

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

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

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

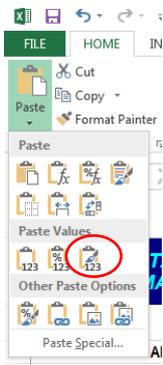
Mapping these additional data sets

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.

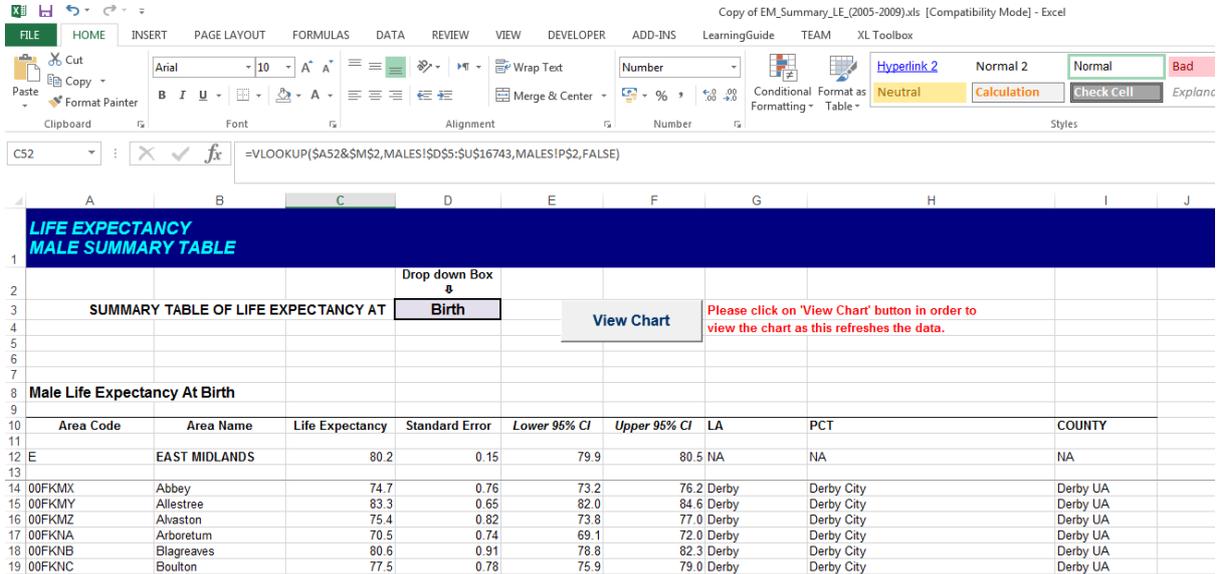
If you head for the 'male summary' worksheet and click on the values in each cell in the 'life expectancy' column, you will see by looking underneath the toolbars that these are based on an Excel formula (VLOOKUP). As we will need to delete some content from this spreadsheet, it is probably best to replace 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:



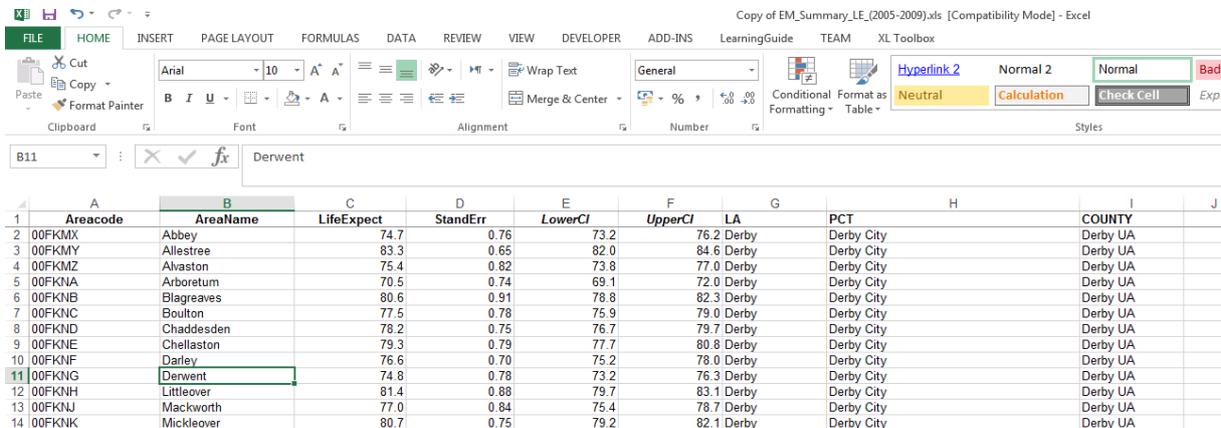
Click on any of the cells containing numbers, and now you should find that the VLOOKUP formula has been replaced with the calculated data value.

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.



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).



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.

[Download the related boundary files](#)

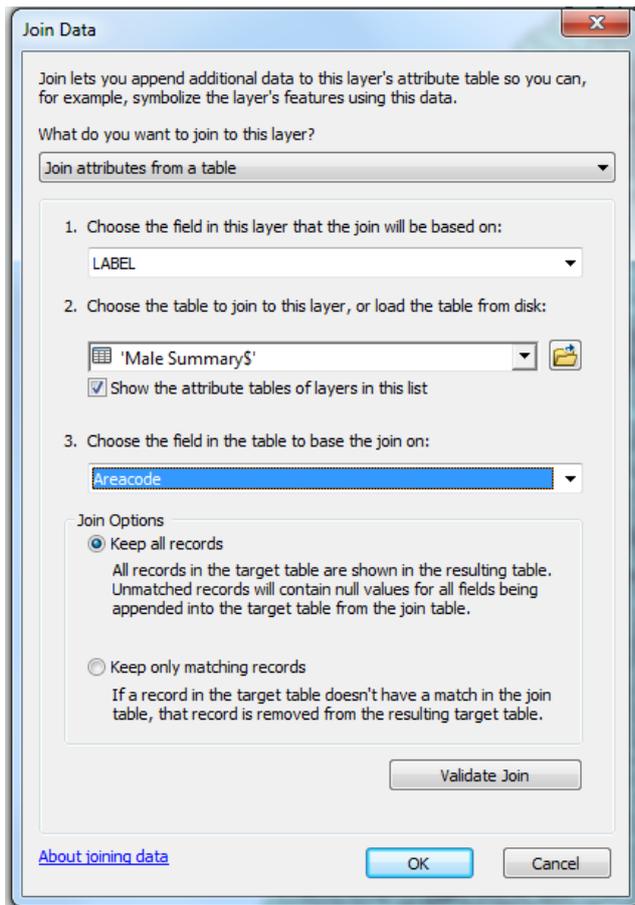
The related boundary files are available via the UK Borders site, a repository that provides administrative boundaries for the UK to the university sector. To access these boundaries, go to the web link for UK Borders: <https://borders.ukdataservice.ac.uk/>. Follow the 'easy download' link:

The screenshot shows a web application interface. On the left, there is a 'QUICK ACCESS TO' sidebar with buttons for 'Easy Download', 'Boundary Data', 'Thematic Mapper', 'Postcode Directory', 'Postcode Data', and 'View Metadata'. Below these is a dark blue button that says 'Hi there.'. The main content area is titled 'DOWNLOAD AND MAKE MAPS FROM UK BOUNDARY DATASETS.' and contains two sections. The first section, 'Boundary Data Selector', features a blue document icon and text: 'Download the boundaries you want, for the area you want, in the format you want. Please read the **Important information on Special Conditions**, below.' The second section, 'Easy Download', features a green download icon and text: 'Quick access to a set of the most frequently requested Census Support boundary datasets. Access is also provided to geographic look up tables. Please read the **Important information on Special Conditions**, below.'

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.

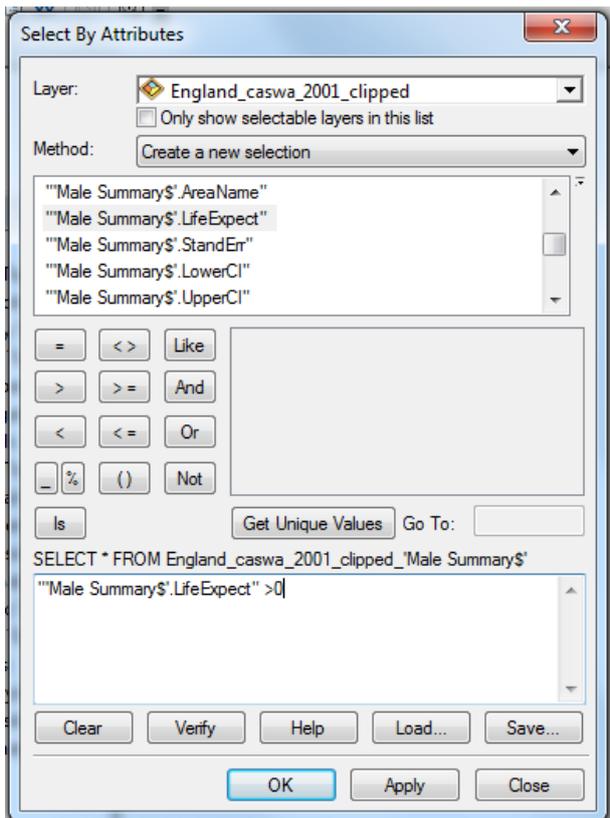
[Link the spreadsheet to the boundaries](#)

In ArcGIS, 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 table* 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 then *joins and relates* to link the spreadsheet data to the boundaries. To link the life expectancy data to the boundaries, choose 'male summary' as the *table to join to this layer* 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 then select OK, 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 *selection* 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 data, 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.