October 2005

The Charts that Excel Cannot Do

John Baker
Natural Maths

The Charts that Excel Cannot Do

Abstract

The purpose of this paper is to explore work-arounds for the charts that, on the face of it, Excel cannot do. The examples given are intended for use by students at two levels of complexity. At the first level, we have suggested spreadsheets that react immediately to user-supplied data, much as a graphing package might do. At the second level, we have supplied only a brief description of how to construct the spreadsheet and chart that will display data in the form required. At this level, the student's involvement is much greater, and hence, their understanding of the construction of these charts is enhanced.

Keywords
charts

The Charts that Excel Cannot Do

John Baker
Natural Maths

Abstract

The purpose of this paper is to explore work-arounds for the charts that, on the face of it, Excel cannot do. The examples given are intended for use by students at two levels of complexity. At the first level, we have suggested spreadsheets that react immediately to user-supplied data, much as a graphing package might do. At the second level, we have supplied only a brief description of how to construct the spreadsheet and chart that will display data in the form required. At this level, the student's involvement is much greater, and hence, their understanding of the construction of these charts is enhanced.

Keywords: spreadsheets, exploratory data analysis

1. Background

Since the development of exploratory data analysis, EDA, and its take up by the educational fraternity, a number of packages that display data in the EDA styles have become available. Indeed, the EDA methods that are based very much on creating a graphical image of data have become the mainstay of statistics teaching in schools. Despite this popularity, the EDA charts such as stem-and-leaf plots, box-and-whisker plots and back-to-back plots are not standard features available within Excel. The histogram, while not strictly from the EDA fold, is another graph type that is very popular in schools, but the histogram add-in available in Excel is both cumbersome and non-interactive. Once a histogram has been constructed with the add-in, changing the data does not alter the chart. Using pivot charts is one work-around, but even there, the steps to create the chart are not intuitively clear, and the chart has to be manually refreshed if the data changes. Thus it seems appropriate to explore whether the omissions in Excel's repertoire can be remedied and to what extent interactivity can be achieved.

A goal in showing how the charts for the above-mentioned EDA topics can be constructed by Excel is to try and leave as much initiative in the hands of the student as possible. Thus we are not trying to make Excel into a package that produces EDA displays automatically; rather our preference would be to have the student construct as much of the chart as is possible. However, the Level 1 examples in which the spreadsheet has the calculations and charts ready to go have been presented in that form because the need to use such charts would precede the student's ability to apply the necessary functions within Excel. For example, to create the stem-and-leaf plot, we use the INT function and this is conceptually much more advanced than the plot itself. Hence it would probably not be appropriate to guide the student through the steps in making the chart. On the other hand, the student is guided through every step of creating a histogram because those using this form of representation are much more likely

to have the sophistication needed to use the INT function in the manner that is called for in the spreadsheet.

Stem-and-Leaf Plot

The instructions below show the student how to use the Stem-and-Leaf Plot spreadsheet that we have prepared. The spreadsheet enables the student to enter any numbers they like into a data area, and the stem-and-leaf plot for that data is created interactively.

1.   Open the Stem-and-Leaf Plot spreadsheet and note that you can delete or add new data to the Data Area. For example, make:

    (C5) 85

Note the change in the diagram that results from this change, and make your own changes to see how the plot changes in response to what you type.

2.   To make use of the simple diagram that the spreadsheet creates, you will most likely want to make a copy of it and then paste your picture into a Word document. To do this, select the cells surrounding the diagram, A20:M33. Now, hold down <Shift> and choose **Edit: Copy Picture**. In the dialogue box that follows, it is best to use the default settings as the grid lines look good in the picture, and the Picture format uses the lesser amount of space.

If you were to open a new word document and paste the resulting diagram into the document, this is what you would see:

Stem	Leaf										
100	X										
90	X	X	X	X	X	X	X				
80	X	X									
70	X	X	X								
60	X	X	X								
50	X	X	X	X							
40	X	X	X	X	X						
30	X	X	X	X	X	X					
20	X	X	X	X	X						
10	X	X	X	X	X	X	X	X	X	X	X
0	X	X	X	X	X	X	X	X			

A number of the rows in this spreadsheet are hidden, with the result that the student need only be concerned with entering their own data and copying the chart into a Word document. If you want to see how the spreadsheet is constructed, then the following steps will reveal them to you.

2

First, to see what's hiding in the hidden rows, select the whole worksheet by clicking in the top left corner of the grid:
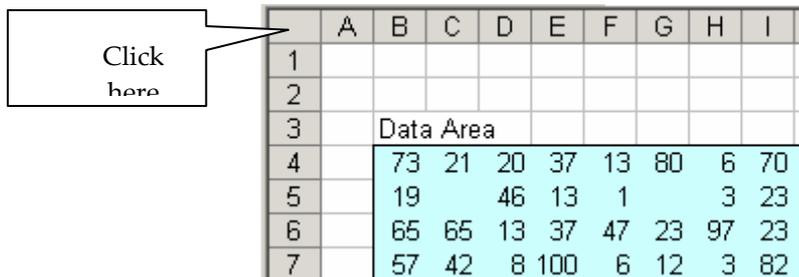
Choose **Format: Row: Unhide** and the table of values used to make the stem-and-leaf plot appear.

Use the drop-down menu of the **Name Box**, located just above cell A1, and select **intervals**. This will take you directly to the cells that have been named *intervals*. Note also how the spreadsheet uses names for all the other cells that are used in the calculations. Using names in this way doesn't take long to do, and it almost always makes the formulae that are used in the spreadsheet much easier to understand.
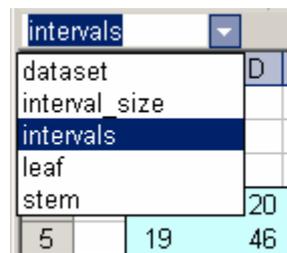
Click on cell B12 to see the formula that it uses. The formula is:

(B12) =IF(dataset <> "", interval_size*INT(dataset/interval_size), "")

In the middle of this formula is the expression that we can think of as $n*INT(x/n)$ where n is replaced by *interval_size* and x by *dataset*. This formula is really useful for finding the interval in which a particular data point lies.

Finally, click on C21 to see the formula used for creating the diagram. It is:

(C21) =IF(COUNTIF(intervals, stem) >=leaf, "X", "")

The COUNTIF function looks through the table of intervals and counts how many of them are equal to the current stem value. If that value is >= the current leaf, then we mark the cell with an X, and use conditional formatting to colour it grey as well.

In this spreadsheet, we have seen how to make a simple diagram from data without using the chart feature of Excel and we also saw that using <Shift> affects the options available on a menu. The spreadsheet also uses names and some key formulae such as $n*INT(x/n)$ and COUNTIF. All these features are ones that we would advise you to share with students in the long term, but to simply make a stem-and-leaf plot, it is not necessary that the students understand how these functions work.

Back-to-Back Charts

To create a back-to-back chart requires no sophisticated formulae and so we have set out full guidelines for how students could create the chart for themselves. The

3

aim of a back-to-back chart is to promote comparison between two datasets. Our example is based on displaying the following temperature data that was taken somewhere in Australia … we leave you to choose where!

1. Open a new spreadsheet and enter the data starting at cell B2 with the heading *Month*. To enter the Month data, all you need to do is o type *Jan* in B3 and use the Fill Handle to fill down to *Dec*.

2. Select all the data and click on the Chart Wizard, choosing the Bar Chart option and the first type of chart. The chart that appears may be too small to show each month on the vertical axis. To fix this, drag the bottom right corner of the chart downwards until all the month names appear.

Month	Max Temp	Min Temp
Jan	26	14
Feb	26	14
Mar	24	13
Apr	20	11
May	17	8
Jun	14	7
Jul	13	6
Aug	15	6
Sep	17	8
Oct	20	9
Nov	22	11
Dec	24	13

The chart is showing the data, but it is not back-to-back. There are a few changes that we need to make in order to fix the problem.

3. We are going to add an extra data column that will make the Min Temp data appear in the opposite direction on the chart. Go to E3 and enter the formula:

    (E3) = -D3

    Double click the Fill Handle to make this formula fill down to E14. Click on the white part of the chart, choose **Source Data** and select the **Series** tab. Select the **Min** series and then the **Values** box. You can now drag over E3:E14 to change to the negatives of *Min Temp*. Although they are negative in value, we at least have the Max and Min temps back-to-back. Click OK to return to the chart.

4. Right click on any of the *Max Temp* bars in the chart and select **Format Data Series**. Select the **Option** tab, which is the last on the right. Set the **Overlap** to 100 (percent) and the **Gap Width** to 20. Click OK to return to the chart.

There are just a few clean-up operations that we need to perform to make the chart look like a real back-to-back chart. These are:

5. Right click the white part of the chart and select **Chart Options**. Give the chart a **Title**, *Temperature Data*, and make the y-axis **label**:
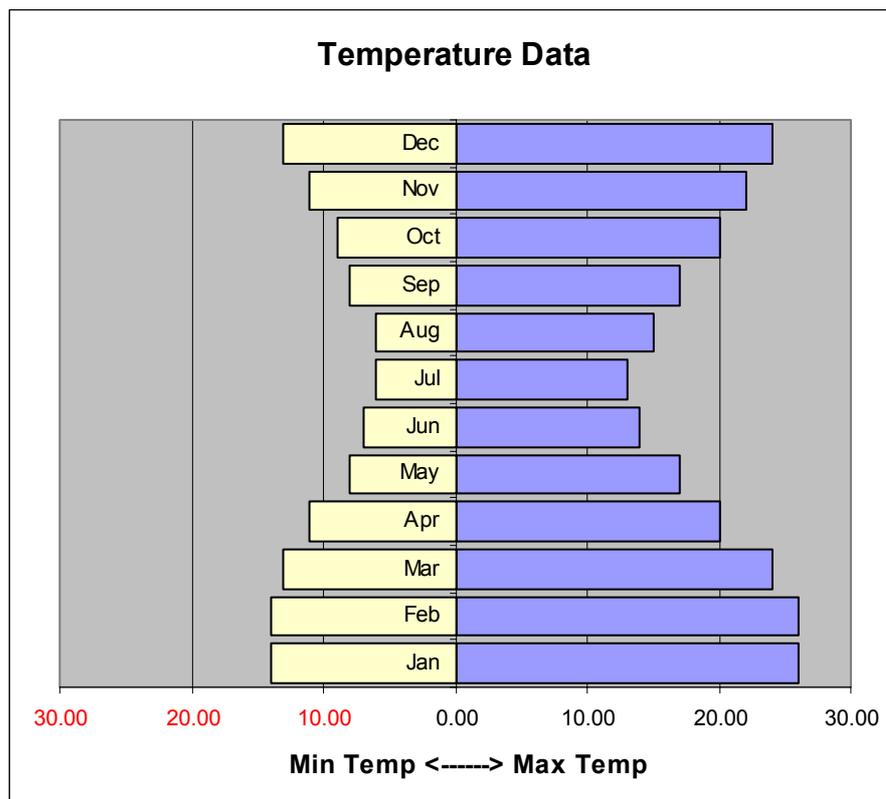
    Min Temp < - - - - - > Max Temp

    Also, remove the **Legend**.

J. BAKER

6. Making sure that the chart is selected, move the mouse into a gap between the bars and over the vertical axis. When a tool tip shows *Category Axis*, right click and choose **Format Axis**. Reduce the font size to 8 pt.

7. The *Min Temp* bars are too dark and are hiding the labels. Move the mouse over the *Min Temp* bars until the *Series …* tool tip shows. Right click and choose *Format Data Series.*, selecting the **Patterns** tab. We chose a light yellow for Area Fill colour … you choose what you like, but make it a light colour so that the axis labels can be clearly read.

8. The Value axis (horizontal) can also be formatted to get rid of the slightly confusing negative values under the *Min Temp* bars. Move the mouse over the horizontal axis until *Value Axis* tool tip appears. Select the **Number** tab, choose the **Number** option and the second option for negative numbers which now show red rather than with a minus sign. Also, select the **Font** tab and format the font to be 8 pt in size.

9. Finally, choose the **Scale** tab and set the minimum to -30 as this will have the effect of centring the axis in the chart. Click OK … we're done! It is attention to small details like this that can make a big difference.

Does your chart look like this?

Temperature Data

	Min Temp <------> Max Temp
30.00 20.00 10.00	0.00 10.00 20.00 30.00

(Chart: back-to-back bar chart showing Min Temp and Max Temp for months Jan through Dec.)

It should do … and if it does then it is a correct back-to-back chart!

These instructions are such that they can be followed by students who need to construct a back-to-back chart. Having complete control over the chart's

5

appearance gives the students a sense of ownership, and undertaking the steps needed to make the chart will enhance their understanding of its meaning.

Box-and-Whisker Plot

The aim of the next spreadsheet is to enable you to create a box-and-whisker plot of data for exploratory, descriptive purposes. But this time, we have provided a skeleton spreadsheet that automatically plots the data provided – it is a ready-to-go spreadsheet. However, we do think that the students will benefit from exploring the functions used and how the chart is based on the data provided.

1. Open the **Box and Whisker** spreadsheet and note that it has three main sections. The data is in cells B5:K7. Intermediate calculations for the range, inter-quartile range and median of the data are in the second section, and the last part for the lines that are drawn on the chart to make the box-and-whisker plot.

2. You should check the formulae used for calculating the range, median and inter-quartile points. The calculations are for the range and median are:

    (C8) = MIN(dataset)

    (C10) = MEDIAN(dataset)

    (C12) = MAX(dataset)

    The lower inter-quartile point is the median of all the data points below the median:

    (C9) = MEDIAN(IF(dataset < C12, dataset, ""))

    The meaning of the formula is "take the median of numbers in dataset only if they are less than the value in C12". Similarly, the formula for C13 is:

    (C11) = MEDIAN(IF(dataset > C12, dataset, ""))

    We recommend that you spend time discussing these formulae with a friend if you have any doubts as to how and why they work. Using some powerful formulae, like MIN, MEDIAN and MAX, we are able to locate the range, inter-quartile range and median of a dataset.

Box and Whisker Plot
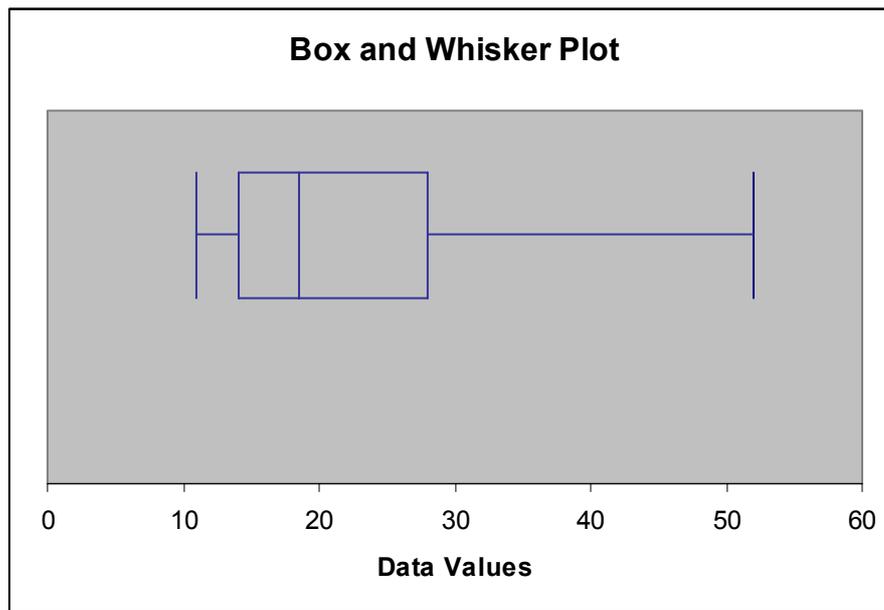
We made the chart using simple line drawing. It may have seemed a bit long-winded, but the process is simple and the result effective. Our chart gives a good representation of the data and has all the features of a box-and-whisker plot. And best of all, once we have made it, we can use it time and time again with different data sets. The long-term payoff looks promising!

On the Data Set 2 worksheet, we have included an overlay of two box-and-whisker plots. This is a minor extension of Dataset 1 and yet one that might be more used than just a single plot.

The Histogram

Here we give the detailed steps needed to create a histogram of data that needs to be displayed on a frequency chart. A histogram is a chart in which the number of data points that fall into intervals of a given size is displayed. For example, if we collect data on the heights of people in the class, we might make a chart of this information that shows how many people have heights between 1m and 1.1m, 1.1m and 1.2m, 1.2m and 1.3m, and so on. The resulting chart would show how the heights are distributed across the range of 1m to 2m.

3. Open the **Histogram** spreadsheet and zoom to 75% in order to get everything that we will be using on the one screen.

4. Give cell C3 the name *isize* as we will use this cell to fix the width of the intervals that we will use in the chart. Remember, it is much easier to enter formulae when names are used in this way.

5. In B7 we are going to enter a most important formula. It will calculate the left end-point of the interval (class) to which the data value in C7 belongs. The formula is:

7

(B7) = isize*INT(C7/isize)

Double click the Fill Handle to add this formula to the cells B8 to B46. Finally, select the cells B7:B46 and give that area the name *dataset*. Note that we don't include the original data in the named area as the data that we want to plot is the newly calculated data in column B rather than the original data which is in Column C.

We can now build the table on which the histogram chart is based.

6. Set E7 to 0 and because we will be exploring the effect of changing the interval size, make:

(E8) = E7 + isize

Drag this formula down to E18. Your spreadsheet should now look like this:

	A	B	C	D	E	F	G
1							
2							
3		Interval Size	5				
4							
5		Data Section			Plotting Section		
6		Interval	Data		Interval	Count	
7		15	18		0		
8		15	19		5		
9		30	32		10		
10		20	24		15		
11		45	45		20		
12		10	14		25		

7. To calculate the number of data points that lie in the interval from 0 to 5, which is the current *isize*, we use the COUNTIF formula. Set F7 to:

(F7) = COUNTIF(dataset, E7)

Note that what the formula does is to count how many times a data point in *dataset* is has an interval equal to 0. When we fill this formula down to F18 the new table give the number of data points in each of the intervals.

8. To make the histogram, we only need to make a column chart of the new data table. Select E6:F18 and click the Chart Wizard. Select the first of the column charts and click **Finish**.

The resulting chart is not what we want, as there are bars for both the Interval and Count data. Here is a good way to fix the problem.

9. Right click the white part of the chart and select **Source Data**. Click on the **Series** tab and, making sure that Series 1 is selected, click on **Remove**. Then click on the range selector button for the **Category (X) axis labels** and drag over E7:E18. Then click on the range selector button to restore the dialogue box. This is what you should see before you click on OK.
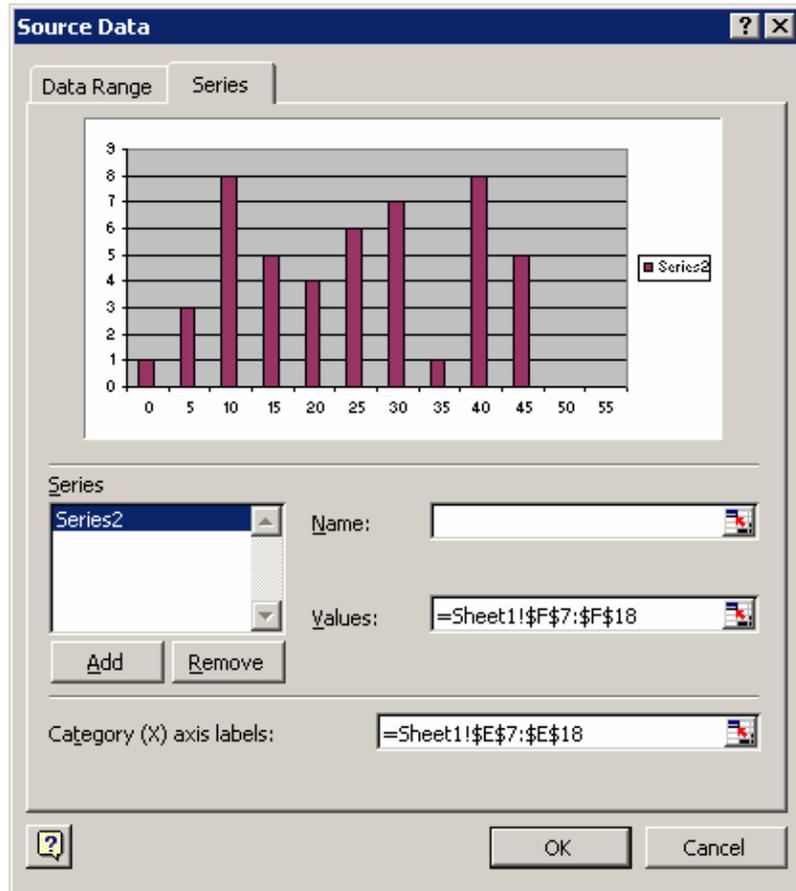
10. Tidy up the chart by removing the **Legend**, adding the **labels** *Interval* and *Frequency* to the x- and y-axes and giving the chart the **title** *Histogram*. Remember that to make these changes you only need to adjust the **Chart Options**, which you get to either from the Chart menu that shows if the chart is selected or by again right clicking on the white part of the chart.

Your chart should now look like this:
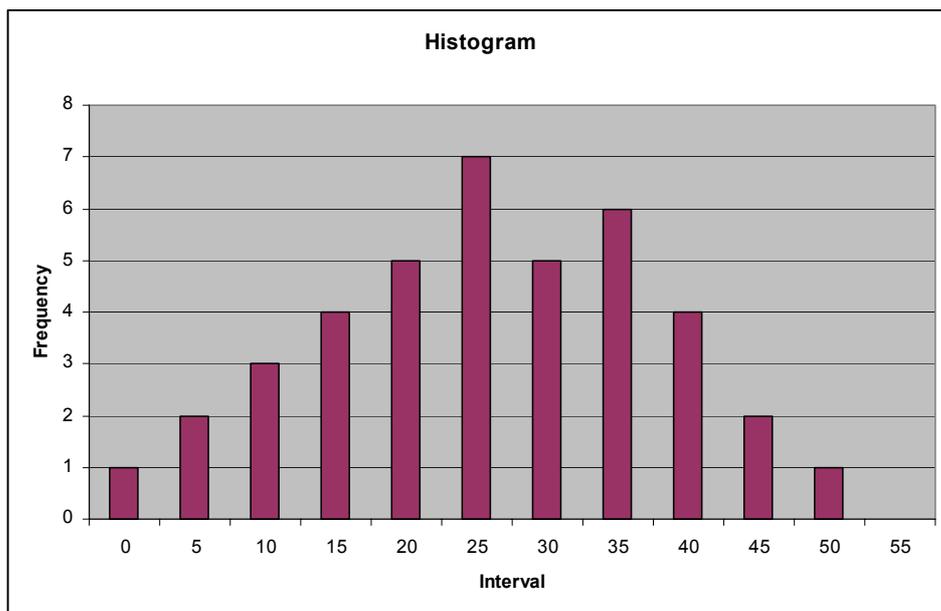
This chart is a good representation of the data in the original dataset, but it can be adjusted to improve its appearance.

First, we can make the chart look more like a histogram in which the columns are usually joined up rather than separated as we have them. Second, we can add a scrollbar that allows us to adjust the interval size so that we can see the impact that it has on the appearance of the histogram.

11.   Right click on one of the columns and choose to **Format Data Series**. The last tab is **Options** which allows us to set the **Gap Width** to 0. This will join up the columns to make the chart into the correct histogram shape.

12.   From the **Forms** toolbar, click on the scrollbar icon, , and add a scrollbar near cell C3. Choose **Format Control**, link the scrollbar directly to C5, make its **Minimum** = 2 and make its **Maximum** = 10. Click OK and then click away from the scrollbar. It will now control the interval size, *isize* or C5, and changes are immediately shown on the chart.

As you alter the scrollbar to values below 5, note that the intervals in the Plotting Section no longer extend far enough to accommodate all the data. This problem could be overcome by extending the plotting section to include more intervals.

So there you have a ready-to-go histogram maker that can be used for any data that you like to add to the data section (C7:C46) and for any interval size (C5).

Summary

Our aim was to show how to create the EDA charts in Excel while at the same time be aware of the constraints of student maturity. As students become more practiced in their use of ICTs we should be guiding them to use a package such as Excel to serve their needs, rather than having to rely on a customized package. And the more we give them control of how the display is produced, the greater their sense of ownership and understanding.