Using Excel to produce graphs - a quick introduction:

This handout presupposes that you know how to start Excel and enter numbers into the cells in a spreadsheet.

Introduction:

The easiest way to produce graphs in Excel is to use the "Chart Wizard". This takes you through the process of producing a graph from start to finish, in a number of simple stages. You can open the Chart Wizard by clicking on the icon on the right-hand side of the toolbar at the top of the screen. (The icon looks a bit like a little house with the sun above it on my Mac; on a PC, it looks like a little bar chart).

In essence, the Chart Wizard will want the numbers that correspond to the values that you want to appear on the graph. That means that, for example, if you want a graph of "average scores for groups A, B and C", you need to provide it with the averages - NOT the raw data.

How to produce a very simple bar chart:

Suppose we want to produce a bar chart that shows the mean performance on a statistics test for three successive years of psychology students. Here are the data (maximum score is 100).:

student year:	1999	2000	2001
mean score:	65	75	80

Step 1:

Enter these data into Excel, in separate cells of the worksheet, exactly as they are here. (You could put them anywhere in the worksheet, but for simplicity's sake, start in the top left-hand corner of the worksheet).

Step 2:

Click and drag from the cell containing "mean score" right-hand cell on the same row (the one containing "80"), to highlight this row of cells. (This involves pressing down the mouse button on the first cell, and keeping it held down while you move the mouse to the last cell. Then release the button).

Step 3:

Click on the Chart Wizard Icon. A new window opens, entitled "Chart Wizard - Step 1 of 4 - Chart Type". Select the kind of chart that you want to produce, by clicking on the relevant icon to highlight it. We want "Column", so make sure that this is highlighted. Then click on "Next".

Step 4:

You should now see a graph, within a window entitled "Chart Wizard - Step 2 of 4 - Chart Source Data". Click on "Next".

Step 5:

You should now see a window entitled "Chart Wizard - Step 3 of 4 - Chart Options.
On the left side of this window are five boxes. You can use these to label the graph, and put labels on the axes of the graph. Just click in a bow, and then type your entry. The "Category (X) Axis" deals with the label that goes along the bottom of the graph, and the "Value (Y) Axis" deals with the label that goes along the graph vertically.

In the "Chart Title" box, type "Statistics performance from 1999 to 2001": don't worry that the box isn't big enough, as the text will scroll along as you type.

In the "Category (X) Axis" box, type "year of study".

In the "Value (Y) axis, type "Frequency (%)".

(You can enter information in these boxes in any order, not just the one recommended here).

(Don't worry about "Second category (X) Axis" and "Second value (Y) Axis" at this stage).

When you have the labels entered, click on "Next".

Step 6:

Your graph should disappear, to be replaced by a window entitled "Chart Wizard - Step 4 of 4 - Chart Location". There are two options: one is for the chart to be opened in a new sheet in the Workbook. The other (the one that it's assumed you want) is for the chart to be placed on the worksheet that you are currently using. We'll keep our chart on the current worksheet, so just click on "Finish".

Step 7:

The chart should now appear, in a white square with little black dots in its corners and halfway along each of its sides. If you click *outside* the white square, the dots will disappear, and you can then go on to do other things with the worksheet (such as save the worksheet, enter more data, etc.)

If you press and hold down the mouse button *within* the square, but *not on the graph itself*, a little hand appears and you can move the graph around the worksheet.

If you press and hold down the mouse button *on the graph itself*, you can move the graph around within the white square. The graph's title at the top should stay put, but the graph and the X and Y labels should move around.

If you press and hold down the mouse button on any of the little black squares, you can resize the graph. The corner squares allow you to resize the graph, keeping its original proportions; the side squares enable you to widen

the graph; and the top and bottom squares enable you to make the graph vertically longer.

Step 8:

If you wanted to, you could now print the graph. There are two options: if the little black squares aren't visible, then "print" (on the "File" menu at the top of the screen) will print the whole worksheet that you can see - in other words, you will get a print-out of all the cells of the worksheet, as well as the graph. If the black squares are visible, then "print" will produce just the graph, and not the rest of the worksheet. You can have a look at what you will get as print-out, by using "print preview" (also on the "File" menu). (To get out of "Print preview" and back to your graph, click on "close" at the top of the screen).

Customising and tinkering with your graph:

As well as the graph, you should see a little toolbar that's placed on the worksheet itself, rather than with the rest of the toolbars. This enables you to alter any aspect of the graph. This is useful, as the graphs that emerge from the steps taken by the Chart Wizard are seldom exactly as you want them.

On the left side of the toolbar is a small menu that has the words "Chart Area" in it. If you click on the small button immediately to the right of this, you can select any part of the graph - the words "Chart Area" will be replaced by other phrases, such as "Category Axis", "Legend", etc. The buttons to the right of the window then enable you to make various changes to the item that you selected in the window - so, if "Chart Area" is selected, you can use the buttons to make alterations to that; if "Legend" is selected, you can use the buttons to change various attributes of the legend; and so on.

For example, pick "Chart Area". Click on the button next to it, which enables you to "Format Chart Area". (A useful tip is that if you move the cursor over an icon and let it dwell there for a few seconds, a little label often pops up to explain what the icon does!) A window opens, containing various options for formatting the chart area. You can alter the border for the graph, and you can also change the colour or pattern of the background to the graph itself. Let's do the latter: below the word "Area" are two buttons. Click on "none", and then click on "OK" at the bottom right of the window. The window should disappear, and your graph should now have blue bars on a white background.

Open the "Format Plot Area" window again, and this time click on "Fill Effects". Yet another window will open. Click on one of the gradient patterns shown, and then click on "OK" to close this window; and then "OK" again, to close the "Format Plot Area" window. You should now have a graph that has a background behind the bars that goes smoothly from dark grey to light grey.

As well as using the toolbar, you can also make changes to a graph by double-clicking on the part of the graph that you wish to alter.

Have a play with these options: try: adding a border to the graph; changing the font size for the labels on the axes; change the font for the graph's title; change the bars from plain blue to pink gingham check!

Other ways of changing the graph:

After having produced your graph, you might decide that you want to make a more fundamental change to it. If you click on the graph so that the little black dots surround it, an option called "Chart" will appear on the toolbar at the top of the screen. There are various options on this.

"Chart Type" enables you to change the graph from whatever type it is now, to some other type. For example, we might decide that we want a line graph instead of a bar chart: if so, click on "Chart Type", and a window appears that allows you to select from the various types of chart that are available. Click on "line" and then click on "OK": your chart will change from a bar chart to a line graph. (You can use "Press and Hold to View Sample" to get a preview of what your graph will look like in its new format, without committing yourself to the change!)

"Source Data" enables you to change the data shown on the graph. For example, at the moment our source data are the three numbers in cells B2, C2 and D2. Enter these numbers in cells B4, C4 and D4: 40, 90, and 50. Now click on "Source Data" (making sure beforehand that the little black dots surround your graph!) Delete the contents of the window entitled "Data range:". You can either type the new data's location into this window, or - much more simply - click and drag over the new data. The values will appear in the window, and you will get a small preview of what the new graph will look like.

"Source Data" also enables you to put category labels on the horizontal (X) axis of the graph. It would be better to label the bars "1999", "2000" and "2001" rather than "1", "2" and "3" as they are at the moment. To do this, click on "Source Data", and then click on the small window to the right of "Category (X) Axis labels". Then click and drag over the cells containing the labels that you want to use - in this case, therefore, click and drag over cells B1, C1 and D1. Click on "OK", and Excel will replace the labels for the bars, on the horizontal axis of the graph.

"Chart options" enables you to change the title of the graph, the scale on the axes, whether or not there is a legend on the graph, and a zillion other things besides. If you click on "gridlines", and then click on the little X in the box next to "Major gridlines", you can get rid of the gridlines on your graph. (I hate those gridlines...)

How to produce a slightly more complicated bar chart:

Here, what we will do is produce a bar chart with two sets of bars on it: one set will show male performance over a three-year period, and the other set will show female performance over the same time-scale.

student year:	1999	2000	2001
male mean score:	78	56	62
female mean score:	65	75	80

Step 1:

Enter these data into Excel as before, in separate cells of the worksheet, exactly as they are here.

Step 2:

Click and drag from the top left cell containing data to the bottom right-hand cell containing data - i.e., from "78" down to "80".

Step 3:

Click on the Chart Wizard Icon. A new window opens, entitled "Chart Wizard - Step 1 of 4 - Chart Type". Select the kind of chart that you want to produce, by clicking on the relevant icon to highlight it, as before. We want "Column", so make sure that this is highlighted. Then click on "Next".

Step 4:

You should now have a graph containing three pairs of bars. Within each pair, one bar represents the male score, and the other bar shows the female score. The remaining steps, and the facilities for altering the graph in any way you care to think of, are all as before.

You'll notice that the legend isn't particularly informative, as it says "Series 1" and "Series 2". To change this to "male" and "female" respectively, do the following. Click on "Chart" on the toolbar. Select "Source Data". In the box beneath the word "Series", highlight "Series 1" (if it isn't highlighted already); click in the box next to "Name", and type "male" into this box. Then highlight "Series 2" and type "female" into the box. Click on "OK": the window will close, and you should find that your legend now says "male" and "female" instead of "Series 1" and "Series 2".

Putting error bars on graphs:

The graphs we have done so far would have bars representing the values of averages (means or medians), or frequencies of occurrence of categories. In the case of graphs showing averages, we usually also want to show some indication of how scores are spread out around the averages. Suppose we wanted to have a graph showing mean reaction time on a task for different ages of participant. We might have, say, four bars representing mean performance at ages 5, 10, 15 and 20 years of age.

Here are the data:

age (years):	5	10	15	20
mean RT (msec):	2400	1800	870	680
standard deviation:	900	780	200	97
standard error:	225	195	50	24
N of participants:	16	16	16	16

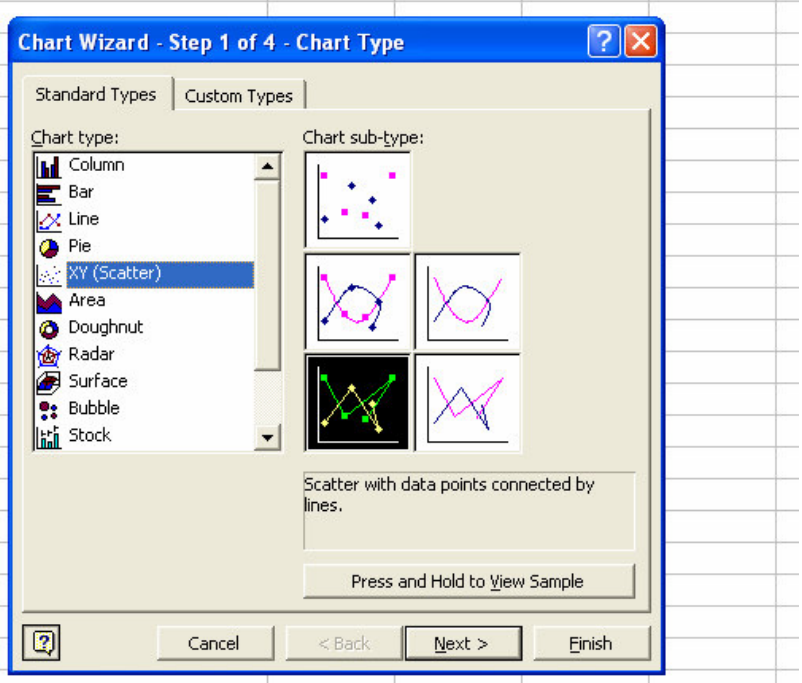
Looking at these data, it's clear that mean RT decreases with age; it's also clear that the spread of the scores also decreases considerably as the participants get older, so we would like to include this information on our graph.

We could include either the standard deviation or the standard error associated with each of these four means. How do we do this?

Step 1:

Enter these data, and obtain a line graph. The best way to do this is as follows. Select the ages and mean RT's, by clicking and dragging. On the first window of the Chart Wizard, select "XY (Scatter)" as the type of graph, and "Scatter with data points connected by lines" as the Chart sub-type. (This forces Excel to produce a line-graph with a correct horizontal axis: you might think that the logical choice of graph was "Line Graph", but if you pick this, Excel doesn't plot the horizontal labels in the right places on the graph!)

age (years):	5	10	15	20
mean RT (msec):	2400	1800	870	680
standard deviation:	900	780	200	97
standard error:	225	195	50	24
N of participants:	16	16	16	16

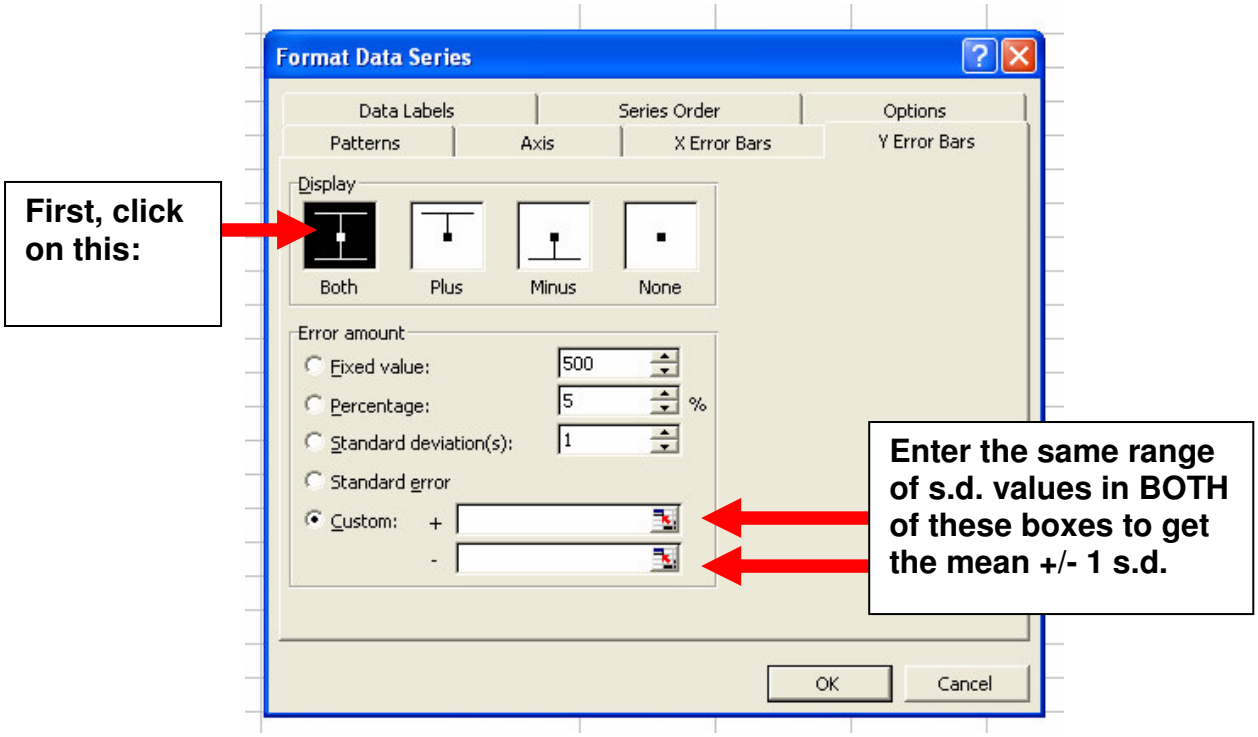


Step 2:

Double-click on the line on the graph (i.e., the line that connects your means together). The window "Format Data Series" pops up. Click on the tab entitled "Y Error Bars". This produces a new window, that shows two things.

The first thing shown is various options for error-bar appearance. You can have no error bars at all; ones that stick up above and below each mean

on the graph; ones that stick up above the mean only; or ones that are shown only below the mean. Click on the far left picture, "both" (the one that shows an error-bar both above and below the mean).



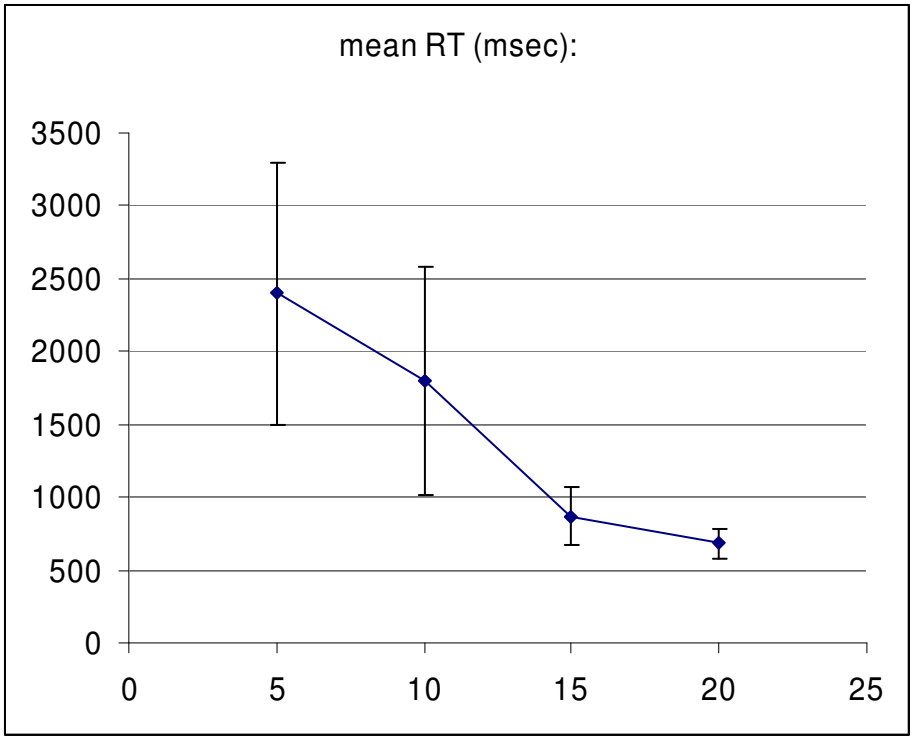
The second set of options that are provided by this window are various choices for what the error bars will actually represent. Ignore all but the bottom one of these, labelled "custom", as apparently Excel calculates the error bars incorrectly if you use the other options!

To produce error-bars on the graph that will show each mean plus and minus its associated standard deviation, do the following. There are two windows to the right of the button labelled "custom". One window has a little plus sign next to it, and the other has a little minus sign. Click in the window next to the plus sign, and then click and drag over the cells in the worksheet that contain the standard deviations. Then click in the window next to the minus sign, and once again click and drag over the cells that contain the standard deviations.

Click and drag over these numbers twice - once to enter them into the "+" box, and again to put them into the -" box:

age (years):	5	10	15	20
mean RT (msec):	2400	1800	870	680
standard deviation:	900	780	200	97
standard error:	225	195	50	24
N of participants:	16	16	16	16

Finally, click on "OK". The "Format Data Series" window will disappear, and your graph should now show each mean together with its error bars.



The procedure in this section will work just the same with a bar chart, if that's what you want rather than a line graph. Note that Excel doesn't "know" that the numbers entered into the boxes are standard deviations: it will simply draw vertical lines above and below the means to represent whatever values are entered into the "+" and "-" boxes. Thus if you wanted, you could have the means plus and minus 2 s.d.'s, plus or minus 3 s.d.'s, plus or minus the range, or plus or minus the standard error. In practice though, most graphs either show the mean plus or minus 1 s.d., or plus and minus 1 s.e. (See next section).

What sort of error-bars should you use?

Different people use different type of error bar. Sometimes you will see means displayed plus and minus their standard deviations, as in the example above. At other times, you will come across means shown plus and minus their standard errors. These are both valid measures of variability in data, but they mean rather different things.

If you use the standard deviation, you are showing each mean plus the range within which 68% of the scores are likely to fall. The bigger the error-bar, the more spread out scores are around the mean. Our 20 year olds in the study above had a mean RT of 680 msec and an S.D. of 97 msec. This means that approximately 68% of these participants had an RT somewhere between 777 and 583 msec (and by implication, the rest had RT's outside of this range).

If you use the standard error, this shows how certain we are about the particular values of the mean that we are displaying. The bigger the error-bar, the less certain we can be that this particular value of the mean would be reproduced on subsequent occasions. In effect, if we use an error-bar that shows the mean plus and minus one standard error, the error bar says "if we re-did this experiment 100 times, then on 68% of those occasions, the mean

would be likely to fall within these limits". In our example, the mean for our 20 year olds is likely to be within the range 704 (680 +24 msec) and 666 (680 -24 msec) 68% of the time.

[Here's some information that may not make much sense until later in the course. If error-bars are based on standard errors, it enables you to get a rough idea of which differences on a graph are likely to be statistically significant from each other: if two data points have non-overlapping error-bars, they are likely to be significantly different from each other, as measured by a t-test].

Thus the interpretation of error bars differs according to whether they are representing standard errors or standard deviations: however, in both cases, note that the larger the error-bar, the less confident we can be about its associated mean as a description of "typical" performance. In the present example, the error bars tell you not only that mean RT decreased with age, but also that it became much less variable (if you use S.D. error-bars) or that measurement of it became much more reliable (if you use S.E. error-bars) - information that you would not have been able to glean from the graph had it not had error bars.

Any questions, contact me - Graham Hole, *grahamh@sussex.ac.uk.*

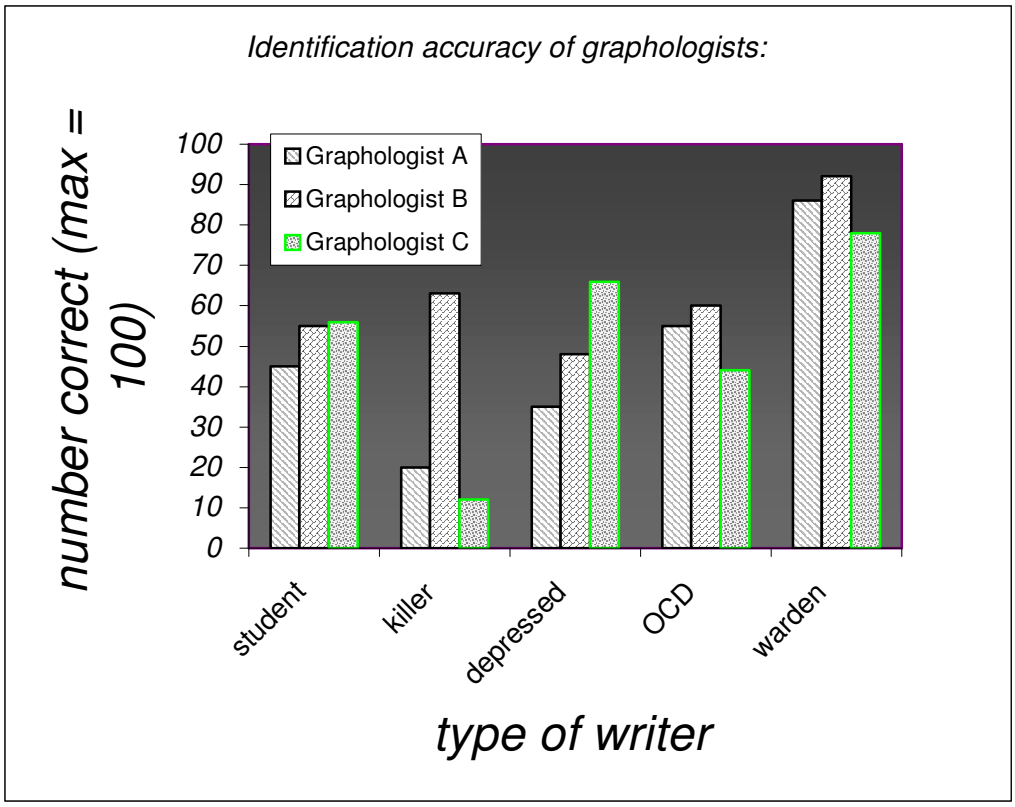
Exercises for you to try:

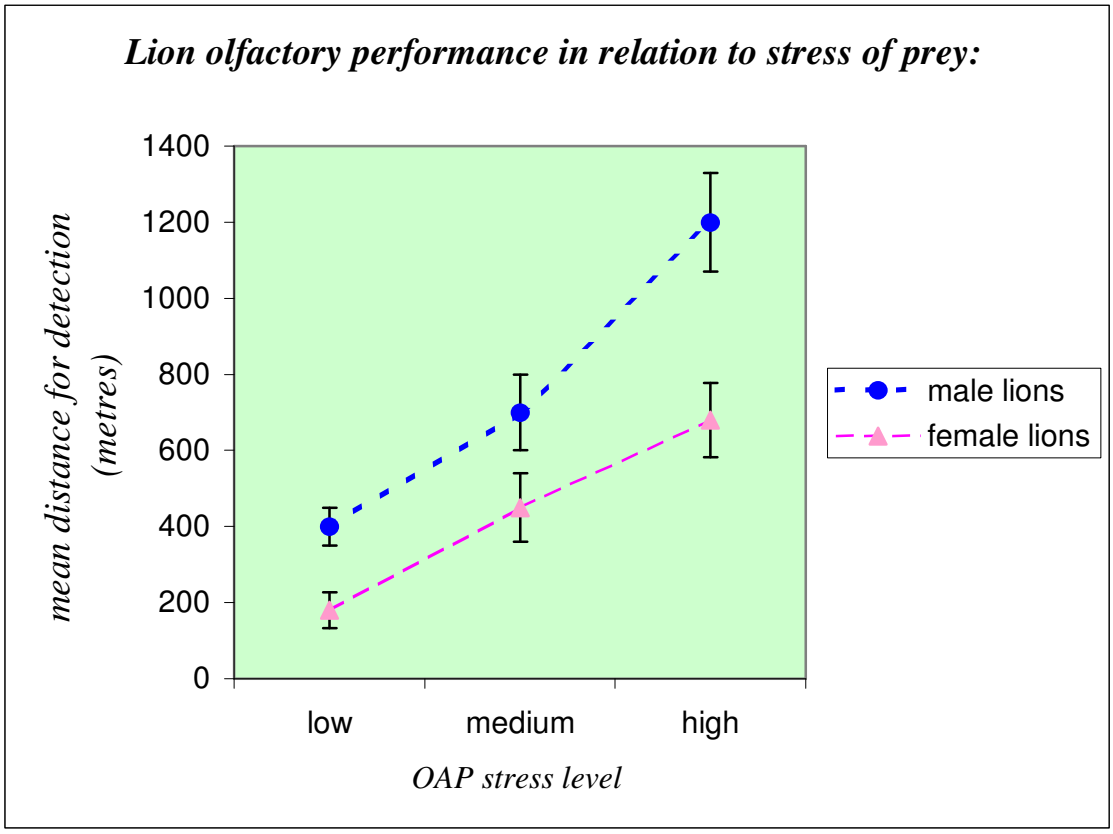
1. How well can lions smell fear? We take two groups of lions (one group of males and one group of females) and expose them to old-age pensioners who are tethered to a post in the zoo, after having experienced either high, medium or low levels of stress. We measure how close each lion has to be to the OAP before they start salivating (the lion that is, not the OAP). (We manipulate the OAPs' stress level by varying the amount of time that they get to cross at a Pelican crossing on a busy road). Graph these data as a line-chart (using the XY scatter option). Give the graph: a light green background; lines with different symbols, patterns and colours; error-bars showing each mean plus and minus one standard error; a title in bold and italic 14-point Times Roman; and appropriately-labelled axes, in 12-point italic Times Roman.

level of OAP stress:	low	medium	high
male lion mean distance (metres)::	400	700	1200
standard error:	50	100	130
female lion mean distance (metres):	180	450	680
standard error:	47	90	98

2. Is graphology a load of squiggle? A psychologist collects 500 samples of handwriting, 100 each for five different groups of people - normal students, serial killers, depressives, obsessive compulsives and traffic wardens. Three graphologists are then shown the samples for each group in turn, and asked to decide whether they are examples of handwriting from that group or not. Each graphologist thus gets a score out of a 100, for each of the five groups, which shows how many of that group they can correctly identify as belonging to that particular group. Thus a graphologist who was accurate would score close to 100% on each group. Graph these data as a bar-chart, with: a purple gradient background; different patterns of bar; 18-point Arial Italic title and label axes; and place the legend just inside the top left corner of the graph itself.

	Student	Serial killer	Depressive	OCD	Traffic Warden
Graphologist A:	45	20	35	55	86
Graphologist B:	55	63	48	60	92
Graphologist C:	56	12	66	44	78


Identification accuracy of graphologists:
number correct (max = 100)
100
90
80
70
60
50
40
30
20
10
0
Graphologist A
Graphologist B
Graphologist C
student
killer
depressed
OCD
warden
type of writer


Lion olfactory performance in relation to stress of prey:
mean distance for detection (metres)
1400
1200
1000
800
600
400
200
0
male lions
female lions
low
medium
high
OAP stress level