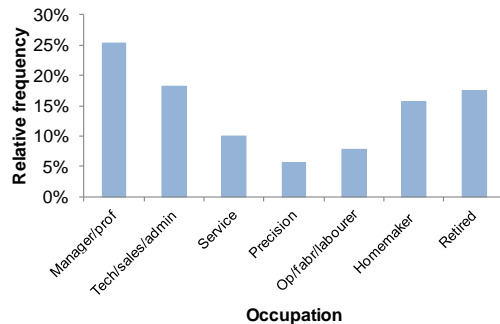


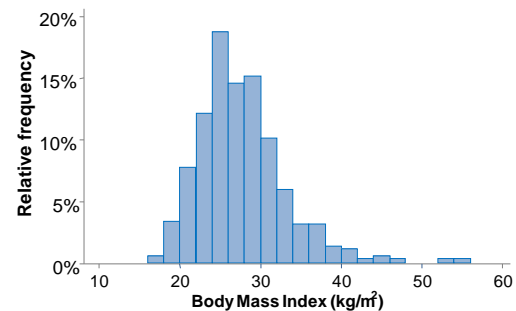
Graphs and Tables

I. Some simple but useful graph formats

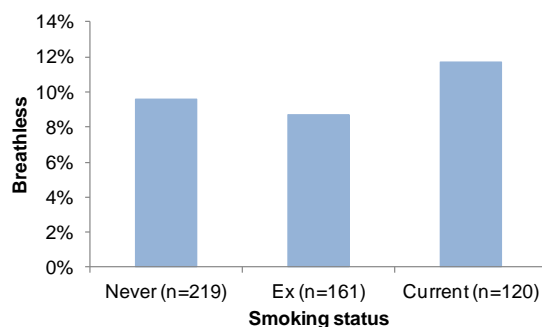
Bar chart showing the distribution of a discrete variable (occupation)



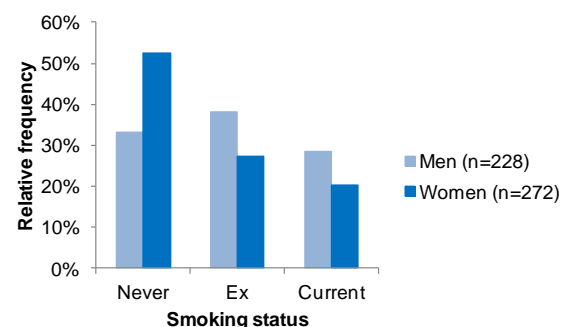
Histogram showing the distribution of a continuous variable (Body Mass Index)



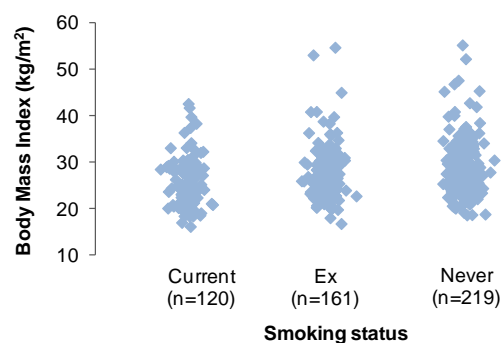
Bar chart showing the distribution of a binary outcome (breathlessness) according to a discrete predictor (smoking)



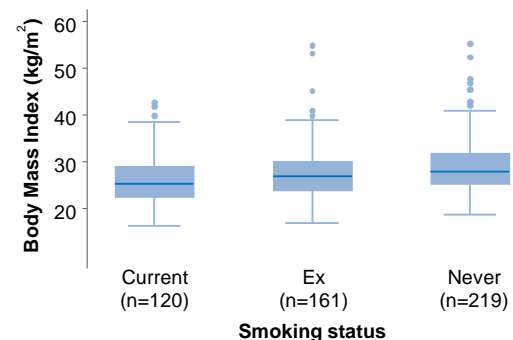
Clustered bar chart showing the distribution of a discrete non-binary outcome (smoking status) according to a discrete predictor (sex)



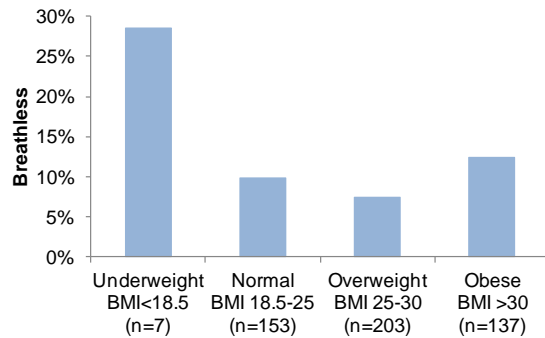
Scatter diagram showing the distribution of a continuous outcome (Body Mass Index) according to a discrete predictor (smoking status)



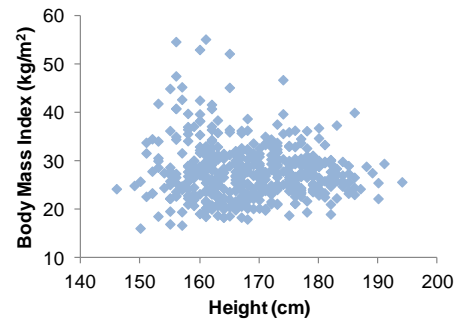
Box-and-whisker plot showing the distribution of a continuous outcome (Body Mass Index) according to a discrete predictor (smoking status)



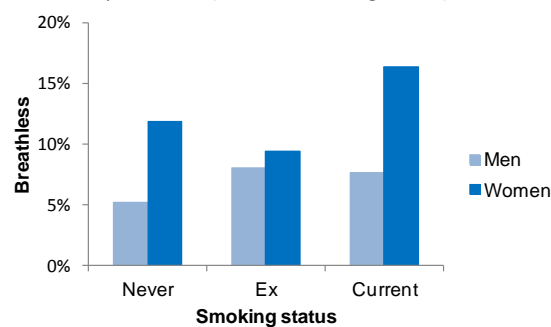
Bar chart showing the distribution of a binary outcome (breathlessness) according to a continuous predictor (Body Mass Index)



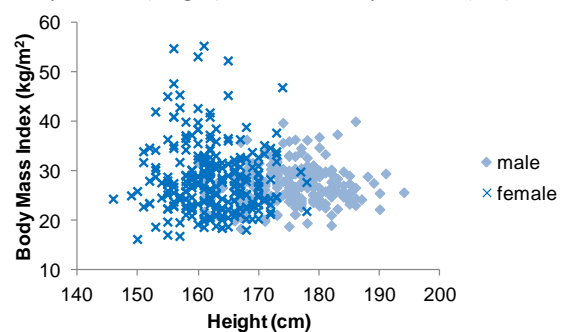
Scatter diagram showing the relationship between a continuous outcome (Body Mass Index) and a continuous predictor (height)



Clustered bar chart showing the distribution of a binary outcome (breathlessness) according to two discrete predictors (sex and smoking status)



Scatter diagram showing the relationship between a continuous outcome (Body Mass Index), a continuous predictor (height) and a discrete predictor (sex)



Tips for drawing graphs

1. Graphs should be designed to stand alone from the text – it shouldn't be necessary to read the text to understand what the graph contains. Use clear titles, and clear labels for axes and legends.
2. On a bar chart, always begin the vertical axis at zero.
3. In a bar chart, leave spaces between bars as illustrated in the examples above. In a histogram the bars should be touching.
3. Where bars show percentages calculated within subgroups, including the total number in each subgroup helps with interpretation (100% might not be so notable if it is out of a total of 2, for example).
4. When producing graphs, you may find yourself limited by what a particular computer package can do, or constrained by the formatting imposed by a particular package. There is no simple answer to this other than experimenting with formatting options and looking at different packages. Do your best not to sacrifice good graphical principles!

II. Tips for presenting tables

1. Tables should be designed to stand alone from the text – it shouldn't be necessary to read the text to understand what the table contains. Use clear titles, and clear labels for rows and columns.
2. It is good practice for the title to explain what kinds of figures are included in the body of the table. For example, if the table contains numbers and percentages, the title might begin "*Numbers and percentages of ...*". If it contains means and standard deviations, the title might begin "*Means and standard deviations of ...*". Following this principle, a table presenting a mixture of summary statistics might have a title such as "*Characteristics of Values are mean (and standard deviation) or median (and interquartile range) as indicated*".
3. Wherever appropriate, add marginal rows and/or columns to the table showing totals or averages over rows/columns.
4. Consider re-ordering the rows or columns of a table in a more meaningful way (or according to a trend in the results revealed by the table) – particularly if the original ordering was arbitrary.
5. If percentages and frequencies appear together in a table, try to distinguish them visually. Sometimes this is most easily done by adding a "%" symbol to each figure. It will usually help to have a total row or column containing the total frequency over which percentages were calculated, and the figure "100%" for the percentage. This makes it much clearer how percentages were calculated.
6. If the table shows the relationship between an "outcome" and a "predictor" in the most general sense, use row percentages when the categories of the predictor are in rows, and column percentages when the categories of the predictor are in columns.
7. Most people find it easier visually to compare columns rather than rows, so if you have two series of figures you want people to compare, try and put them in columns rather than rows.
8. Use spacing and layout to make tables more readable and to bring related elements of a table closer. If there are lots of rows in a table, you can make it more readable by introducing a gap every five rows or so. Consider combining rows or columns where frequencies are small or distinctions are unimportant.
9. Right-justify figures, or use a right tab-stop/decimal point tab-stop so that digits/decimal points line up correctly.
10. Don't use too many digits, since this makes figures harder to read and may imply a spurious level of accuracy. People have different views on what is an appropriate number of digits. One simple rule of thumb is to use only two "variable digits", where these are the digits that vary from one entry of the table to another. So, if the table shows a series of figures 181.63, 182.80, 189.13, 186.24, 191.15 – then the first digit (1) doesn't vary, and the figures written with two variable digits would be 182, 183, 189, 186, 191 (*i.e.* rounded to the nearest whole number in this case). The appropriate number of digits also depends on how accurately a figure is determined. A good (and a different) rule of thumb for percentages, for example, is to write them as a whole number of percent if they are calculated from a total frequency less than 100, and with one decimal place if they are calculated from a total frequency of 100 or greater (because such percentages will be determined with greater accuracy). NB you may need to keep more digits for calculations than you include in the table: if you throw digits away before doing calculations you can introduce inaccuracies. Sometimes, because of rounding, a total in a table may not equal the total of the separate, rounded figures you've also included – this is fine.
11. Where figures in the table are taken from an external source, include a footnote with an appropriate reference, or include the source in the title.

Example table (adapted from Table 5.1, Chapter 5 of Bland, *An introduction to medical statistics*):

Numbers and percentages of deaths from different causes, by sex, in England and Wales, 1989

ICD code	Type of disease	Males		Females		Total	
		Number	%	Number	%	Number	%
VII	Circulatory system	127,400	45.6%	137,200	46.6%	264,600	46.1%
II	Neoplasms (cancers)	75,200	26.9%	69,900	23.8%	145,100	25.3%
VIII	Respiratory system	33,500	12.0%	33,200	11.3%	66,700	11.6%
IX	Digestive system	7,900	2.8%	10,800	3.7%	18,700	3.3%
XVII	Injury and poisoning	11,100	4.0%	6,400	2.2%	17,500	3.1%
V	Mental disorders	4,500	1.6%	9,200	3.1%	13,700	2.4%
VI	Nervous system and sense organs	5,500	2.0%	6,000	2.0%	11,500	2.0%
III	Endocrine, nutritional and metabolic diseases and immunity disorders	4,400	1.6%	5,800	2.0%	10,200	1.8%
X	Genitourinary system	3,600	1.3%	4,200	1.4%	7,800	1.4%
XIII	Musculo-skeletal system and connective tissues	1,200	0.4%	4,100	1.4%	5,400	0.9%
XVI	Signs, symptoms and ill-defined conditions	1,600	0.6%	3,100	1.0%	4,700	0.8%
I	Infections and parasites	1,200	0.4%	1,300	0.4%	2,500	0.4%
IV	Blood and blood-forming organs	1,000	0.4%	1,400	0.5%	2,400	0.4%
	Other diseases	1,300	0.5%	1,600	0.5%	2,900	0.5%
	Total	279,400	100.0%	294,200	100.0%	573,600	100.0%

Source: OPCS 1991, DH2 No. 10

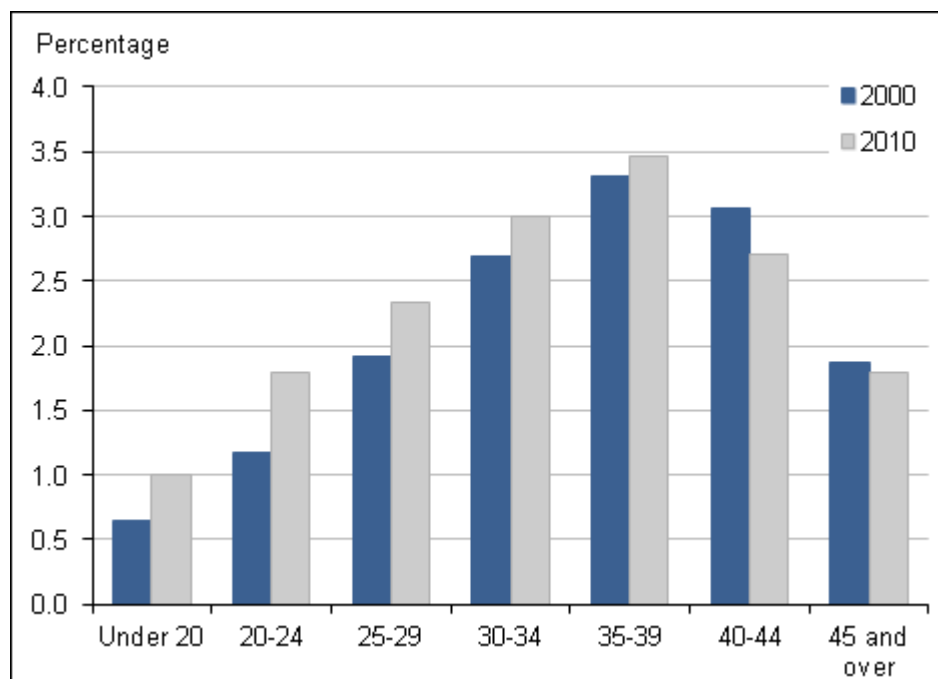
III. Producing graphs: Microsoft Excel exercise using UK National Statistics

Many UK national statistics can be downloaded in Microsoft Excel format. If you think you will be using Excel for your project or other work, try the following example.

Have a look at the ONS Statistical Bulletin on live births in England and Wales by characteristics of birth, 2010, which is available online at http://www.ons.gov.uk/ons/dcp171778_241936.pdf.

Skip down to the following graph in the report:

Maternities taking place at home, by age of mother, 2000 and 2010, England and Wales



Source: Office for National Statistics

Notes:

1. A maternity is a pregnancy resulting in the birth of one or more children, including stillbirths. The chart therefore shows the percentage of women giving birth at home rather than the number of babies born at home. The percentage of women giving birth at home is reported within each age group. For example in 2010, 3.0 per cent of all maternities to women aged 30–34 took place at home.

Click on the link in the pdf to download this chart data into XLS format. Have a go at re-creating the bar chart from the report in Excel, either by playing around yourself or by following the instructions overleaf (these instructions are for Excel 2007).

1. Select the following range of the spreadsheet by clicking and dragging:

Under 20	20-24	25-29	30-34	35-39	40-44	45 and over
0.6	1.2	1.9	2.7	3.3	3.1	1.9
1.0	1.8	2.3	3.0	3.5	2.7	1.8

2. Select the “Insert” menu tab, click on “Column” in the “Charts” group, and select a 2D clustered column chart.

3. Right-click somewhere in the chart, and choose “Select data” from the menu. Click on “Series 1”, then click “Edit”. An “Edit Series” dialogue box opens up, with the cursor flashing in the “Series name” box. Click on the spreadsheet in the cell containing “2000” (this is the label to be used for the first series), and click “OK”. Now click on “Series 2” and “Edit” and select the label “2010”, and click “OK” twice to exit all the dialogue boxes.

4. You can edit or add elements to the chart by right-clicking different parts of the chart and choosing “Format xxx”, or using the menu choices in the “Chart Tools, Layout” menu tab.

NB It might seem that you could do the labelling of the series in one go by selecting the following range of the spreadsheet in Step 1:

Year	Under 20	20-24	25-29	30-34	35-39	40-44	45 and over
2000	0.6	1.2	1.9	2.7	3.3	3.1	1.9
2010	1.0	1.8	2.3	3.0	3.5	2.7	1.8

but this doesn’t work as expected (try it and see).