

Interpreting Geographical Data

Social Analysis
Part IV
with
Microsoft
Excel



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 IV focusses on functions to enable you to present good quality tables.

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
Exploring Ethnicity and Place	4
Further resources.....	9
Appendix 1 - Dataset.....	10
Local Authority.....	10
Social Class	10
Ethnicity	11
Highest Level of Qualification	12

Tables and Figures

Table 1 – use of Text and Concatenate function	7
Table 2 – local authority variables	10
Table 3 – social class variables.....	11
Table 4 – ethnic group population.....	11
Table 5 – highest level of qualification categories	12

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
- ... work with complex datasets, preparing worksheets and variables for a series of analyses
- ... analyse interactions between two variables and produce pivot tables
- ... 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 Excel formulae. Please refer to the previous workbooks if you are unsure of any of these.

Exploring Ethnicity and Place

In this section we are going to develop your skills so that you can produce good quality tables to support your analysis and reporting. This is a common task – having, produced some data in Excel or SPSS it needs to be *pretty printed*, making it presentable as a table for a report.

In this exercise the aim is to produce a table, showing the counts and percentages of each ethnic group in Local Authorities characterised by different urban and rural settings.

So you know what you're aiming for, here is the first row of the final table:

Ethnicity	Urban/Rural Setting of Local Authority		
	Mainly/Largely Urban (%)	Other Urban, Town & Fringe (%)	Village, Hamlet or dispersed (%)
White British	17,721,127 (41.9%)	14,133,402 (33.4%)	10,424,707 (24.7%)
...

To begin:

- Download the file 'LA.xlsx' from Moodle and open it
- Select the worksheet <Ethnicity Table> and review what has already been done to get you started.

Currently the ethnicity variables have been collated in a pivot table, drawing the data from the <Ethnic Diversity> worksheet. The contents of the pivot table have been transposed and, moving to the right, some of the groups have been aggregated.

The next task is to finish preparing the data:

- Starting at row 14, complete column M, combining the populations of the ‘Village’ and ‘Hamlet or dispersed’ categories.
- Complete columns N to P, calculating the row percentages.
 - For example row 14 is the wb (White British) group so cell N14 will contain the proportion of the whole White British Population living in a Largely Urban setting.
 - Once the proportions have been calculated they can be formatted as percentages to 1 decimal place.

Presenting Text pt I – the VLOOKUP function

The area where the final table should be built starts at cell R14. Up to now the ethnic groups have been described with short codes e.g. bcar for ‘Black Caribbean’ or ban for ‘Bangladeshi’. As with the other exercises, the full description of the codes is given on the <lookup> worksheet.

Whilst we use short codes as a convenient shorthand during the analysis, they won’t mean a lot to anyone else! Presenting a good finished table requires proper descriptions. The VLOOKUP function helps, using a lookup table to translate the codes to something more legible. It is much less painful than having to re-type all the descriptions every time you produce a table.

Noting that:

- Our first code to lookup is ‘wb’ in cell J14
 - The lookup table is on the <lookup> worksheet in cell range A13:B26
- In cell R14 enter the formula `=VLOOKUP(J14, lookup!A13:B26, 2, 0)`

Note in the formula above, ‘lookup!’ directs Excel to retrieve the values from the <lookup> worksheet.

When the formula is copied down the location of the lookup table must remain constant. The formula can be fixed by:

- Select cell R14 and edit it by pressing <F2> (or double click in cell R14).
- Move the cursor to the reference to cell A13 and press <F4>

Excel for Social Analysis part IV

- And do the same for the reference to cell B26
- Finish the edit by pressing <return>

The <F4> shortcut changes cell references so they are static, referring to a fixed location. The formula should now read:

```
=VLOOKUP(J14, lookup!$A$13:$B$26, 2, 0)
```

- Copy the formula down as far as row 28 to lookup the descriptions for all the ethnic groups.
- The final row is for the population as a whole so enter 'Total Population' in cell R29.
- Highlight R29 in Bold

Sensitivity of VLOOKUP – why isn't it working?

You will notice that the VLOOKUP function has not worked for a few of the cells (J19 and J22). As with data you encounter in the real world, these data are not perfect, and although the ethnic group code in these cells look the same as the rest of the column, the cells contains a trailing space after the 'ind' and 'chi'.

Cleaning text using the TRIM function

In order to easily remove trailing or leading characters from a cell that will lead to errors, you can use the TRIM function.

- Update the formulae to include the TRIM function on the cells that are currently not providing a match, e.g. =VLOOKUP(TRIM(J19), lookup!\$A\$13:\$B\$26, 2, 0)

Approximate or exact matching with VLOOKUP

We have specified in the original VLOOKUP argument that we want a perfect match in the lookup table, as the last argument is 0 for approximate matching, as follows:

```
=VLOOKUP(J14, lookup!$A$13:$B$26, 2, 0)
```

Another option to use is 1, or TRUE, which allows the VLOOKUP function to return an approximate match if your search criteria does not have to be exact. This option can also be used in our case, to extract the full names of each ethnic group – however approximation will not always be appropriate to use with your data.

Presenting Text pt II – the TEXT and CONCATENATE functions.

For the columns describing the populations for the different urban/rural settings we are aiming to present the population followed by the percentage in brackets. So the first cell will contain both:

Excel for Social Analysis part IV

- The total White British population residing in a Mainly/Largely Urban setting
- The percentage of the group in this setting (surrounded by brackets)

This first cell, S14, will contain a nicely formatted text string like this:

17,721,127 (41.9%)

Assembling content like this in a cell requires the use of some of Excel's functions for manipulating text:

Text String Functions

We can put together values from different places to provide a formatted string of text. In the next step we are going to create a format for the first cell and then copy it to the others. This will combine the use of the two functions

`TEXT(cell, format)` which converts a number to the format specified in the second parameter.

`CONCATENATE(text1, text2, ...)` which adds together two or more strings

Table 1 has some examples of the functions usage. For the purpose of these examples, cell A1 contains the value 12345 and cell A2 has the value 0.1.

Table 1 – use of Text and Concatenate function

Function usage	Resulting Text
<code>=TEXT(1.5, "£0.00")</code>	£1.50
<code>=TEXT(A1, "#,###")</code>	12,345
<code>=TEXT(A2, "0.0%")</code>	10.0%
<code>=CONCATENATE("Hello", " World")</code>	Hello World
<code>=CONCATENATE("(", TEXT(A1, "#,###"), ")")</code>	(12,345)

- Write a formula for cell S14 so the data is presented in the required format. i.e. the cell should say:

17,721,127 (41.9%)

- The formula is more complicated than anything covered so far. It might help to build it up in small steps rather than trying to get the whole thing right first time.
- The formula can then be copied to all the table cells (S14:U28)

Excel for Social Analysis part IV

The table contents are now complete. All that remains to be done is some tidying up to make the table more presentable i.e. making column headings bold, aligning cell contents and adding lines around the cells.

- Double clicking on the line between the column letters will automatically make the column adjust to fit the largest contents
- Select the table headers and click **B** on the home ribbon to make the headings bold
- Selecting the numerical data (S14:U28), these will look better if the cell contents are aligned to the right. On the home ribbon you can change the alignment of this text.
- Finally add borders to the table using the 'Borders' tool on the home ribbon.

As a reminder, the first row of the final table should look like this:

	Urban/Rural Setting of Local Authority		
Ethnicity	Mainly/Largely Urban (%)	Other Urban, Town & Fringe (%)	Village, Hamlet or dispersed (%)
White British	17,721,127 (41.9%)	14,133,402 (33.4%)	10,424,707 (24.7%)
...

You may like to experiment with copying and pasting the table into Word. Excel and Word don't always agree about the width required for some text so this might require some final further adjustments to the column widths.

Well done, that's the final exercise completed in this section.

Summary

This section has involved working with data and using a range of functions within Excel to present tables. This completes the main tools in Excel. Two further workbooks, *Sources of Data* and *Advanced Graphical presentation* contain the technical skills needed to complete the first assignment.

Further resources

Vlookup issues

Vlookups can be very useful, but as we touched on earlier they are prone to sensitivities. See here for troubleshooting tips: <http://howtovlookupinexcel.com/13-common-problems-with-vlookups/>

Appendix 1 - Dataset

The dataset being used provides information on population, social class, ethnicity and highest level of qualification achieved for each local authority in England. The tables below describe how this data is organised, the main categories used and the population which different variables are drawn from.

Local Authority

Table 2 shows the variable names and, for the regions and urban rural classification, the possible values that are given to each local authority.

Table 2 – local authority variables

Variable name	Description
LA	Local authority
Region	English region <i>Possible values</i> East East Midlands London North East North West South East South West West Midlands Yorkshire and Humberside
Urban-rural	Urban-rural classification <i>Possible values</i> Major urban Largely urban Other urban Town and fringe Village Hamlet or dispersed

Social Class

The measure of social class is derived from the occupation of the respondent. It is therefore only applied to those identified as of working age which is 16-74 in the census data. Table 3 shows the nine categories used in the analysis together with a way of combining these to higher level categories if required.

Table 3 – social class variables

Variable	Description
nssec1	Higher managerial, administrative or professional
nssec2	Lower managerial, administrative or professional
nssec3	Intermediate
nssec4	Small employer and self-employed
nssec5	Lower supervisory and technical
nssec6	Semi-routine
nssec7	Routine
nssec8	Never worked and long-term unemployed
nssec9	Full time students

Ethnicity

The number of members of each ethnic group is provided as a separate population count for each local authority. The count reflects the number by ethnicity for the whole population. Table 4 provides these counts as well as the proportion of the population in the local authority who come from each ethnic group.

Table 4 – ethnic group population

Variable	Description
wb	White British
wi	White Irish
wgt	White Gypsy or Irish Traveller
wo	White Other
mix	Mixed
ind	Indian
pak	Pakistani
ban	Bangladeshi
chi	Chinese
ao	Asian other
bafr	Black African
bcar	Black Caribbean
bo	Black other
oth	other ethnic group

Additionally further columns use these variable names but prefixed by **pr**, giving the proportion/percentage of the ethnic group relative to the total population in that Local Authority.

Highest Level of Qualification

Details of the translation of the highest level of qualification to the categories are shown in table 5.

Table 5 – highest level of qualification categories

Variable	Description
Noqual	No qualifications
L1qual	1-4 O Levels/CSE/GCSEs (any grades), Entry Level, Foundation Diploma, NVQ level 1, Foundation GNVQ, Basic/Essential Skills
L2qual	5+ O Level (Passes)/CSEs (Grade 1)/GCSEs (Grades A*-C), School Certificate, 1 A Level / 2-3 AS Levels/VCEs, Intermediate/Higher Diploma, Intermediate Diploma, NVQ level 2, Intermediate GNVQ, City and Guilds Craft, BTEC First/General Diploma, RSA Diploma
Appqual	Apprenticeship
L3qual	2+ A Levels/VCEs, 4+ AS Levels, Higher School Certificate, Progression/Advanced Diploma, NVQ Level 3; Advanced GNVQ, City and Guilds Advanced Craft, ONC, OND, BTEC National, RSA Advanced Diploma
L4qual	Degree (for example BA, BSc), Higher Degree (for example MA, PhD, PGCE), NVQ Level 4-5, HNC, HND, RSA Higher Diploma, BTEC Higher level, Foundation degree
Othqual	Vocational/Work-related Qualifications, Qualifications gained outside the UK (Not stated/level unknown)