Creating a choropleth map from census data in QGIS.

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***

England
- Regions (9 areas)
- Counties (35 areas)
- ☑ Local Authorities (324 areas)
- Wards and Electoral Divisions (7678 areas)
- Middle Super Output Areas and Intermediate Zones (6791 areas)
- Lower Super Output Areas and Data Zones (32844 areas)
- Output Areas and Small Areas (171372 areas)
- Merging Local Authorities (4 areas)
- Merging Wards and Electoral Divisions (18 areas)
- Workplace Zone Layer (50868 areas)

Northern Ireland
- ☑ Local Authorities (26 areas)
- Wards and Electoral Divisions (582 areas)
- Lower Super Output Areas and Data Zones (890 areas)
- Output Areas and Small Areas (4537 areas)

Scotland
- ☑ Local Authorities (32 areas)
- Wards and Electoral Divisions (353 areas)
- Middle Super Output Areas and Intermediate Zones (1279 areas)
- Lower Super Output Areas and Data Zones (6976 areas)
- Output Areas and Small Areas (46351 areas)

Wales
- ☑ Local Authorities (22 areas)
- Wards and Electoral Divisions (868 areas)
- Middle Super Output Areas and Intermediate Zones (410 areas)
- Lower Super Output Areas and Data Zones (1909 areas)
- Output Areas and Small Areas (10036 areas)
- Merging Wards and Electoral Divisions (25 areas)
- Workplace Zone Layer (2710 areas)

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

Filters   Clear filters

1 topic combination(s) found for topic(s) (Industry, Sex)

Page 1 of 1

Topics

- [ ] Accommodation type
- [ ] Accommodation, adaptation of [N]
- [ ] Activity the week before the census [E][W]
- [ ] Adult lifestage (alternative adult definition)
- [ ] Adults in household, number of
- [x] Age
- [ ] Armed Forces [E][W]
- [ ] Armed Forces, household reference and associated

- Age
- Economic activity
- Industry
- Sex

Select

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**

Showing 6 topic combinations for topic(s) (Industry)

Age

Age is derived from the date of birth question and is a person's age at their last birthday, at 27 March 2011. Dates of birth that imply an age over 115 are treated as invalid and the person's age is imputed. Infants less than one year old are classified as 0 years of age.

Economic activity

Economic activity relates to whether or not a person who was aged 16 to 74 was working or looking for work in the week before census. Rather than a simple indicator of whether or not someone was currently in employment, it provides a measure of whether or not a person was an active participant in the labour market.
A person's economic activity is derived from their 'activity last week'. This is an indicator of their status or availability for employment - whether employed, unemployed, or their status if not employed and not seeking employment. Additional information included in the economic activity classification is also derived from information about the number of hours a person works and their type of employment - whether employed or self-employed.
The census concept of economic activity is compatible with the standard for economic status defined by the International Labour Organisation (ILO). It is one of a number of definitions used internationally to produce accurate and comparable statistics on employment, unemployment and economic status.

Industry

The industry in which a person aged 16 to 74 works relates to their main job, and is derived from information provided on the main activity of their employer or business. This is used to assign responses to an industry code based on the Standard Industrial Classification 2007 .

Sex

The classification of a person as either male or female.

Unit

The unit is for a particular count (e.g. people or households)

Previous

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 **Males** and **Females**

- Age
  - ☐ Age 16 to 19
  - ☑ Age 16 to 74
  - ☐ Age 25 to 29
  - ☐ Age 30 to 34
  - ☐ Age 35 to 39
  - ☐ Age 40 to 44
  - ☐ Age 45 to 49
  - ☐ Age 50 to 54
  - ☐ Age 55 to 59
  - ☐ Age 60 to 64

- Economic activity
  - ☑ In employment the week before the census

- Industry
  - ☑ Total: Industry
  - ☐ A, B, D, E Agriculture, energy and water
  - ☑ C Manufacturing
  - ☐ F Construction
  - ☐ ....G, I Distribution, hotels and restaurants
  - ☐ ....H, J Transport and communication
  - ☐ ....K, L, M, N Financial, real estate, professional and administrative activities
  - ☐ ....O, P, Q Public administration, education and health
  - ☐ R, S, T, U Other

- Sex
  - ☐ Total: Sex
  - ☑ Males
  - ☑ Females

Click **Add**

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

Add	Remove	Remove All

317 - Age : Age 16 to 74 - Economic activity : In employment the week before the census - Industry : Total: Industry - Sex : Males - Unit : Persons
318 - Age : Age 16 to 74 - Economic activity : In employment the week before the census - Industry : Total: Industry - Sex : Females - Unit : Persons
329 - Age : Age 16 to 74 - Economic activity : In employment the week before the census - Industry : C Manufacturing - Sex : Males - Unit : Persons
330 - Age : Age 16 to 74 - Economic activity : In employment the week before the census - Industry : C Manufacturing - Sex : Females - Unit : Persons

You have selected 4 category combinations

Previous   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**

Download

Show Guidance

Selected Category Combinations

- 317 - Age : Age 16 to 74 - Economic activity : In employment the week before the census - Industry : Total: Industry - Sex : Males - Unit : Persons
- 318 - Age : Age 16 to 74 - Economic activity : In employment the week before the census - Industry : Total: Industry - Sex : Females - Unit : Persons
- 329 - Age : Age 16 to 74 - Economic activity : In employment the week before the census - Industry : C Manufacturing - Sex : Males - Unit : Persons
- 330 - Age : Age 16 to 74 - Economic activity : In employment the week before the census - Industry : C Manufacturing - Sex : Females - Unit : Persons

Selected Areas

- All Local Authorities (22 areas) in Wales
- All Local Authorities (26 areas) in Northern Ireland
- All Local Authorities (32 areas) in Scotland
- All Local Authorities (324 areas) in England

File reference | AGE_ECOACT_INDUST_SEX_UNIT |

Download Data

Previous **Get the Data**

A purple **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 F318 and insert Female_all
- In cell H1, delete F329 and insert Male_manuf
- In cell I1, delete F330 and insert Female_manuf

Now delete Row 2.

The GEO_CODE column contains geographical identifier codes that uniquely identifes the census geography instance which the statistics in each record is associated with. Most of these codes are provided as a 9 digit code. The exception are the codes for the local authorities in Northern Ireland. These are all provided as a 4 digit code starting with 95. In the CSV that was opened in Excel, these contain trailing spaces that will cause problems later so lets get rid of these trailing spaces.

- In J2 type the function =TRIM(B2)
- Hold down the CTRL key, click on cell J2 and then drag down to cell J27
- Then select all cells from J2 to J27, right-click >> Copy
- Then select all cells from B2 to B27, right-click >> Paste-Special, change to Values and click OK
- Now delete all cells in Column J

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	Ballymena	Local Auth	LA	15788	14132	3474	952	
6	56	95EE	Ballymone	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	95IJ	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. Enter in cell J1: Male_man_p
2. Enter in cell K1: 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	Ballymone	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	Castlereag	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 and	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

QUICK ACCESS TO

- Easy Download
- Boundary Data Selector
- Thematic Mapper
- Postcode Directory
- Postcode Data Selector

You have selected the following dataset to download: **InFuse Local Authorities, 2011**

You can download an individual file using your web browser by clicking on its name in the table. If you use Firefox/Chrome, the best way to do this is to click with the right mouse button and go to **Save Link As**. With Internet Explorer you should go to **Save Target As** after clicking with the right mouse button.
Warning: Some of these boundary files are very large. Please make sure you have plenty of space on your PC or server to enable you to download and uncompress these files

Help on .... Using Easy Download | Data Formats | What is Generalisation? | Zip and Tar Gzip Files

InFuse Local Authorities, 2011

Download attributes in CSV format as ZIP file

Download features in KML format as ZIP file

Download features in MapInfo TAB format as ZIP file

Download features in Shapefile format as ZIP file

InFuse Local Authorities, 2011, Clipped

Download attributes in CSV format as ZIP file

Download features in KML format as ZIP file

Download features in MapInfo TAB format as ZIP file

Download features in Shapefile format as ZIP file

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 3.16.0)

2. QGIS defaults to displaying data using a global coordinate system. To avoid complications before opening any of our UK census data in QGIS we will tell QGIS to display data using the local British National Grid projected coordinate system. At the bottom right hand corner of the main QGIS window, the coordinate system will currently be set to EPSG:4326. EPSG:4326 is a shorthand code for the global coordinate system WGS84. Click on EPSG:4326
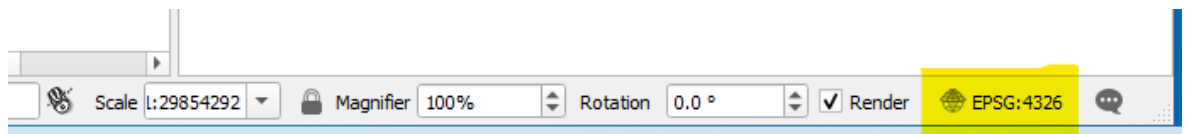
The Project Properties – CRS dialogue will be shown.

In the Filter box type in "British National Grid"

Under Predefined Coordinate Reference Systems, click on

OSGB 1936 / British National Grid EPSG:27700
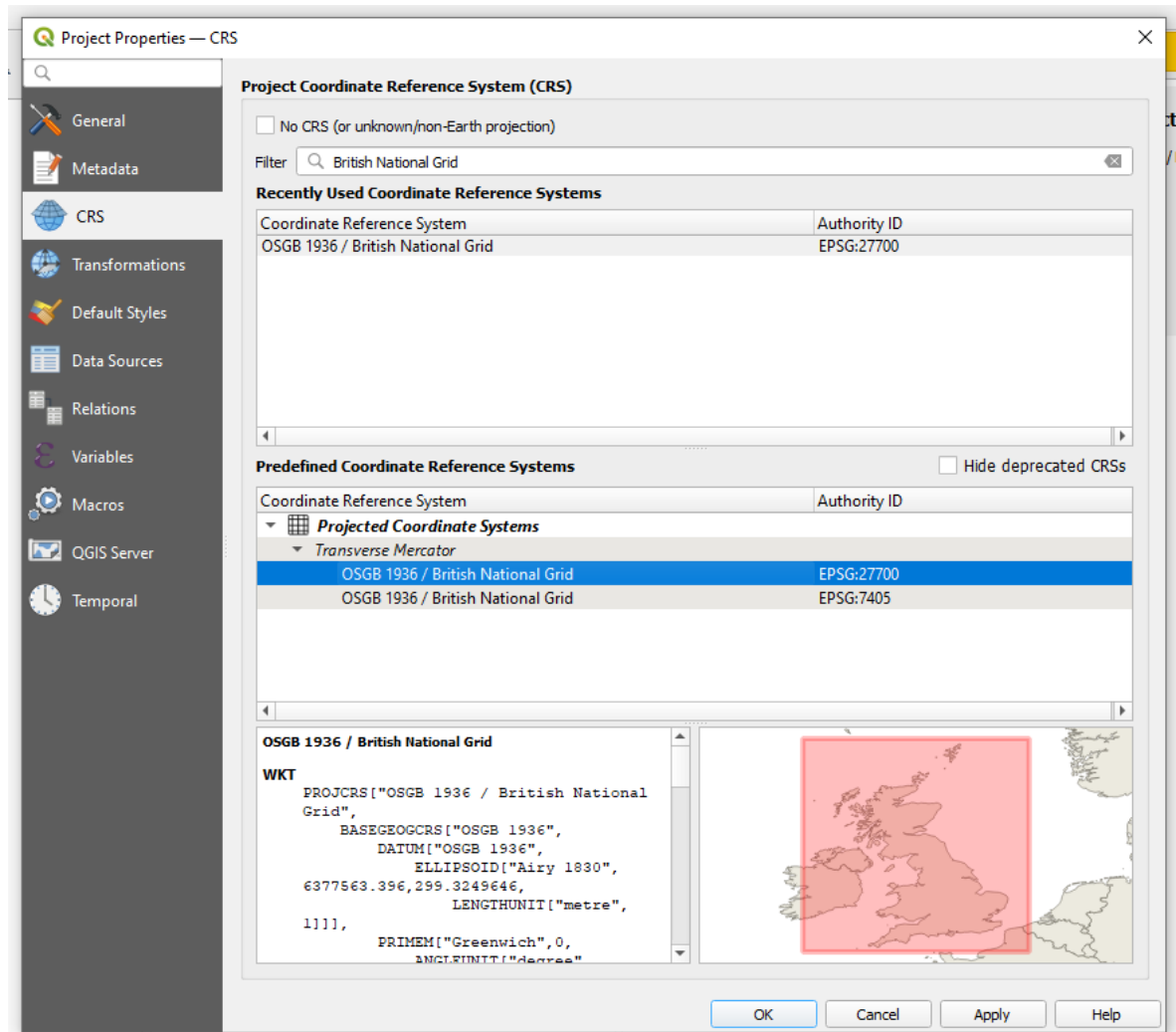
so that it is selected and click the Apply button followed by the OK button.

The status bar should update to show that the default coordinate system that QGIS will use to display data has been set to EPSG: 27700
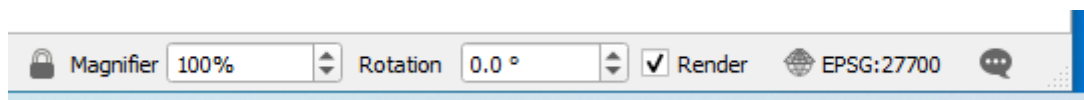
3. Click the **Open Data Source Manager** icon
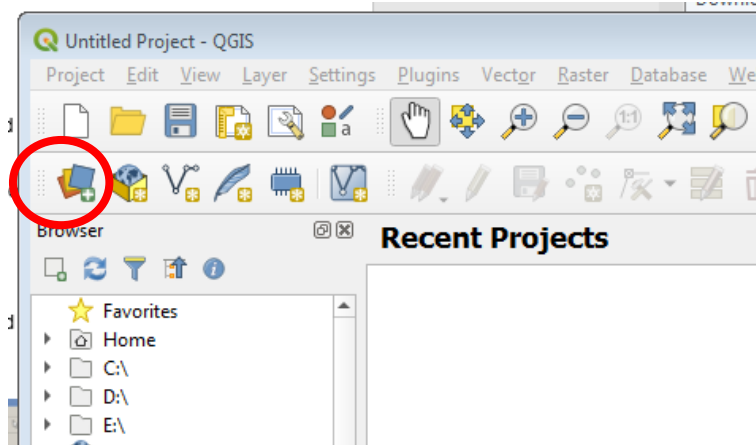
4. This will open the Data Source Manager dialogue which allows you to add datasets to QGIS. From the types of data list at left, click on Vector and then click on the […] button.
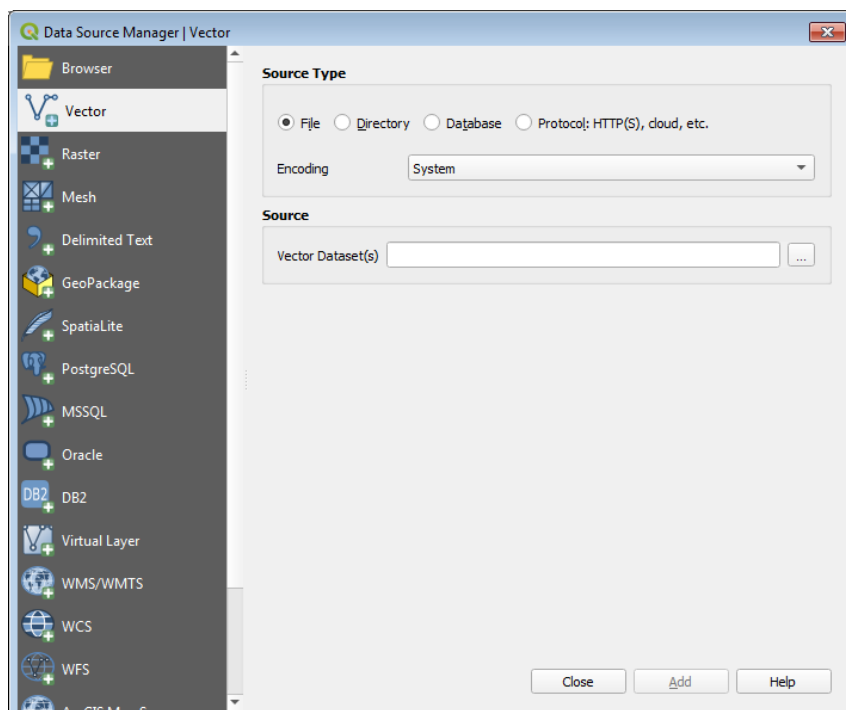

5. Browse to the boundary .shp file data that you downloaded earlier (i.e. infuse_dist_lyr_2011_clipped.shp) and click open, then Add. Then click Close to close the Data Source Manager dialogue.

6. You will now see an outline of the UK
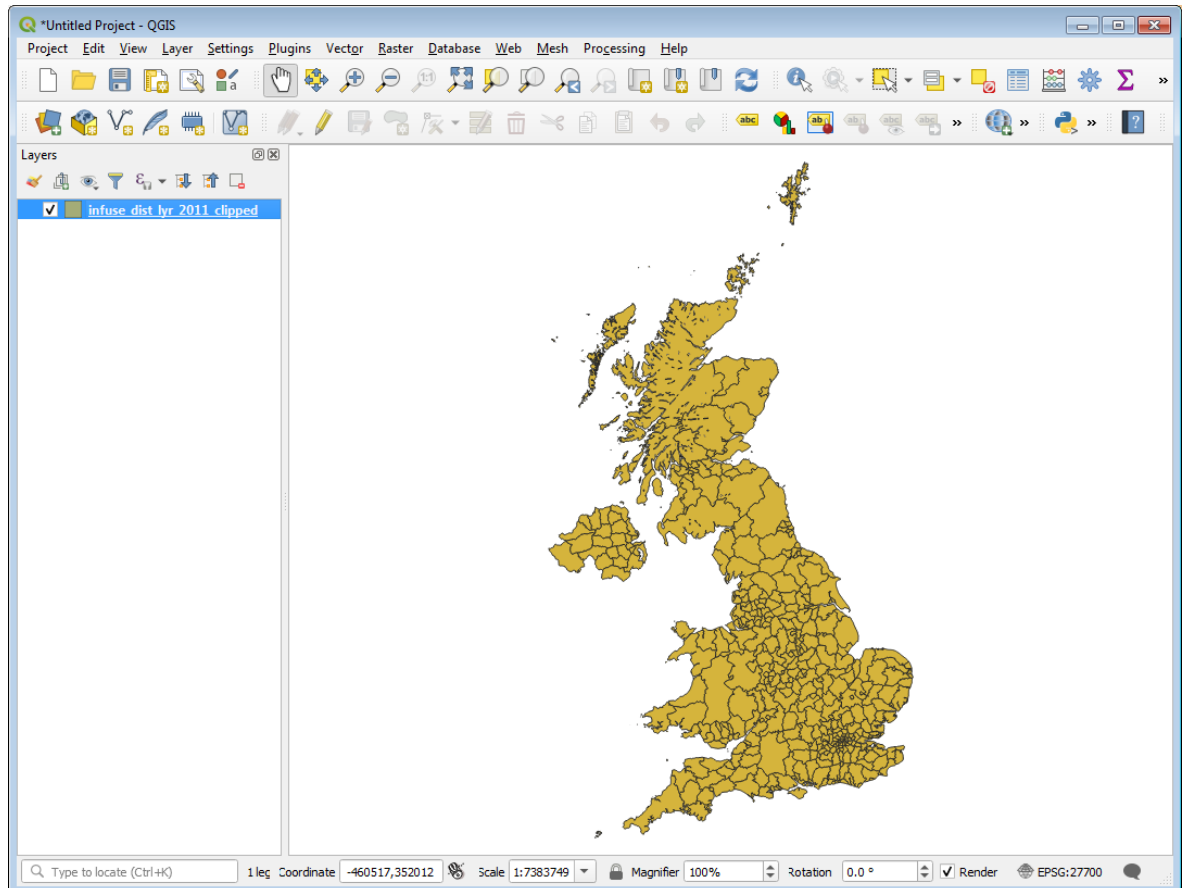
7. We have added the boundaries, now we need to add the aggregate data that we downloaded from InFuse. As before click the Open Data Source Manager icon. This time from the types of data at left, click on Delimited Text
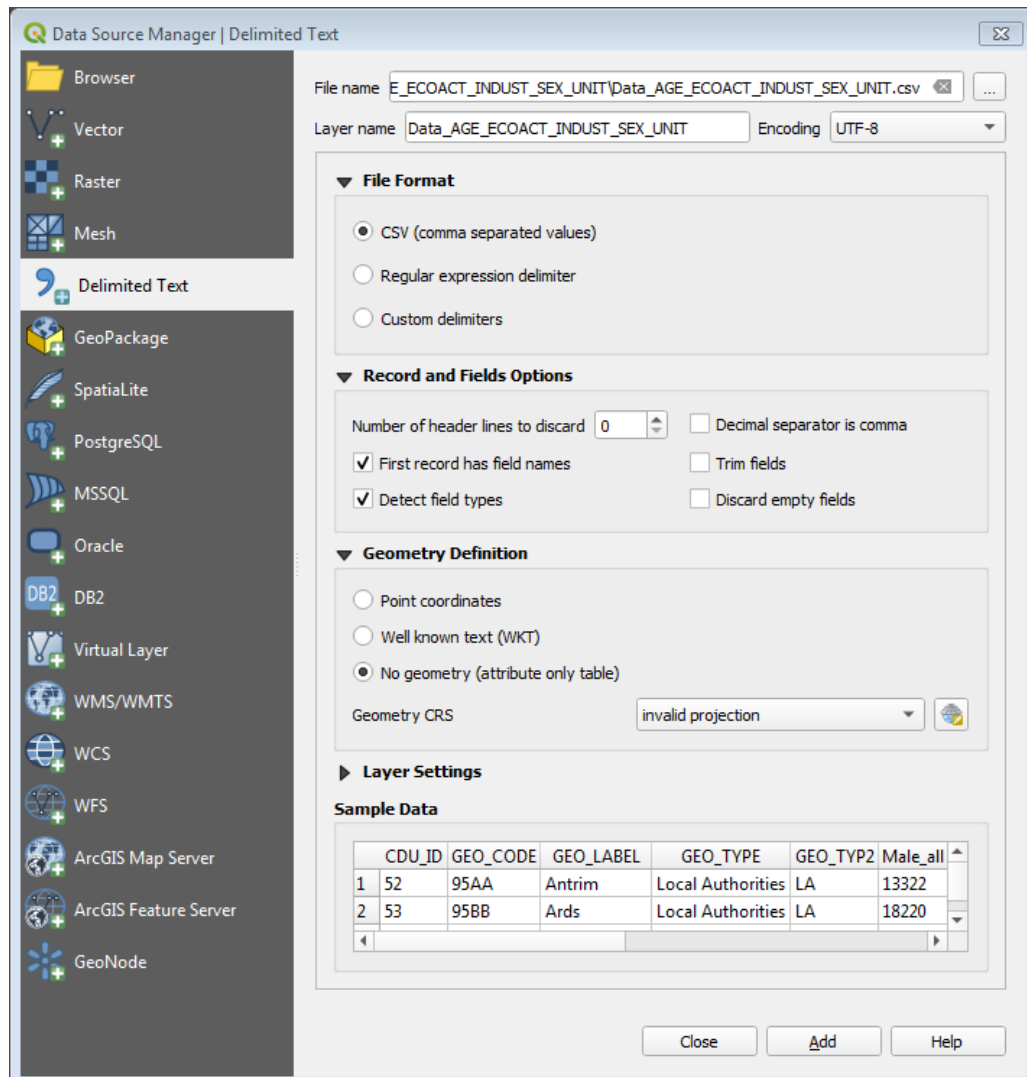
8. Click on […] and then browse for the Data_AGE_ECOACT_INDUST_SEX_UNIT.csv file and select csv for the file format. Ensure that under Geometry Definition, **No geometry (attribute table only)** is set, then click Add, followed by Close to close the Data Source Manager dialogue.

9. We now need to join these two layers together

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

11. Click *joins*, and tick the green + symbol
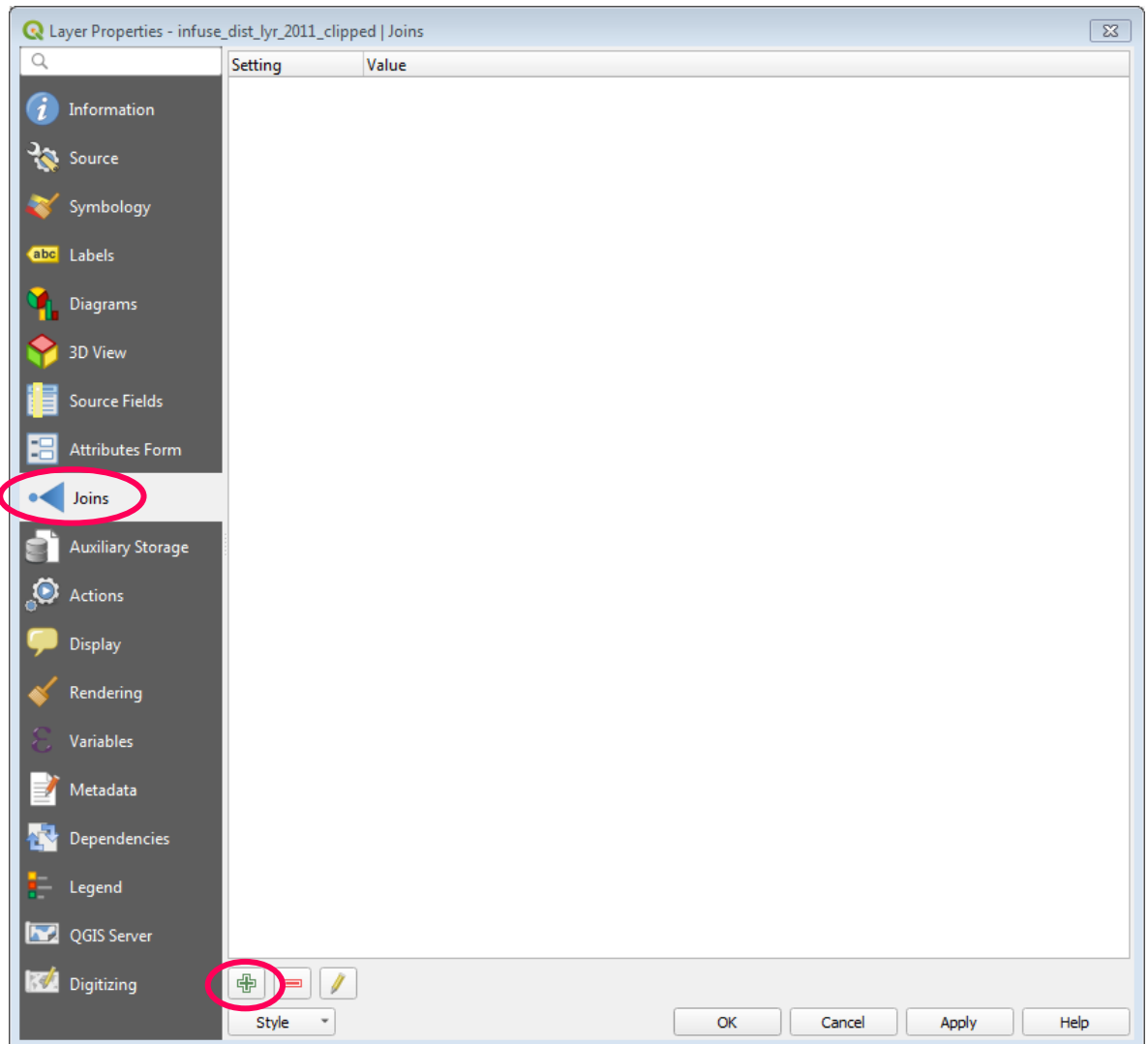
12. **Select GEO_CODE as the join field and geo_code as the target field** click *OK* then *Apply* and then *OK* again

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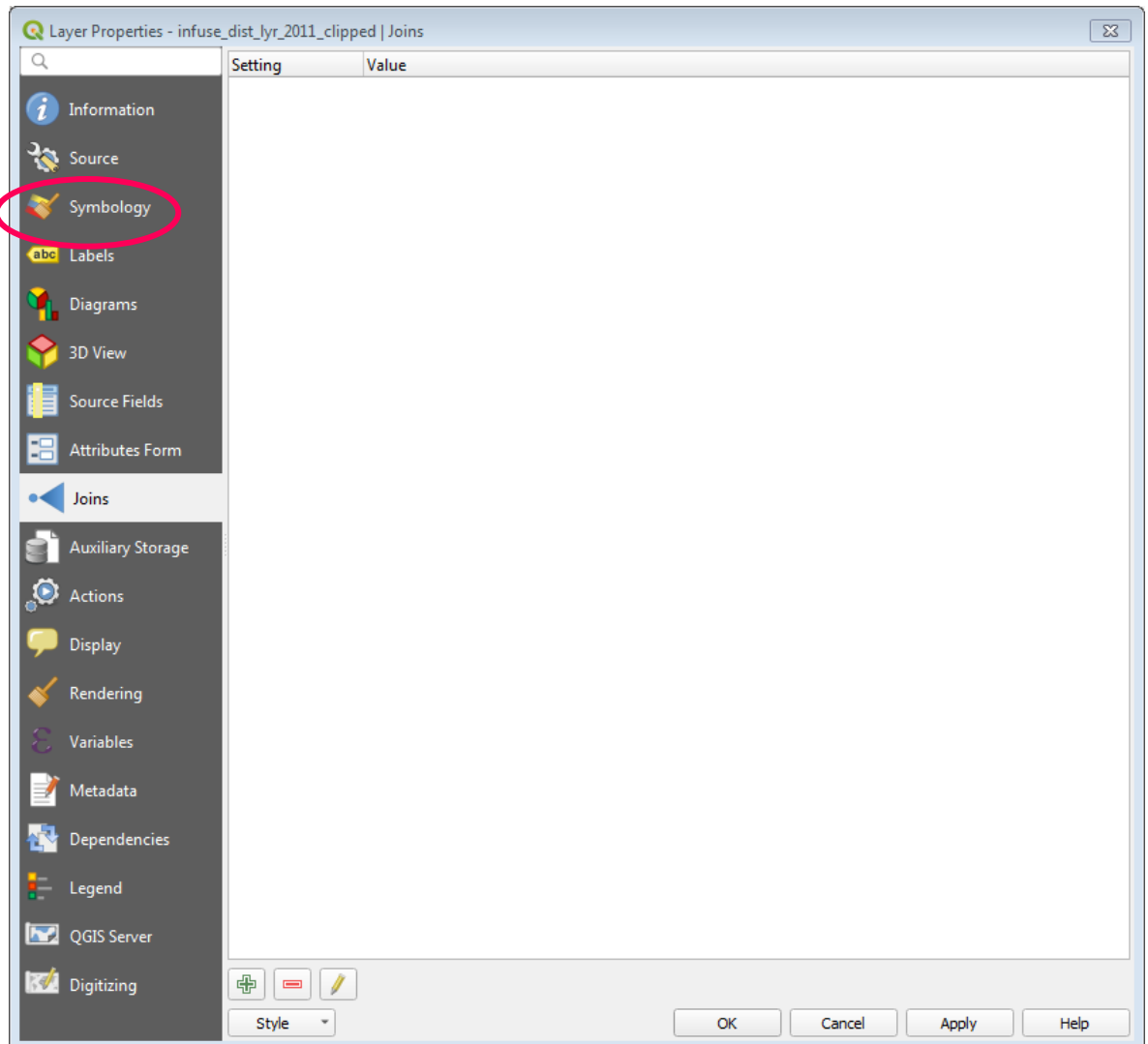
2. Click **Symbology**, you have a few choices. In the top dropdown menu, click **Graduated.** The colour ramp picker allows you to choose the colours to be used to represent the percentages; you can change this if you wish

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 **Data_AGE_ECOACT_INDUST_SEX_UNIT_Male_man_p. Note: QGIS renames the**

column by prefixing the data source name to the name of the column as it appears in that data source which is why the column is listed as Data_AGE_ECOACT_INDUST_SEX_UNIT_Male_man_p rather than just Male_man_p as we named the column when we edited the CSV file in Excel earlier.

4. Click on *Classify* and you can see that 5 classes appear. The number of classes can be changed but it is recommended to stick to between 5 and 7 classes. The Mode dropdown will allow you to specify the classification method that will be used to classify the data. There are a few options; we would recommend using "natural breaks (Jenks)".

5. We can check on how the classification has performed by examining the distribution of the data as a histogram. To do so click on the Histogram tab and click the Load Values button
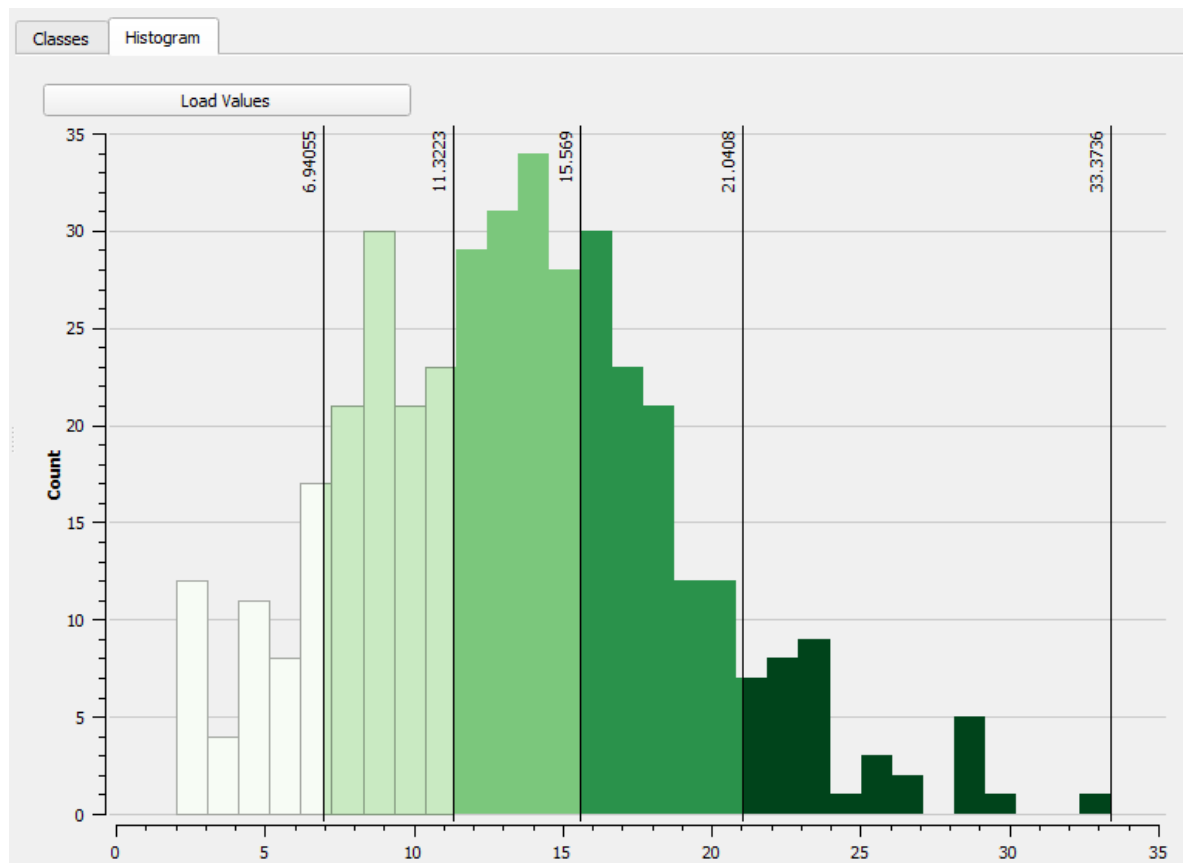

6. Change these options as you wish, and then click Apply and OK
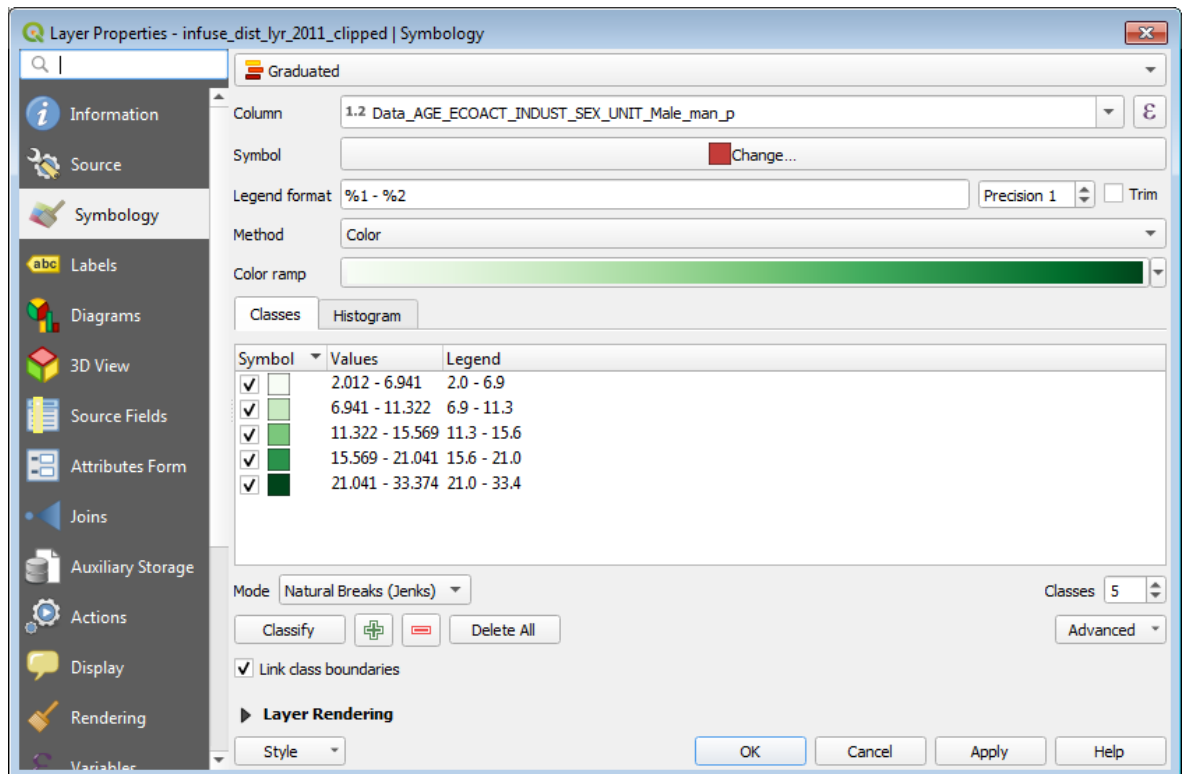
7. You will now have a map, looking something like this:

8. 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

9. To format your map for printing go to *Project > New Print Layout*, and insert a title. You can change the size of the page by right-clicking on the map canvas and choosing **Page Properties…**
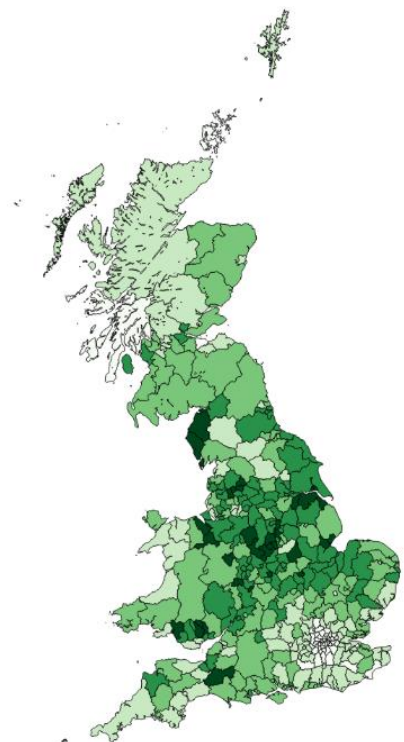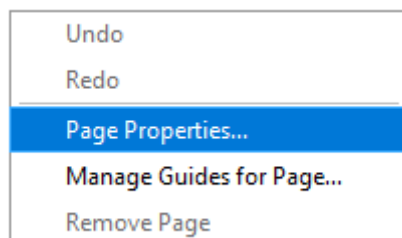
10. If you want to, change the page size using the panel on the right hand size (composition tab). It makes sense to leave the page size as A4, but to change the orientation to portrait
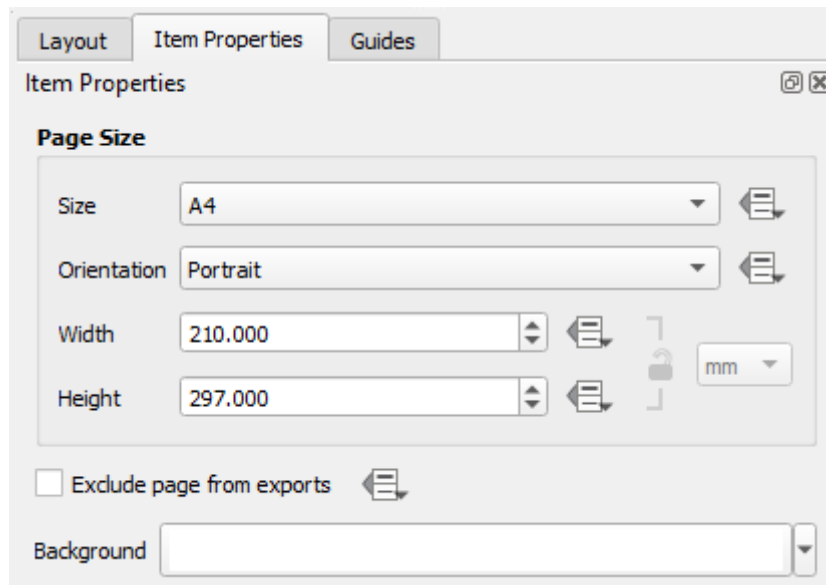
11. To insert the map, click the [icon] icon and drag the new map item across the page with your mouse.

12. Adjust the size. You can do this by clicking the arrows around the map, but there is a lot of extra white space. Or use the options within the *item properties* tab on the right if the screen (for example you can specify scale, try setting to 5000000)

13. Insert a label. Click the [icon] icon and drag out the new label item across the map with your mouse. 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

14. Insert a Legend. Click the [icon] icon and drag out the new legend item across the map with your mouse.

 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.
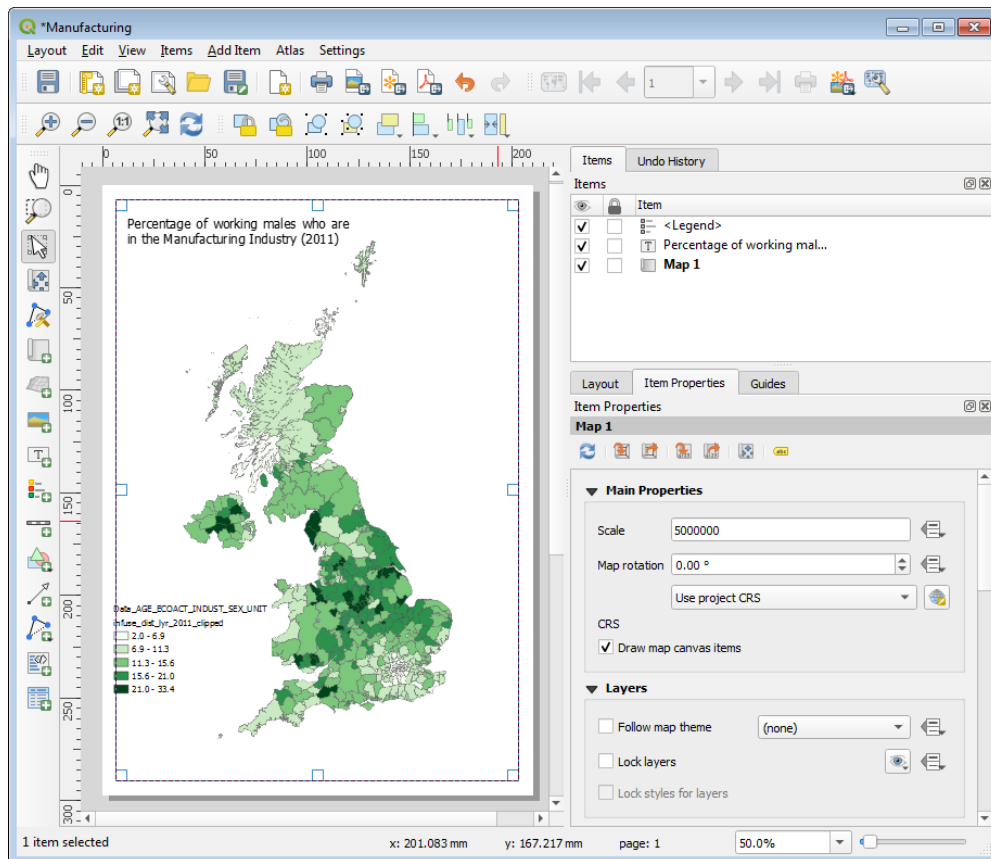
Useful links

- 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/

UK Data Service

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