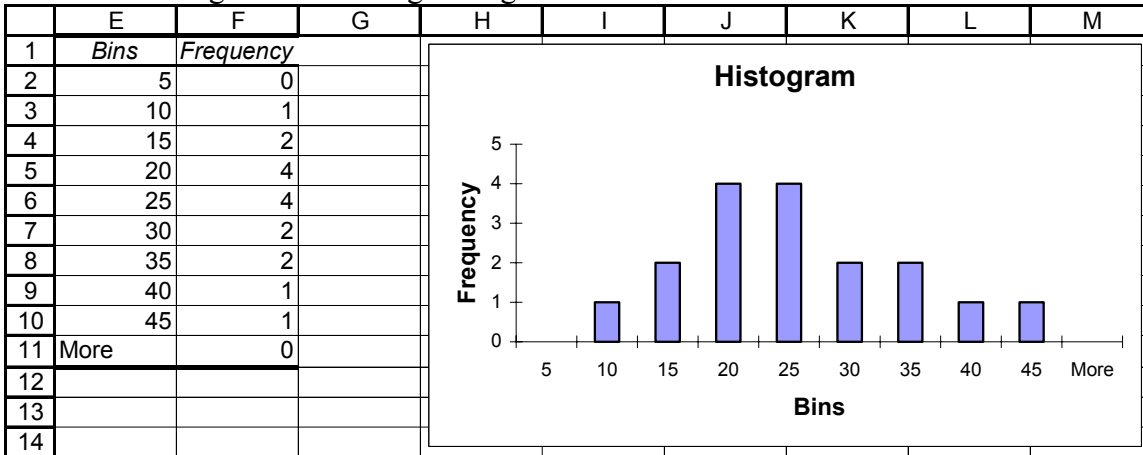


Charts may be used for both analytic and presentation purposes. If you are doing your own analysis that will not be shared with others, you may not be concerned about properly labeling the axes. However, for formal presentations involving continuous-valued data, you may want to take the extra steps to create a histogram with horizontal axis labels aligned directly under the tick marks separating each bar.

Step-by-Step Instructions

The following steps start with the chart produced by Excel's data analysis Histogram tool. To see the initial histogram chart more clearly, I make the entire chart larger by dragging the chart fill handles, delete the legend, and use a smaller font on the horizontal and vertical axes.

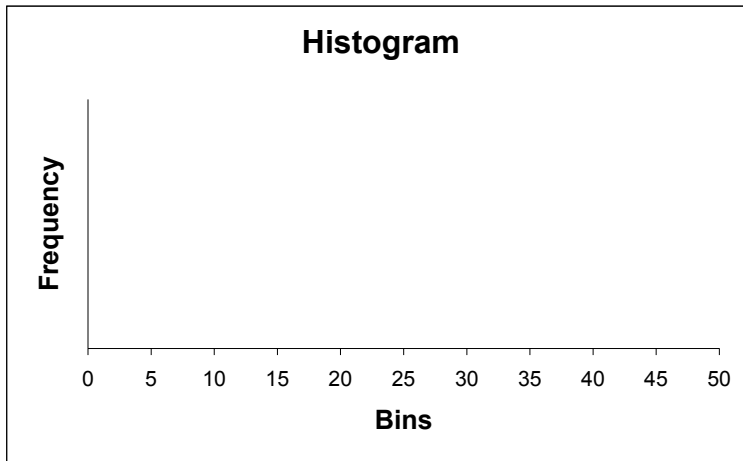
Basic Formatting of Chart Using Histogram Tool



If you prefer not to have empty intervals at each end, select any chart object, and from Excel's main menu choose Chart | Source Data | Series to change the Values range and the Category (X) Axis Labels range. Alternatively, in recent versions of Excel, the cells of the data ranges are highlighted on the worksheet, so you can click and drag the fill handles to modify the two ranges.

The primary task is to replace the horizontal axis with one that has a number aligned directly under each tick mark separating the columns. The desired axis is shown below.

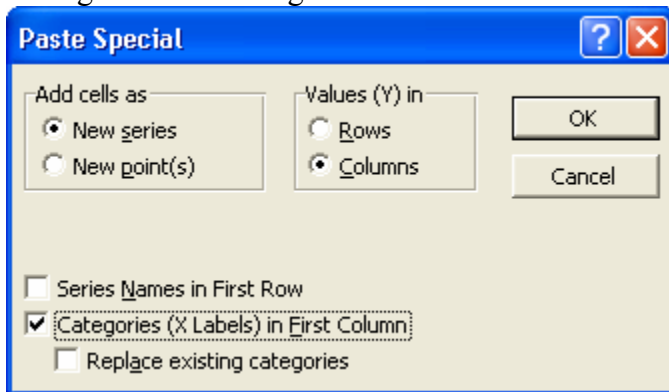
Desired Horizontal Axis



In the Histogram tool frequency distribution output labeled Bins and Frequencies, select two rows by two columns of numbers somewhere in the middle of the output. (Any adjacent two-by-two selection in the Bins and Frequencies columns is satisfactory; the exact values are not important.) Edit | Copy. For this example I selected E4:F5.

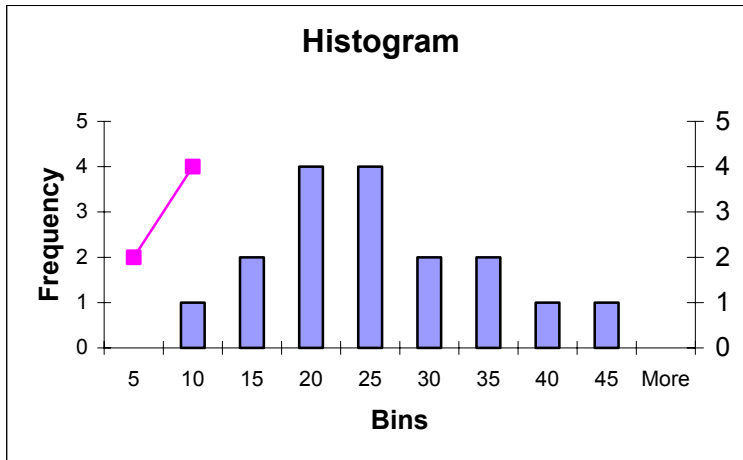
Single click just inside the outer border of the chart to select the chart. From Excel's main menu, choose Edit | Paste Special. In the Paste Special dialog box, verify Add cells as New series and Values (Y) in Columns; check only the box Categories (X Labels) in First Column, as shown below.

Dialog Box for Adding New Data Series to Chart



Click OK, and the chart appears with the new data series, as shown below.

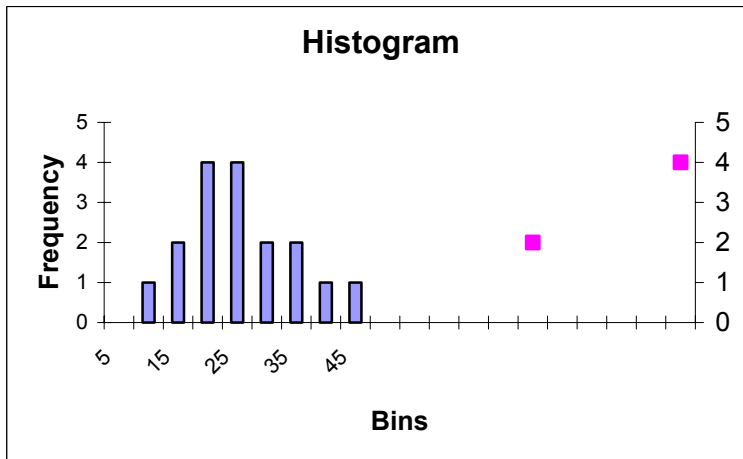
New Data Series on Chart



Select the new data series on the chart. (If it's difficult to select the data series using the mouse, first select any object on the chart, and then use the up and down arrow keys to cycle through the main chart objects until the new data series is selected. The left and right arrow keys cycle through all detailed chart objects.) The formula bar shows =SERIES(...,2).

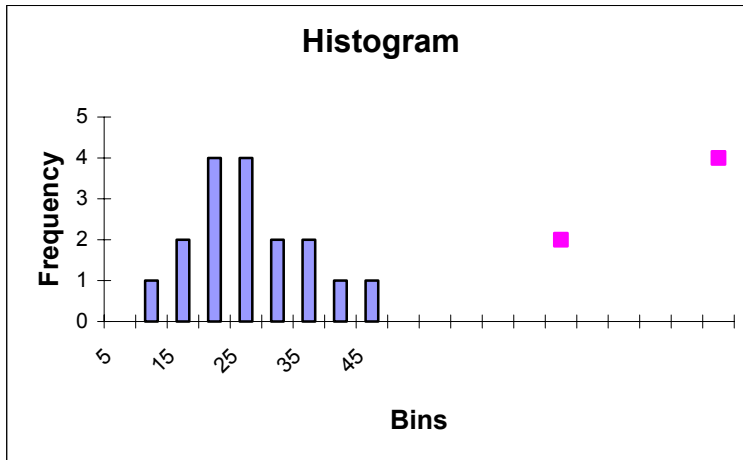
With the new data series selected, from the Chart menu, choose Chart Type. In the Chart Type dialog box, select the Standard Types tab. Select XY (Scatter) as the Chart type, and check the Options box for Apply to selection. Click OK. The result is shown below.

New Data Series as XY (Scatter) Type



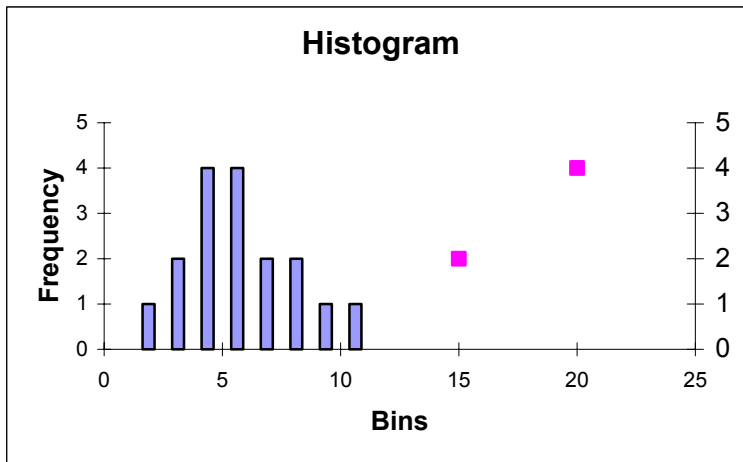
With the new data series selected, choose Format Selected Data Series. Select the Axis tab. Select Primary Axis, and click OK. The result is shown below.

New XY Data Series Using Primary Axis



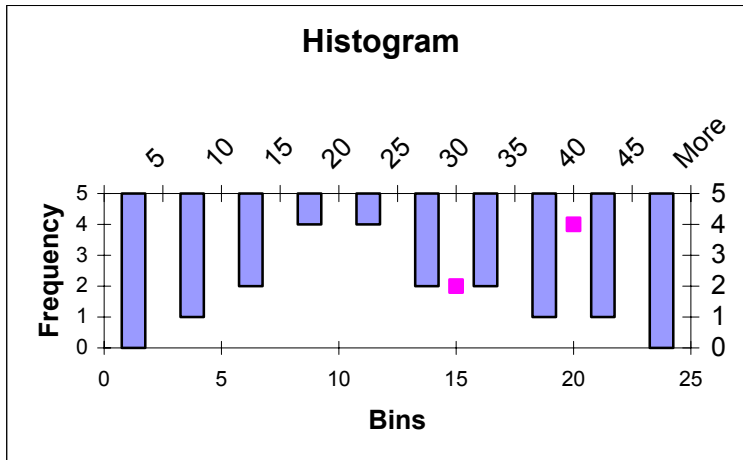
Select the column data series, and choose Format Selected Data Series. Select the Axis tab. Select Secondary Axis, and click OK. The result is shown below.

Column Data Series Using Secondary Axis



With any chart object selected, from the Chart menu, choose Chart Options and select the Axes tab. Check each of the four check boxes, and select Automatic for each X axis. Click OK. The result is shown below.

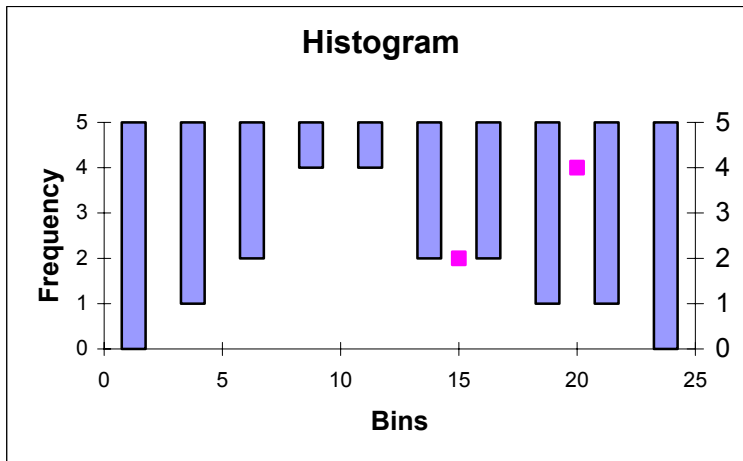
Chart Showing All Four Axes



Select the horizontal Secondary Category Axis at the top of the chart. If it's difficult to select the axis, first select any object on the chart (e.g., the plot area), and then use the left and right arrow keys to cycle through the detailed chart objects. The Name Box shows Secondary Category Axis.

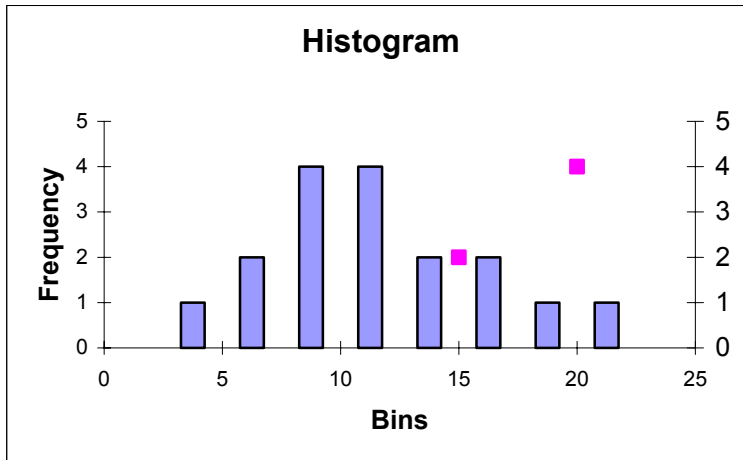
With the Secondary Category Axis selected, from the Format menu, choose Selected Axis and select the Patterns tab. Click None for Lines (or for Axis), None for Major tick mark type, None for Minor tick mark type, and None for Tick mark labels. Click OK. The result is shown below.

Chart Not Showing Secondary Category Axis



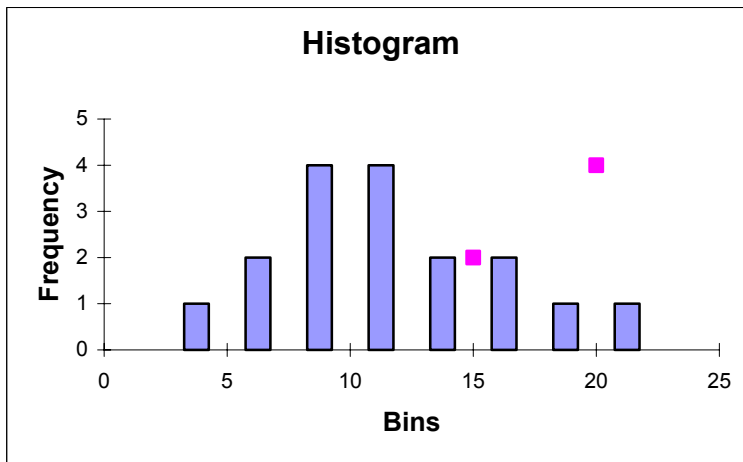
Select the vertical Secondary Value axis at the right side of the chart. From the Format menu, choose Selected Axis and select the Scale tab. Clear the check box for "Category (X) axis crosses at," type 0 (zero) in the edit box, and click OK. The result is shown below.

Column Secondary Value Axis Crossing at Zero



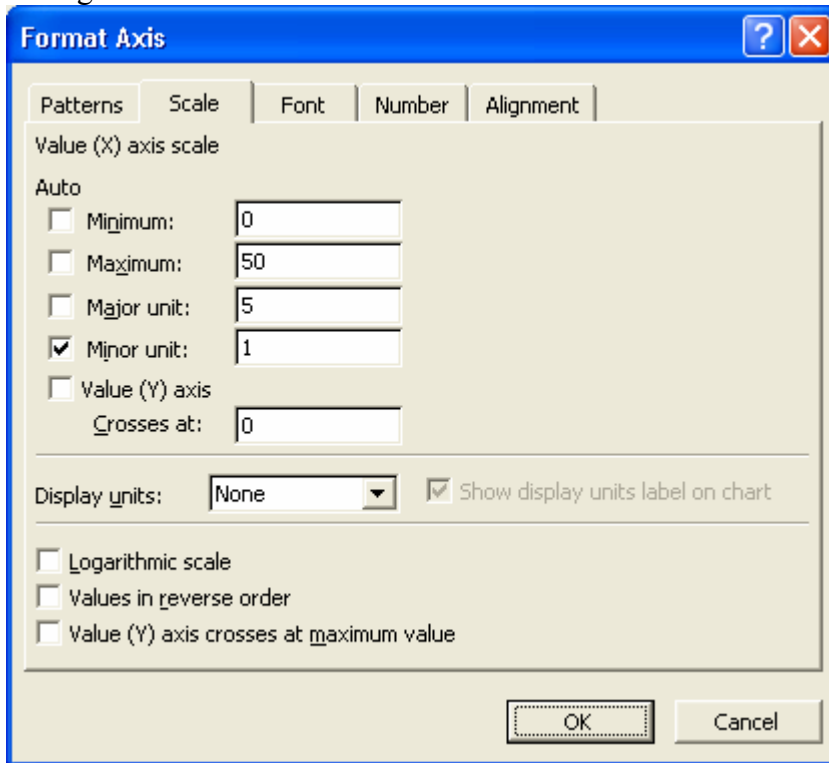
With any chart object selected, from the Chart menu, choose Chart Options and select the Axes tab. At the bottom of the dialog box, for the Secondary axis, clear the check box for Value (Y) axis. The other three boxes should have check marks. Click OK. The result is shown below.

Chart Not Showing Column Secondary Value Axis



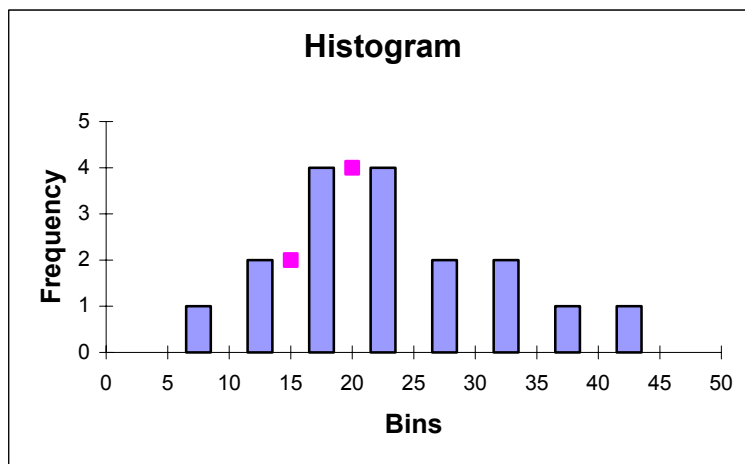
Select the horizontal Value (X) axis at the bottom of the chart. From the Format menu, choose Selected Axis and select the Scale tab. Click Minimum and type the value for the leftmost tick mark in the edit box. (If your original chart was developed from the Histogram tool, the Minimum value is equal to the first bin minus the interval width.) Click Maximum and type the value for the rightmost tick mark in the edit box. Click Major unit and type the value of the interval width in the edit box. Click "Value (Y) axis Crosses at" and type the Minimum value in the edit box.

Dialog Box for XY Horizontal Axis Scale



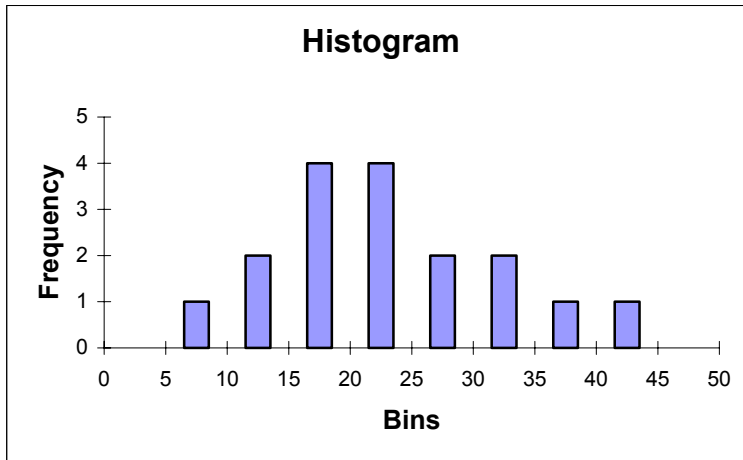
Click OK. The result is shown below.

Chart Showing Desired Horizontal Axis



Click one of the XY data points on the chart to select the new XY data series. From the Format menu, choose Selected Data Series, and click the Patterns tab. Select None for Line and None for Marker. Click OK. The result is shown below.

Chart Not Showing XY Data Series

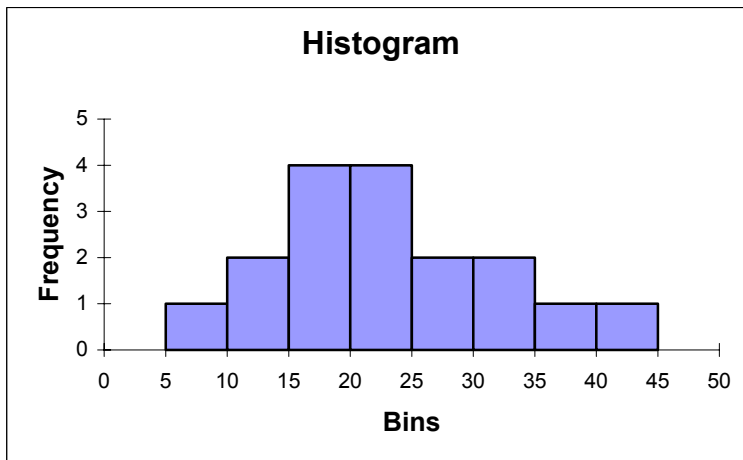


To resize the tick mark labels, select the axis, from the Format menu choose Selected Axis, click the Font tab, and enter a font size. If you plan to resize the entire chart and you don't want the font size to change, uncheck the Auto scale check box. Click OK.

The Usual Histogram Embellishments

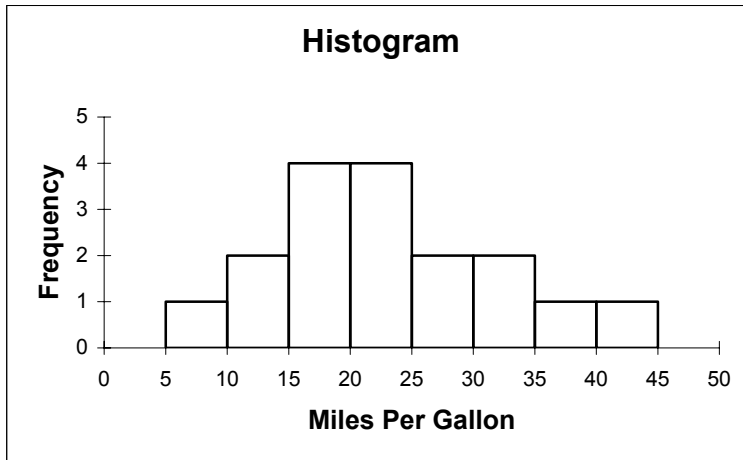
Click on one of the columns of the chart to select the data series. From the Format menu choose Selected Data Series. Optionally, on the Patterns tab, select Automatic for Border and None for Area. Most importantly, on the Options tab, change the Gap width to 0 (zero). Click OK. The result is shown below.

Chart Showing Zero Gap Width



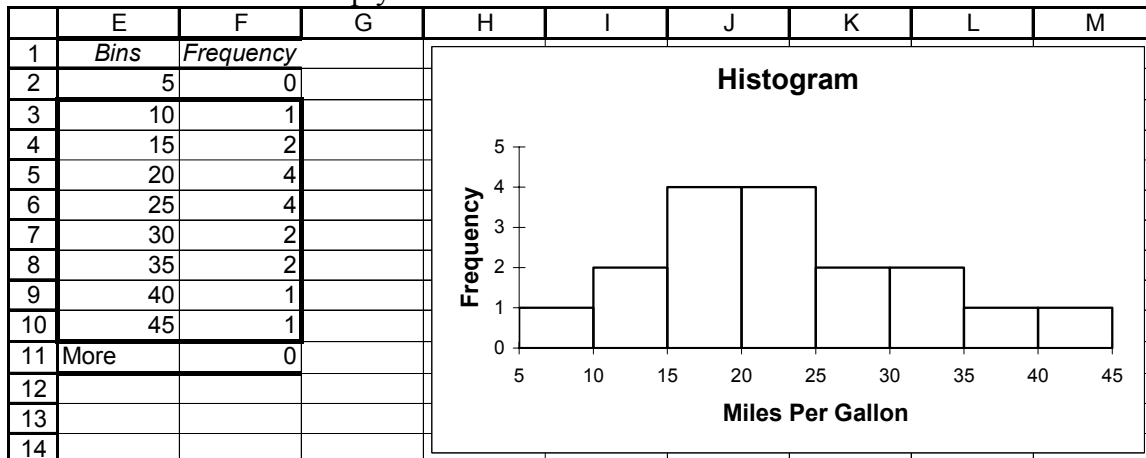
Optionally, click above the columns of the chart to select the Plot Area. From the Format menu choose Selected Plot Area. On the Patterns tab, select None for Border, select None for Area, and click OK. Select the horizontal axis title, and type a descriptive title. The result is shown below.

Desired Chart



If you prefer not to have empty intervals at each end, click one of the histogram vertical bars to select the Frequency data series of the Column chart, and change the reference from E2:F11 to E3:F10. Also, select the horizontal axis, and format the axis scale to be Minimum 5, Maximum 45, and Major Unit 5. The result is shown below.

Desired Chart Without Empty Intervals



The charts shown here are embedded chart objects on Excel worksheets, as opposed to a chart on a separate chart sheet. I prefer the formatting options that are only available for an embedded chart on a worksheet. If you want to include such a chart in a Word document, you may want to explore various methods for copying the Excel chart. Select the entire chart by clicking just inside its outer border. In addition to using a simple Edit | Copy command, try holding down the Shift key while you select the Edit menu. The options available from Shift Edit | Copy Picture may yield a higher quality chart when you paste into Word.

Conclusion

The chart produced by Excel's data analysis Histogram tool can be easily modified for discrete-valued data and personal analysis purposes. The extensive modifications described here are most appropriate for developing a histogram of continuous-valued data for presentation purposes. Possible extensions involve superimposing a normal curve on the histogram and using VBA to automate creating the histogram.

References

[1] Middleton, M. R. Data Analysis Using Microsoft Excel: Updated for Office XP, Duxbury (2004).

Web Sites of Excel Chart Gurus

www.tushar-mehta.com
peltiertech.com/index.html
www.j-walk.com/ss
www.bmsltd.ie/excel
www.appspro.com/Utilities/ChartLabeler.htm

Tushar Mehta's Tutorials and Add-Ins
Jon Peltier's Excel Chart Pages
John Walkenbach's Spreadsheet Page
Stephen Bullen's Excel Page
Rob Bovey's XY Chart Labeler

Excel Chart Newsgroup

Account: msnews.microsoft.com
Newsgroup: microsoft.public.excel.charting

Updates for this paper

Primary web site: <http://www.treeplan.com>
Secondary sites: <http://www.mikemiddleton.com>
<http://www.usfca.edu/~middleton>

Reference for VBA Add-In

John Walkenbach, Excel 2002 Power Programming with VBA, M&T Books, 2001.