

Guidelines for good statistical graphics in Excel

Table of contents

1	Introduction.....	1
1.1	Chart Basics	1
1.2	Location	1
1.3	The Chart Object.....	2
2	Exploratory graphs.....	2
2.1	Scatterplots.....	3
2.2	Line plots	4
2.3	Factor plots.....	5
2.4	Pivot charts.....	6
3	Clear and immediate graphs.....	7
3.1	Aspect ratio	7
3.2	Axis ranges / intercept	8
3.3	Comparable ranges.....	9
3.4	Joining dots with lines	9
4	Tidying graphs	10
4.1	Background.....	10
4.2	Gridlines.....	10
4.3	Colours: black & white printing and photocopying.....	10
4.4	Legends	11
4.5	Titles	11
5	Adding to an existing plot.....	11
5.1	Adding series	11
5.2	Adding trendlines.....	11
5.3	Adding trendlines by series.....	12
6	Chart templates	12
6.1	Making your own user defined templates	12
6.2	A scatterplot template	12
6.3	A line plot template.....	12
6.4	A factor plot template	12
6.5	Where files with user-defined template charts reside	13
6.6	What cannot be saved in a template.....	13
6.7	Setting a user defined chart template as default chart.....	13
7	Adding a chart button to the standard toolbar.....	14

1 Introduction

This guide aims at making the most of Excel's inbuilt features for the production of graphics, especially for creating and customising graphics that are appropriate for statistical work. Thus this guide can be a companion for anyone who is involved with routine statistical tasks, especially producing graphics.

In this guide we explore a subset of Excel's **Standard** and **Built-in Custom chart types**, which are relevant and appropriate to create statistical graphics. We also show how to set up your own **User-defined chart templates**, which can be used to format charts easily in your chosen styles.

We assume that you have basic skills in using Excel and in producing graphs. Complete beginners could also look at the following:

- The University of Reading ITS online training material for Excel: www.rdg.ac.uk/SerDepts/su/Index/spreadsh.html
- The SSC booklet on "Informative Presentation of Tables, Graphs and Statistics": www.rdg.ac.uk/ssc/dfid/booklets/toptgs.html#a3graphsandcharts
- An introductory tutorial, freely downloadable (after registration): www.leeds.ac.uk/iss/documentation/tut/tut49.pdf

1.1 Chart Basics

The quick way to make a chart is by pressing the **Chart wizard** button on the standard toolbar. In step 1, Excel offers two collections of charts types: **Standard types** and **Custom types**. The latter type has an extra choice via radio buttons: **Built-in** and **User defined chart types**. You then continue to use the Wizard to produce the chart of your choice.

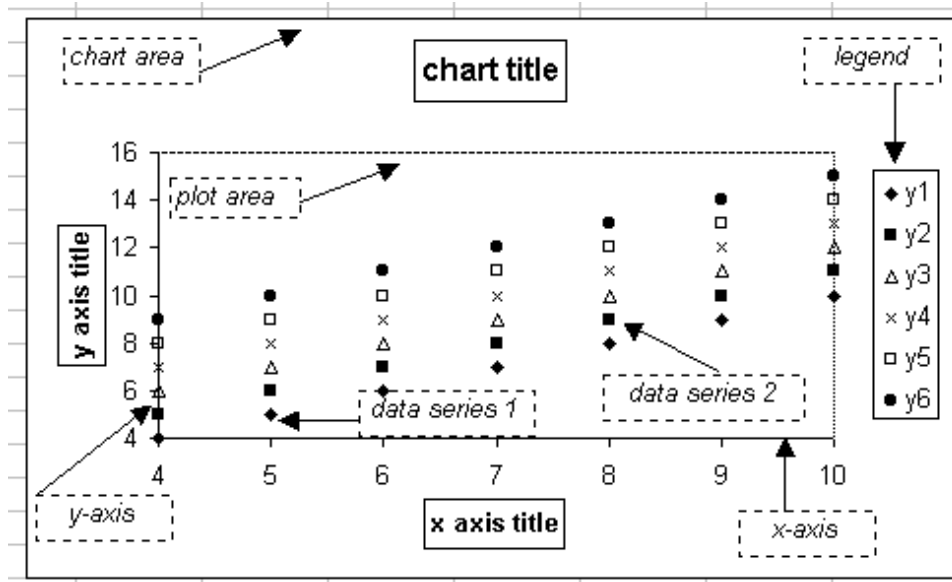
1.2 Location

In the final step of the **Chart Wizard** you choose where to place a chart, either as a new object in an existing sheet, or as a new sheet in the same workbook. Once the chart has been created, you can modify your choice by right clicking on the chart, selecting **Location** and changing the option in the **Chart Location** dialog box.

We recommend you place a chart in the same worksheet as the source data, so that you can print the charted data and the chart together. Should you wish to print a magnified version of the chart without the source data you can print the chart while it's selected.

1.3 The Chart Object

In Excel, a chart is just an object made up of several components, all individually accessible for formatting. In using Excel you have probably already selected an axis, a data series, a title or a legend for customising. We will often refer to the chart area and the plot area, which are shown in the chart below:



2 Exploratory graphs

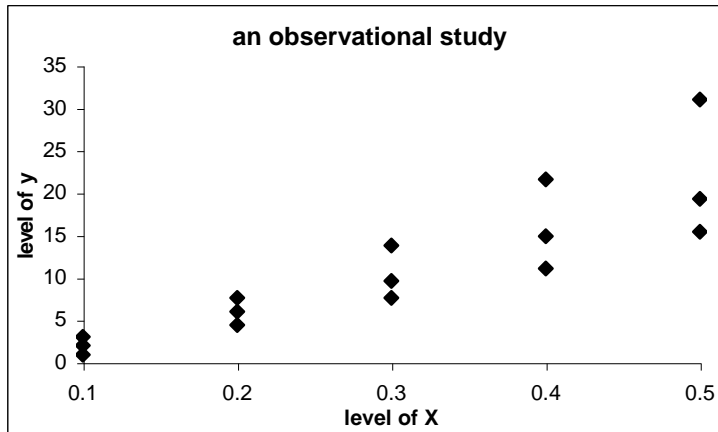
It is useful to distinguish between two uses of graphs which are common in statistical work. Exploratory graphs are designed to help the user see and understand the data during the analysis. Presentation graphs are designed to show the results of the analysis to others.

There are three basic types of plots that are often used in statistical work, especially at the preliminary stage of analysis known Exploratory Data Analysis (EDA). These are: scatter plots, line plots and factor plots. All three can be created using Excel's Standard chart types. Note that Excel's **Line plot** is in fact a form of factor plot, and that statistical line plots are best produced using a version of the **XY(scatter) plot** to give a numeric x-axis.

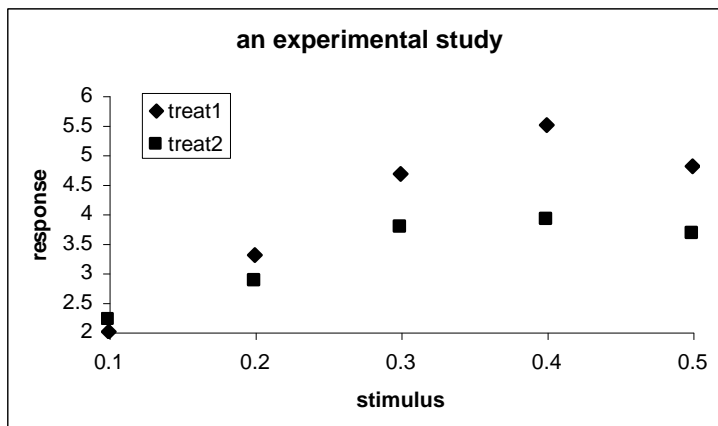
Excel can also be used to plot summary statistics with **Pivot Charts**, which is the plotting tool of **Pivot Tables**. If you have not used **Pivot Tables** before then see our guide on Excel and Pivot Tables at <http://www.rdg.ac.uk/ssc/dfid/booklets/topxfs.htm>.

2.1 Scatterplots

These are also called point plots. They provide a plot of y values against x values for the same observation, showing just a symbol for each datapoint. They are useful both in observational and experimental studies. They are good at showing relationships between the two variables and may suggest which summary model is most plausible for the dependence of y on x. With replicates at each x level or nearby, a scatterplot can indicate whether y needs a transformation to stabilise its variance over the range of x.



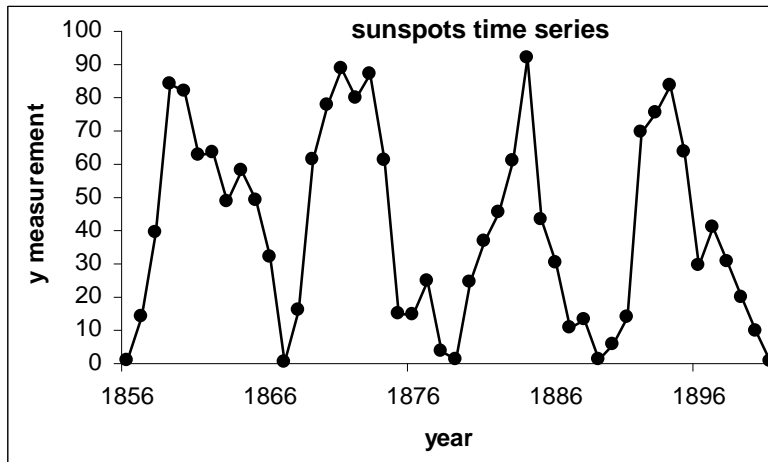
A straight line model is plausible but only after transformation of y to stabilise its variance, which is higher for higher values of x



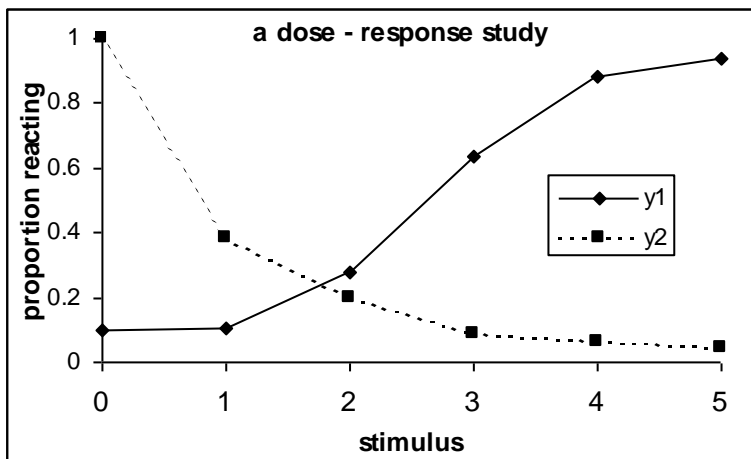
A quadratic model, or other curved model, is required. Also the rate of response to stimulus is different for the 2 treatments

2.2 Line plots

Line plots are useful to show trends or other patterns in the relationship between y and x . They are particularly useful for time series data, where they help identify trends, periodicity or other interesting features. Straight line interpolation between the points is preferable to other interpolation techniques, to describe the underlying association between x and y , and especially to decide on recognisable functional forms.



In this time series there is a cycle roughly every 10 to 12 years, there is a blip towards the end of the cycle, and each cycle rises more quickly than it subsides



y1 looks like a sigmoid curve,
y2 looks like an exponential decay

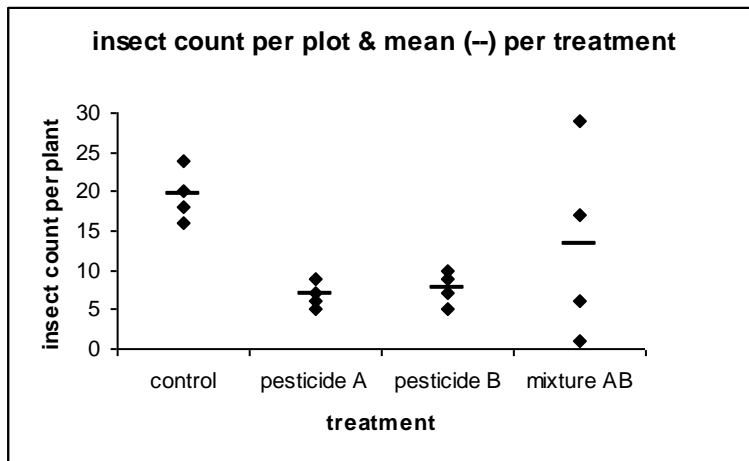
Note that these charts are plotted as XY scatterplots in Excel using a numeric x-axis. Interpolating lines are added by selecting a suitable chart sub-type in the **Chart Wizard** or by editing the series format on a scatterplot.

2.3 Factor plots

It is possible to use scatterplots at the exploratory stage when dealing with designed experiments with qualitative treatments with no inherent ordering. These are often known as factor plots. We can plot the raw values to check that their spread within each treatment is not very different [assumption of equal variances]. To plot the treatment means, you can compute these in an additional column using the AVERAGE() worksheet function and add the new data-series to the chart using a different plotting symbol.

Factor plots differ from scatterplots in that the x-axis consists of a number of categories, rather than a numerical scale. For a factor with no inherent ordering, we strongly recommend that the plotted points should not be joined by lines. Lines imply a sequential relationship between the categories, which may be highly misleading.

To obtain a factor plot, start from a **Line Chart** to obtain a chart with a categorical x-axis with the appropriate labels. Eliminate the joining lines and format all data series [except for the means] with the same plotting symbol, as shown below:

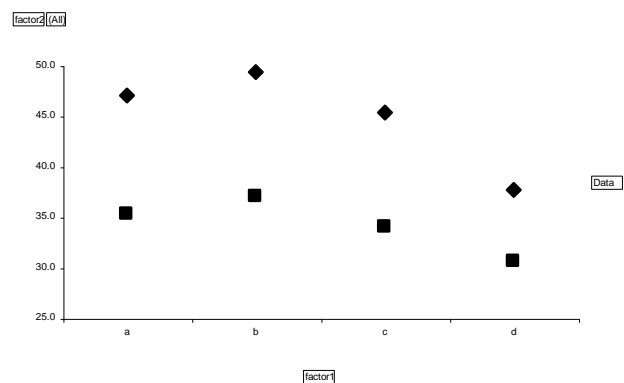


It is clear that the insect count for the mixture has a much wider variability than the remaining three treatments and that either pesticide greatly reduces the insect count.

2.4 Pivot charts

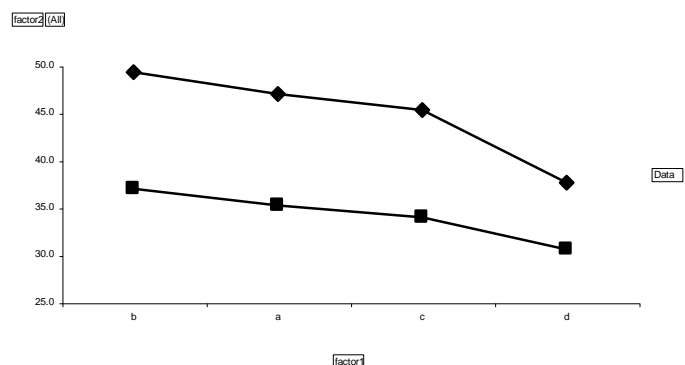
If you are already familiar with the use of Excel's statistical functions to compute summary statistics for plotting, you can take this a step further by using **Pivot Charts**. Discipline in data entry and management is required here as the data source must be in a **list** format for the **Pivot Tables** to work (to learn about the list format use the Help system). Once you've obtained the desired summary statistic in the **Pivot Table** you can press the **Pivot Chart** button on the **Pivot Chart** toolbar and a system default bar chart is created as a new sheet (the number of summary statistics available in **Pivot Tables** is limited, for example percentiles and standard errors are not directly computable). You can now edit the pivot chart, change chart type (though you cannot make it a scatterplot, since **Pivot Charts** are essentially **factor plots**), customise lines and symbols, etc.

factor2	(All)	
	Data	
factor1	Average of y1	Average of y2
a	47.0	35.3
b	49.3	37.0
c	45.3	34.0
d	37.7	30.6



Notice that the x-axis label and the legend label in the **Pivot Chart** are active field buttons, with a pull down list from which you can choose to hide levels on the x-axis or entire series on the y-axis. As the chart is dynamically linked to the pivot table containing the data source, you can change the order of the labels on the x-axis by moving the corresponding label row by dragging and dropping it within the pivot table. For example the default order for a text x-axis is alphabetical but you may want to rearrange it in order of decreasing value on the y-axis, as shown below:

factor2	(All)	
	Data	
factor1	Average of y1	Average of y2
b	49.3	37.0
a	47.0	35.3
c	45.3	34.0
d	37.7	30.6



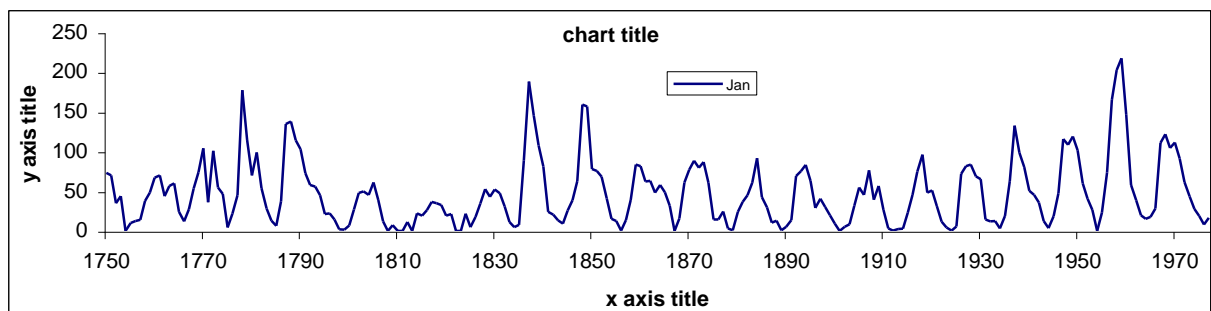
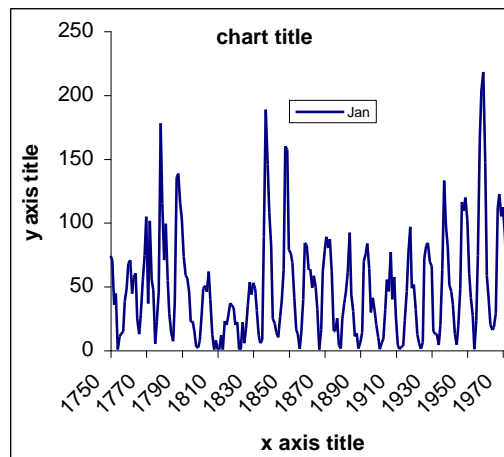
Remember though that pivot tables are not dynamically linked to the source data. Should your raw data change you need to update your **Pivot Table** by using the Refresh data button with the ! symbol on the **Pivot Table toolbar** and the **Pivot Chart** will be updated automatically.

3 Clear and immediate graphs

There are a few aspects concerning the chart shape, size, axis range and datapoint interpolation that can help bring out and clarify patterns in the data.

3.1 Aspect ratio

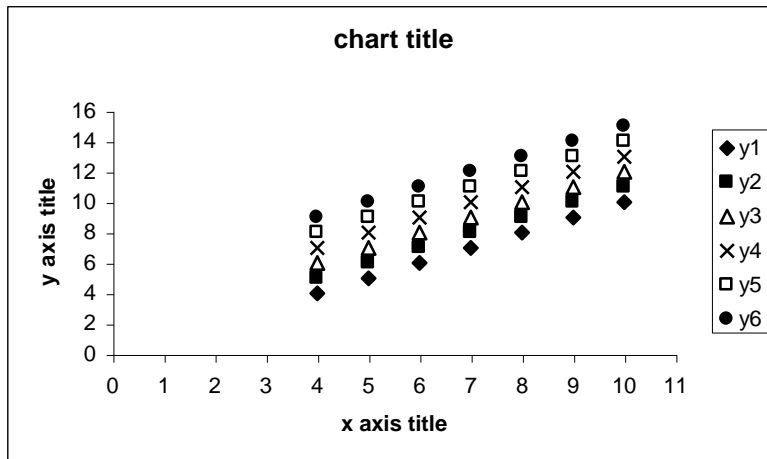
The height/width ratio of the two axes in the plot area is a critical factor, which determines how clearly a chart displays the patterns in the data. This is especially true when graphing series that exhibit cycles. Adjusting the ratio is very simple in Excel as both the chart area and the plot area can be resized once either area is selected. You might want to consider resizing not only to copy charts into reports but also to emphasize patterns in the data.



The top line plot with an aspect ratio of 1 only shows the difference in height of the cycles. However, the bottom line plot with an aspect ratio of 0.25 also shows that cycles rise more rapidly than they fall, though this trend is less pronounced for medium peaks and doesn't hold for the lowest peaks. Note also that for this long series of 220 observations we have not shown each dot but just the lines joining them.

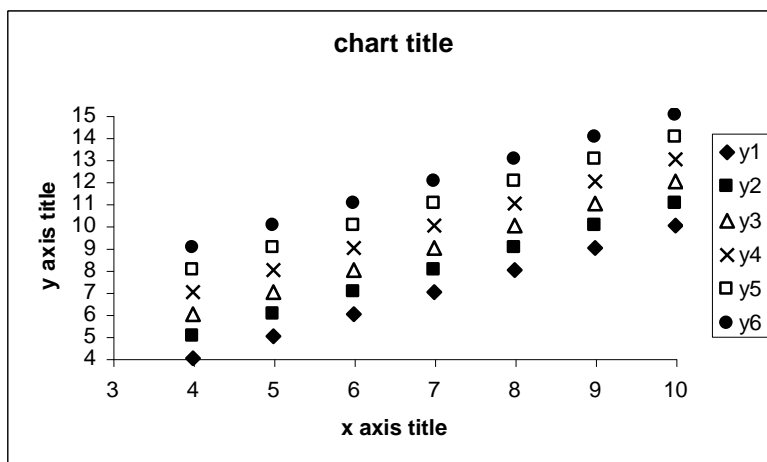
3.2 Axis ranges / intercept

Often the default scatterplot leaves a lot of blank space below and to the left of the plotted series since the x and y-axes start at zero. However, the information conveyed by the chart is only in that part which contains the actual data, i.e we need to see this variation clearly. To make full use of the plot area and to emphasize patterns in the data it is best to make both axes start at the minimum value of the plotted datapoints, or some other sensible reference value. To do this, select the plot area, select the x-axis, click twice on it to make the Format axis dialog appear, select the Scale tab and set the minimum and maximum yourself.



Default x-axis range is 0 to 11 and default y-axis range is 0 to 16.

This leaves too much blank space in the plot area.

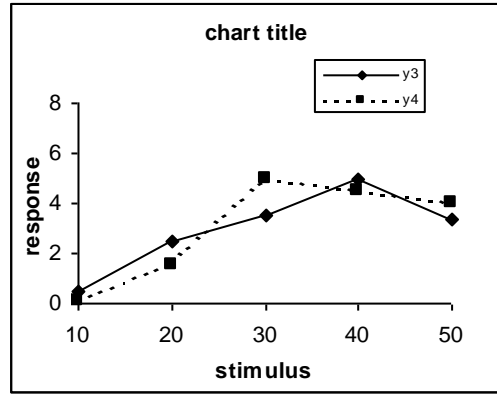
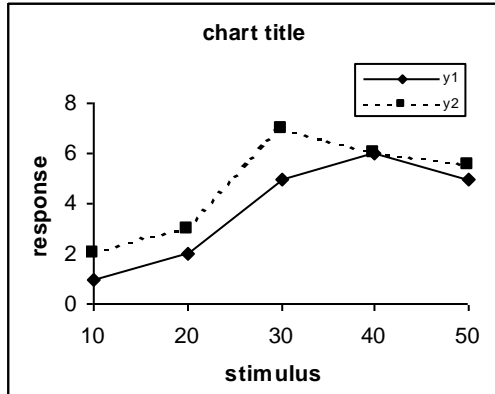


The x-axis starts at 4 and ends at 10, the y-axis starts at 4 and ends at 15.

The blank space within the plot area is minimised and the variability in the data is emphasized.

3.3 Comparable ranges

If you have so many series to plot that would make the chart too cluttered to be informative you can split these into two or more charts, keeping the same variable on the x-axis. Y-series plotted on separate charts with the same x-axis are then still comparable provided that the range of both x and y-axes is the same for all charts.

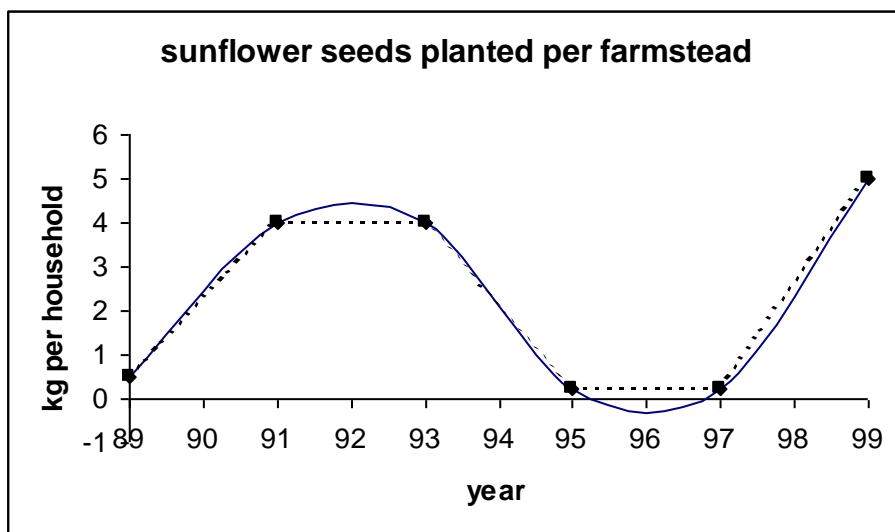


Although all four series show a similar pattern of response, clearly y3 and y4 start responding to the x stimulus only at 20 units and do not reach as high as y1 and y2.

If you have a lot of data you may need to divide the display into many such charts. The modern way of dealing with this type of display is called Trellis graphics. These are not yet available in Excel.

3.4 Joining dots with lines

One of the inbuilt sub types of the **Scatterplot Chart** shows datapoints connected by smoothed lines. These are interpolation lines that are forced to pass through the datapoints and are not based on any specific statistical technique. The graph below shows smoothed solid line. We discourage their use, as they can be misleading when interpreting a trend.



The smoothed line suggests that the quantity of seeds planted in 1992 exceeded both 1991 and 1993, whereas it is logical to assume it has remained unchanged. Also it implies that in 1996 households used a negative quantity of seeds, which does not make sense. We recommend giving the estimates for the intervening years by straight interpolation shown by the dashed line. Notice also that the solid interpolation line deviates slightly from the dashed line at other places, caused by the fitting of a smooth, but not necessarily sensible, curve.

4 Tidying graphs

The default settings in Excel produce a plot area with a cluttered grey background that distracts the eye from seeing the pattern in the plotted datapoints. We should create a graph with only those elements that convey useful information and omit unnecessary or redundant features. Some of these defaults can be altered by changing the defaults on the **Color tab** in the **Options dialog box** under the **Tools menu** choice. These **Options** are presented below.

4.1 Background

A dark grey background distracts the eye from focusing on the plotted datapoints and will result in a dark print. To produce a tidy graph select the plot area, right click on it and choose **Clear**.

4.2 Gridlines

The default settings for most chart types add solid gridlines in the background. Unless needed for a specific purpose these should be turned off or at least made less conspicuous by making the lines dashed or grey (or both). Double click on the gridlines to make the **Format gridlines** dialog box appear.

4.3 Colours: black & white printing and photocopying

Even if you have a colour printer, the default colours assigned to the plotted series are sometimes difficult to distinguish or even see, e.g. the system default colour for the third line is bright yellow. You can reset them by the menu selection **Tools, Options** and selecting the **Color tab**. The **Chart lines** choice shows the colour for the first 8 lines and symbols: to avoid colouring, make all of them black by selecting each square in turn and pressing the **Modify** button. You can do the same with the first eight bar chart fills by alternating black, grey and white. However, you cannot customise the bar fill pattern or the line style in this dialog box. These changes take place immediately, i.e. they change active graphs (where the colours have not been changed from defaults) in the current session.

Thinking carefully about the use of colours – i.e. what is being shown by them, as with other graphs characteristics – will lead to better graphs.

These custom defaults are not available across workbooks. One way of transferring custom colour defaults to a newly opened workbook is in the same **Color tab** of the **Options dialog box**; in the “**Copy colors from**” frame choose the appropriate workbook from the pull down list (this workbook must be open).

An alternative is to save the customised workbook as a template workbook (*.xlt extension). To find out more about workbook templates, use the **Excel Help** system:

type the key word “**template**” in the **Index tab** and press **Enter**: the relevant information appears as the very first topic in the help list.

4.4 Legends

A legend is often necessary to identify the series in the plot area and Excel offers a choice of 5 placements. However, all 5 locations are outside the plot area. To make the most of the plot area a legend is best moved inside the plot area whenever this contains enough blank space for the legend not to partially hide important information. Most chart types show a legend by default; like other chart objects, you can delete a legend by selecting it and press delete. If you want to add a legend when it is not there, the quickest way is to bring up the **Chart toolbar** (menu selection **View, Toolbars**, check **Chart toolbar**) and press the **Legend button** while the chart is selected. An alternative way is to right click on the **Chart area**, select **Chart options** and the **Legend tab**.

4.5 Titles

The font used in the spreadsheet view, inherited by axis labels and chart title, is set on the General tab in the **Options dialog box** from the **Tools menu** choice. Changes to defaults for font styles and size done this way will take place in the next Excel session. If you want to customise axis labels and chart titles in the current session you have to do it in an active chart.

The default settings make titles resize automatically when you resize the chart. To turn off this feature, right click on the selected chart and choose **Format chart area**: in the **Font tab** uncheck the **Auto scale** checkbox and reset the font size to your choice.

5 Adding to an existing plot

5.1 Adding series

If the graph object is embedded in the same worksheet as the data table, simply highlight the block containing the new series (including the row with the series name) and “drag and drop” it in the chart area. Then add a legend if necessary. If the chart is in a separate worksheet then highlight the series, press the copy button, switch to the chart sheet, select the chart area and press the **Paste button**. Alternatively you can right click on the selected chart, choose **Source data**, and expand the data range to include the new series. The latter method works in either chart location.

5.2 Adding trendlines

You can add a trendline based on statistical models by selecting a series, then from the **Chart menu**, select **Add trendline**, and in the **Type tab** choose between a straight line and a curve, that is a polynomial of order 2 (we discourage the use of polynomials of higher orders). You can also choose to show the regression equation and R-square value in the **Options tab**.

5.3 Adding trendlines by series

If you have more than one series in the same chart then you can add separate trendlines for each series just by selecting each series before you choose **Add trendline** from the **Chart menu**.

6 Chart templates

It is a very repetitive task to format titles, gridlines, background and line styles for every new chart in order to obtain a tidy chart as described above. If you are likely to use the same type of chart repeatedly, it is worth thinking about setting up user defined chart templates. These allow you to store re-usable pre-formatted charts.

6.1 Making your own user defined templates

Firstly create a tidy chart. Once you are satisfied with it, make this type of chart available as a user-defined template by making the finished chart active and selecting the **menu Chart, Chart type, Custom types, User defined**. Press the **Add button**, then give it a short name and a meaningful description for future use. Next time you want to use one of these charts simply select your data range then choose the chart from the **Custom types, User defined gallery** in step 1 of the **Chart wizard**. Some examples are described below.

6.2 A scatterplot template

Make an **XY scatterplot** with one y series, then clear the background, delete the gridlines, and format the axis labels and chart title as you like. If you have followed section 5.3 on colours, the symbols for the first eight y-series will all be black, thus distinguishable without a colour printer. After pressing the **Add button** name the template “**ClearScatter**” and describe it as “**Clear background, no gridlines, up to 8 series in black**”.

6.3 A line plot template

Make a copy of the above graph, paste it in the same worksheet, right click on it and change **Chart type** to **Scatterplot** with the datapoints connected by straight lines. Add at least 4 more y series and format each line style differently to make them distinguishable. The reason for using so many series is that should you start with, say, one y series, you can then add up to 4 more on the same chart before you hit Excel’s internal line style formatting. After pressing the **Add button**, call it “**ClearLines**” and describe it as “**Clear background, no gridlines, up to 5 series in black**”.

6.4 A factor plot template

Make an **Excel Line** plot with symbols for each datapoint for at least five y-series (equivalent to five replicates per treatment), change the data range to row or column as necessary to obtain the treatment labels on the x axis, clear the plot area and gridlines, delete the joining lines of all series and format all series to show the same symbol (once the first series has been done, just select the next series and press F4). After pressing the **Add button**, call it “**FactorPlot**” and describe it as “**Shows up to 5 replicates in black but treatment means need adding**”. Once this chart is added to the user-defined templates, notice how its icon shows that this is a **Line plot**.

6.5 Where files with user-defined template charts reside

On your PC, use **Find Files** to find the location of a file called XLUSRGAL.XLS. Open it to see that it contains your three user-defined templates. Should you want to take these templates with you to use on another PC, just save the XLUSRGAL.XLS workbook on a diskette and copy it to the relevant system folder on another PC.

6.6 What cannot be saved in a template

There are some customisations that are better not, or cannot be, saved in template charts, as they need specific series to be active:

- Axis ranges are data dependent: if you format the axis scale to be within the range 1-10, then in the new chart these values will be retained. Consequently if your new y-series starts at 11, your chart will be blank and you will have to reset the axis scale.
- Added trendlines [with displayed equations and R square] and error bars: they will simply disappear from the fresh chart.

6.7 Setting a user defined chart template as default chart

When you create a new chart, Excel will normally create a **Column Chart** by default. This will inherit the black, grey & white color scheme you set in section 4.3, but its background will still be cluttered and it is probably not the type of chart that you wanted. You can make a new chart inherit the appearance of one of your tidy templates by setting that template as the default chart. Select any active chart, then from the **Chart menu** select **Chart type, Custom type tab**, select from **User defined**, select the **CleanScatterplot icon**, then at the bottom of the dialog click the **Set as default chart** button. If you are successful, notice that the top icon, whose caption reads “default”, has changed from a bar chart to a scatterplot (this action can be undone by clicking the **Delete button** while the new default chart is selected). Now new charts created by the **Chart wizard** will be of the **CleanScatterplot type** by default, but if you choose a different **Chart type**, say a **Column chart**, it will inherit the black and white colour scheme and clean background of the **CleanScatterplot** default chart.

WARNING: this action has a side effect: setting a **Scatterplot** as **default chart** will prevent you from producing **Pivot Charts**. When you press the **Chart Wizard** button, Excel only returns an error message and the Chart Wizard dialogue box is not opened, so you cannot request a different **Chart type**. Thus if you often create **Pivot Chart** it is a better idea to set a **Factor plot** as a default chart.

7 Adding a chart button to the standard toolbar

If you often use the same user defined template, it makes sense to have a shortcut to access it in fewer mouse clicks than when using the **Chart wizard**: this can be done by adding a button on the standard toolbar. From the **Tools menu** select **Customize, Commands tab**, select the **Charting category**, locate and select the **Default chart command** in the list in the right hand pane of this dialog box, drag and drop it on the standard toolbar on the right of the **Chart wizard button**. Then press the **Modify selection button** and in the **Name box** add the template chart name so that the tooltip reads something like “**Default chart: CleanScatterplot**”. Finally close this dialog: you are now able to produce this chart type with just one mouse click after selecting the data source.

As a quicker alternative, you can just highlight the block of cells containing the source data and press F11. This creates a chart of the default type and locates it onto a new worksheet.