



University of
Nottingham

UK | CHINA | MALAYSIA

Interpreting Geographical Data

Social Analysis Part I with Microsoft Excel



Version with Apple
MAC keys



Purpose

This document has been prepared to support the development of skills in the use of Excel for social research. It provides a series of exercises to develop your skills in using Excel for social analysis. Part I introduces ways of manipulating, presenting and reporting on a dataset that holds information about the nine regions of England.

Author(s) and funding

This document has been produced by Dr Nigel de Noronha. Material included in this guidance comes from work funded by the Nuffield Foundation as part of the Q-Step funding to the University of Warwick and initial work by the author with colleagues Richard Conibere and Rebecca Bromley to develop courses for Q-Step students at the University of Manchester.

Contents

Purpose	2
Author(/s) and funding	2
Using Excel for Social Analysis	4
Does the highest level of qualification differ between regions in England?	4
How does the population differ between regions in England?	9
How does ethnic diversity vary between English regions?	11
How does social class differ between English regions?	15
Further resources	18

Tables and Figures

Figure 1 – percentage of the population with degree or higher qualification by region	8
Figure 2 – percentage of non-white British compared to the regional average	13
Table 1 – Occupational social class	15

Using Excel for Social Analysis

The exercises in this part of the toolkit are designed to help you develop a number of skills in Excel. At the end of this section you will be able to ...

... load existing data and carry out a range of analysis of individual variables to answer a range of research questions

... present your findings in professional looking tables and graphs

These exercises begin on the assumption that you are familiar with opening Excel documents, and are able to work with simple Excel formulae. Please refer to the introductory exercises if you are unsure of any of these.

- First download the spreadsheet 'Region.xls' from Moodle and open it
- Click on the 'Region' tab (bottom left of the Excel window) to move to the Region data worksheet



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

This worksheet contains population information of UK regions, including detailed information about qualifications.

For this exercise you will examine differences between UK regions in terms of qualifications held by the total population. Data you receive is not always ready to use, and as in this case, the worksheet contains more information than you need to work with. From this worksheet, you will need the headings and total populations in each region (rows 1 & 2) and the counts of people with different levels of qualifications within each region (rows 11 to 17). Some datasets are likely to contain information that is not relevant to for your research question and it doesn't always arrive in a convenient format. In this case we need to 'transpose' the data.

Transposing a set of cells swaps the layout around so that the rows become the columns and vice versa - so in this example, the rows will represent Regions in order for us to easily answer our questions.

Using Excel for Social Analysis (part I)

- To transpose the data you need, start with the active cells in the heading and total region population 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
 - Hold down the <shift> key, using the cursor keys to move to J
- On the <Home> ribbon select <Copy> (on the left hand side of the ribbon)
- Click on <Sheet 3> 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 <region> worksheet and repeat for rows with information about qualifications – 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 3> 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.

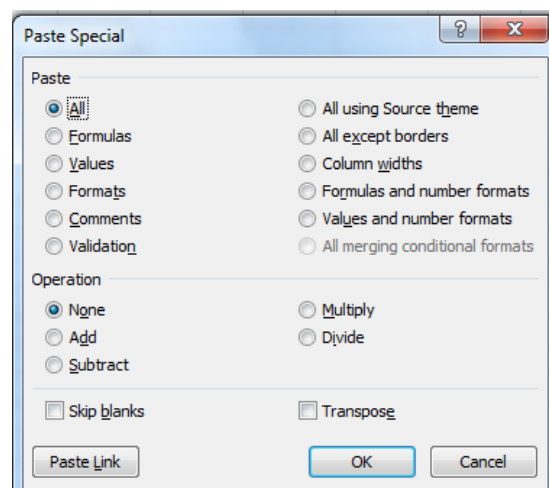
To give <Sheet 3> a more meaningful name:

- Double click on <Sheet 3> 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. (Alternatively you can right-click and select rename).
- Type in `qualifications` and press <Enter>.

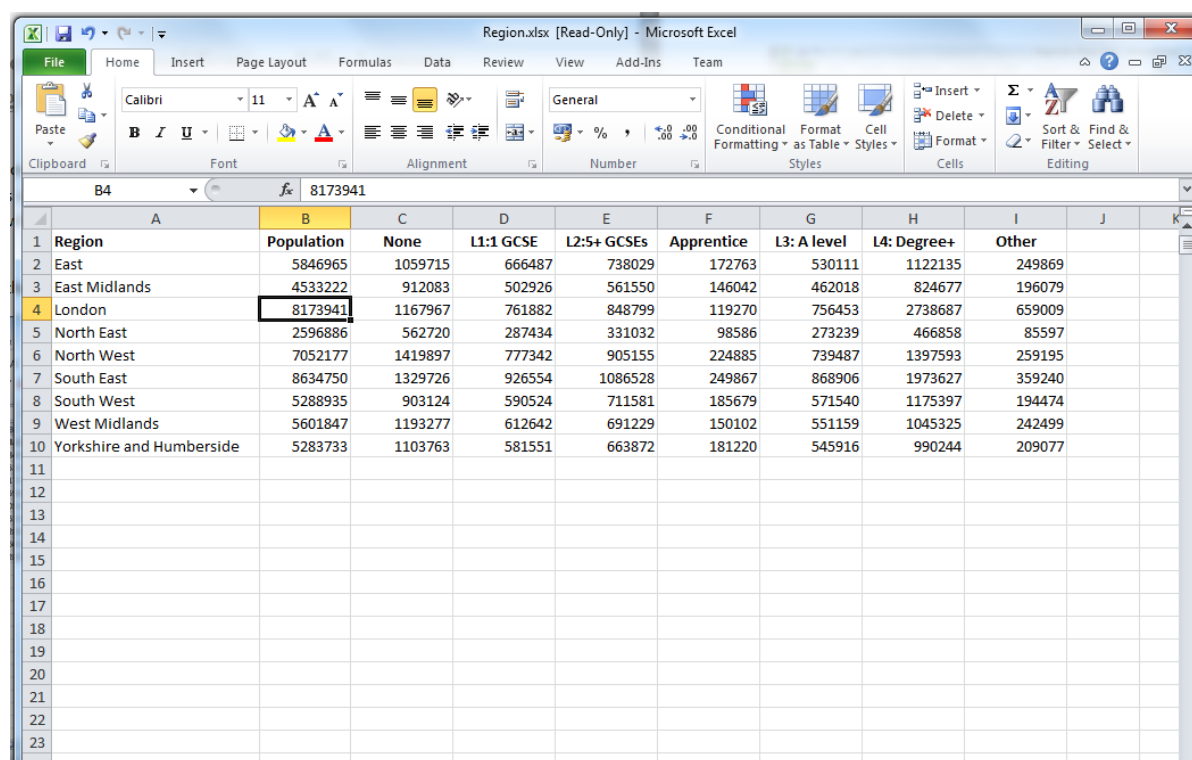
You might like to make the data more presentable by:

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

When done the sheet should look like this:



Using Excel for Social Analysis (part I)



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:

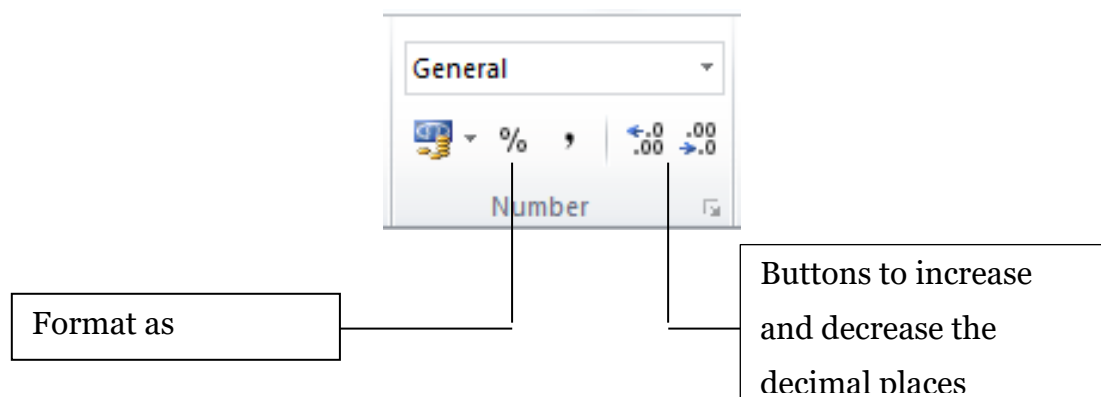
- First delete the columns that aren't needed, keeping the region, population, no qualifications and level 4 qualification columns.
- In a new column first calculate the proportion of the population with no qualifications in the East region. The formulae is $=c2/b2$
- Copy the formula down to produce 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.

Using Excel for Social Analysis (part I)

When done your data should look like this:

	A	B	C	D	E	F	G	H	I	J	K	L
1		Population	None	L4: Degree+	None %	L4: Degree+ %						
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?

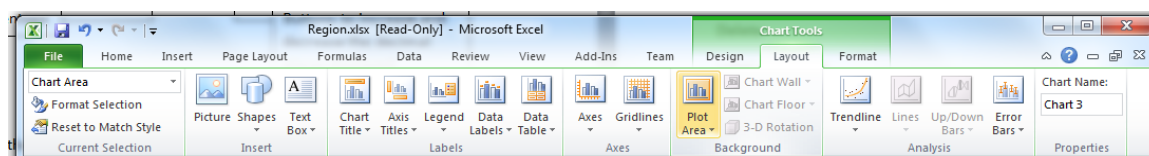
Using Excel for Social Analysis (part I)

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:

- 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.

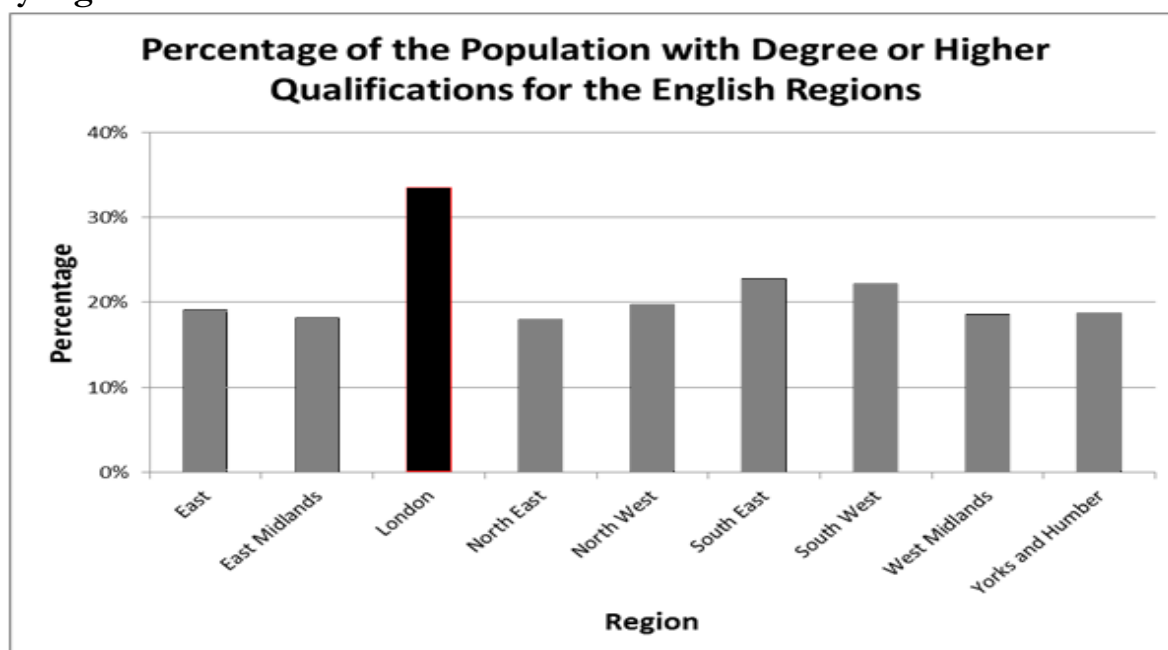
Then to produce the graph:

- 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 Figure 1. Some hints on how this was done are included below.

Figure 1 – percentage of the population with degree or higher qualification by region



- 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 pressing <Shift-F10> to select them and choosing the appropriate menu option. For example, press <Shift-F10> on the y axis and choosing Format Axis opens a dialog box that includes options to change the decimal places of the percentage labels. To format a single column you need to first <click it> which selects all columns, then <click it> again which selects the individual column. You can then change its appearance using the <Shift-F10> options displayed.

How does the population differ between regions in England?

To help us answer the question we will extract this information from the 'Region' worksheet and put it on a separate worksheet. For this exercise you will need to copy the region headings and the rows detailing total populations for households, establishments, working age and over 16 populations, and male and female populations (rows 1 to 8).

- Return to the 'Region' worksheet, select the required data and <copy> it.
- Move to worksheet <Sheet 4> and, as before use the <paste 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 4> 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 three 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.

Using Excel for Social Analysis (part I)

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?

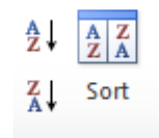
You have created two worksheets showing the highest level of qualification and a number of population characteristics for the English regions. Do the same for ethnic diversity as measured by non-white British. This time you will need to add a new worksheet to hold the data.

- Next to the existing worksheets along the bottom, click on the folder symbol with a star sign and a new worksheet will be created (or Excel 2013 shows a '+' sign).
- 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:

My data has headers

- 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.

Using Excel for Social Analysis (part I)

	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.).

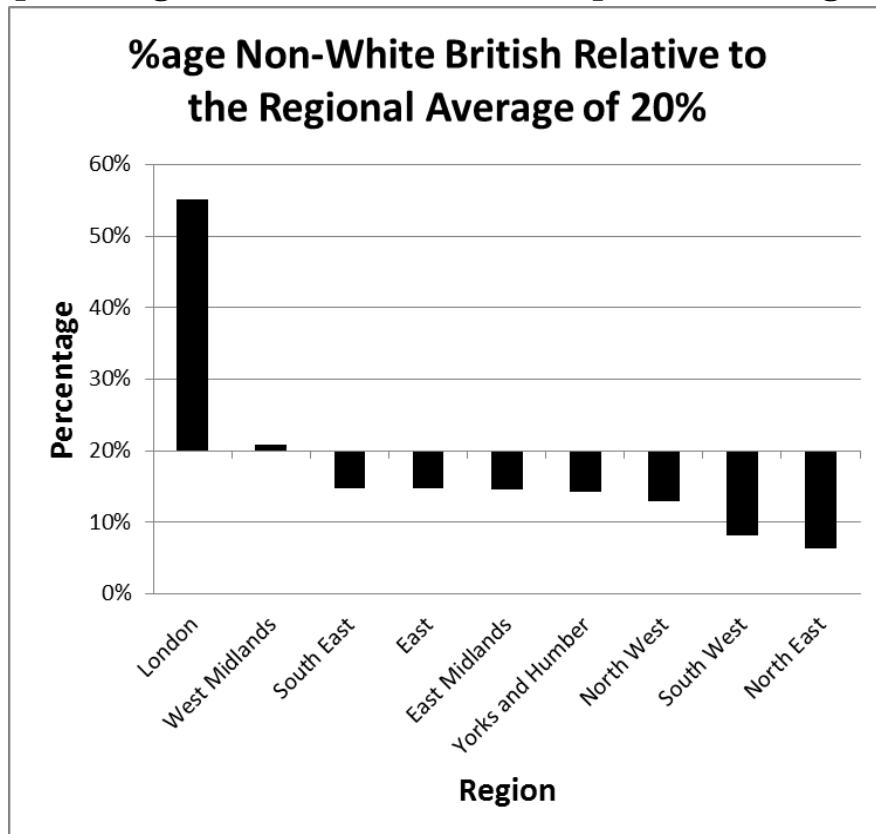
- Press <Shift-F10> on the Y axis (on the left of the chart). A dialogue box opens.
- Select <Format axis> and at the bottom click <Axis value> under the heading *Horizontal axis crosses*.
- Enter 0.2 (i.e. 20% expressed as a proportion) and press <OK>.
- Additionally set the Axis Labels to <Low> so 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 figure

Figure 2 – percentage of non-white British compared to the regional average



What does this analysis tell you?

How does social class differ between 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). The NS-SeC provides an indication of socio-economic position based on occupation. To assign a person (aged 16 to 74) to an NS-SeC category, their occupation title is combined with information about their employment status, whether they are employed or self-employed and whether or not they supervise other employees. Full-time students are recorded in the 'full-time students' category regardless of whether they are economically active or not.

- In the open spreadsheet, move to the 'NS-SeC' worksheet, next to the 'Region' worksheet, which contains the following:

region	sec1	sec2	sec3	sec4	sec5	sec6	sec7	sec8	NC
North East	144,419	353,248	249,525	132,767	153,727	315,484	276,289	126,073	172,674
North West	461,159	1,007,834	678,759	442,941	381,432	785,174	640,250	320,712	465,955
Yorkshire and The Humber	329,084	730,033	472,161	339,876	293,094	591,628	516,359	244,917	358,067
East Midlands	302,960	646,714	410,363	299,635	258,312	505,171	453,849	163,199	296,329
West Midlands	355,020	763,395	502,429	354,864	293,691	629,013	524,140	278,493	366,074
East	461,118	919,307	592,101	438,669	298,430	599,391	447,158	173,726	315,644
London	807,936	1,410,785	723,354	575,331	305,781	633,790	453,923	506,290	700,292
South East	789,267	1,466,583	850,558	643,058	409,969	799,485	557,852	234,881	522,688
South West	394,860	834,208	492,794	435,470	281,682	571,727	407,663	131,735	306,576

The official NS-SeC categories have long names, so they are shortened in this dataset to make them easier to work with. Here is a reference table of our Excel column headers and their full definition is shown in table 1.

Table 1 – Occupational social class

Variable	Description
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
sec8	8. Excluded from labour market i.e. Never worked or long-term unemployed.
NC	Not Classified (inc. full-time students)

- Next, to be consistent with the previous exercises, sort the data by region so it is listed alphabetically.

- When looking at new data, it's good practice to confirm that it is what you think it is, and that it matches your existing knowledge. For example, you could sum the NS-SeC rows for each region, and check that these region totals are the same as the 'Working age' population figures given on the 'Data' worksheet.

Recoding the NS-SeC into 3 classes

When analysing data it is often convenient and sensible to aggregate, or combine, groups together. For example, if we were considering housing tenure, people living in local authority rented properties – “council housing” - and people who rent property through housing associations 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 the classes 1-7 (the categories of active types of work) into a hierarchy of three larger groups.

These three groups and their component classes are:

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

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

Further 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 further analysis tell you?

Summary

This section has involved working with data and using a range of functions within Excel to present tables and charts. The next toolkit will use a larger dataset of the local authorities in England.

Further resources

Graphs

Many (many) online resources exist showing different techniques to use when presenting data. For some general creative ideas, see

<https://sites.google.com/site/e90e50charts/> (not all are useful but interesting nonetheless!).

Recommended book for graph design and data presentation: Show Me the Numbers, Stephen Few, 2nd edition 2012, Atlantic Press.