Task 4.6 Building and mapping a small area deprivation index for health needs assessment

The purpose of this exercise is to construct and map a health-oriented deprivation index for a study area in southern England. The exercise demonstrates the multiple stages of data retrieval and manipulation in the creation of such indices, compared to the relative ease of mapping the results. The index to be constructed uses the same variables as the Carstairs index but will not be standardized and calibrated using exactly the same reference data.

Data:

Provided within this compressed file is a Shapefile **HantsOA01** containing a boundary dataset for 2001 census output areas in the Hampshire, Southampton and Portsmouth local government areas in southern England, extracted from data freely distributed on DVD by the Office for National Statistics. The approximate area of the map can be explored online at http://local.live.com/default.aspx?v=2&cp=51.046574~-1.198883&style=r&lvl=10&tilt=-90&dir=0&alt=-1000&scene=4318724 The license documents **Conditions of supply of 2001 Census Output Area Boundaries.doc** and **OS_OA_License.doc** specify the conditions of use of these datasets.

Exercise:

Getting to grips with the Carstairs deprivation index:

The paper by Morgan and Baker (2006) describes the construction of a Carstairs deprivation index for areas known as wards in England and Wales by the Office for National Statistics. The analysis described in the paper uses raw, unpublished census data that are not available to researchers outside the statistical agencies. In this exercise, we shall produce a Carstairs-based index for census output areas using published census data from the Neighbourhood Statistics website. Before commencing this exercise, you should read the paper by Morgan and Baker (2006).

Task 1: What are Morgan and Baker's (2006) reasons for selecting a Carstairs deprivation index and the types of health-related analysis which they were intending to perform?

Note the four census variables which make up the Carstairs index.

See last page for answers.

Accessing census data for the Carstairs index

Census data need to be retrieved from the Nomis website https://www.nomisweb.co.uk/census/2011/data_finder . The site allows spreadsheet tables to be downloaded which contain census variables for a single topic, with variables as the columns and geographical areas as the rows.

Make sure you select 'Output Area' as the Geography to work with:

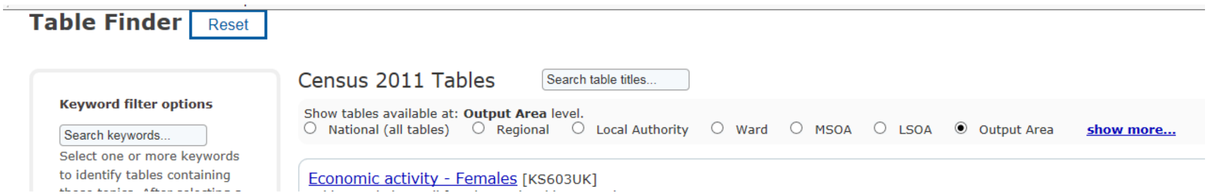


Table Finder Reset

Keyword filter options

Search keywords...

Select one or more keywords to identify tables containing

Census 2011 Tables

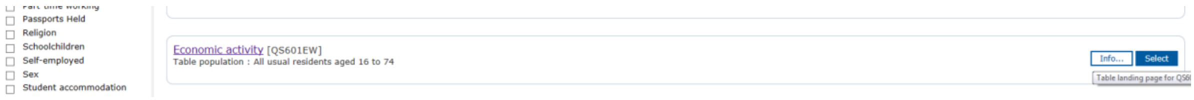
Search table titles...

Show tables available at: **Output Area** level.

○ National (all tables) ○ Regional ○ Local Authority ○ Ward ○ MSOA ○ LSOA ● Output Area show more...

Economic activity - Females [KS603UK]

Next, from the left-hand side, select the topic 'Economic Activity' and then hit the *select* button to the right of table 'Economic Activity (QS601EW)' – economic activity in residents aged 16 to 74 in England and Wales:



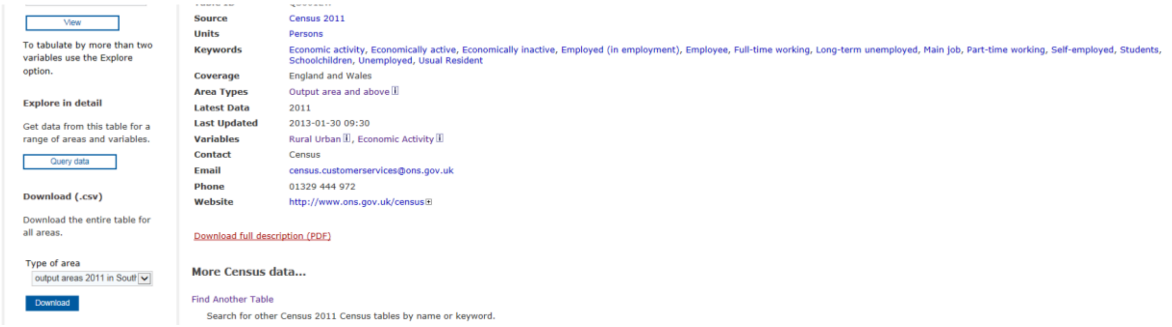
□ Passports Held
□ Religion
□ Schoolchildren
□ Self-employed
□ Sex
□ Student accommodation

Economic activity [QS601EW]
Table population : All usual residents aged 16 to 74

Info... Select

Table landing page for QS6

[Note: Morgan and Baker actually use male unemployment, but given the nature of the modern workforce, we will look at overall unemployment in both males and females]. Having done this, on the next screen, we can now download the relevant data. In the bottom-left corner, under *type of area*, select *output areas in south east*, then press the *Download* button.



View

To tabulate by more than two variables use the Explore option.

Explore in detail

Get data from this table for a range of areas and variables.

Query data

Download (.csv)

Download the entire table for all areas.

Type of area

output areas 2011 in South

Download

Source	Census 2011
Units	Persons
Keywords	Economic activity, Economically active, Economically inactive, Employed (in employment), Employee, Full-time working, Long-term unemployed, Main job, Part-time working, Self-employed, Students, Schoolchildren, Unemployed, Usual Resident
Coverage	England and Wales
Area Types	Output area and above
Latest Data	2011
Last Updated	2013-01-30 09:30
Variables	Rural Urban, Economic Activity
Contact	Census
Email	census.customerservices@ons.gov.uk
Phone	01329 444 972
Website	http://www.ons.gov.uk/census

Download full description (PDF)

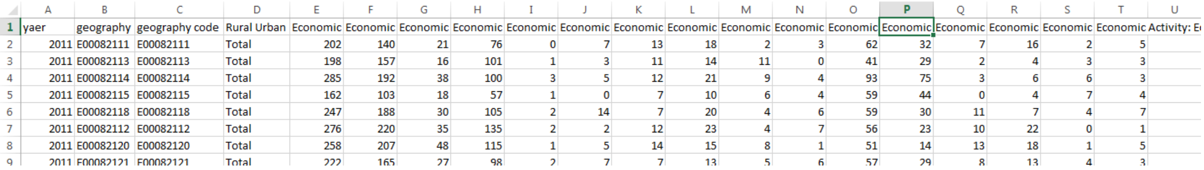
More Census data...

Find Another Table

Search for other Census 2011 Census tables by name or keyword.

You should now be able to download a comma-separated values file, in which each row is an output area and the various columns provide counts of the population in different economic activity categories (e.g. students; retired; unemployed, etc). Office for National Statistics use different polygon layers for distributing census data, sometimes referred to as census geographies. Output areas are the smallest of the polygon layers used – you can find about these and other polygons used (in fact, University of Southampton helped design many of these polygon layers!) here: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography.

If you open this .csv file up in Excel (note you may need to select ‘all files’ rather than ‘Excel files’ to do so), you can see its structure:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	yaer	geography	geography code	Rural Urban	Economic	Economic	Economic	Economic	Economic	Economic	Economic	Economic	Economic	Economic	Economic	Economic	Economic	Economic	Economic	Economic	Activity: E
2	2011	E00082111	E00082111	Total	202	140	21	76	0	7	13	18	2	3	62	32	7	16	2	5	
3	2011	E00082113	E00082113	Total	198	157	16	101	1	3	11	14	11	0	41	29	2	4	3	3	
4	2011	E00082114	E00082114	Total	285	192	38	100	3	5	12	21	9	4	93	75	3	6	6	3	
5	2011	E00082115	E00082115	Total	162	103	18	57	1	0	7	10	6	4	59	44	0	4	7	4	
6	2011	E00082118	E00082118	Total	247	188	30	105	2	14	7	20	4	6	59	30	11	7	4	7	
7	2011	E00082112	E00082112	Total	276	220	35	135	2	2	12	23	4	7	56	23	10	22	0	1	
8	2011	E00082120	E00082120	Total	258	207	48	115	1	5	14	15	8	1	51	14	13	18	1	5	
9	2011	E00082121	E00082121	Total	222	165	27	98	2	7	7	13	5	6	57	29	8	13	4	3	

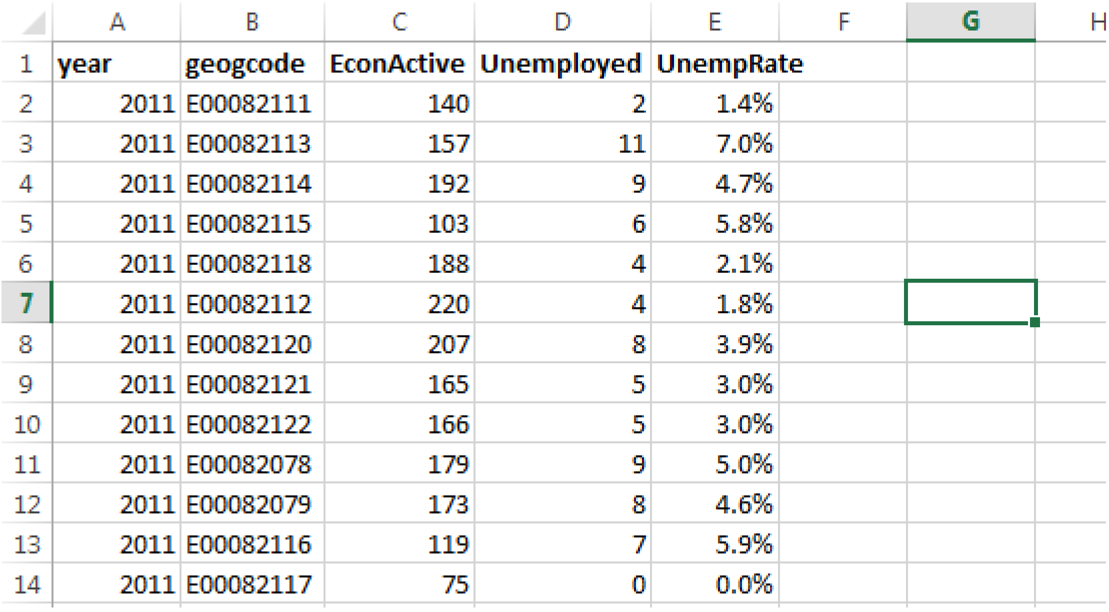
The first column tells us the year that the data relate to; the second and third contain unique ID numbers for each output area (the ‘E’ as the first character tells us that they are English output areas). We then see that the figures are totals for both rural and urban unemployment, rather than having separate figures for rural versus urban areas. There is then a lot of information here about the types of economic activity that those aged 16 to 74 undertake.

From our perspective in pulling together our deprivation index, we can simplify this information greatly. Of all the information in the file, there are three fields that particularly interest us (possibly also retaining the ‘year’ field): Column B or C (the output area polygon IDs); column F, the count of the economically active population aged 16-74 years (‘Economic Activity: Economically active: Total; measures: Value’); and column M, the count of those unemployed (‘Economic Activity: Economically active: Unemployed; measures: Value’). The economically active population excludes certain groups, such as the retired, so is seen as a better denominator when calculating unemployment rates.

If we are to import this file in to ArcGIS, it would probably help at this point to simplify it, by:

- Removing all fields, except those we are interested in.
- Renaming the column headers, so that they are shorter and do not contain spaces or special characters (e.g. ‘:’), as required by ArcGIS for import
- Calculating the unemployment rate as the count of employed divided by the economically active population.

The result should look something like this:



	A	B	C	D	E	F	G	H
1	year	geogcode	EconActive	Unemployed	UnempRate			
2	2011	E00082111	140	2	1.4%			
3	2011	E00082113	157	11	7.0%			
4	2011	E00082114	192	9	4.7%			
5	2011	E00082115	103	6	5.8%			
6	2011	E00082118	188	4	2.1%			
7	2011	E00082112	220	4	1.8%			
8	2011	E00082120	207	8	3.9%			
9	2011	E00082121	165	5	3.0%			
10	2011	E00082122	166	5	3.0%			
11	2011	E00082078	179	9	5.0%			
12	2011	E00082079	173	8	4.6%			
13	2011	E00082116	119	7	5.9%			
14	2011	E00082117	75	0	0.0%			

You may wish to save your data file as a new .csv file.

Producing standardised scores

Following Morgan and Baker (see Table 2), we now have one final task to do – namely to convert our unemployment rate into Z scores. A Z score takes a value from a distribution (e.g. the unemployment rate for one output area from the distribution of unemployment rates for all output areas in a wider study area) and expresses it as standard deviations above or below the mean of that distribution. Thus, a Z score of -1 would indicate that the unemployment rate for a given output area is one standard deviation lower than the average across all output areas; a Z score of 0 would mean that unemployment in the output area equalled the average rate across all output areas; whilst a Z score of +2 would mean that the unemployment rate was two standard deviations above the average across all output areas. In other words, the Z-score tells us about deprivation relative to a reference population, rather than in absolute terms. Here, for simplicity, we will use south-east England as our reference population.

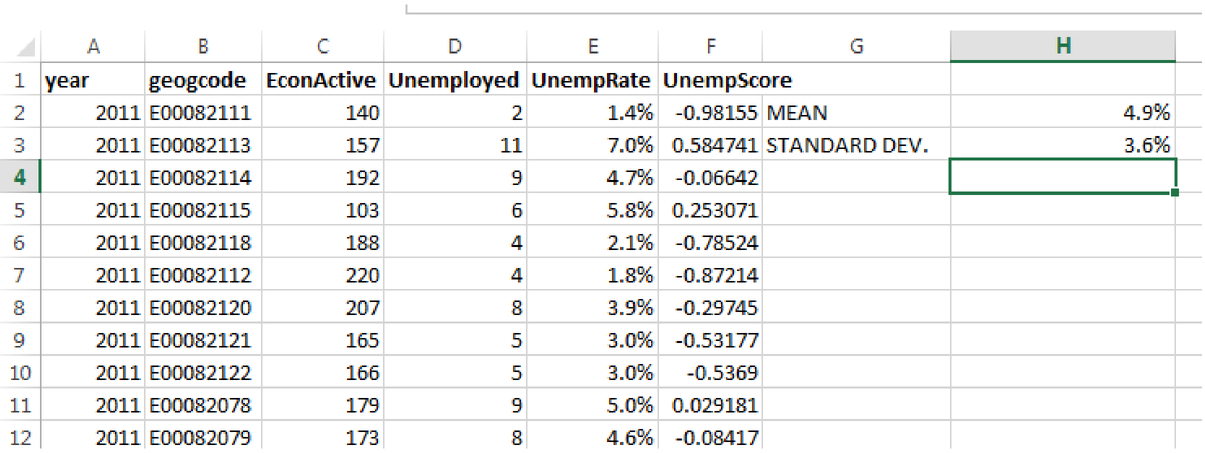
To calculate Z-scores for unemployment, we need to do the following:

- Work out the mean unemployment rate for all output areas in south-east England. In an empty cell away from your main data table, you can do this either by entering the Excel formula '=AVERAGE(E2:E27639)', referencing the cells (in my case E2:E27639) that contain your unemployment rates, or you can do this by heading for the *formulas* menu, choosing *insert function,* then searching for the *average* function.
- Work out the standard deviation of unemployment rates for all output areas in south-east England. Again, in an empty cell away from your main data table, you can do this either by entering the Excel formula '=STDEV(E2:E27639)', referencing the cells (in my case E2:E27639) contain your unemployment rates,

or you can do this by heading for the *formulas* menu, choosing *insert function,* then searching for the *stdev* function.

- Compute standardised scores. This involves us subtracting the mean from the unemployment rate for each output area, then dividing by the standard deviation. Probably the easiest way to do this is to enter '=(E2-H2)/H3', where H2 is the cell containing your mean, and H3 is the cell containing your standard deviation. The H2 notation in Excel is what is known as an absolute cell reference. It refers to cell H2, but if you drag this formula down to other cells in your spreadsheet, for example the one immediately below where you typed in your formula, it will always reference cell H2. If you did this with a relative reference, e.g. 'H2', it would update to refer to cell H3, which would be unhelpful with this particular calculation.

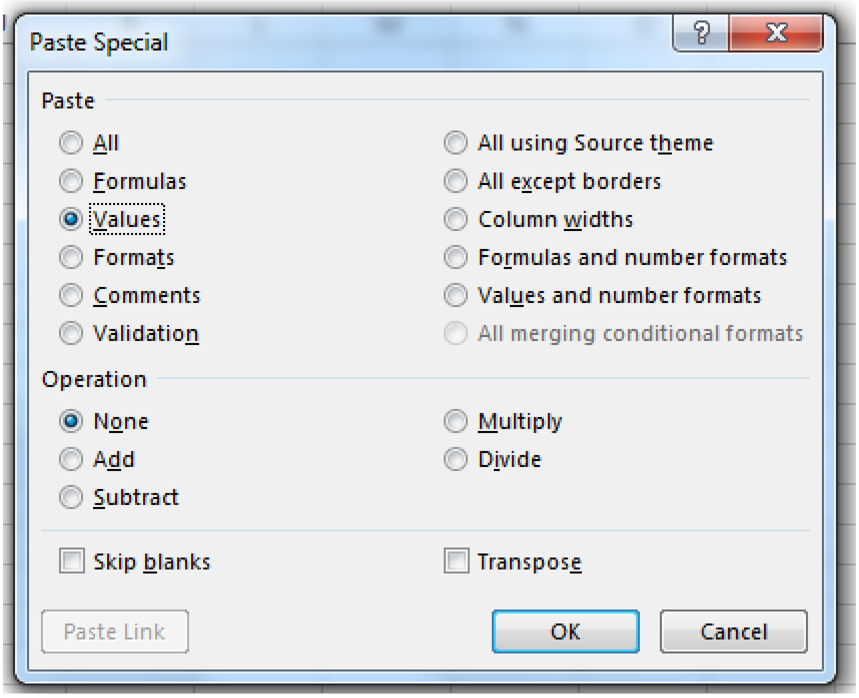
If things go to plan, you should have a spreadsheet that looks something like this:



	A	B	C	D	E	F	G	H
1	year	geogcode	EconActive	Unemployed	UnempRate	UnempScore		
2	2011	E00082111	140	2	1.4%	-0.98155	MEAN	4.9%
3	2011	E00082113	157	11	7.0%	0.584741	STANDARD DEV.	3.6%
4	2011	E00082114	192	9	4.7%	-0.06642		
5	2011	E00082115	103	6	5.8%	0.253071		
6	2011	E00082118	188	4	2.1%	-0.78524		
7	2011	E00082112	220	4	1.8%	-0.87214		
8	2011	E00082120	207	8	3.9%	-0.29745		
9	2011	E00082121	165	5	3.0%	-0.53177		
10	2011	E00082122	166	5	3.0%	-0.5369		
11	2011	E00082078	179	9	5.0%	0.029181		
12	2011	E00082079	173	8	4.6%	-0.08417		

Finally, we need to tidy up our calculations a little for import into ArcMap. Having used the mean and standard deviation in our calculations, we now need to remove them. To do this:

- Highlight your column of standardised scores, then copy this to the clipboard.
- Click on 'paste', then 'paste special', and in the next dialog box, select 'values':



Paste Special

Paste

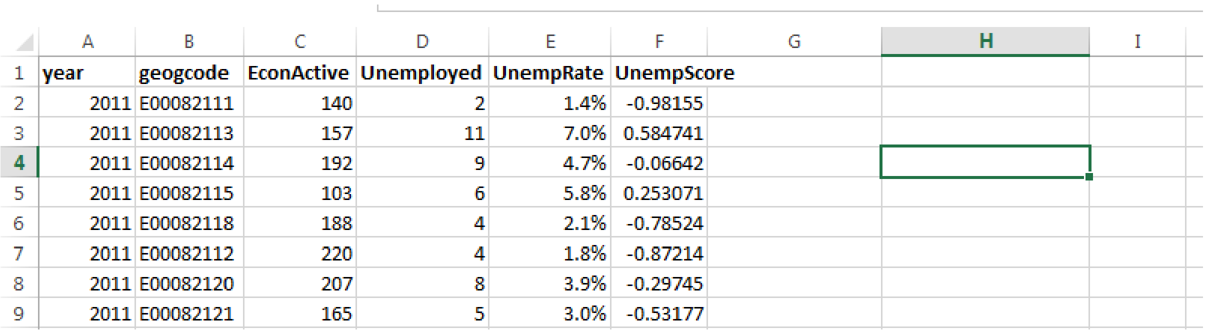
○ All
○ Formulas
● Values
○ Formats
○ Comments
○ Validation
○ All using Source theme
○ All except borders
○ Column widths
○ Formulas and number formats
○ Values and number formats
○ All merging conditional formats

Operation

● None
○ Add
○ Subtract
○ Multiply
○ Divide

□ Skip blanks
□ Transpose

- This will replace your formulas with the calculated values. We can now safely delete the cells with our mean and standard deviation, without the index values being affected. Your spreadsheet should now consist only of the table of information about the output areas:



	A	B	C	D	E	F	G	H	I
1	year	geogcode	EconActive	Unemployed	UnempRate	UnempScore			
2	2011	E00082111	140	2	1.4%	-0.98155			
3	2011	E00082113	157	11	7.0%	0.584741			
4	2011	E00082114	192	9	4.7%	-0.06642			
5	2011	E00082115	103	6	5.8%	0.253071			
6	2011	E00082118	188	4	2.1%	-0.78524			
7	2011	E00082112	220	4	1.8%	-0.87214			
8	2011	E00082120	207	8	3.9%	-0.29745			
9	2011	E00082121	165	5	3.0%	-0.53177			

- If you save this as a .csv file, we can now import it into ArcGIS.

Map your unemployment scores in ArcGIS

Now we can map our unemployment scores. To do this, first head for https://borders.ukdataservice.ac.uk/. Click on ‘easy download’ and select ‘English output areas 2011’ as the file to download. It is probably best to select the version that is ‘English Census Output Areas, 2011, Clipped and Generalised (simplified polygon geometry)’, choosing shape file format. Alternatively, you can download these boundaries from here: https://data.gov.uk/dataset/output-areas-oa-boundaries.

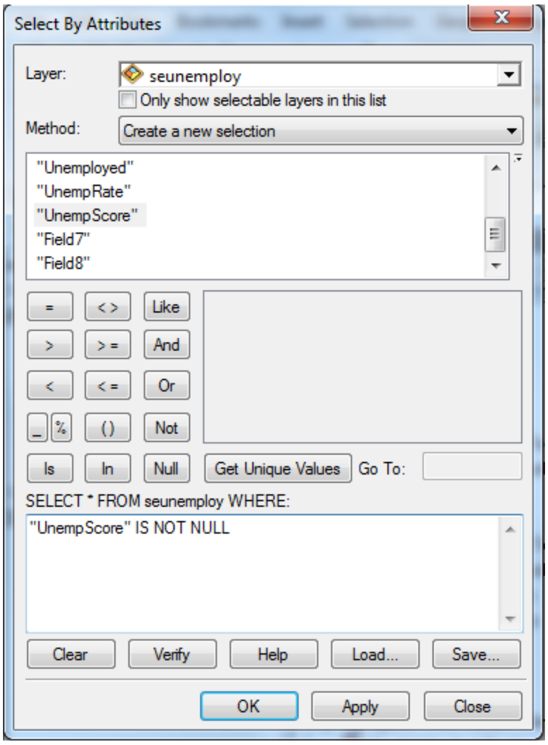
To join your unemployment scores to this file, open up ArcMap and add this layer, having a look at its attribute table. Open up your newly created .csv file of unemployment data in ArcMap as well. Next, right-click on the layer in the table of contents panel, choose *joins and relates* and then *join.* If you choose *join attributes from a table,* you should be able to choose the **code** field under *choose the field in this layer that the join will be based*, then under *choose the table to join to this layer…*, choose your csv file. You should then be able to select the field with the IDs for each output area under *choose the field in the table to base the join on* – in my case, I called this **geogcode**.

Finally, if you 'validate join', you can check whether the join will work – given the size of the files, this may take some time.

> A note about reserved words when validating joins: A reserved word is one that should not be used in file or field names with a given piece of software, because it has a specific purpose within that piece of software. In my case, I had a field called 'year'. Because ArcGIS has a 'year' function, this is a reserved word, so when I pressed 'validate join' I had a non-fatal warning message that one field name was a reserved word. More generally, use of reserved words can sometimes produce errors in ArcGIS.

Once the validation process is complete, press 'OK' to join your unemployment statistics to the boundaries. As we are only interested in southeast England, it may be helpful at this point to remove the remaining output areas from the shape file. To do this:

- Choose *select by attributes* from the *selection* menu.
- You can then choose to select those output areas where your unemployment scores are not null, as shown below:



Select By Attributes

Layer: seunemploy

☐ Only show selectable layers in this list

Method: Create a new selection

"Unemployed"
"UnempRate"
"UnempScore"
"Field7"
"Field8"

SELECT * FROM seunemploy WHERE:

"UnempScore" IS NOT NULL

Finally, if you right-click on your output areas layer in the left-hand table of contents panel and choose *data*, then *export data*, you should be able save just the selected output areas. Note that this will also permanently join your unemployment fields to the attribute table. Again, this may take a little time given the size of the files.

You should now be in a position to create a choropleth map of your unemployment scores for southeast England (see the Morgan and Baker paper for an example visualisation).

Creating the deprivation index

This is the basic workflow for just one of the four indicators. To construct the Carstairs index, we need to repeat this workflow for the remaining three indicators. On the Nomis site that we used, these remaining three indicators are:

- Car or van, then 'Car or van availability' [KS404EW]
- Occupancy rating, then 'Occupancy rating (rooms)' [QS408EW] – a modern version of overcrowding.
- Ns-SeC, the National Statistics Socio-economic Classification. Note that this has changed somewhat by 2011 relative to the earlier version of the Ns-SeC shown in Table 3 of Morgan and Baker. We suggest you calculate the proportion of 'routine' or 'semi-routine' occupations among those in employment, so excluding students and the unemployed from the denominator.

A few more tips:

- Particularly for variables like 'occupancy rating', it is worth clicking on 'download full description' in Nomis, so that you understand how the variable has been calculated.
- The Carstairs index is meant to calculate rates based on population rather than households. However, without a lot of work, you may find it difficult to do this using the standard census tables for some indicators (e.g. car ownership), so it is probably easiest to work with rates for households.
- When downloading shape files in particular, do check the accompanying 'terms and conditions' file for data attribution statements to be used.

You will need to join repeatedly each of the three indicator tables to the output areas shape file for southeast England. When you have done this for the three remaining indicators, you will need to create a new attribute field for your overall Carstairs deprivation index, and then use the *field calculator* within the attribute table to calculate the sum of the four standardised indicator scores.

When you are done, try mapping your output Carstairs deprivation scores in a manner similar to Morgan and Baker.

Task 2: Based on the discussion in Morgan and Baker (2006), what relationships might you expect to find between your map and patterns of health in this study area?

See last page for answers.

Reference:

Morgan, O. and Baker, A. (2006) Measuring deprivation in England and Wales using 2001 Carstairs scores *Health Statistics Quarterly* 31, 28-33 http://www.ons.gov.uk/ons/rel/hsq/health-statistics-quarterly/no--31--autumn2006/measuring-deprivation-in-england-and-wales-using-2001-carstairs-scores.pdf

Suggested answers to tasks:

Task 1: Morgan and Baker give several reasons for using the Carstairs index. One reason was that by using such a census-based measure, they could calculate deprivation index values for both England and Wales. Another reason was that one of the alternative possible indicators, the Index of Multiple Deprivation (IMD, released for various years starting with 2004 through to 2015), contains a health domain as one of its component parts. There is something circular in the logic of looking at the relationship between health and deprivation, when the deprivation measure itself is in part based on health data.

Task 2: The authors suggest in their conclusions that 'Carstairs deprivation index has been shown to perform well in explaining variations in health measures' – in other words, one would expect to find a significant positive relationship between the Carstairs index and measures such as infant mortality. They suggest that this relationship may be stronger than for other variables, such as social class. Even though not all households living in a deprived area will be deprived, nonetheless, there is often an 'area effect' whereby even better-off households in deprived areas are at greater risk of ill health.