Mapping Census data in QGIS

UK Data Service

Contents

Contents

1. Introduction

Mapping census data requires three ingredients:

- census data for a given set of geographic areas (such as local authorities, wards or super output areas)
- boundary data for the same set of areas
- mapping software which brings these first two elements together so that the data can be mapped

This guide shows you how to create your own *choropleth* map which will show the percentage of males who work in the Manufacturing sector using the QGIS package. We will map the percentage of males who are working in Manufacturing in each local authority within the United Kingdom.

1.1. Census data

Traditional census outputs are aggregated counts, or other summary statistics, for a set of geographical areas. Because data are collected for the entire population these geographical areas can be very small. For example, it is possible to obtain a wide range of statistics for areas as small as *output areas* which are census areas containing approximately 300 persons.

1.2. Boundary data

Boundary data are the digital information about the polygons that describe the shape and location of the geographical units so that they can be mapped. We will be using boundaries in the form of a shape file which is actually a format created by Esri; it is a folder containing a number of inter-related files. These boundaries can be obtained from the UK Data Service.

1.3. Mapping software

QGIS is an open source mapping package which can be downloaded for free from www.qgis.org it has a good range of functionality and is straightforward to use. Similar commercial packages are available which could be used to undertake the same task, such as ARCGIS from http://www.esri.com/ or mapinfo from http://www.pitneybowes.com/uk.

Other approaches which do not involve using conventional desktop include:

- mapping data with a statistical package such as R
- mapping data using an online service such as Google Maps

Guides to mapping census data using these two approaches are also available as part of this guide series.

2. Obtain your census data

You are interested in finding out how many Males and Females work in the **Manufacturing** industry for all Local and Unitary Authorities in the United Kingdom.

Start

Go to: **infuse.ukdataservice.ac.uk**

- Click the ***2011 Census data***
- Not all census data are available at all geographical levels, you have a choice to pick Geography or Topics first
- click ***Geography*** to search by geography first

Step 1: Geography

In this step we can pick the areas for which you want data by using the + and - buttons to expand or collapse nested areas.

Note the geographies that are displayed are dependent on the categories you have selected. Not all data is available at all geographical levels.

Geography types containing large numbers of areas (Districts within the UK, for example) can be chosen, but not expanded, to avoid displaying huge lists. Individual areas for these geography types can be displayed by expanding within higher level geography types (Counties, for example).

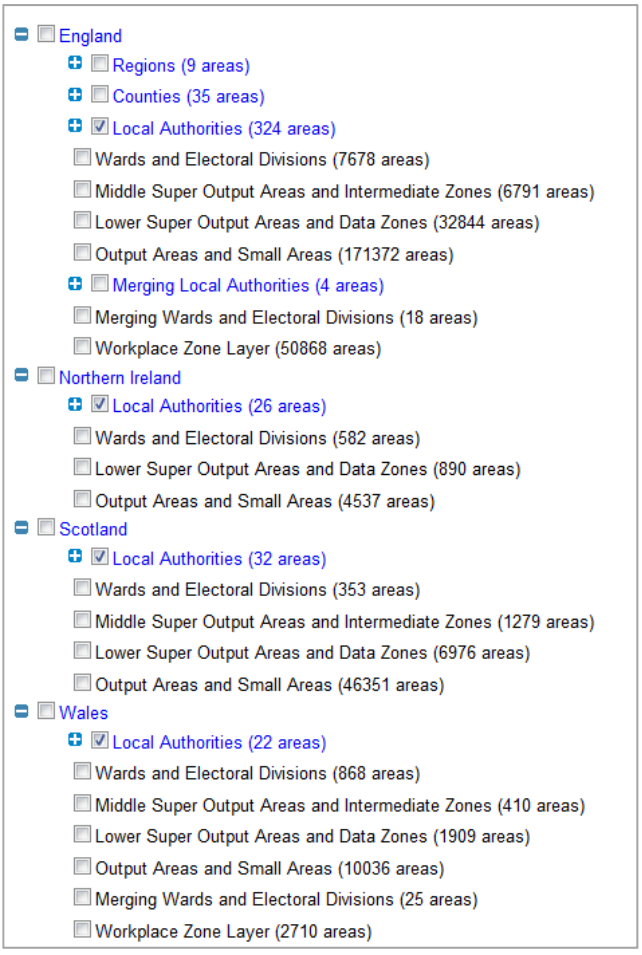
The Geography picker reflects the geographies for which data are released, so you may see that some units are available both separately and in combination.

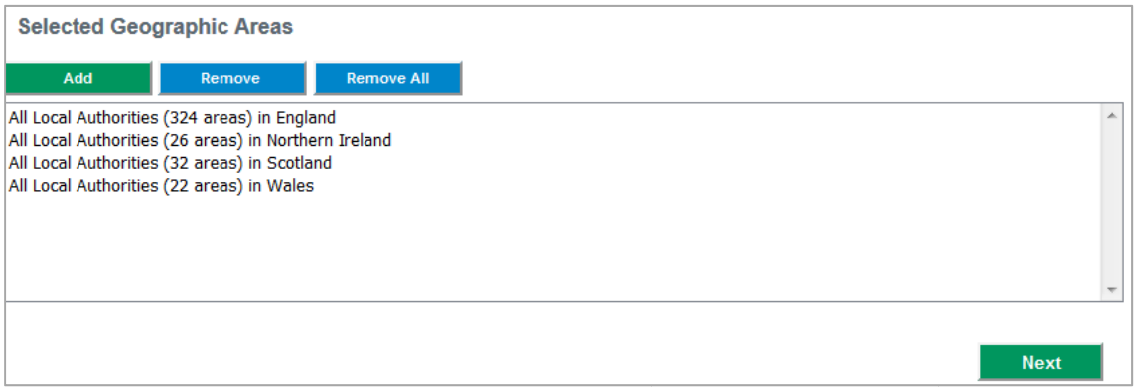
We would like all Local and Unitary Authorities in the UK (which in InFuse are referred to collectively as 'Local Authorities'). So expand:

- Tick ***Local Authorities*** for England, Northern Ireland, Scotland, and Wales

- Click ***Add***

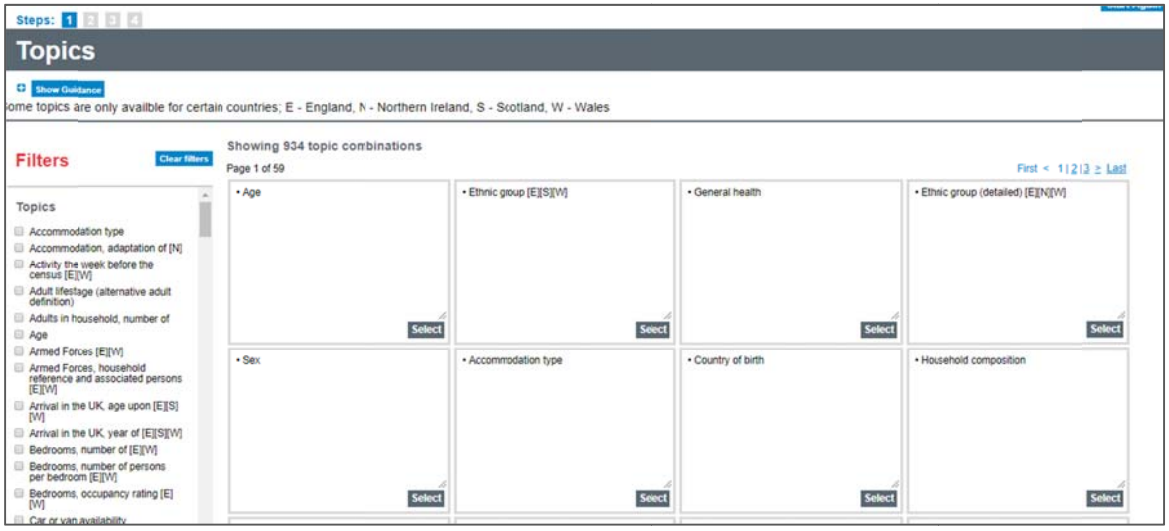
A summary of your selections will now be displayed, click ***Next***





Step 2: Topics

Census data comes in predefined topic combinations. These can be found in the grid.



On the left hand side of the screen you can filter the list.

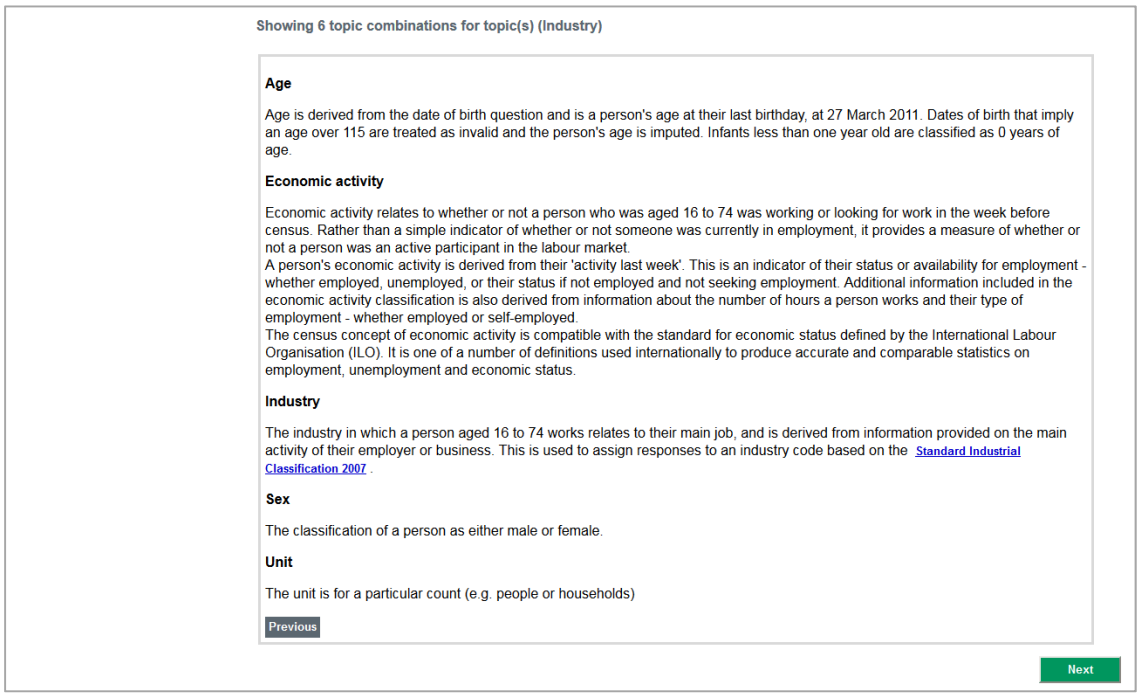
- Click ***Sex*** (notice that the amount of combinations have reduced to 18 topic combinations)
- Click ***Industry***

We now just have a single topic combination.

Click ***Select*** for the *Age* | *Economic activity* | *Industry* | *Sex* combination

You will now see another screen giving you the definitions of the topics you have selected. Have a read through, in particular the *industry* definition.

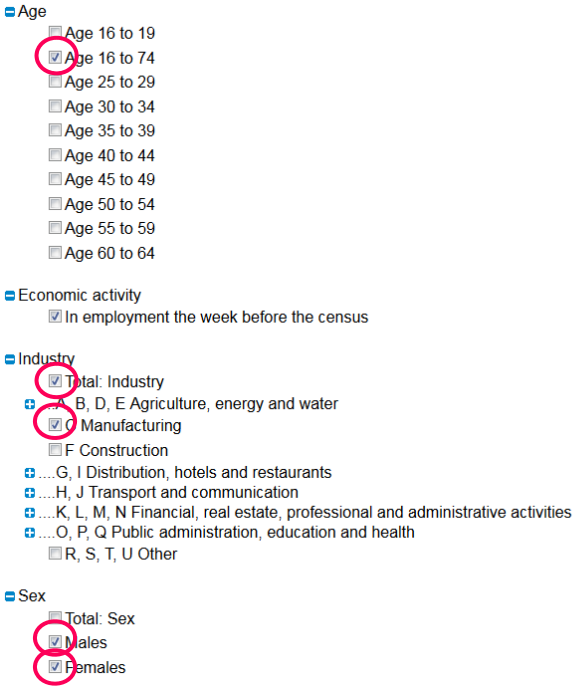
Click ***Next***



Step 3: Categories

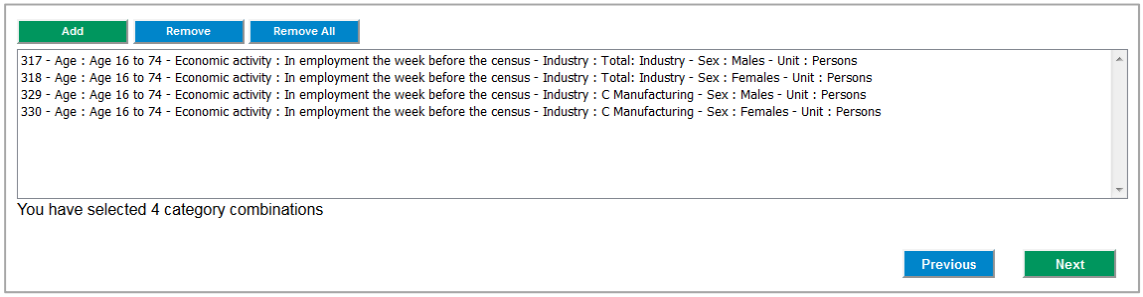
In this step you pick the categories to make up your combination. Items with a single category come preselected.

Age: Pick ***Age 16-74***
Industry: Pick ***Total: Industry*** and ***C Manufacturing***
Sex: Pick ***Male*** and ***Female***



Click ***Add***

A summary of your selections will now be displayed, click ***Next***

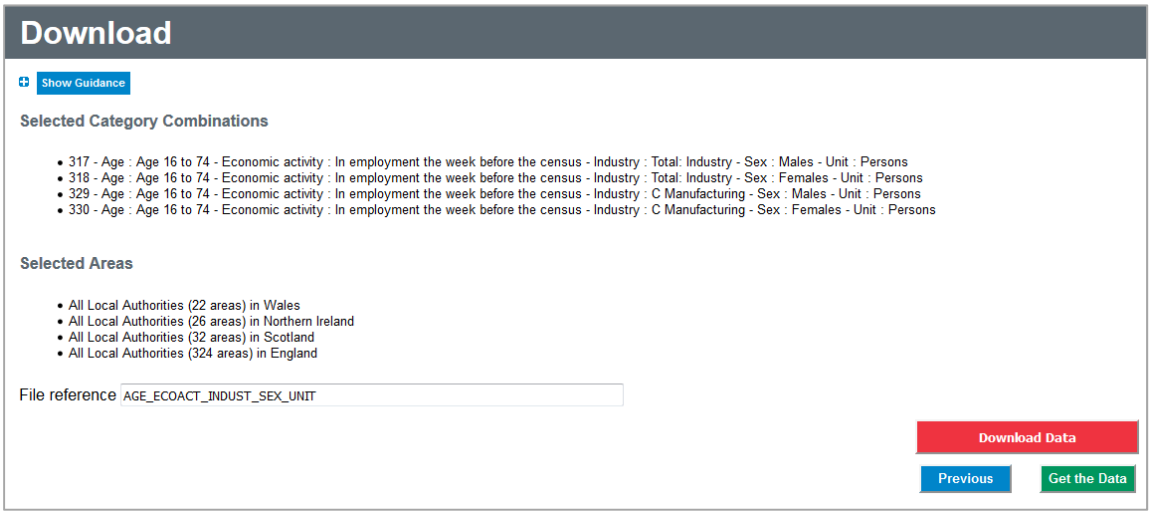


By selecting all persons with an industry as well as those in manufacturing, we will be able to use the former count to calculate the percentage of males and females within each LA who were in work in the week preceding the census who were working in manufacturing.

Step 4: Download

You can now see a summary of your selections along with a file reference (which can be changed).

Click ***Get the data***



An orange ***Download Data*** button will appear, click it and you are given the choice to open/save the zipped file as normal. Save this file to your working folder

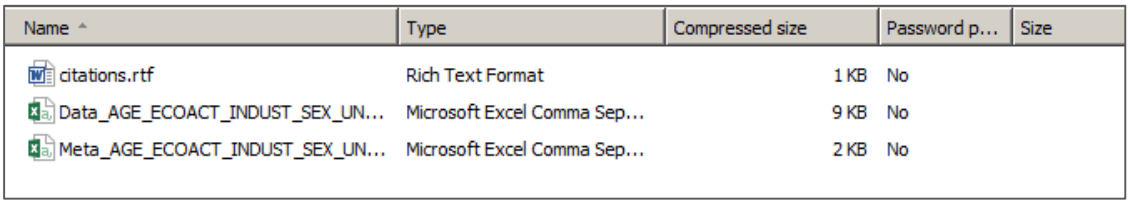
Results

The output comes in a Zip file format and will be named with a date stamp followed by the file reference (e.g.: 201842410383133_AGE_ECOACT_INDUST_SEX_UNIT)

The zip file consists of three files:

- Citation File in rtf format, simply tells you how to cite the data
- Meta file: This file gives you further information about the data you have chosen. It includes full definitions for the components of your topic combination
- Data file in csv format. This contains the data you have chosen

Unzip the file, and ensure you save to your working folder. The exact method you need to unzip your data will depend on your unzipping package and your computer set up. For example, you may find that you can unzip by right clicking the file and selecting *Extract All*.



Name	Type	Compressed size	Password p...	Size
citations.rtf	Rich Text Format	1 KB	No	
Data_AGE_ECOACT_INDUST_SEX_UN...	Microsoft Excel Comma Sep...	9 KB	No	
Meta_AGE_ECOACT_INDUST_SEX_UN...	Microsoft Excel Comma Sep...	2 KB	No	

Open up the data file and look at the results. Note that the second row contains the full label of the data you downloaded.

3. Formatting your data file

You should now have a file called: Data_AGE_ECOACT_INDUST_SEX_UNIT.csv

We need to modify this file slightly so QGIS will understand it. The second row shows the full category titles, we could abbreviate these. For example:

- In cell F1, delete F317 and insert Male_all
- In cell G1, delete F321 and insert Female_all
- In cell H1, delete F318 and insert Male_manuf
- In cell I1, delete F322 and insert Female_manuf

Now delete Row 2. Your file should now look like this:

	A	B	C	D	E	F	G	H	I	J
1	CDU_ID	GEO_COD	GEO_LABE	GEO_TYPE	GEO_TYP2	Male_all	Female_all	Male_manuf	Female_manuf	
2	52	95AA	Antrim	Local Auth	LA	13322	12273	1912	633	
3	53	95BB	Ards	Local Auth	LA	18220	17523	2265	557	
4	54	95CC	Armagh	Local Auth	LA	14128	12458	2183	731	
5	55	95DD	Ballymen	Local Auth	LA	15788	14132	3474	952	
6	56	95EE	Ballymon	Local Auth	LA	7171	6255	1320	325	
7	57	95FF	Banbridge	Local Auth	LA	11945	10676	1740	558	
8	58	95GG	Belfast	Local Auth	LA	59097	59657	5455	1451	
9	59	95HH	Carrickfer	Local Auth	LA	9189	8915	1463	347	
10	60	95II	Castlerea	Local Auth	LA	16055	16061	1858	446	

Visualising the data by the total count of people working in an industry may not be entirely useful on its own because it does not take into account the population of each Local Authority. When we come to map the data we will need to use a proportion. So let's create two extra columns and name them:

1. Male_man_p
2. Female_man_p

And create a percentage using the other columns:

- In J2 type the function =(H2/F2)*100
- And copy this function down the column to calculate the percentage of male workers in each local authority who are in manufacturing
- In K2 type the function = (I2/G2)*100 and copy this function down the column to calculate the equivalent percentage for females

Your data should now look like this:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	CDU_ID	GEO_COD	GEO_LABE	GEO_TYPE	GEO_TYP2	Male_all	Female_all	Male_manuf	Female_manuf	Male_man_P	Female_man_P		
2	52	95AA	Antrim	Local Auth	LA	13322	12273	1912	633	14.35219937	5.157663163		
3	53	95BB	Ards	Local Auth	LA	18220	17523	2265	557	12.43139407	3.17867945		
4	54	95CC	Armagh	Local Auth	LA	14128	12458	2183	731	15.4515855	5.867715524		
5	55	95DD	Ballymena	Local Auth	LA	15788	14132	3474	952	22.00405371	6.736484574		
6	56	95EE	Ballymon	Local Auth	LA	7171	6255	1320	325	18.40747455	5.195843325		
7	57	95FF	Banbridge	Local Auth	LA	11945	10676	1740	558	14.56676434	5.226676658		
8	58	95GG	Belfast	Local Auth	LA	59097	59657	5455	1451	9.230587001	2.432237625		
9	59	95HH	Carrickfer	Local Auth	LA	9189	8915	1463	347	15.92121014	3.892316321		
10	60	95II	Castlerea	Local Auth	LA	16055	16061	1858	446	11.57271878	2.776913019		
11	61	95JJ	Coleraine	Local Auth	LA	13122	12150	1684	488	12.83340954	4.016460905		
12	62	95KK	Cookstow	Local Auth	LA	8660	7336	1959	616	22.62124711	8.396946565		
13	63	95LL	Craigavon	Local Auth	LA	21615	19705	4926	1813	22.78972935	9.20071048		
14	64	95MM	Derry	Local Auth	LA	20649	19963	2870	697	13.89897816	3.4914592		
15	65	95NN	Down	Local Auth	LA	15918	14992	1398	457	8.782510366	3.048292423		
16	66	95OO	Dungannc	Local Auth	LA	13955	11691	3608	1396	25.85453243	11.94080917		
17	67	95PP	Fermanag	Local Auth	LA	14749	12825	2114	752	14.33317513	5.863547758		
18	68	95QQ	Larne	Local Auth	LA	7695	7166	1370	450	17.80376868	6.279653921		
19	69	95RR	Limavady	Local Auth	LA	7010	6207	887	232	12.65335235	3.737715483		
20	70	95SS	Lisburn	Local Auth	LA	28211	26751	3607	1130	12.78579278	4.224141154		
21	71	95TT	Magherafe	Local Auth	LA	11010	8973	1967	543	17.86557675	6.051487797		
22	72	95UU	Moyle	Local Auth	LA	3793	3228	386	105	10.17664118	3.252788104		
23	73	95VV	Newry an	Local Auth	LA	22034	19429	3210	1323	14.5683943	6.809408616		
24	74	95WW	Newtown	Local Auth	LA	20151	19818	2872	692	14.25239442	3.491775154		
25	75	95XX	North Dov	Local Auth	LA	18887	18029	1963	489	10.39339228	2.712296855		
26	76	95YY	Omagh	Local Auth	LA	11603	10073	1662	377	14.32388175	3.742678447		
27	77	95ZZ	Strabane	Local Auth	LA	8153	6847	1045	314	12.81736784	4.585950051		
28	78	E06000001	Hartlepoo	Local Auth	LA	19511	18256	3656	754	18.73814771	4.130148992		

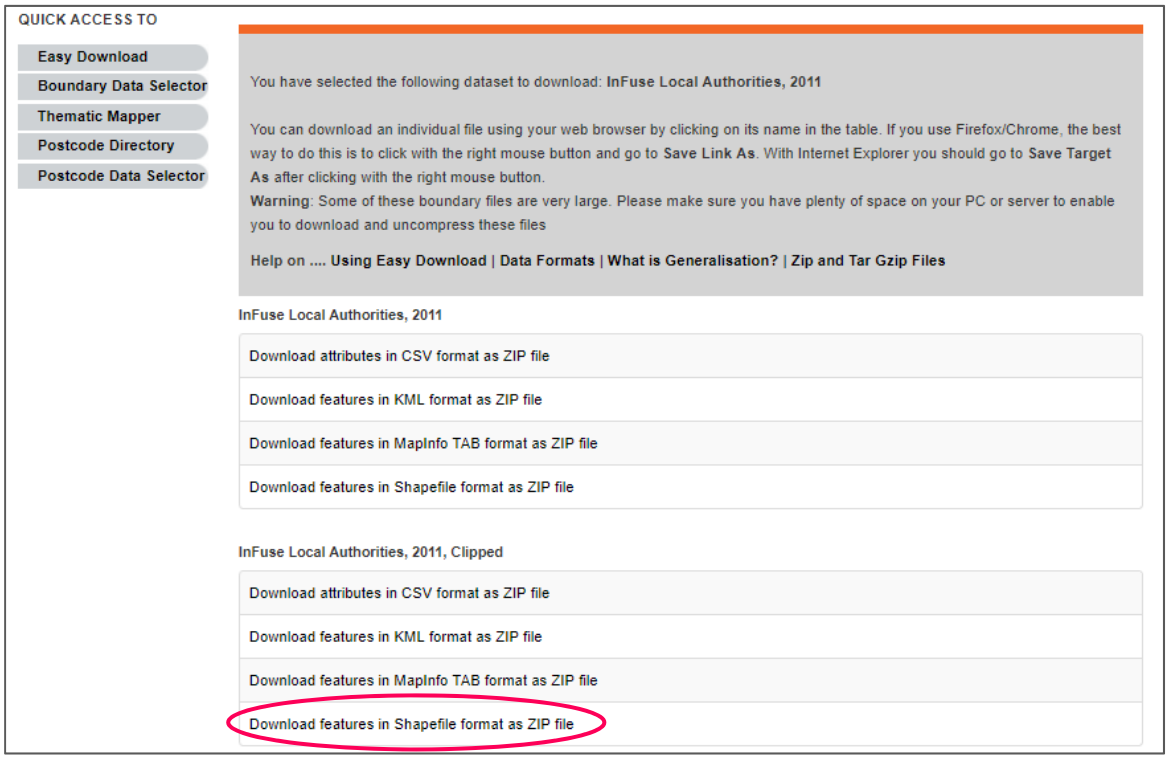
Save the file to your folder (in .csv format).

4. Downloading boundary data

1. Go to: **http://census.ukdataservice.ac.uk/** and go to:
2. **Get census data** > **Boundary data** > **Easy download**
3. Click the ***InFuse* United Kingdom** tab
4. Click ***InFuse Local Authorities, 2011***
5. Under the ***InFuse 2011 Local Authorities, Clipped*** click ***Download features in Shapefile format as ZIP file***

6. This will start to download the boundary dataset. Save the Zip file to your computer in the folder you created earlier – it will be named by default as Infuse_dist_lyr_2011_clipped

7. Navigate to the zip file and right click and select ***Extract all*** to unzip the files, ensure that you know the location of these files

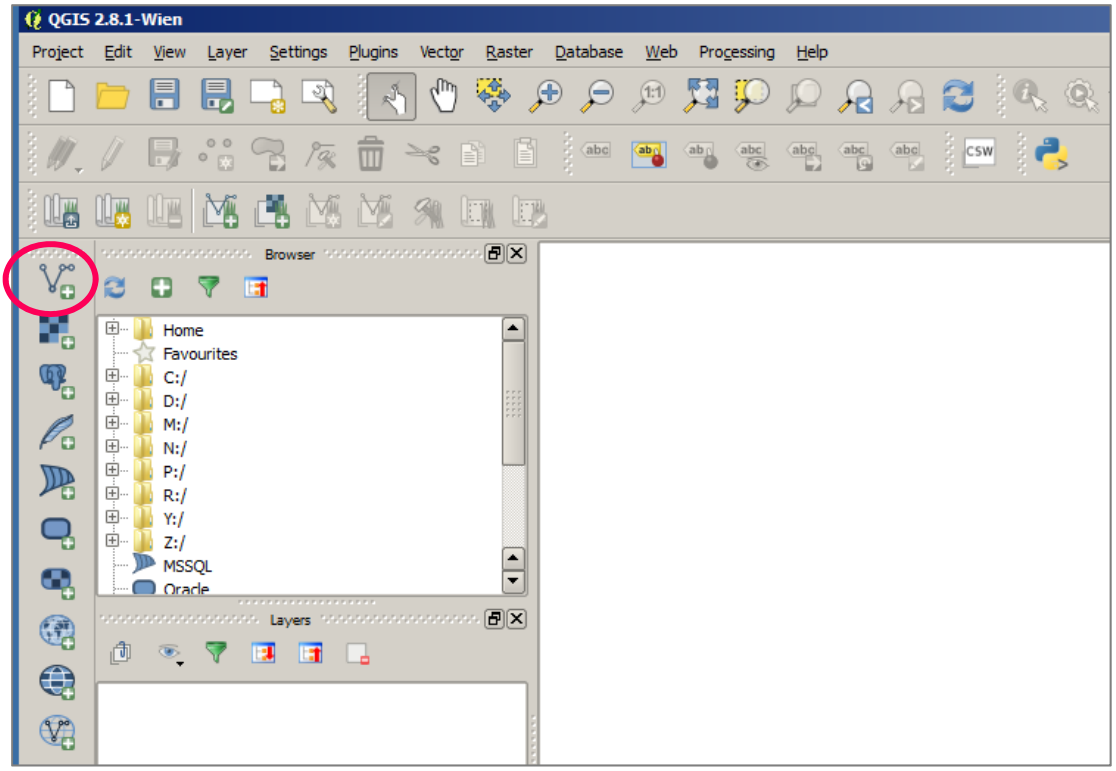


5. Matching data in QGIS

We can now map the census aggregate data we downloaded from InFuse and combine it with the census boundary data.

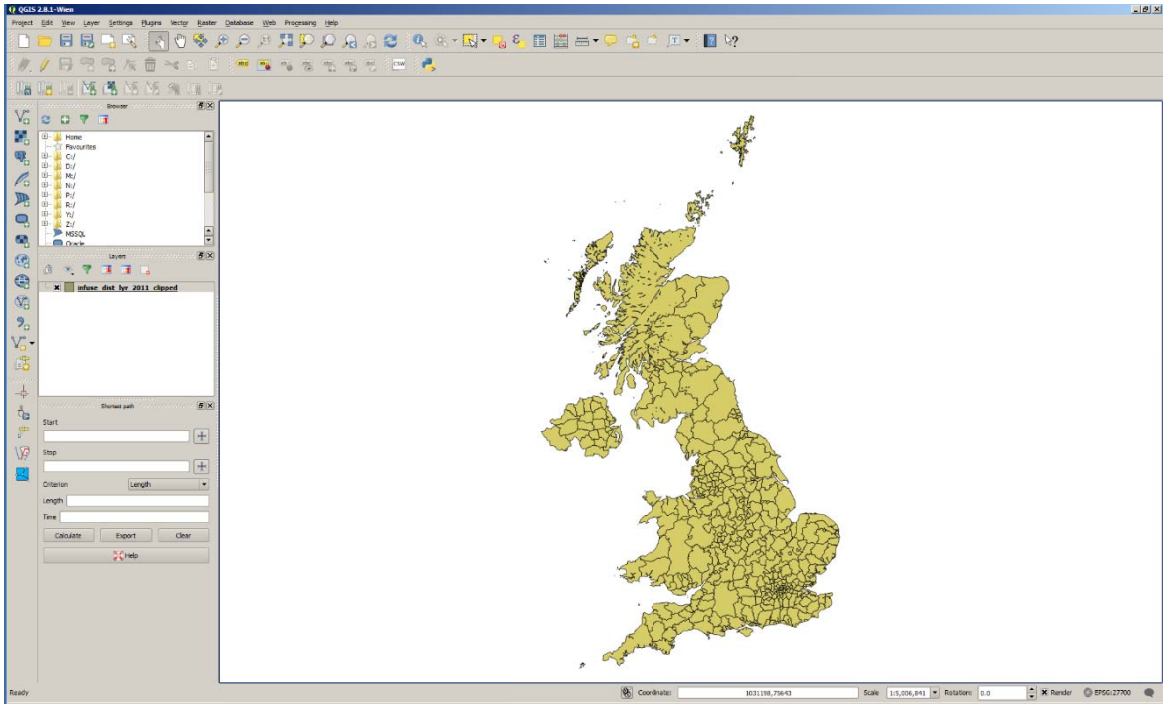
1. Open QGIS Desktop (the following instructions are based on version 2.8.9)

2. Click the ***Add Vector Layer*** icon



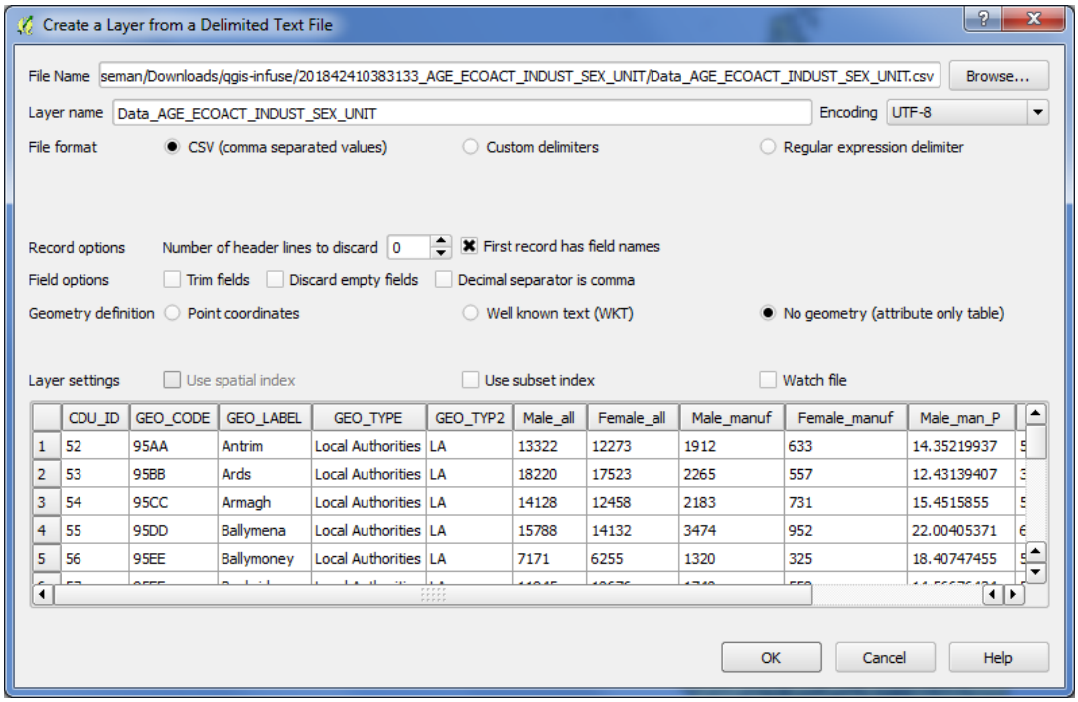
3. Browse to the boundary .shp file data that you downloaded earlier (i.e. infuse_dist_lyr_2011_clipped.shp) and click open

4. You will now see an outline of the UK



5. We have added the boundaries, now we need to add the aggregate data that we downloaded from InFuse. Click the ADD Delimited Text Layer icon  in the left hand menu of the screen

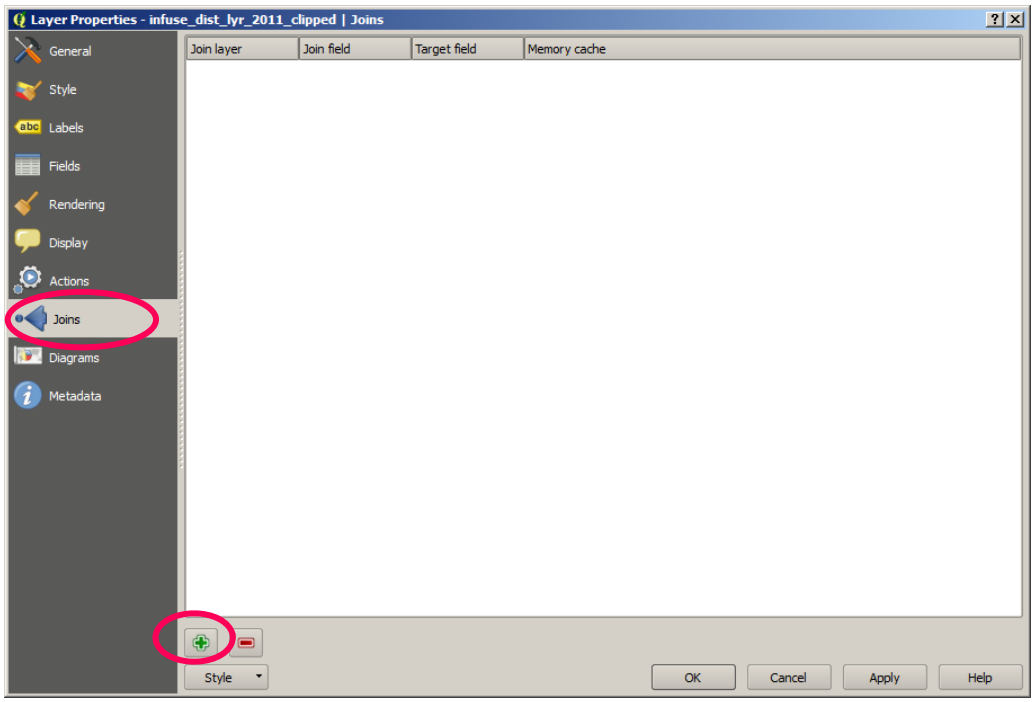
6. Browse for the Data_AGE_ECOACT_INDUST_SEX_UNIT.csv file and select csv for the file format and ***no geometry (attribute table only)***, then click ***OK***



7. We now need to join these two layers together

8. Double-click the *infuse_dist_lyr_2011_clipped* layer in the layer box on the right hand side

9. Click ***joins***, and tick the green + symbol



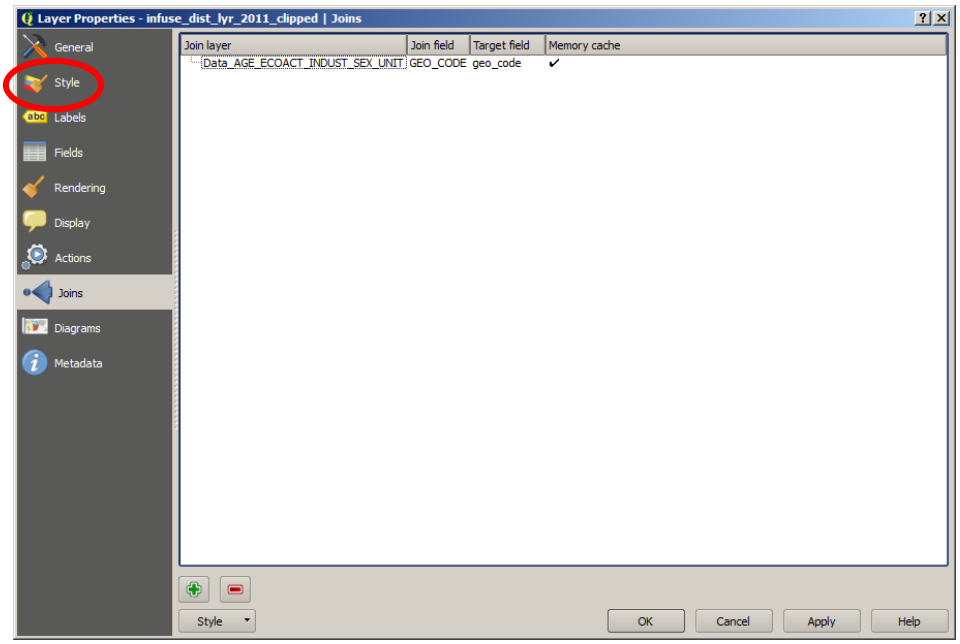
10. **Select GEO_CODE as both the join field and target field** click ***OK*** then ***Apply*** and then ***OK*** again

11. The ***Open Attribute Table*** icon  on the top menu bar allows you to see all of the data in table form. The geo_code fields in both tables are identical; this is how QGIS knows which Local Authority boundary belongs to which Local Authority data

6. Visualising data in QGIS

In this task we will create a choropleth map in QGIS.

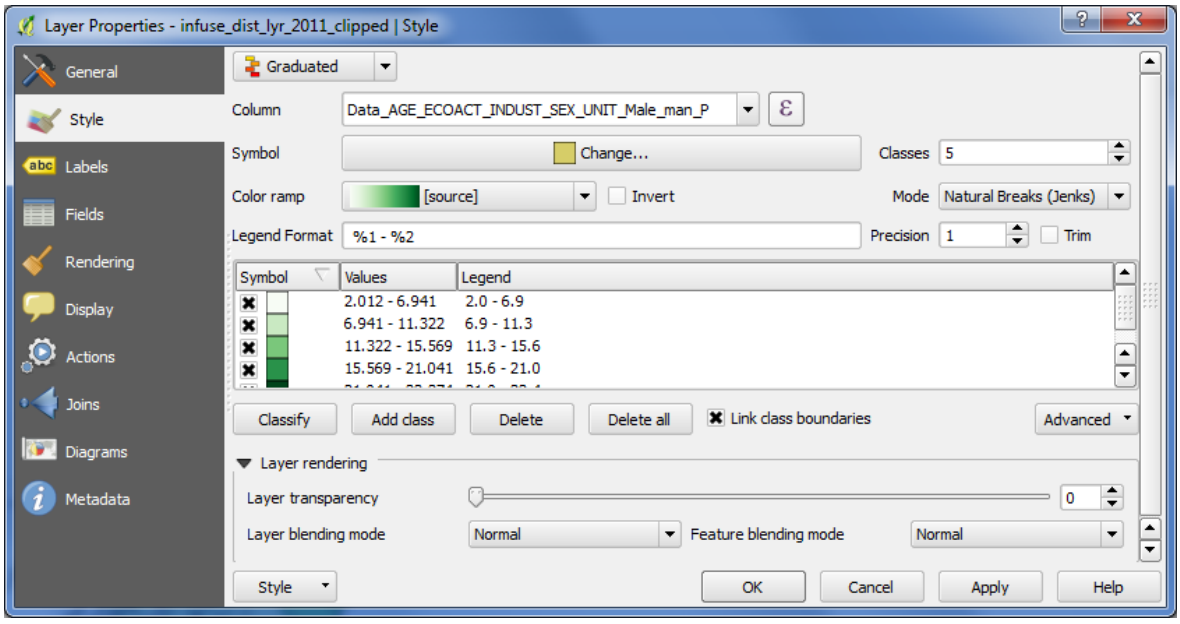
1. Double-click on the *infuse_dist_lyr_2011_clipped* layer in the right hand layer panel. This will bring up the layer properties box



2. Click ***Style***, you have a few choices. In the top dropdown menu, click ***Graduated.*** The colour ramp picker allows you to choose the colours to be used to represent the percentages; you can change this if you wish

3. We can now select what data we want to be displayed on the map. Earlier we created two percentages columns, select one of these from the column dropdown. In this example I have selected ***Male_man_p***

4. Click on *Classify* and you can see that 5 bands appear. You can change the number of these by changing the number in the *classes'* box. There are a few options for *mode*; we would recommend using "natural breaks (Jenks)". Change these options as you wish, and then click Apply and OK



5. You will now have a map, looking something like this:

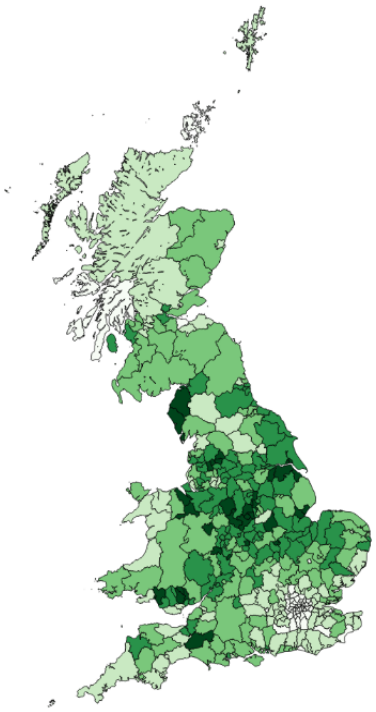
6. There are a lot of things you can do but some basic things would be to create a map to be printed, change the size, add a label and add a legend

7. To format your map for printing go to ***Project > New Print Composer***, and insert a title. You can change the size of the page by clicking the composition tab on the right hand side of the screen

8. If you want to, change the page size using the panel on the right hand size (composition tab). It makes sense to leave the page size as A4, but to change the orientation to portrait

9. To insert the map, click ***Layout > Add map***, and drag your mouse across the page

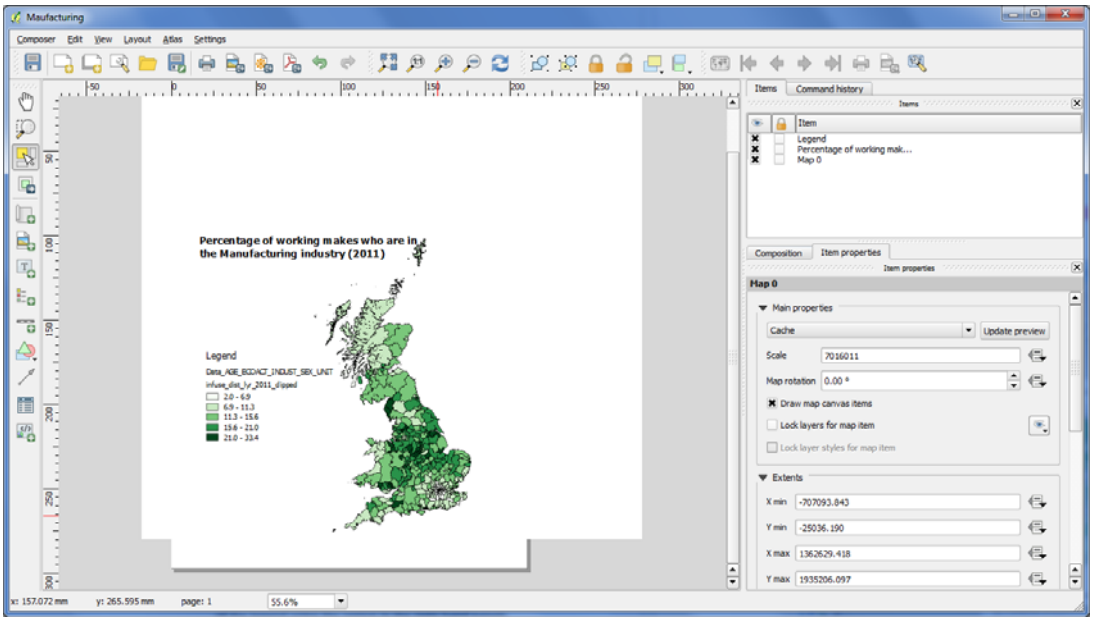
10. Adjust the size. You can do this by clicking the arrows around the map, but there is a lot of extra white space. Go to ***Layout > Move item***, and ***Layout > Move content*** to adjust the size and scale of the map. While using ***Layout > Move content*** you can use your mouse's scroll wheel to adjust the size of the map. Or use the options within the ***item properties*** tab on the right if the screen (for example you can specify scale, the example map is set to 35000)



11. Insert a label. Go to ***Layout > Add Label***, and click and move your cursor on the map. Type in the right hand side box. You can also adjust the size and font. Add a title and the citation information which you can find at: census.ukdataservice.ac.uk/use-data/citing-data

12. Insert a Legend. Go to ***Layout > Add Legend***. You can modify some aspect of the legend using the option in the right hand panels

This map can be saved as a project to allow further editing. It can also be exported as a .svg, .png or .pdf file for publication.



Useful links

- UK Data Service – Census Support: census.ukdataservice.ac.uk
- Get data page: census.ukdataservice.ac.uk/get-data
- Use data page: census.ukdataservice.ac.uk/use-data
- InFuse: infuse.ukdataservice.ac.uk
- Casweb: casweb.ukdataservice.ac.uk
- Boundary data: census.ukdataservice.ac.uk/get-data/boundary-data
- Census forms:

census.ukdataservice.ac.uk/use-data/censuses/forms

- Census definitions:

census.ukdataservice.ac.uk/use-data/censuses/definitions

- Datashine: datashine.org.uk
- QGIS: www.qgis.org/

T +44 (0) 1206 872143
E help@ukdataservice.ac.uk
W ukdataservice.ac.uk

The UK Data Service delivers quality social and economic data resources for researchers, teachers and policymakers.

UK Data Service

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL