

Interpreting Geographical Data

Social Analysis

Part III

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 III introduces pivot tables which allow you to produce tables from larger datasets. They provide a quick and easy way of exploring the relationship between categorical variables.

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
Introducing Pivot Tables	4
Appendix 1 - Dataset.....	10
Local Authority.....	10
Social Class	10
Ethnicity	11
Highest Level of Qualification	12

Tables and Figures

Table 1 – local authority variables	10
Table 2 – social class variables.....	11
Table 3 – ethnic group population.....	11
Table 4 – 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 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.

To begin open the workbook:

- Download the file 'LA.xlsx' from Moodle and open it

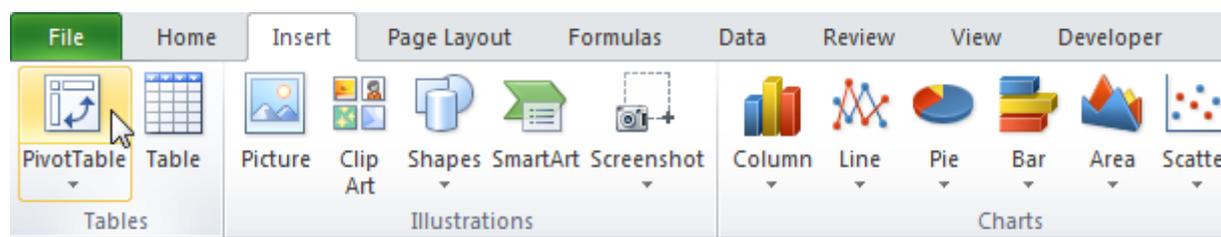
Introducing Pivot Tables

In this section we will be using Pivot Tables, a powerful feature of Excel that allows data to be grouped and aggregated. A way to view the wood rather than the individual trees. The first exercises aim to introduce a variety of useful features for pivot tables. The final exercise works with the census NS-SeC measure of occupation based social class, aiming to distinguish the occupational roles in which men and women predominate.

Exercise 1 – A Simple Pivot Table

The first task is to count the number of Local Authorities in each Region.

- Don't use the 'Social Class' worksheet yet. Instead click on a cell in the 'Qualifications' worksheet.
- On the Insert tab, click PivotTable.



- Click OK in the dialog box. Excel has selected all the data and will, by default, place the pivot table in a new worksheet. Rename the worksheet for your future reference.
- Excel produces a blank pivot table with a panel on the right hand side of the screen for controlling it. We would like a list of regions with the count of LAs in each so, working on the panel, drag the <Region> field into the Row Labels box and the <LA> field should be dragged into the Σ Values box.

The pivot table and control panel should then look like this:

Row Labels	Count of LA
East	47
East Midlands	40
London	33
North East	12
North West	39
South East	67
South West	37
West Midlands	30
Yorkshire and Humberside	21
Grand Total	326

We have quickly produced a summary of the data, counting the number of local authorities in each district. Without pivot tables the process would have been much more painful, sorting the data by region and then counting by hand the number of rows for each.

Example 2 – Pivot Table Grouping by Many Variables

Now we will progress to the ‘Social Class’ data and try a more complicated example. Suppose you have been asked to identify the region with the largest rural population and to give a breakdown of this population by sex. The rural population are those living in either the ‘Village’ or ‘Hamlet and dispersed’ categories of the Urban-rural classification variable.

As before the Regions will be the rows in our pivot table. The results needs to be Filtered to just some of the urban-rural classes and the breakdown by sex requires separate Columns for each region. The Value to be counted is the total population.

- Produce a new pivot table on a new worksheet to the spec given above.

When done the table and panel should look like this:

Excel for Social Analysis (Part III)

Urban-rural	(All)		
Sum of Pop	Column Labels		
Row Labels	Female	Male	Grand Total
East	2971158	2875807	5846965
East Midlands	2298729	2234493	4533222
London	4140652	4033289	8173941
North East	1327183	1269703	2596886
North West	3587492	3464685	7052177
South East	4395452	4239298	8634750
South West	2698327	2590608	5288935
West Midlands	2838660	2763187	5601847
Yorkshire and Humberside	2685655	2598078	5283733
Grand Total	26943308	26069148	53012456

Drag fields between areas below:

Report Filter	Column Labels
Urban-rural	Sex
Row Labels	Σ Values
Region	Sum of Pop

The results can be limited to the Rural Areas as follows:

- Click the down arrow on the Urban-rural filter at the top of the pivot table:

- Select the Multiple Items check box at the bottom of the menu.
- Adjust the urban-rural tick boxes so that only 'Village' and 'Hamlet or Dispersed' are selected.

The pivot table will then be limited to the local authorities that are mostly of a rural character.

Finally the results can be sorted so that the region with the greatest population appears at the top of the list:

- Right click on one of the numbers in the Grand Total column and, on the context menu, select <Sort> and then <Largest to Smallest>.

Q1 Returning to the original research question, which region has the greatest population within predominantly rural Local Authorities?

Q2 What is the rural population of men and women in this region?

If you have it right, the South West is the region of interest with a total rural population of 2,445,003 people.

The intention here was mainly to illustrate the process of creating and configuring a pivot table. An analysis like this could be interesting to those involved in planning the provision of council services. Delivery of services like refuse collection will tend to be more expensive in rural areas because of transport costs. However, in reality, a Local Authority may contain a combination of rural and urban areas. Therefore, a more sophisticated analysis would use smaller units; electoral wards for example.

Example 3 – Collating Pivot Data for Analysis

The table gives data from 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. Investigating sexual inequality in employment, research questions are:

- The highest class position is nssec1 – the higher managerial, administrative and professional workers. To what extent is this class skewed towards male workers?
- Ignoring the long term unemployed/those who have never worked and full-time students, which of the 7 classes are mostly composed of female workers.

The basic procedure for the analysis is to use a pivot table to collate the required data and this is used as a basis for calculating proportions for each class.

- Select the worksheet <Social Class>
- Insert a pivot table, creating a new worksheet.
- Put the Sex variable in Row Names
- Put all of the nssec variables in Σ values

By default the pivot table is showing total counts i.e. the number of males and females in each class group. The counts can be changed to percentages so doing this for the nssec1 variable will be sufficient for answering the first research question:

- Select one of the cells in the pivot table that collates nssec1 e.g. cell B4
- Right click and on the context menu select <Show Values As> and then click <% of Column Total> on the submenu.

So returning to the research question:

Q3 The highest class position is nssec1 – the higher managerial, administrative and professional workers. To what extent is this class skewed towards male workers?

It would be convenient to convert all the variables nssec1...nssec7 to give column percentages. However an inconvenient feature of Excel is that it insists this is done one variable at a time and, additionally, it will not calculate row totals and percentages across multiple variables. So an alternative is to copy the pivot data to make a table describing each of the social class groups.

- <Copy> the descriptions of the NS-SeC classes 1 to 7 from the Lookup worksheet and <Paste> them in column A beneath the Pivot Table.
- Make Column A wider, as appropriate, to show the full class descriptions.
- Select the pivot table data including the Row Labels Grand Total row.
- Do a <Paste Special> - <Transpose> to place the data next to the class descriptions.
- Calculate two further columns, working out the proportion of males and females in each class.
- Format the proportions as percentages to 1 d.p.

To check your progress, the sheet below shows the work mostly done, with only the percentage columns remaining:

	A	B	C	D	E	F	G	H
1								
2								
3		Sum of nssec1	Sum of nssec2	Sum of nssec3	Sum of nssec4	Sum of nssec5	Sum of nssec6	Sum of nssec7
4	Female	1407660	4404270	3676483	1089293	771637	3390047	1815652
5	Male	2638163	3727837	1295561	2573318	1904481	2040816	2461831
6	Grand Total	4045823	8132107	4972044	3662611	2676118	5430863	4277483
7								
8								
9								
10		Female	Male	Total	% Female	% Male		
11	Higher managerial, administrative or professional	1407660	2638163	4045823				
12	Lower managerial, administrative or professional	4404270	3727837	8132107				
13	Intermediate	3676483	1295561	4972044				
14	Small employer and self-employed	1089293	2573318	3662611				
15	Lower supervisory and technical	771637	1904481	2676118				
16	Semi-routine	3390047	2040816	5430863				
17	Routine	1815652	2461831	4277483				

Finally, with the percentages calculated, it will now be possible to answer the second research question:

Q4 Ignoring the long term unemployed/those who have never worked and full-time students, which of the 7 classes are mostly composed of female workers?

Summary

This section has involved working with data using the pivot table tool to explore data. The next toolkit explores ways to enhance the quality of tables produced in Excel.

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

Table 1 – 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 2 shows the nine categories used in the analysis together with a way of combining these to higher level categories if required.

Table 2 – 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 3 provides these counts as well as the proportion of the population in the local authority who come from each ethnic group.

Table 3 – 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 4.

Table 4 – 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)