



The University of Manchester

Faculty of Humanities

Microsoft Excel Basic training



**A step-change in
quantitative social
science skills**

Funded by the
Nuffield Foundation,
ESRC and HEFCE

The University of Manchester
is one of 15 universities taking
part in the Q-Step programme



Contents

Contents

Aims.....	3
About this workbook	3
Getting started.....	4
Getting familiar with the worksheet.....	5
Introductory Exercise 1 – Entry, Edit and Basic Formulae.....	6
Introductory Exercise 2 – Working with Columns of Data	8
Practical exercises.....	11
Does the highest level of qualification differ between regions in England?	11
How does the population differ between regions in England?	16
How does ethnic diversity vary between English regions?	18
How does social class differ between the English regions?	21
Further resources.....	26

Aims

At the end of the session / workbook you should have acquired basic skills in using Excel. You will be able to ...

- ... open an Excel workbook and enter data and simple formulae
- ... load existing data and carry out a range of analysis of individual variables to answer a range of research questions
- ... present your findings in simple graphs
- ... download data from Nomis and prepare it for analysis

About this workbook

This workbook is designed to enable you to work at your own pace through the exercises. During the taught session tutors will be available to help you with issues as they arise.

Instructions to follow are broken down into bullet points marked with arrows and choices are enclosed in <> markers e.g.

- Select the <File> option from the top menu
- Select <Open> from the drop down menu displayed
- ...

Where you are required to type in information it is shown in Courier font e.g. type
`=sum (...)`.

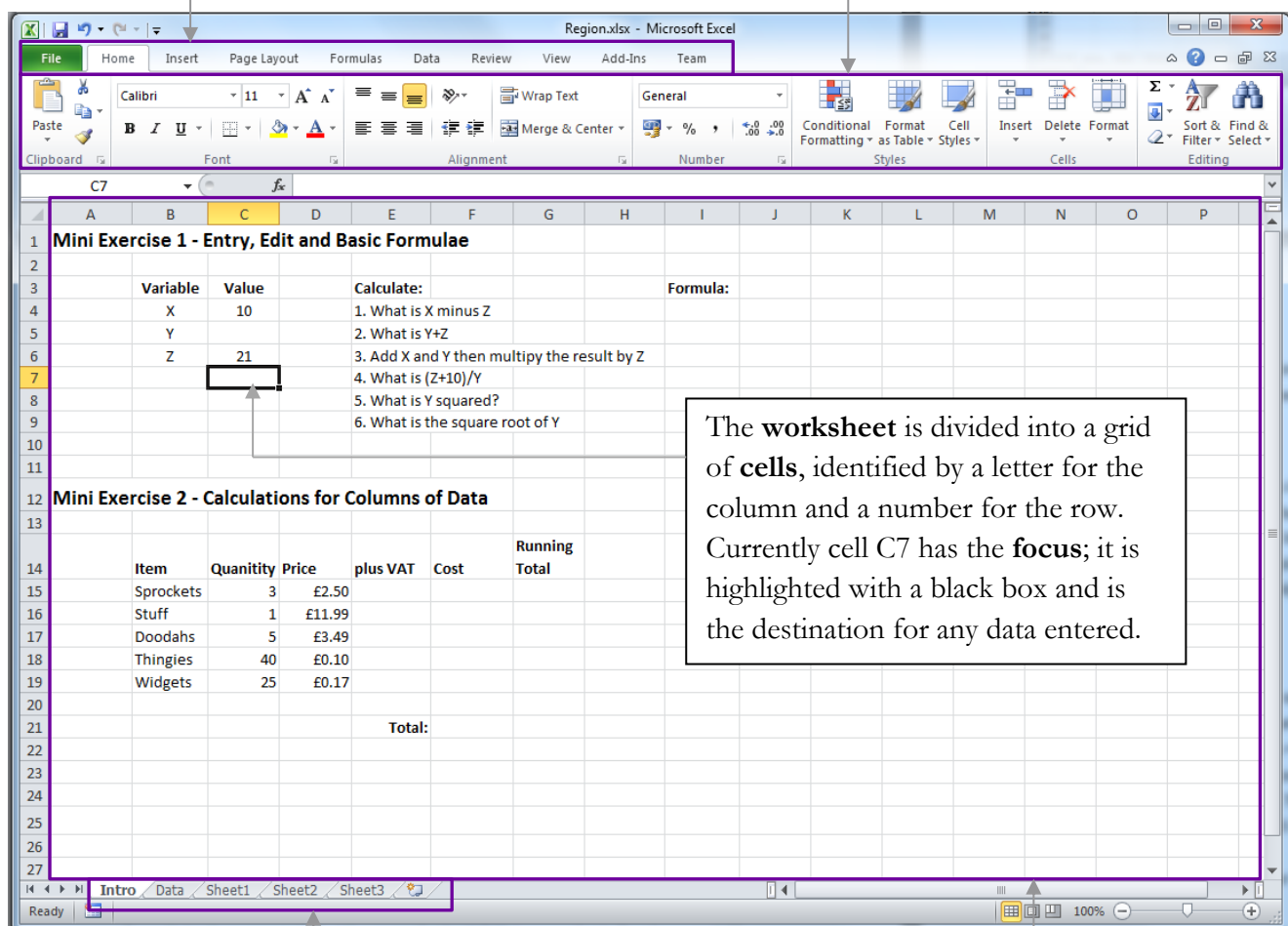
This workbook has been prepared by Richard Conibere and Nigel de Noronha for the Q-Step programme at the University of Manchester.

Getting started

Below is a screenshot showing the first exercises in the session. The different sections of the screen are described in the text that points to them.

The top line provides a high level **menu** to navigate around the tools available within Excel. Click on a couple of them to see how the screen changes.

The **ribbon** changes as you move between high level menu items. Each item on the ribbon carries out an Excel command. Hovering the mouse cursor over an item will reveal what it does (and tells you the keyboard shortcut for doing it quicker)



At the bottom are set of tabs which let you move between worksheets. Each worksheet can contain different information.

The main part of the window, the **worksheet**, is where you can put in data – numbers, text and dates. You can also add formulae and functions to manipulate the data.

On the main worksheet you can move around using the arrow keys or the mouse. The other areas are selected by clicking on the item you want.

Getting familiar with the worksheet

This section introduces basic skills to enter and edit data and to calculate values with formulae. Whilst some of this may be familiar Excel often allows you to do something in several different ways so the simple exercises aim to illustrate these different options.

We will be using one workbook for all the exercises in the session. So to begin open the workbook:

- Select the <File> option from the top menu and then <Open> from the dropdown menu.
- Choose the file 'workbook.xlsx' which is available in the 'C:\Work' folder.

We will be working on the 'Intro' worksheet as shown on the 'Getting Started' page. As a reminder, worksheets consist of a grid of cells. Each cell is specified with a letter for the column and a number for the row. So, for example, cell B4 contains some text naming a variable 'X' whilst the value of X, 10, is given in cell C4. Additionally a cell may contain a formula, calculating a value based on other cells.

All the exercises are designed to be carried out at your own pace. The tutors are available to help if you are not clear about what to do or why. Please let us know if you need our help.

Introductory Exercise 1 – Entry, Edit and Basic Formulae

Data Entry Fundamentals

To begin with some simple data entry, the value for Y should be 7 but it is currently missing.

- Select cell C5 by either moving the focus to the cell with the cursor keys or clicking on it directly. Type 7 and press <return>

Also the value for Z had been mistyped; it is meant to be 12.

- Select cell C6 and enter the correct value.

A few other handy tips for data entry are:

- When inputting a value finishing with <return> moves the focus down to the next row. Finishing with <tab> moves the focus right to the next column – useful for entering a row of values.
- Sometimes, rather than entering a new value in a cell from scratch, it is useful to edit what is already there. To edit a cell you can either double click it or press F2.

The Formula

Now we introduce formulae i.e. calculations based on the values of other cells. In Excel a formula is always preceded with an equals sign '='. At its most basic Excel can work like a calculator. So if you were to enter =2+2 in a cell, the result of 4 would be displayed. The cell C4 is actually an example of this. If you move the focus to C4 this shows that the value of the cell is the formula '=3+7', revealed in the formula box f_x above the worksheet.

		C4		f_x =3+7						
	A	B	C	D	E	F	G	H	I	J
1	Mini Exercise 1 - Entry, Edit and Basic Formulae									
2										
3		Variable	Value		Calculate:				Result:	
4		X	10		1. What is X minus Z					
5		Y			2. What is Y+Z					
6		Z	21		3. Add X and Y then multiply the result by Z					
7					4. What is (Z+1)/Y					

Now proceeding with some basic formulae, work on the calculations 1-6 placing the results in the I column. Each formula will need to reference the appropriate values in the C column. So, for the first calculation:

- Move the focus to cell I4
- Noting that the value of X is in cell C4 and Z is in C6, enter the formula `=C4-C6`
- Alternatively in cell I4 start by entering '=' then move the focus to cell C4 then type '-' and finally move to cell C6 before pressing <return> .
- Now continue down the I column, writing the formulae for calculations 2 to 6. See the hints below for some help.

A few tips for the formulae:

- The star character '*' is the operator for multiplication. Use '/' for division.
- For power calculation the hat character '^' does the job. So 6 squared could be calculated with the formula `=6^2`.
- Excel prioritises multiplication and division over addition and subtraction unless brackets specify otherwise. The first time it matters is for calculation 3 – the addition of X and Y must be enclosed in brackets to ensure that it happens before the multiplication.
- For calculation 6 press F1 for help and search for 'square root' to find the SQRT function.

Final answers should look like this:

I9		fx		=SQRT(C5)							
	A	B	C	D	E	F	G	H	I	J	
1	Mini Exercise 1 - Entry, Edit and Basic Formulae										
2											
3		Variable	Value		Calculate:				Formula:		
4		X	10		1. What is X minus Z				-2		
5		Y	7		2. What is Y+Z				19		
6		Z	12		3. Add X and Y then multiply the result by Z				204		
7					4. What is (Z+10)/Y				3.143		
8					5. What is Y squared?				49		
9					6. What is the square root of Y				2.646		
10											

To finish this first exercise, here are a couple of questions:

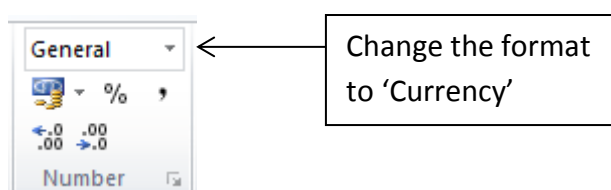
Q1 Why would the formula `=C6+10/C5` give the wrong answer for calculation 4?

Q2 Having completed the formulae, what happens if you, say, edit C5 to change the value of Y to 2

Introductory Exercise 2 – Working with Columns of Data

The data for Exercise 2 shows a fictitious order for a variety of items, the quantity required and their price before VAT (sales tax) is added. The aim of the exercise is to calculate the total cost of the order and a running total for each order line.

- Firstly the order details for the last item are missing so they need to be added. Twenty five 'widgets' are required and they cost 17 pence each.
- If the £ sign is missing on the price, the formatting of the cell needs to be adjusted. To fix it, select the cell and then, on the home ribbon, find the formatting tools:



Now, assuming, that the VAT rate for all items is 17.5% the cost would need to be multiplied by 1.175 to get the full plus VAT cost.

- In cell E15 enter a formula to calculate the cost of sprockets with VAT included. If you have it right the cost should be 2.94.
- As before the cell will need to be formatted as 'Currency'.

Copying a Formula for a Column

The next task is to calculate the plus VAT cost of the remaining order items. This could be done by repeating the step above but this would be rather tedious, especially with an order of 10s or 100s of items. So here is one way to apply the formula in E15 in the other cells:

- Select cell E15
- Copy the formula, either by choosing the <copy> button on the left hand side of the home ribbon, by right clicking on the cell and choosing <copy> from the context menu or by using the keyboard short-cut <Ctrl-C>.
- Holding down the <Shift> key, move the cursor to extend the selection to the last order line, cell E19. Alternatively the range of cells can be selected with the mouse by clicking and dragging.
- Paste the formula into all the cells, either by choosing the <paste> button on the left hand side of the home ribbon, by right click and selecting <paste> or with the short-cut <Ctrl-V>.

Q1 Selecting cells E15 to E19 in turn, look at the formulae in the f_x box. How do they change from row to row?

It should be apparent that when the formula was copied it was updated to work for each row.

Now that the item costs including VAT are sorted it is possible to work out the full cost for each order line:

- To begin, in cell F15 enter a formula to calculate the quantity multiplied by the full plus VAT cost.
- As before, copy the formula so it applies to all of the order lines. Note that a slightly quicker way of doing this is to select the range of cells from F15 to F19 and then copy the formula down with the short-cut <Ctrl-D>.

With the currency formatting applied to all results then, at this point, your spreadsheet should look like this:

	A	B	C	D	E	F	G
12	Mini Exercise 2 - Calculations for Columns of Data						
13							
14		Item	Quantity	Price	plus VAT	Cost	Running Total
15		Sprockets	3	£2.50	£2.94	£8.81	
16		Stuff	1	£11.99	£14.09	£14.09	
17		Doodahs	5	£3.49	£4.10	£20.50	
18		Thingies	40	£0.10	£0.12	£4.70	
19		Widgets	25	£0.17	£0.20	£4.99	


Totals and Running Totals

Now that the cost for each order line has been calculated, we can return to the original aims of the exercise:

- Calculating the total cost of the order
- Calculating a running total for the order lines.

For the total cost we require the sum of the cells from F15 to F19. The painful way to do this would be with a formula like =F15+F16+F17+... but that's not very practical. Instead Excel has a SUM function that works across a range of cells – the range F15:F19 in this case. So:

- In cell F21 type the formula =SUM(F15:F19)
- Press Ctrl-Z to undo this and alternatively, start by entering =SUM(
- Then select the required range of cells (with the mouse or keyboard)
- Finish by typing) and pressing <return>
- As a final way of summing a column, first delete your result in cell F21.

- Select the range of cells from F15 but extending to F21 and then click the sum button  towards the top right of the home ribbon.

Finally, a running total will accumulate the cost of the order for each order line. For the first order line it is just the cost of the 'sprockets' but the next line adds the cost of the 'Stuff' and so on.

- Try calculating the running total.
- Finally the cost of the 'Sprockets' is discounted to £1.50 each (before VAT). Update the price.

Q2 What is the total cost of the order?

Q3 What had the running total reached when 'Doodahs' are included in the order.

Answers:

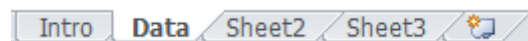
Initially the total cost of the order was £53.10 but this is reduced to £49.57 when the 'Sprocket' discount is applied. The running total reaches £39.88 for the 'Doodah' order line.

If the answers given above don't match your spreadsheet have a look at your formulae and see if you can find and fix any problems. Also, please let us know if you need any help.

Practical exercises

The initial introductory exercises were designed to highlight many of the fundamentals of using Excel. From now on we will be working with real data, with exercises to answer research questions. No more accountancy and more social science...

- To begin, first click on the 'Data' tab (bottom left of the Excel window) to move to the Data worksheet



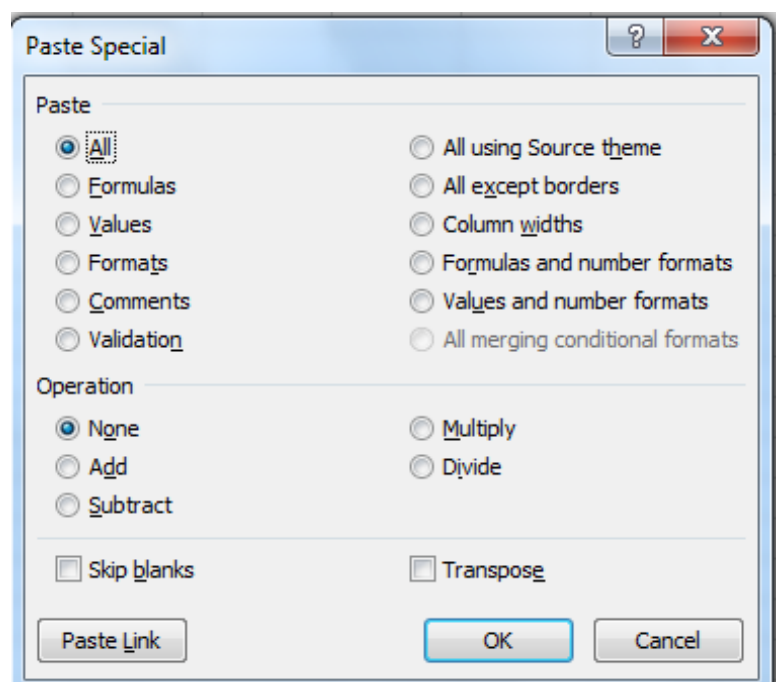
Does the highest level of qualification differ between regions in England?

For this exercise you will need to copy the headings and the total populations (rows 1 & 2) and the counts of people with different levels of qualification (rows 11 to 17). Data doesn't always arrive in a convenient format. In this case we need to 'transpose' the data, swapping things around so that the rows become the columns and vice versa:

- Select rows 1 and 2. Start by clicking on cell A1 i.e. the start of row 1. Then either:
 - Hold the mouse click and drag to cell J2, the end of row 2.
 - Hold down the <shift> key, using the cursor keys to move to J2.
- On the <Home> ribbon select <Copy> (on the left hand side of the ribbon)
- Click on <Sheet 2> on the tab at the bottom of the screen where you can select a worksheet.
- Click the arrow under <Paste> on the left of the ribbon and select <Paste Special>.

A dialogue box is opened enabling you to transform the data as shown.

- Click the box next to <Transpose>
- Click on <OK>



Your headings will now be on the new worksheet with the regions down the left hand side and the variables across the top.

- Return to the <data> worksheet and repeat for rows 11 to 17.
- Click on the left column on row 11, then drag the cursor to the end of row 17 and select <Copy>
- Return to the <Sheet 2> worksheet and select cell C1 (so Excel knows that this is where you want to paste the data).
- As before choose <Paste Special> and tick the dialog box <Transpose> option.

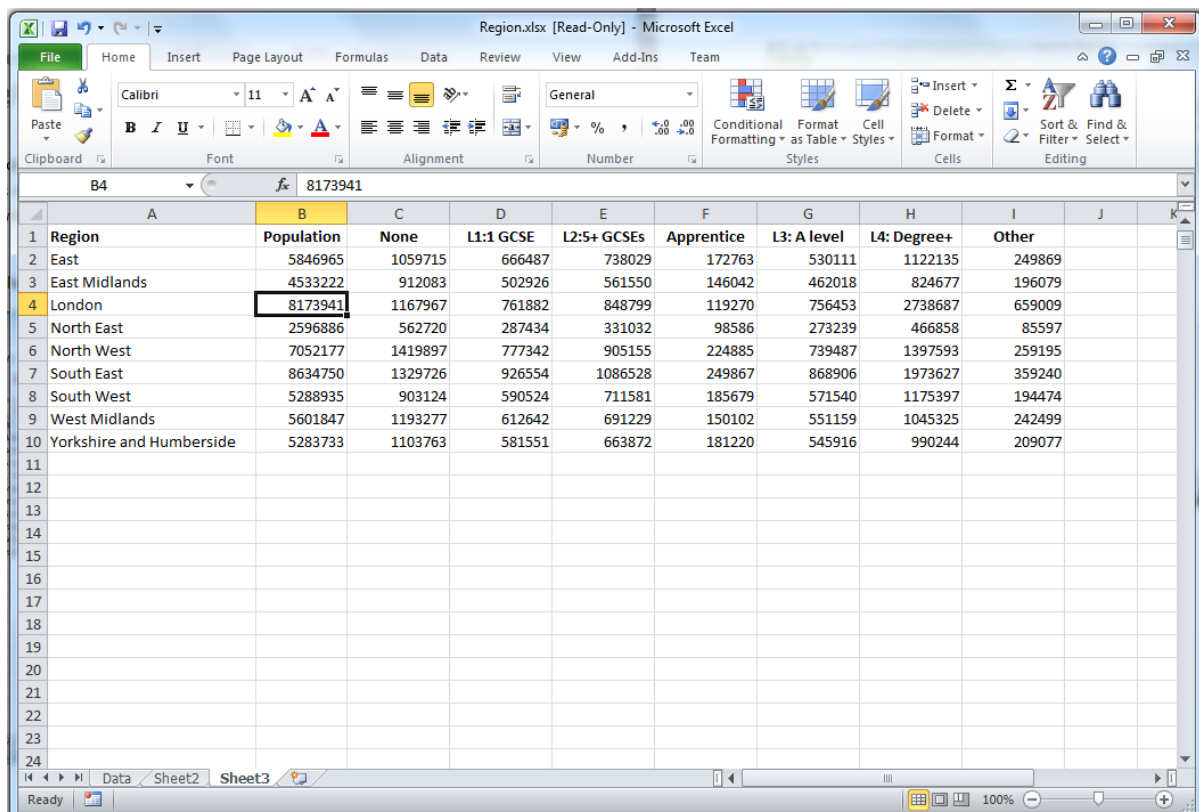
To give <Sheet 2> a more meaningful name:

- Double click on <Sheet 2> on the tab at the bottom of the screen where you can select a worksheet. You can now put in a name for this worksheet.
- Type in *qualification* and press <Enter>.

You might like to make the data more presentable by:

- Adjusting the width of the columns
- Simplifying the column labels
- Emphasise the column labels in Bold

When done the sheet should look like this:



Region	Population	None	L1:1 GCSE	L2:5+ GCSEs	Apprentice	L3: A level	L4: Degree+	Other
East	5846965	1059715	666487	738029	172763	530111	1122135	249869
East Midlands	4533222	912083	502926	561550	146042	462018	824677	196079
London	8173941	1167967	761882	848799	119270	756453	2738687	659009
North East	2596886	562720	287434	331032	98586	273239	466858	85597
North West	7052177	1419897	777342	905155	224885	739487	1397593	259195
South East	8634750	1329726	926554	1086528	249867	868906	1973627	359240
South West	5288935	903124	590524	711581	185679	571540	1175397	194474
West Midlands	5601847	1193277	612642	691229	150102	551159	1045325	242499
Yorkshire and Humberside	5283733	1103763	581551	663872	181220	545916	990244	209077

Q1 Which region has the lowest number of people with no qualifications?

Q2 Which region has the highest number of people with level 4 (degree or higher) qualifications?

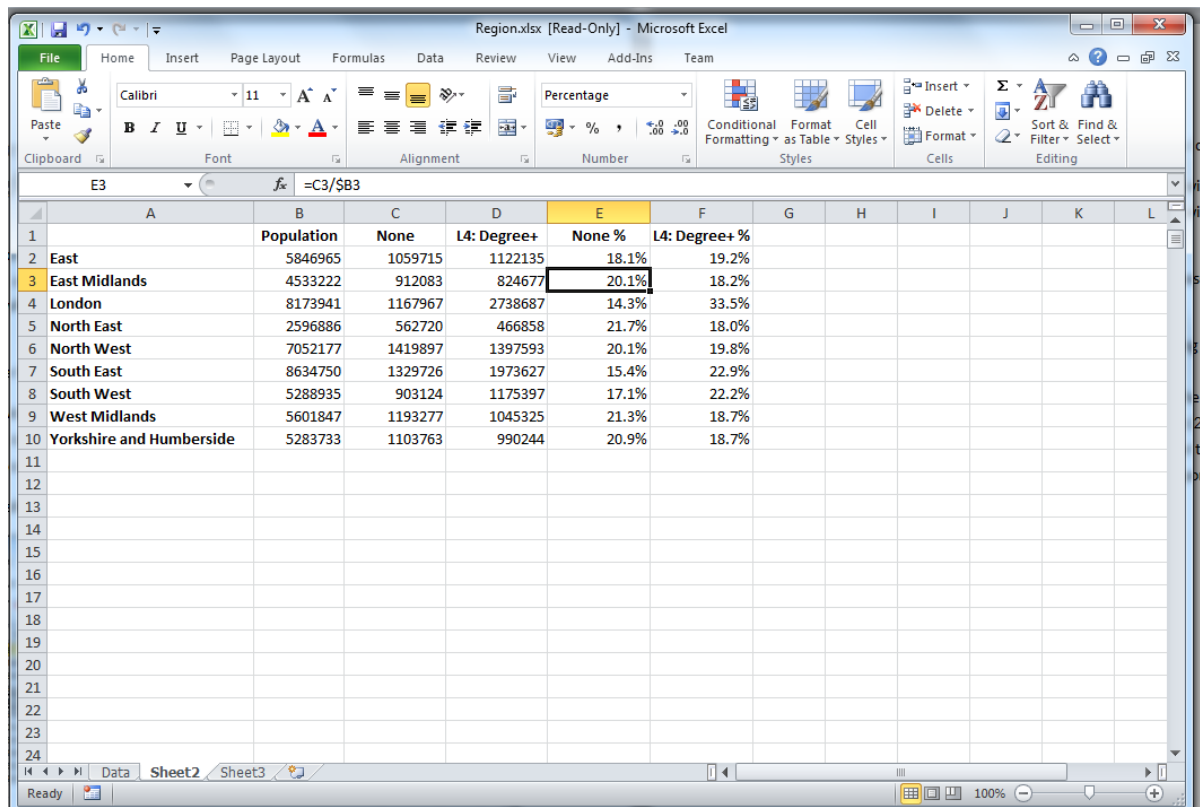
Suppose you are a researcher and you have been asked to identify:

- The region with the smallest proportion of people with no qualifications.
- The region with the greatest proportion of people with level 4 (degree or higher) qualifications.

Note that this asks for proportions (i.e. percentages) so these will need to be calculated as follows:

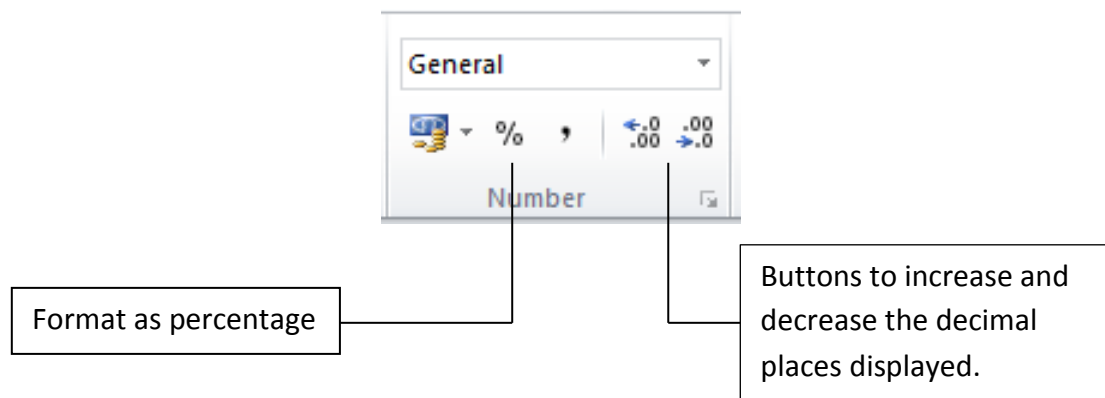
- First delete the columns that aren't needed, keeping the region, population, no qualifications and level 4 qualification columns.
- In a new column first calculate the proportion of the population with no qualifications in the East region. The formulae is $=c2/b2$
- Copy the formula down to produce percentages for the other regions.
- Similarly in a new column calculate the proportions with a level 4 qualification
- Format the proportions as percentages to 1 decimal place. The method for doing this is similar to the formatting of currencies in the mini exercises. See the next page for a hint if it isn't obvious.

When done your data should look like this:



	A	B	C	D	E	F	G	H	I	J	K	L
		Population	None	L4: Degree+	None %	L4: Degree+ %						
1												
2	East	5846965	1059715	1122135	18.1%	19.2%						
3	East Midlands	4533222	912083	824677	20.1%	18.2%						
4	London	8173941	1167967	2738687	14.3%	33.5%						
5	North East	2596886	562720	466858	21.7%	18.0%						
6	North West	7052177	1419897	1397593	20.1%	19.8%						
7	South East	8634750	1329726	1973627	15.4%	22.9%						
8	South West	5288935	903124	1175397	17.1%	22.2%						
9	West Midlands	5601847	1193277	1045325	21.3%	18.7%						
10	Yorkshire and Humberside	5283733	1103763	990244	20.9%	18.7%						
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												
24												

Values can be formatted as percentages to 1 decimal place using some buttons on the Home ribbon:



Q3 Which region has the lowest proportion of people with no qualifications?

Q4 Which region has the highest proportion of people with Level 4 (Degree or higher) qualifications?

Q5 Are the answers the same as for Qs 1 & 2? How do you explain any differences?

Finally we will produce a bar graph to illustrate the exceptional proportion of higher qualifications held by residents of London. This will require the Regions (column A) and the level 4 proportions (column F). It's easier to produce graphs if all the relevant data is grouped together on the worksheet so:

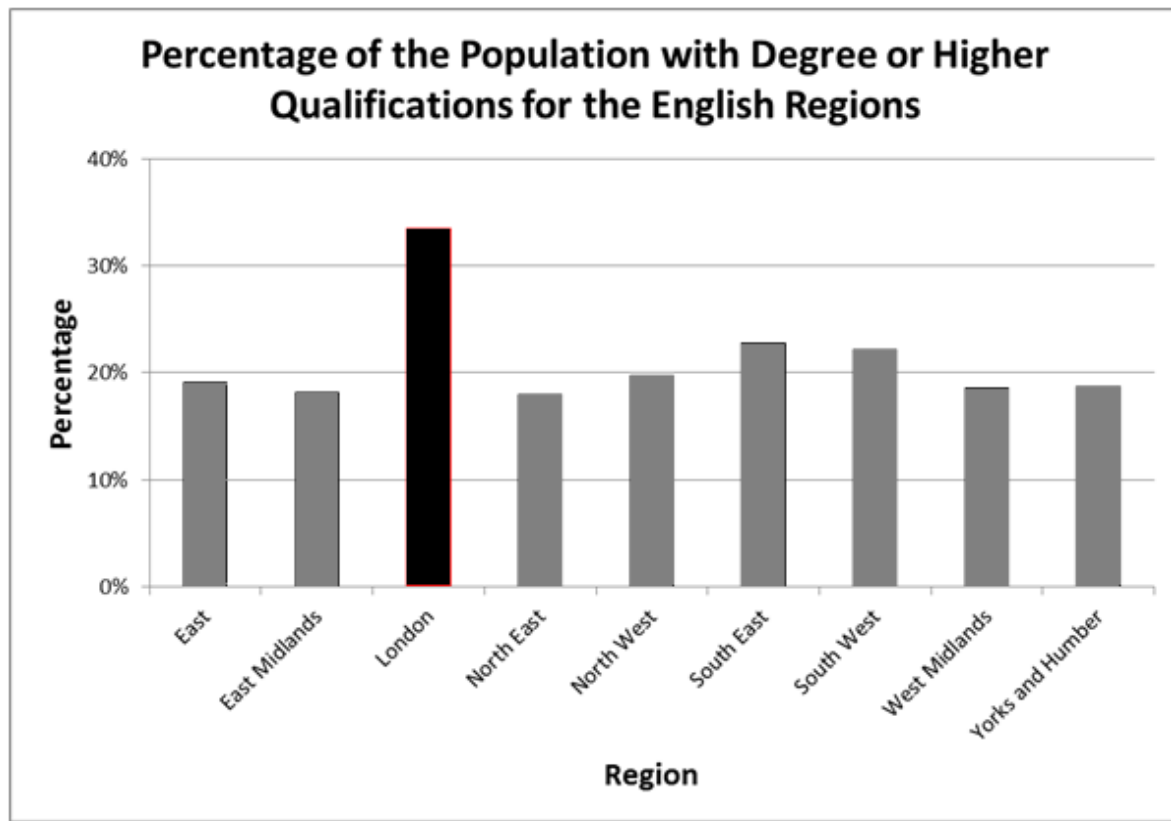
- Right click on the top bar of column F and choose Insert to add a new column.
- Copy and paste the region names into the new column so they are next to the percentage values.

Then to produce the graph:

- Select the region names and percentages data.
- On the Insert ribbon click Column chart -> 2D column
- Once the chart has appeared. The menus extend to include the Chart Tools. Select the Layout Ribbon to access the tools for adjusting the graph e.g. to add an axis title.



Try to edit your graph so it looks like the one below. Some hints on how this was done are included below.



- Tools on the chart tools Layout ribbon were used to add the chart and axis titles and to switch off the legend.
- On the worksheet the region name for 'Yorkshire and Humberside' has been edited to make it shorter.
- Other aspects of the graph were edited by right clicking on them and choosing the appropriate menu option. For example, right clicking on the y axis and choosing Format Axis opens a dialog box that includes options to change the decimal places of the percentage labels.

How does the population differ between regions in England?

To help us answer the question we will extract this information from the spreadsheet and put it on a separate worksheet. For this exercise you will need to copy the headings and the first 7 rows of total populations (rows 1 to 8).

- Return to the 'Data' worksheet, select the required data and <copy> it.
- Move to worksheet <Sheet 3> and, as before use the <past special> <transpose> option when pasting so that you have rows of regions and columns for the different types of population count.
- Double click on <Sheet 3> on the tab at the bottom of the screen where you can select a worksheet
- Type in `population` and press <Enter>.

The first thing we will do is add in a row to show the totals for each column for England.

- In the cell below Yorkshire and Humberside type *England*.
- In cell B11 calculate the column totals for the overall population, using one of the 3 methods covered in Introductory Exercise 2

The total population of England should be displayed. You can now copy this formula along the row as Excel will adjust the list of cell numbers to the relative position.

- Select the total and press <Copy> on the ribbon
- Highlight the next cells along by clicking and holding the mouse
- Press <Paste> and the totals for England are now displayed.

Have a look at the formula for each cell to see how it has changed.

The definition of a communal establishment includes a nursing or residential home, an armed forces base, a prison or student halls. What questions does the total by region suggest to you?

Why do you think there are more women than men?

We can derive some further information by including a formula at the end of the column. We know the total population, the population aged 16 or over and the working age population of 16 to 74 year olds.

- Add columns to show the number of children and the number of people aged 75 or over (*Hint: do the formulas for the first region then copy and paste to complete for the other regions*).

The completed spreadsheet should look like this:

	A	B	C	D	E	F	G	H	I	J
1	Region	Population	Household	Communal	Working Age	Over 16	Male	Female	Under 16	Aged 75+
2	East	5,846,965	5,748,605	98,360	4,245,544	4,738,333	2,875,807	2,971,158	1,108,632	492,789
3	East Midlands	4,533,222	4,442,192	91,030	3,336,532	3,694,767	2,234,493	2,298,729	838,455	358,235
4	London	8,173,941	8,073,700	100,241	6,117,482	6,549,173	4,033,289	4,140,652	1,624,768	431,691
5	North East	2,596,886	2,550,818	46,068	1,924,206	2,134,449	1,269,703	1,327,183	462,437	210,243
6	North West	7,052,177	6,927,820	124,357	5,184,216	5,727,629	3,464,685	3,587,492	1,324,548	543,413
7	South East	8,634,750	8,446,500	188,250	6,274,341	6,992,666	4,239,298	4,395,452	1,642,084	718,325
8	South West	5,288,935	5,175,084	113,851	3,856,715	4,359,257	2,590,608	2,698,327	929,678	502,542
9	West Midlands	5,601,847	5,509,535	92,312	4,067,119	4,507,405	2,763,187	2,838,660	1,094,442	440,286
10	Yorkshire and Humberside	5,283,733	5,185,677	98,056	3,875,219	4,285,941	2,598,078	2,685,655	997,792	410,722
11	England	53,012,456	52,059,931	952,525	38,881,374	42,989,620	26,069,148	26,943,308	10,022,836	4,108,246

What does this analysis of the population tell you?

How does ethnic diversity vary between English regions?

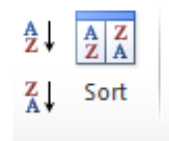
You have created two worksheets showing the highest level of qualification and a number of population characteristics for the English regions. Do the same for ethnic diversity as measured by non-white British. This time you will need to add a new worksheet to hold the data.

- Click on the folder symbol with a star sign and a new worksheet will be created.
- Put your data in the new worksheet – this time you will need rows 1-2 (region names and population) and rows 9-10 (the counts by ethnic group).
- Calculate column totals in a new row labelled England.

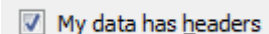
In the previous exercise we used the population counts for our analysis. In many cases this might not be the best measure as the total size of the population varies between regions. A simple solution is to calculate the proportion of a category to the total.

- Add a new column and calculate the proportion of non-white British for each region and for England (*Hint: make sure you divide by the total number in the population i.e. use the total population or add white British and non-white British*).

Which is the most diverse region as measured by the proportion of the population who are not white British? In this example the small number of regions allows you to answer that question by scanning the values. If there were more cases there are other ways to help you answer the question.



- Select the regional data, excluding the column totals i.e. the range from cell A1 to cell E10
- Click to get the <Data> ribbon from the main menu at the top of the screen.
- Press the <Sort> button and a dialogue box is opened. Ensure that the box in the top right corner of the dialog box is ticked so Excel knows that the first row gives names for the variables:



- You can select the column of your data you want to sort by (*choose Proportion non-white British*), the type of sort (*default values is fine*) and the order (*choose Largest to Smallest*). Click <OK>. The regions are now organised in descending order by the proportion of non-white British.

Now see the next page for an illustration of what the worksheet should look like now.

	A	B	C	D	E
1	Region	Population	White British	Not white British	%age not WB
2	London	8,173,941	3,669,284	4,504,657	55.11%
3	West Midlands	5,601,847	4,434,333	1,167,514	20.84%
4	South East	8,634,750	7,358,998	1,275,752	14.77%
5	East	5,846,965	4,986,170	860,795	14.72%
6	East Midlands	4,533,222	3,871,146	662,076	14.60%
7	Yorkshire and Humberside	5,283,733	4,531,137	752,596	14.24%
8	North West	7,052,177	6,141,069	911,108	12.92%
9	South West	5,288,935	4,855,676	433,259	8.19%
10	North East	2,596,886	2,431,423	165,463	6.37%
11					
12	England	53,012,456	42,279,236	10,733,220	

We can now produce a graph of this data.

- Select the regional data (rows 1 to 10)
- Select the <Insert> item on the top menu and from the <Column> graph option choose the first <2D Column> option.

This produces a graph of all the data in the worksheet. It needs to be changed to include only the Proportion of non-white British. When the graph is selected the ribbon gives you a number of options.

- Select the graph by clicking on it and click <Select data> towards the left of the ribbon. Alternatively right-click on the graph and choose <Select data> from the context menu.

A dialogue box shows the columns displayed on the graph.

- Remove those that you do not need.

You can adjust the colours, remove the legend and change the title if you want to.

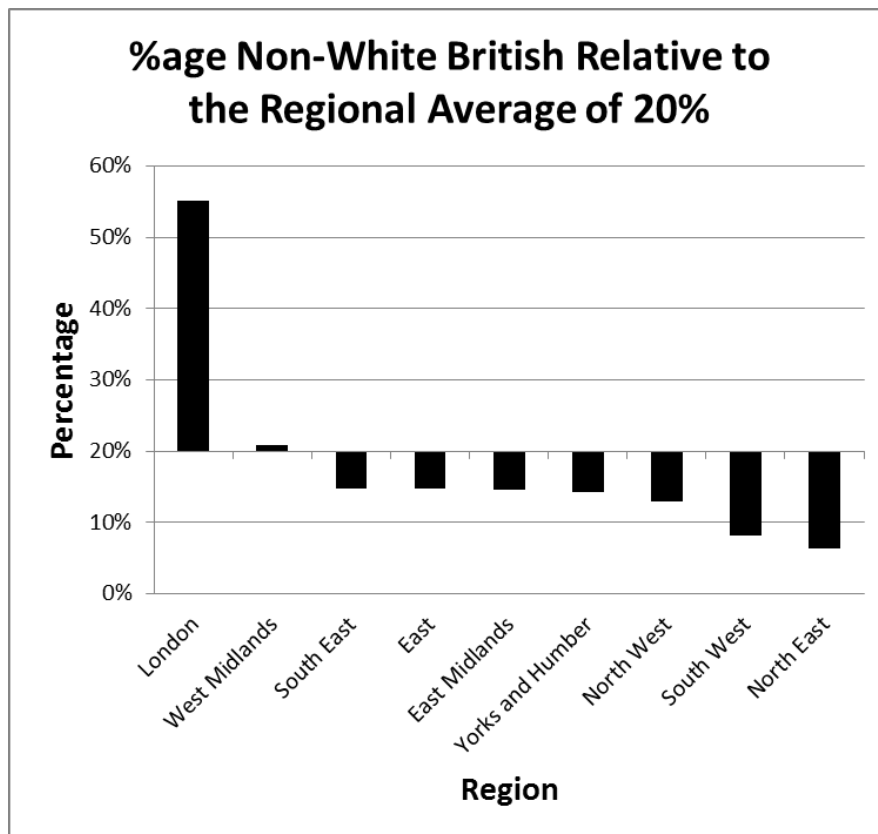
Finally we can adjust the axis to show the regions relative to the average level of diversity for England. In the table of data the average value for England is 20% (approx.).

- Right click on the Y axis (on the left of the chart). A dialogue box opens.
- Select <Format axis> and at the bottom click <Axis value> under the heading *Horizontal axis crosses*.
- Enter 0.2 (i.e. 20% expressed as a proportion) and press <OK>.
- Additionally set the Axis Labels to <Low> so that they don't overlap with the data.

You now have a graphical representation of the relative diversity of the English regions.

When working with graphs the options at the top of the screen enable you to change the appearance of your graph. Have a look at the options available when you select the different toolbars <Design>, <Layout> and <Format>.

A fairly well polished graph might look something like this:



What does this analysis tell you?

How does social class differ between the English regions?

In this exercise we will be looking at counts of different social classes, based on the National Statistics Socio-economic Classification (NS-SeC). This measure assigns a class to people based on their occupation, whether they are self-employed and the size of the organisation where they work. Unlike the previous exercises, here you will be downloading the data from scratch. Acquiring data like this is a useful skill, vital for many tasks you are likely to encounter in the workplace.

Stage 1: Downloading the Data

- Go to ... <https://www.nomisweb.co.uk/>
- Click on the link to '2011 Census Data on Nomis'
- Follow the link to 'Quick Statistics (QS)'
- Then find and select table QS607EW for the NS-SeC.

This table gives the Ns-SeC classification for the working-age population; Everybody aged between 16 and 74. As before we will be working with the English regions, so this geography needs to be specified. Also, by default, the Ns-SEC consists of a hierarchy of over 50 different categories whilst we only require the social classes. Therefore the data must be filtered appropriately.

- To begin the process of filtering the data select the 'Query Data' button in the panel on the left hand side of the screen.

Query data

☐ Guide me step-by-step

Make selections:

[Geography](#)

[NS-SeC](#)

[Rural - Urban](#)

Review selections:

[Summary Of Selections](#)

Get your data:

[Format / Layout](#)

[Download Data](#)

The screen now includes the panel shown on the left.

- First select '**Geography**' and then choose 'All' for the 'regions' (at the bottom of the 'commonly used' list).
- Next select '**NS-SeC**' from the panel.
- As shown below, untick the first and tick the second of the boxes for selecting the categories. This limits the data to the level of the hierarchy that we need i.e. just the social classes.

NS-SeC

☐ ☒ ☐ ☐ Tick to select all items in the column

- Selecting 'Format/Layout' from the left-hand panel choose to receive the data as a csv file.



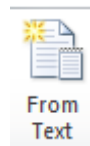
Comma separated values (.csv)

- Select 'Download Data'.
- NOMIS will then take a moment to prepare the data and will, finally, offer you a link to 'Download data (.csv)'. Click the link and save the file somewhere you can find it again.

Admittedly, rather than choosing a csv file it would have been possible to download an Excel xlsx file directly. However doing this the hard way allows us to look at Excel's import features. Data are very frequently delivered as csv files so it's good to know how to handle them.

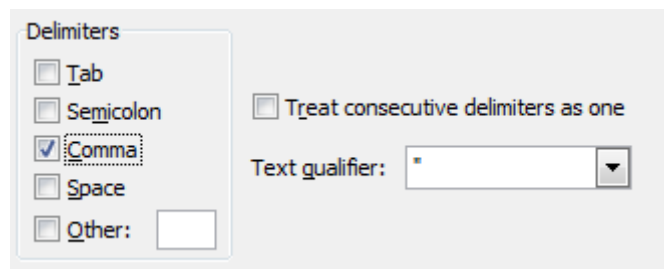
Stage 2: Importing a csv into Excel

- Leave the web browser and return to Excel.
- As in the previous exercise, create a new worksheet to receive the data.
- Give the new worksheet a meaningful name like 'Social Class'
- From the top-level menu choose 'Data', selecting the 'From Text' button on the command ribbon.
- Find and 'Open' the downloaded file.



Excel then brings up a dialog box, working through the necessary stages to import a text file.

- At the first step ensure that 'Delimited' rather than 'Fixed width' is selected and then click 'Next'
- At the second step specify that the variables are separated by commas before proceeding to 'Next'.



- The third pane allows you to force columns to be imported as text or dates. Not an issue here so just click 'Finish'
- When prompted, agree that the data can be placed in cell A1 of the current worksheet.

Expect the data to look like this:

	A	B	C
1			
2	Q5607EW - NS-SeC		
3	ONS Crown Copyright Reserved [from Nomis on 8 June 2015]		
4	Population :	All usual residents aged 16 to 74	
5	Units :	Persons	
6	Date :		2011
7	Rural Urban:	Total	
8			
9	region	1. Higher managerial, administrative and professional occupations	2. Lower managerial, administrative
10			
11	North East		144419
12	North West		461159
13	Yorkshire and The Humber		329084
14	East Midlands		302960
15	West Midlands		355020
16	East		461118
17	London		807936
18	South East		789267
19	South West		394860
20	Wales		174943
21			
22			
23			
24			
25		In order to protect against disclosure of personal information, records	
26		have been swapped between different geographic areas. Some counts will	
27		be affected, particularly small counts at the lowest geographies	
28		be affected, particularly small counts at the lowest geographies	
29			

Coming from a third party – NOMIS – the data contains various extra leading and trailing rows and the column names are so verbose that only a few fit on the screen. So as a final stage the data needs to be formatted so it is easier to work with. To ensure that the columns can be given simpler names, here is a reference table of new column headers and their full definition:

sec1	1. Higher managerial, administrative and professional occupations
sec2	2. Lower managerial, administrative and professional occupations
sec3	3. Intermediate occupations
sec4	4. Small employers and own account workers
sec5	5. Lower supervisory and technical occupations
sec6	6. Semi-routine occupations
sec7	7. Routine occupations
exLab	8. Excluded from labour market i.e. Never worked or long-term unemployed.
NC	Not Classified (inc. full-time students)

In no particular order, steps to tidy up the data are:

- Delete rows 1-8 to drop the header and also the rows beyond Wales to get rid of the footer.
- Rename the columns as outlined in the table above and adjust the column widths so the data fits on the screen.
- As all the analyses so far have been for the English regions, the Wales row can also be deleted.

- To be consistent with the previous exercises, sort the data by Region so it is listed alphabetically.
- It's good practice to confirm that the data is what you think it is! For example, you could sum the rows and verify that these totals are the same as the 'Working age' population figures given on the 'Data' worksheet.

Job done:

K2 fx =SUM(B2:J2)											
	A	B	C	D	E	F	G	H	I	J	K
1	Region	sec1	sec2	sec3	sec4	sec5	sec6	sec7	exLab	NC	Total
2	East	461118	919307	592101	438669	298430	599391	447158	173726	315644	4245544
3	East Midlands	302960	646714	410363	299635	258312	505171	453849	163199	296329	3336532
4	London	807936	1410785	723354	575331	305781	633790	453923	506290	700292	6117482
5	North East	144419	353248	249525	132767	153727	315484	276289	126073	172674	1924206
6	North West	461159	1007834	678759	442941	381432	785174	640250	320712	465955	5184216
7	South East	789267	1466583	850558	643058	409969	799485	557852	234881	522688	6274341
8	South West	394860	834208	492794	435470	281682	571727	407663	131735	306576	3856715
9	West Midlands	355020	763395	502429	354864	293691	629013	524140	278493	366074	4067119
10	Yorkshire and The Humber	329084	730033	472161	339876	293094	591628	516359	244917	358067	3875219
11											

Stage 3: Recoding the NS-SeC into 3 classes

When analysing data it is often convenient and sensible to aggregate groups together. For example, if we were considering housing tenure, people living in local authority rented properties – “council housing” - and housing association rentals might be grouped together because of their shared position as occupants of social housing. In the case of the NS-SeC there is a pre-defined aggregation that collapses classes 1-7 into a hierarchy of 3 groups. The 3 groups and their component classes are:

- 1. Managerial/Professional** = sec1 + sec2
- 2. Intermediate** = sec3 + sec 4
- 3. Routine/Manual** = sec5 + sec6 + sec7

- In 3 empty columns, create variables for the 3 group class hierarchy by summing the appropriate NS-SeC classes.

Stage 4: Analysis

Explore the data – think about how the different categories of social class vary and whether this tells a story. You might:

- Calculate percentage variables as a proportion of the working-age population.
- Ask which regions contain the largest and smallest proportions of the different class groupings and those who are excluded from the labour market.
- Both occupation based social class and levels of education are indicators of socio-economic status. So how do the results contrast with those seen in the first practical?
- Produce a stacked bar chart showing the composition of each region.

What does this analysis tell you?

Further resources

- Whatever you are trying to do remember that pressing F1 to get help is always an option.
- There are good online training materials like those at Excel Easy:
<http://www.excel-easy.com/>

This training session introduced various Excel features in the context of work to look at research questions across a range of social data. The approach at Excel Easy is much more skill-focussed, providing specific sections on basic skills like 'Printing' and 'Formatting a cell' through to more advanced topics like how to deal with 'Formula errors' or the creation of 'Pivot tables'.