

## Additional data sets for public health geovisualisation

This practical sheet steps you through how to link spreadsheet format data for small areas in the UK to related boundaries. These data were originally uploaded here:

[http://www.empho.org.uk/THEMES/SMALL\\_AREA\\_DATA/smallarea.aspx](http://www.empho.org.uk/THEMES/SMALL_AREA_DATA/smallarea.aspx)

...but as of this year, much of the data have migrated to here:

<http://www.localhealth.org.uk/>

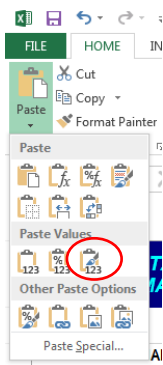
### Mapping these additional data sets

To map these additional data sets, you will first need to be able to link them to relevant boundary data. This exercise summarises how this can be done, taking an example data set (male life expectancy in the East Midlands) as an example.

### Download the health or health-related data in spreadsheet format

Here, let us take some figures for ward-level male life expectancy data and use this as an example, including in the accompanying spreadsheet of life expectancy data. As with any data set, be sure to look at the meta-data worksheet, describing how the data were constructed and their characteristics.

If you head for the 'male summary' worksheet and click on the values in each cell in the 'life expectancy' column, you will see by looking underneath the toolbars that these are based on an Excel formula (VLOOKUP). As we will need to delete some content from this spreadsheet, it is probably best to replace these formulae with actual data values. To do this, mark up the numeric values in columns C through F, being sure to mark up all of the rows with numbers in these columns (starting with row 14). Next, copy these values to the clip board. Then, if you choose *file* and *paste*, you can *paste values* to replace the formulae with the calculated life expectancy numbers:



Click on any of the cells containing numbers, and now you should find that the VLOOKUP formula has been replaced with the calculated data value.

## GIS for Analysis of Health

If you look at the spreadsheet at the moment, you will see that there are lots of header rows. However, ArcGIS will expect a single header row, followed by rows of data. Furthermore, to avoid potential problems with ArcGIS, it is best to remove spaces and non-standard characters (e.g. '%') from column headings. Row 12 also contains data for the study area (the East Midlands) as a whole: we do not need this summary row of data.

Copy of EM\_Summary\_LE\_(2005-2009).xls [Compatibility Mode] - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW DEVELOPER ADD-INS LearningGuide TEAM XL Toolbox

Clipboard Font Alignment Number Styles

Clipboard Font Alignment Number Styles

Drop down Box

SUMMARY TABLE OF LIFE EXPECTANCY AT Birth

View Chart

Please click on 'View Chart' button in order to view the chart as this refreshes the data.

Male Life Expectancy At Birth

Area Code	Area Name	Life Expectancy	Standard Error	Lower 95% CI	Upper 95% CI	LA	PCT	COUNTY
E	EAST MIDLANDS	80.2	0.15	79.9	80.5	NA	NA	NA
00FKMX	Abbey	74.7	0.76	73.2	76.2	Derby	Derby City	Derby UA
00FKMY	Allestree	83.3	0.65	82.0	84.6	Derby	Derby City	Derby UA
00FKMZ	Alvaston	75.4	0.82	73.8	77.0	Derby	Derby City	Derby UA
00FKNA	Arboretum	70.5	0.74	69.1	72.0	Derby	Derby City	Derby UA
00FKNB	Blagreaves	80.6	0.91	78.8	82.3	Derby	Derby City	Derby UA
00FKNC	Boulton	77.5	0.78	75.9	79.0	Derby	Derby City	Derby UA

To fix these problems, in Excel, delete rows 1 to 9 and 11 to 13. Change the names of columns so that they do not contain spaces and non-standard characters, so as to enable data to be loaded into ArcGIS (see below).

Copy of EM\_Summary\_LE\_(2005-2009).xls [Compatibility Mode] - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW DEVELOPER ADD-INS LearningGuide TEAM XL Toolbox

Clipboard Font Alignment Number Styles

Clipboard Font Alignment Number Styles

Derwent

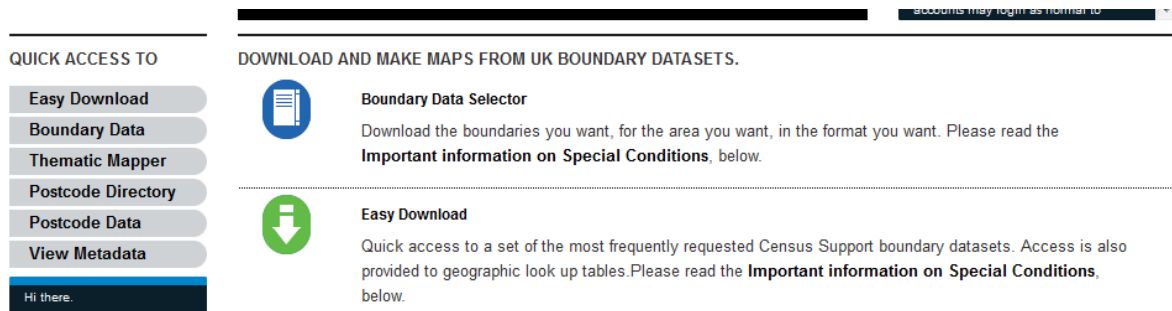
Areacode	AreaName	LifeExpect	StandErr	LowerCI	UpperCI	LA	PCT	COUNTY
00FKMX	Abbey	74.7	0.76	73.2	76.2	Derby	Derby City	Derby UA
00FKMY	Allestree	83.3	0.65	82.0	84.6	Derby	Derby City	Derby UA
00FKMZ	Alvaston	75.4	0.82	73.8	77.0	Derby	Derby City	Derby UA
00FKNA	Arboretum	70.5	0.74	69.1	72.0	Derby	Derby City	Derby UA
00FKNB	Blagreaves	80.6	0.91	78.8	82.3	Derby	Derby City	Derby UA
00FKNC	Boulton	77.5	0.78	75.9	79.0	Derby	Derby City	Derby UA
00FKND	Chaddesden	78.2	0.75	76.7	79.7	Derby	Derby City	Derby UA
00FKNE	Chellaston	78.3	0.79	77.7	80.8	Derby	Derby City	Derby UA
00FKNF	Darley	76.6	0.70	75.2	78.0	Derby	Derby City	Derby UA
00FKNG	Derwent	74.8	0.76	73.2	76.3	Derby	Derby City	Derby UA
00FKNH	Littleover	81.4	0.88	79.7	83.1	Derby	Derby City	Derby UA
00FKNJ	Mackworth	77.0	0.84	75.4	78.7	Derby	Derby City	Derby UA
00FKNK	Mickleover	80.7	0.75	79.2	82.1	Derby	Derby City	Derby UA

If you scroll down the columns to the right of your data (K to M), you may also notice some hidden cell content here, showing in a very faint font. To avoid problems with import, we suggest highlighting these columns and deleting this content too.

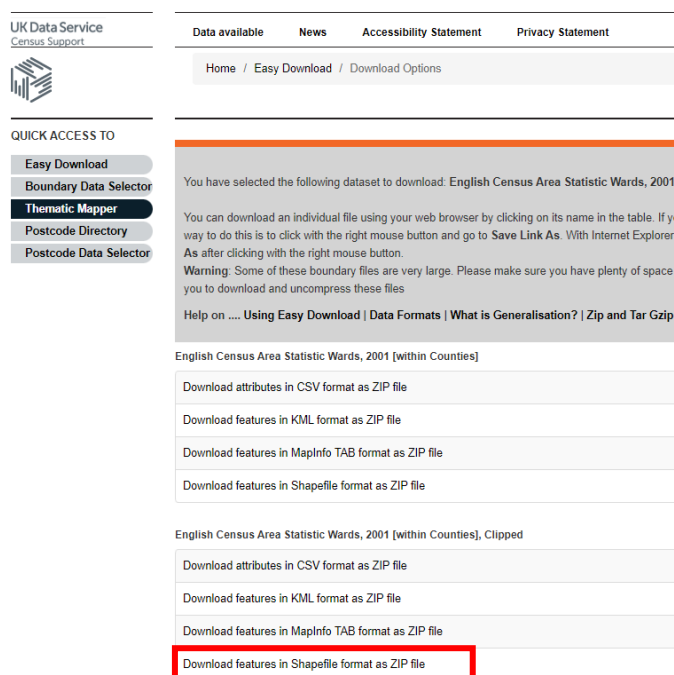
Finally, save this new file in Excel format as **malelifeexp** (again: no spaces in the file name to avoid potential problems in ArcGIS). If you look again at the 'metadata' worksheet, you will see that the area codes relate to English statistical (CAS) wards. What we will need to do next is to track down these boundaries. Close down your new spreadsheet file so that is no longer locked by MS-Excel and can be opened in ArcGIS. We also recommend heading for *file* and *save as* and saving the file in CSV (comma delimited) format rather than Excel format in doing so, which can sometimes be easier to load into ArcGIS. We suggest giving a different name to the CSV format file.

## Download the related boundary files

The related boundary files are available via the UK Borders site, a repository that provides administrative boundaries for the UK to the university sector. To access these boundaries, go to the web link for UK Borders: <https://borders.ukdataservice.ac.uk/> . Follow the 'easy download' link:



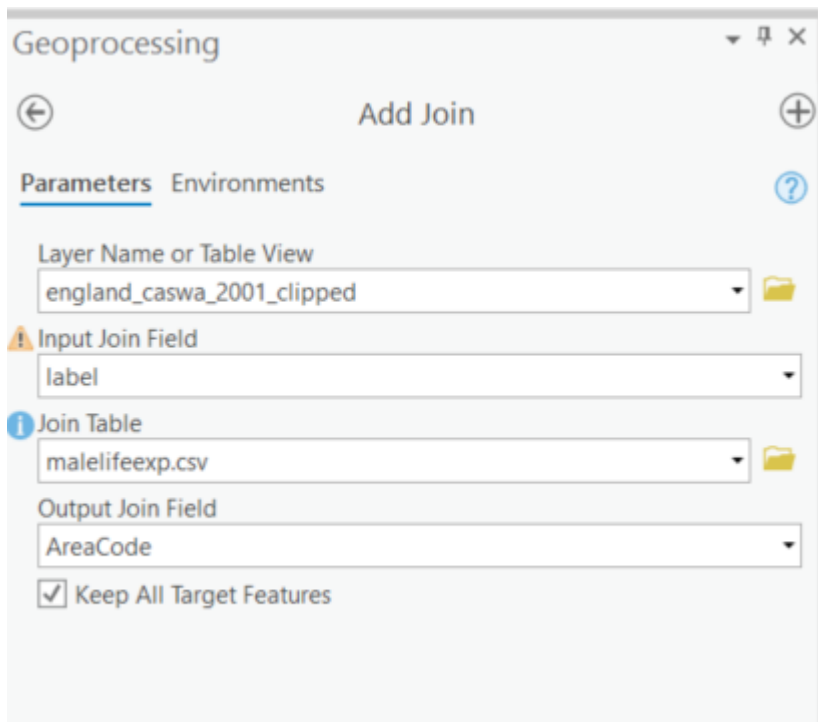
Recall that we require English CAS wards. These date from the 2001 census. If you scroll down through the list of available administrative boundaries, under '2001 boundaries', you should see a link to 'English CAS wards'. Follow this link and download the 'Clipped version' as a shapefile in zip format (clipped meaning that the boundaries have been cropped to the England coastline and shape file being an ArcGIS compatible format). After downloading the data, unzip the contents of the file and then open up the wards map layer in Excel.



## Link the spreadsheet to the boundaries

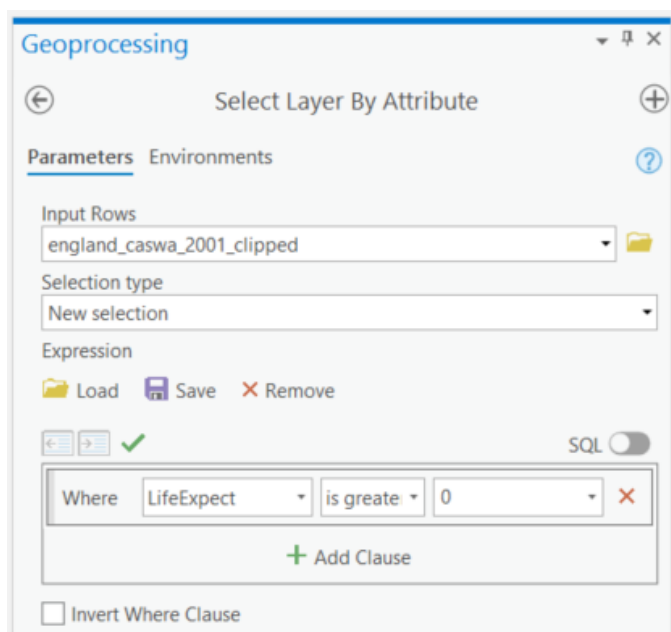
In ArcGIS PRO, open up your 'male summary' worksheet within the **malelifeexp** spreadsheet. You may want to right-click on it in the left-hand table of contents and choose *open* to check that it has imported correctly. Next, in the left-hand table of contents, click on your wards map layer, then choose *joins and relates* and then *add join* to link the spreadsheet data to the boundaries. To link

the life expectancy data to the boundaries, choose 'male summary' as the *join table* and enter **label** and **areacode** as the fields in the wards attribute table and life expectancy worksheet that containing matching codes for wards:



If you Run, you should find that the life expectancy data are joined to the attribute table for the wards. You can check this by right-clicking on the wards layer and selecting *open attribute table*.

Because we have only downloaded life expectancy data for the East Midlands, many rows will likely be filled with null values (missing data). To restrict our map layer only to the East Midlands, we can head for the *map* menu and choose *select by attributes*. If we then pick out just those rows where male life expectancy is greater than zero, that should select out our East Midlands study area:



If we now right-click on our map layer and choose data, then export features, we can now save a copy of the East Midlands polygons only, with the life expectancy data attached.



You should be able to apply this same process to map the other health-related data sets at the links provided, though of course you may need to select different boundary map layers, depending on the specific data set that you choose.

Optional: You can continue by styling your map.

