

# Creating a Flow map from Census data in QGIS



# Contents

1. Introduction .....	3
2. Exploring Census Flow data in QGIS .....	3
3. Filtering data in QGIS.....	8
4. Styling Census Flow Data in QGIS.....	12
5. (Extra) Appendix 1 - Accessing Census Flow data at the UK Data Service .....	23
6. (Extra) Appendix 2 - Processing UK Data Service Census Flow Data in PostgreSQL and PostGIS to create flow line geometries. ....	24

## 1. Introduction

In this exercise we will use QGIS to create a flow map visualisation from a 2011 Census Origin-Destination table that shows at Local Authority level, flows between place of residence and place of work.

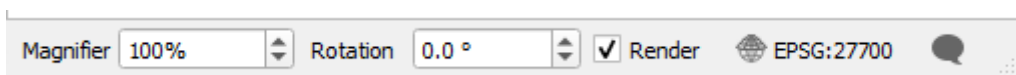
The visualisation of Origin-Destination data requires data pre-processing to create line geometries between the Origins and Destinations which can be time consuming for large UK wide datasets. Given the time constraints available for this exercise, prepared pre-processed data is provided to you, to enable the exercise to concentrate on the visualisation of this data rather than the processing of it.

Details on the pre-processing steps that could be applied to other Census 2011 Origin-Destination datasets and information on how to obtain Census Flow data from the UK Data Service are provided in the Appendices at the end of this workbook.

## 2. Exploring Census Flow data in QGIS

If you do not already have QGIS open, start QGIS.

Make sure that the QGIS project Coordinate Reference System is set to EPSG:27700



By default QGIS has the project Coordinate Reference System set to EPSG:4326 i.e. the Global WGS84 CRS. We are using data provided in EPSG:27700 – British National Grid so to avoid complications set the QGIS project Coordinate Reference System to EPSG:27700 if it is not set to this already. Help on how to do this is provided in Exercise 1.

Open the Data Source Manager from the main QGIS toolbar



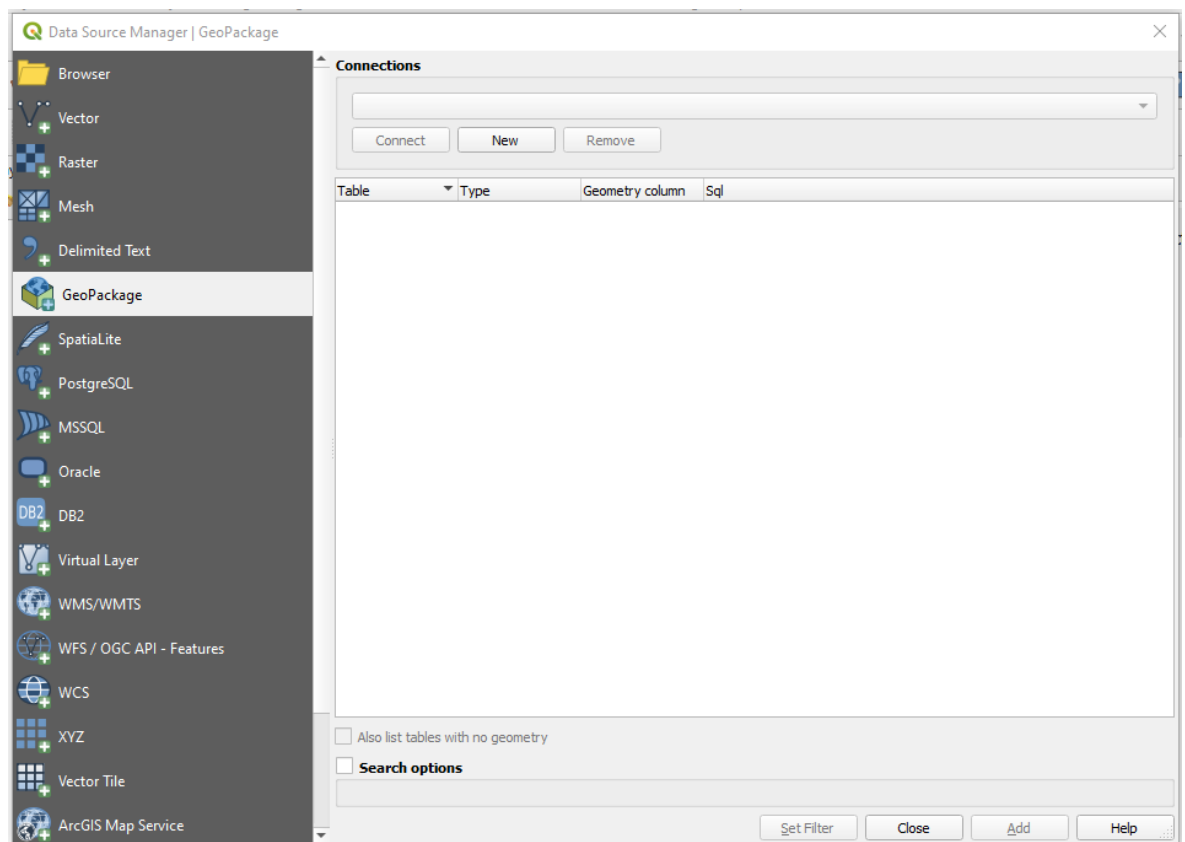
The QGIS Data Source Manager will open. The Data Source Manager enables you to add different types of data to QGIS by clicking on the data type at left.

In other exercises we added data to QGIS that was held in a shapefile or a CSV file. In this exercise we will add Census Flow data that has been downloaded from

the UK Data Service WICID application, processed into line geometries connecting places of residence and workplaces and placed into a GeoPackage dataset alongside a polygon boundary dataset of UK Local Authorities.

A GeoPackage is a newish form of geospatial data. Unlike a data format like shapefile where only a single layer of information can be stored, the GeoPackage data format enables multiple layers of information to be held in the same file.

Click on GeoPackage at left. Before we can access the layers of Census flow data in a GeoPackage we need to establish a connection to the GeoPackage. The list of current connections to GeoPackages that QGIS is aware of is shown in the top dropdown under Connections



Click on the New button.

The Open GeoPackage dialogue will be shown. Navigate to the location of

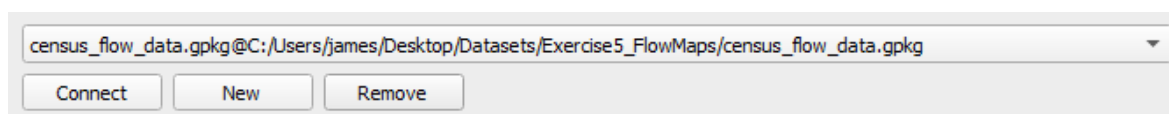
**Census\_Training\_Data\FlowMaps\census\_flow\_data.gpkg**

and click the Open button.

The QGIS Data Source Manager should list a connection to the census\_flow\_data.gpkg GeoPackage. Note that if you have added different GeoPackages to QGIS before QGIS will retain knowledge of these connections and will include them in the list of available GeoPackage connections.

If this is the case, QGIS might have defaulted to listing these other GeoPackages rather than the GeoPackage we need for this exercise.

Make sure that the connection to the census\_flow\_data.gpkg GeoPackage is shown something like this:



Then click the Connect button. QGIS will show what layers are available within the GeoPackage that we just connected to.



In this case our GeoPackage contain 2 layers:

The wu01uk\_v2\_no\_quasi\_workplaces\_w\_geom layer contains line features that connect Local Authority places of residence with Local Authority workplaces. Each line has information on the volume of flows between these origins and destinations.

The \_no\_quasi\_workplaces bit just comes from the fact that when the dataset was pre-processed for this exercise we have excluded all records where place of work was a special quasi-workplace such as offshore workplace or people were working from home.

The uk\_census\_merged\_local\_authorities\_2011 layer contains a polygon layer of the Local Authorities themselves.

Add both of these layers to QGIS.

To do this, click on both layers so that they are highlighted and then click the Add button at the bottom right of the Data Source Manager.

QGIS will display something like this:



Try turning the top most line layer off for now.

The `uk_census_merged_local_authorities_2011` layers is a polygon dataset showing all of the UK's 404 individual Local Authorities for example in London, Camden Borough Council and in Scotland, the City of Edinburgh council.

If you open the attribute table of the `uk_census_merged_local_authorities_2011` polygon dataset (either click on `uk_census_merged_local_authorities_2011` so it is selected in the list of layers, left of the main QGIS map window and then from the main QGIS menubar choose Layer >> Open Attribute Table OR right-click on

uk\_census\_merged\_local\_authorities\_2011 in the list of layers and choose Open Attribute Table) you can see that it has 4 columns. Two of these (fid and gid) can be ignored.

The geocode and geolabel columns provide for each local authority a geographic identifier code (geocode) that uniquely identifies a particular local authority amongst all of the other local authorities in the UK and the placename (geolabel) associated with that local authority.

Scroll through the table and see if you can find the name of the local authority in which you are currently living and take a note of the geocode for this local authority.

For example the City of Edinburgh Scottish Council Area has the geocode of S12000036.

Turn the wu01uk\_v2\_no\_quasi\_workplaces\_w\_geom line layer back on and likewise open its attribute table also. As you can as well as the fid and gid columns which we can ignore, the other columns provided are as follows:

COLUMN_NAME	DESCRIPTION
area_of_residence	Geocode of Local Authority place of residence
area_of_workplace	Geocode of Local Authority place of workplace
male_and_female	Total number of people that travel for work between places
male	Total number of male people that travel for work between places
female	Total number of female people that travel for work between places

If we look at a few of the records in the attribute table of this layer you will see something like this:

	fid	gid	area_of_residence ^	area_of_workplace	male_and_female	male	female
100315	107897	107897	S12000035	S12000023	13	7	6
100316	107898	107898	S12000035	S12000038	460	241	219
100317	107899	107899	S12000035	S12000039	1225	468	757
100318	3545	3545	S12000036	95WW	1	1	0
100319	3546	3546	S12000036	E41000023	11	6	5
100320	3547	3547	S12000036	E41000031	1	1	0
100321	3548	3548	S12000036	E41000064	6	5	1

Each record describes flow between 2 different places where people live (area\_of\_residence) and people work (area\_of\_workplace).

So in this specific example, record 100317 indicates that 1225 people (468 male; 757 female) travel between area\_of\_residence S12000035 and area\_of\_workplace S12000039. From the polygon dataset that describes the list of local authorities in the UK we can determine that S12000035 is the Argyll & Bute local authority whereas S12000039 is the West Dunbartonshire local authority, hence this data indicates that the 2011 census indicates that 1225 people travel to work from Argyll & Bute to West Dunbartonshire.

As we saw above there are 404 Local Authorities in the UK. If we had a flow between each Local Authority and every other Local Authority this would come to  $404 * 404$  or 163216. Even if we exclude cases where the area\_of\_residence is the same as area\_of\_workplace this still leaves up to a maximum of 162812 flows.

If you look at the title bar of the attribute table of the wu01uk\_v2\_no\_quasi\_workplaces\_w\_geom dataset



QGIS tells us how many records/features are present. 108142 is smaller than 162812 by some margin. So what has happened to some of the flows. Well there will not be a travel to work flow between every area\_of\_residence and every area\_of\_workplace since people tend to work close to where they live. The exceptions to this being people that travel between big cities or conurbations.

### 3. Filtering data in QGIS

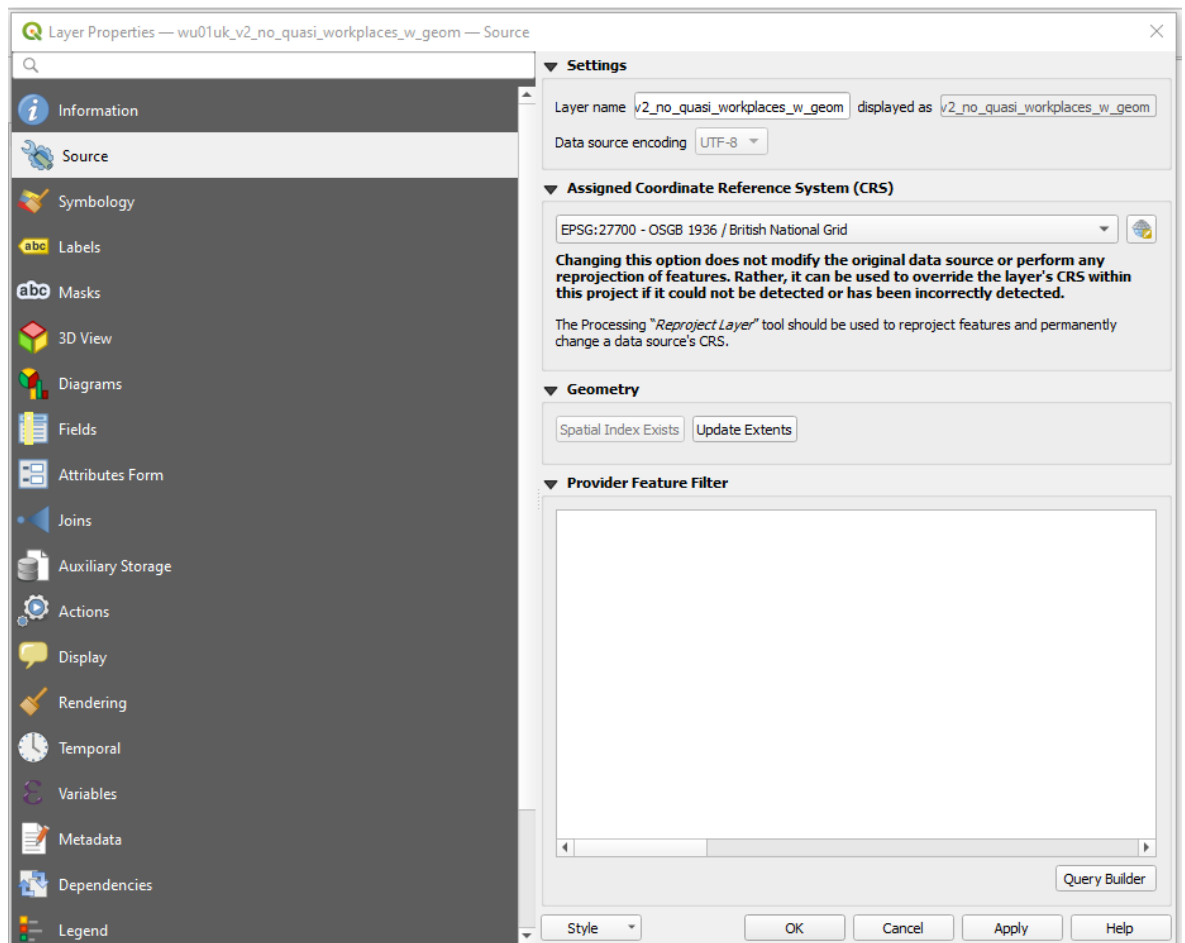
Rather than creating a flow map that maps the flow from each zone to every other zone we will create a flow map that shows the travel to work flow for a single Local Authority.

We want to apply a filter to the lines layer to show only those lines that represent flows from our Local Authority of interest. One way of doing this in QGIS would be to select such line features and save them to a new dataset. This has the downside that it is a new dataset that we have to manage and if we wanted to look at another Local Authority we would have to create another new dataset.

An alternative is to tell QGIS to apply a filter to the *data source* (in this case the wu01uk\_v2\_no\_quasi\_workplaces\_w\_geom) layer and display only those features that we want.

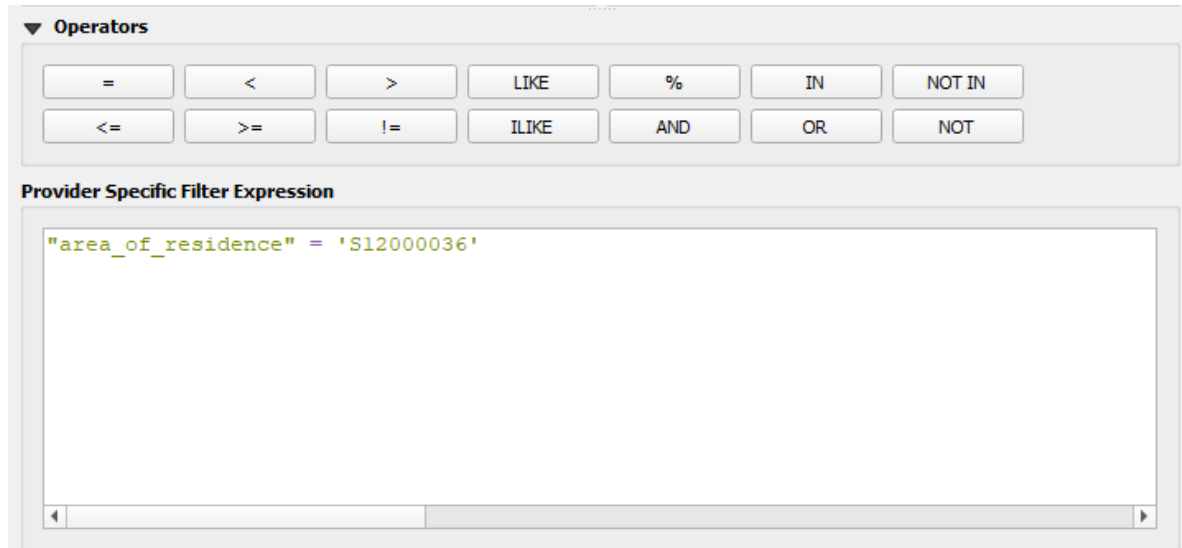
Right-click on the wu01uk\_v2\_no\_quasi\_workplaces\_w\_geom layer in the QGIS layer list and select Properties... from the pop-up-menu. Click on Source at left:





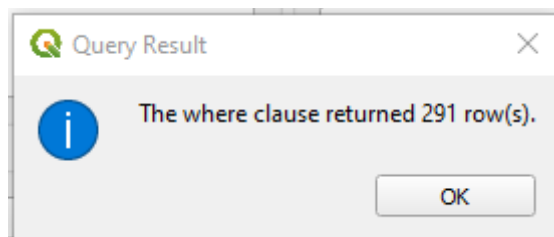
By providing a Feature Filter we will be able to filter the line features displayed in QGIS. To do so click the Query Builder button.

By typing expressions into the Provider Specific Filter Expression box at the bottom of the Query Builder dialogue we can set the criteria which QGIS can use to restrict (filter) the features displayed. Types something like this:



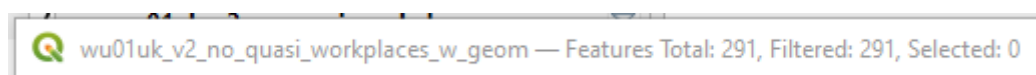
In this case I am setting the filter such that QGIS will select all records whose `area_of_residence` column is equal to S12000036, the geocode for the City of Edinburgh that I noted earlier. In your case enter the geocode for the local authority in which you reside that you noted earlier or just stick with this one for Edinburgh.

Having entered an expression, click the Test button.



QGIS will test evaluate the expression. If there is a problem with the syntax of the expression, QGIS will report that the 0 row(s) will be returned. Assuming the expression tests OK, click OK to return to the Layer Properties dialogue and then click Apply followed by OK to return to the main QGIS window. QGIS will redisplay the line layer with this time much fewer lines displayed, in fact only those lines radiating out from the source local authority that we filtered the dataset on.

Open the attribute table of the `wu01uk_v2_no_quasi_workplaces_w_geom` layer again. As expected QGIS reports that there are now only hundreds rather than thousands of records visible in the table.

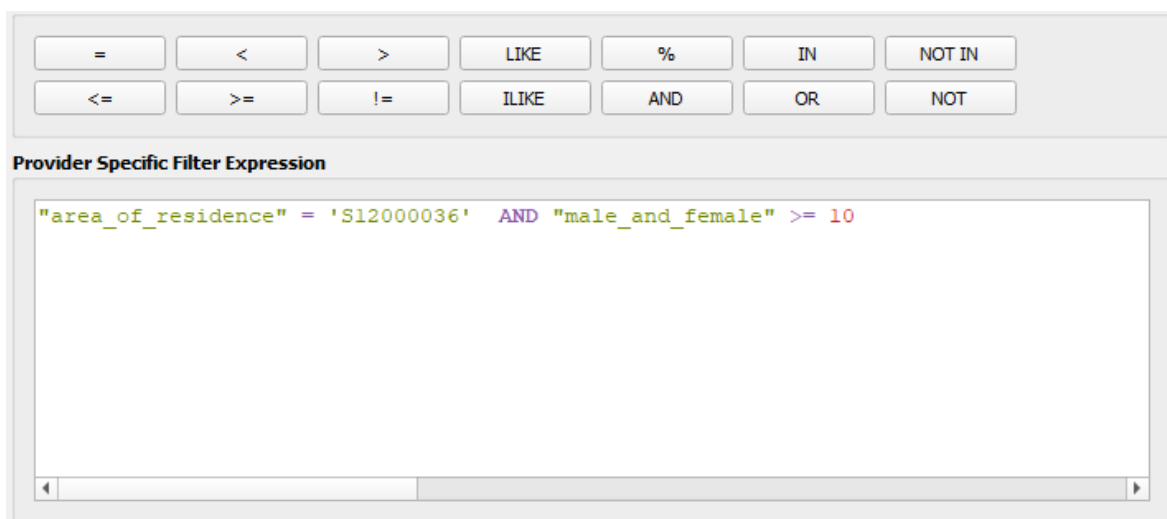


If you click on the `male_and_female` column header this will sort the values in ascending order from low values to high values. If you then scroll through the records you will probably see that there are lots of cases where the value of the `male_and_female` column is very small i.e. 1 or 2 etc. Such small values are outliers in the dataset which represent atypical travel flows between local authorities where people live and where they work. Although these values are small enough to be insignificant they are still displayed as a line in the dataset so they will get in the way of more important values in the dataset and add visual clutter / noise. We can exclude these low values by modifying the Feature Filter so that as well as filtering on local authority it also excludes these low values.

So open up Layer Properties again and click on Source. Our current Feature Filter expression should still be visible. We want to modify this though. So click on the Query Builder button and modify the Provider Specific Filter Expression by adding to the end of the current expression:

AND "male\_and\_female" >= 10

so that it reads something like this:



Again, the first `area_of_residence` bit will vary depending on what you chose.

Test the expression again, and once happy click OK to return to the Layer Properties and then Apply followed by OK to return to the QGIS main window once more.

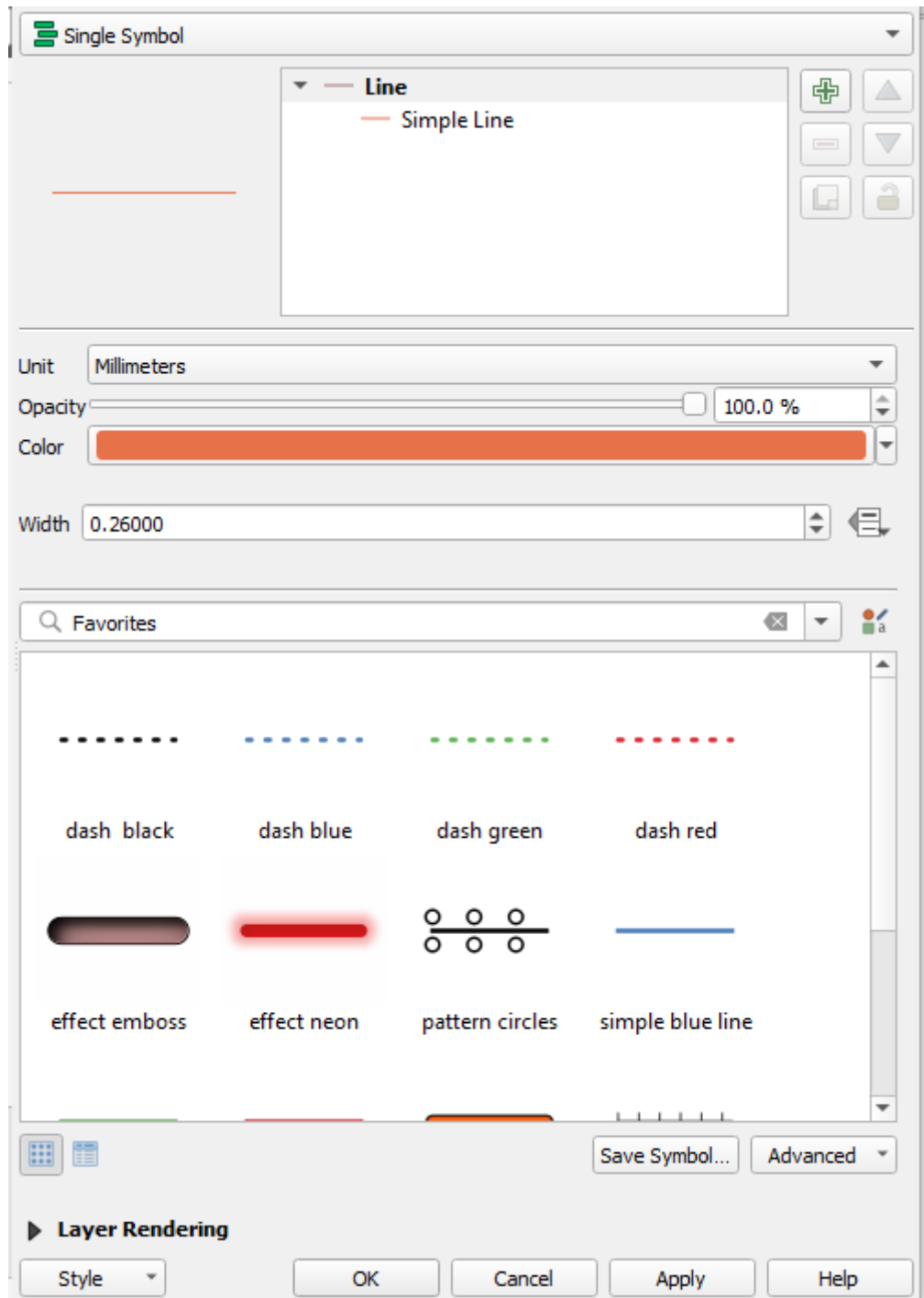
Even more of the flow lines radiating out from the source local authority should now have been turned off and only the important flows are now shown.

## 4. Styling Census Flow Data in QGIS

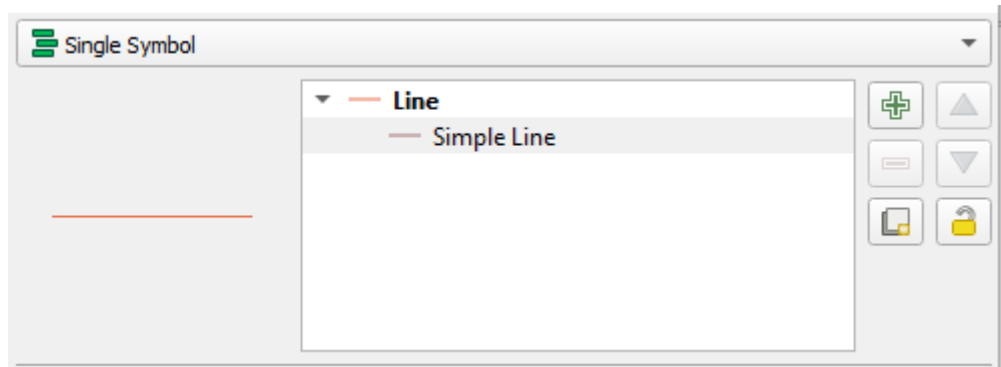
In the same way that in Choropleth Maps we change the colour of the polygon depending on the variable being displayed and in Cartograms we may change the shape of the geometry according to the variable, in flow maps we can change the symbology of the flow lines themselves depending on the variable being displayed.

Normally flow lines are display as arrows with the arrow pointing in the direction from the Origin (area\_of\_residence) to the Destination (area\_of\_workplace) and the width which the arrow is drawn with is altered depending on the volume of the flow. So large flows are drawn as thick arrows; smaller flows are drawn as thinner arrows. We will use QGIS`s symbology options to apply this kind of styling to our filtered flows.

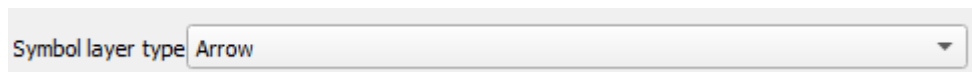
Open Layer Properties for the wu01uk\_v2\_no\_quasi\_workplaces\_w\_geom layer and click on Symbology. Currently features in the layer are drawn using a Simple Line symbology:



In the list box at top under Line, click on Simple Line



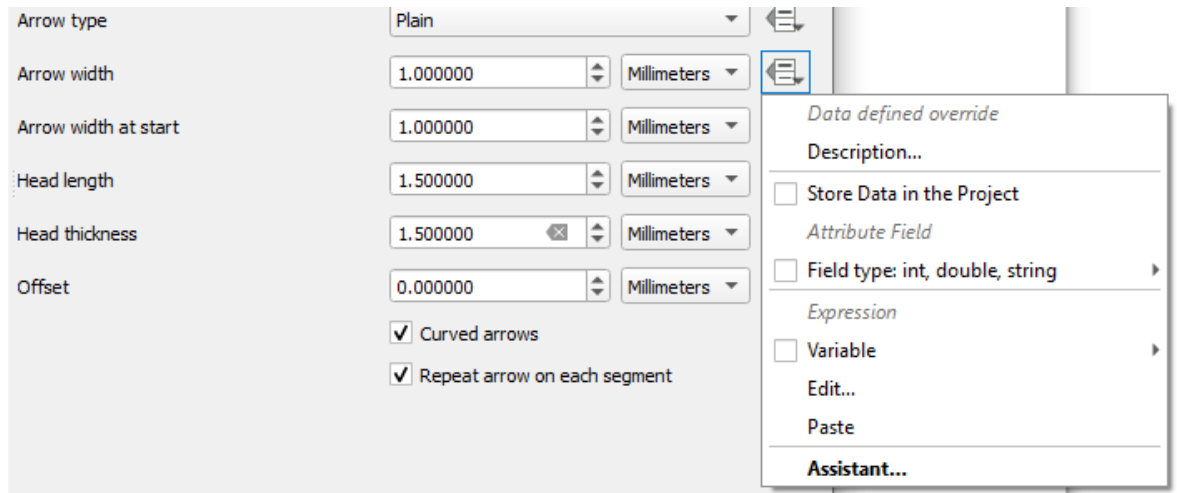
and then change the Symbol layer type dropdown from Simple Line to Arrow



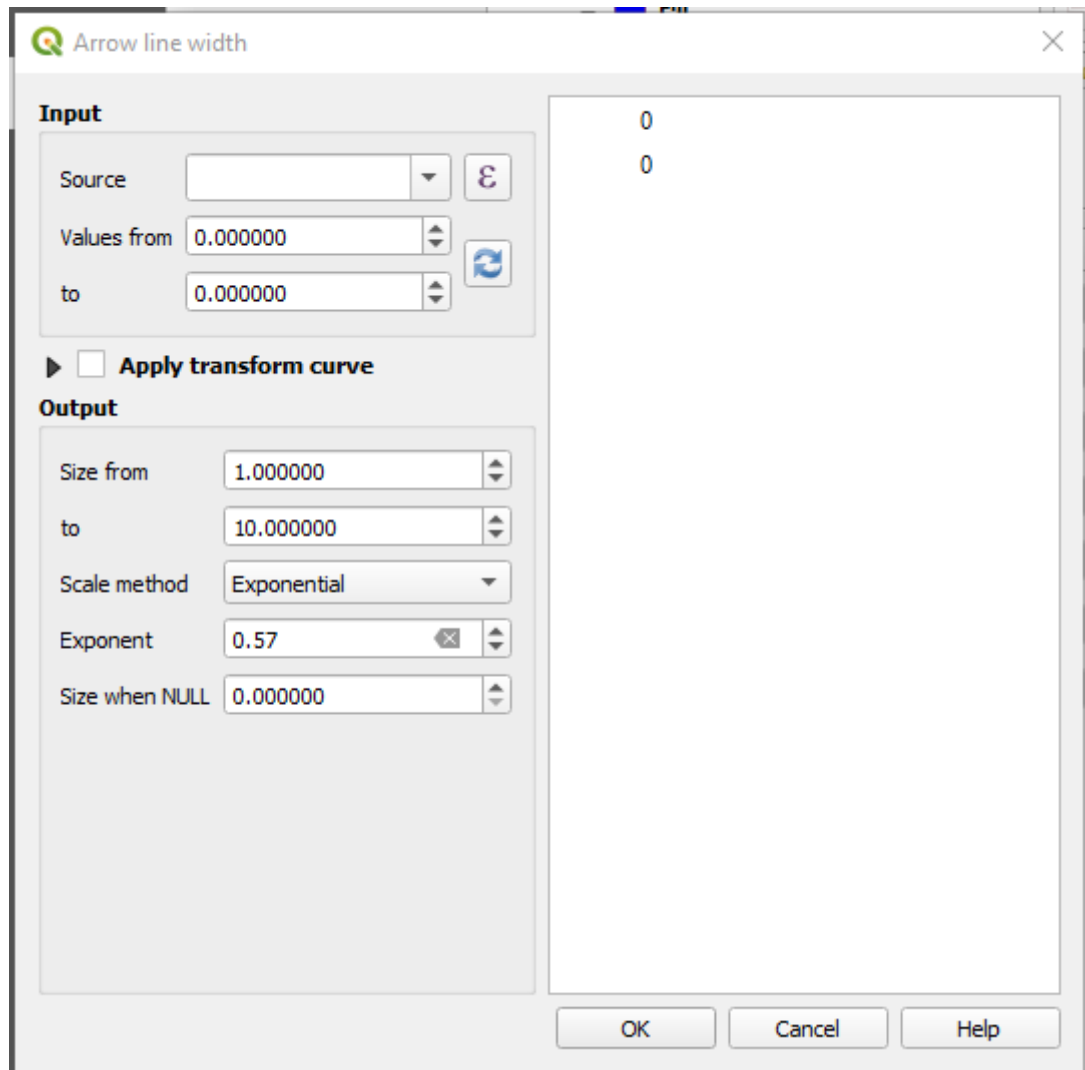
The dialogue will change to show the specific options that we have at our control to modify the appearance of the Arrow symbology.

There are 3 different properties of the arrow symbology which we will set. We will instruct QGIS to alter the width of the arrows based on the value of the `male_and_female` column so that high values are drawn with thicker arrows that smaller values. Since we are changing the width of the arrows we will also tell QGIS to similarly change the width of the arrowheads otherwise we will end up with thick arrows with tiny arrowheads!. Lastly we will change the colour of the arrows.

To instruct QGIS to alter the width of the drawn arrows based on the value of the `male_and_female` column, to the right of the Arrow width section is a dropdown which when clicked will bring up the pop-up menu shown below. From this menu choose the Assistant option



This will open an Arrow line width dialogue like this:



This dialogue consists of 2 parts there is an Input section and an Output section which allow us to control how an input values for a particular data value, in our case the total flow of people between 2 local authorities result in a particular width of arrow.

Under the source dropdown, select the male\_and\_female column, then click the blue Fetch value range from layer button



**Input**

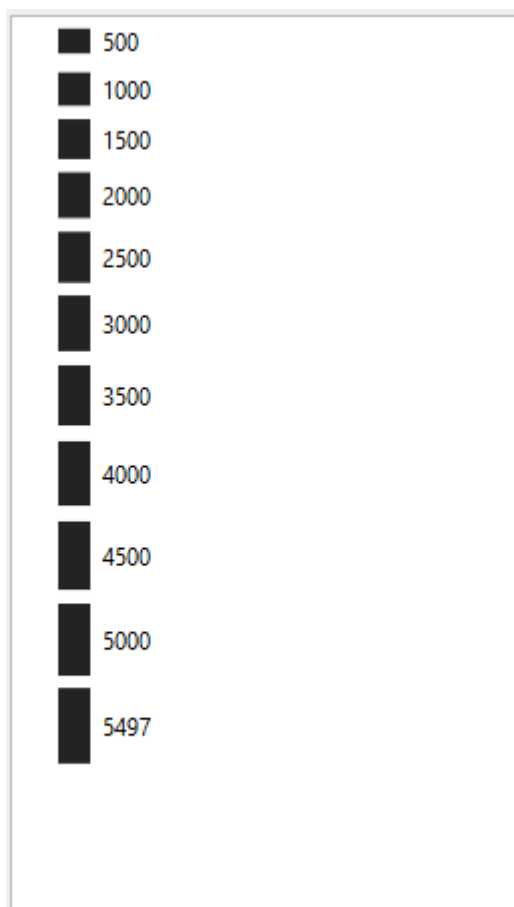
Source

Values from

to

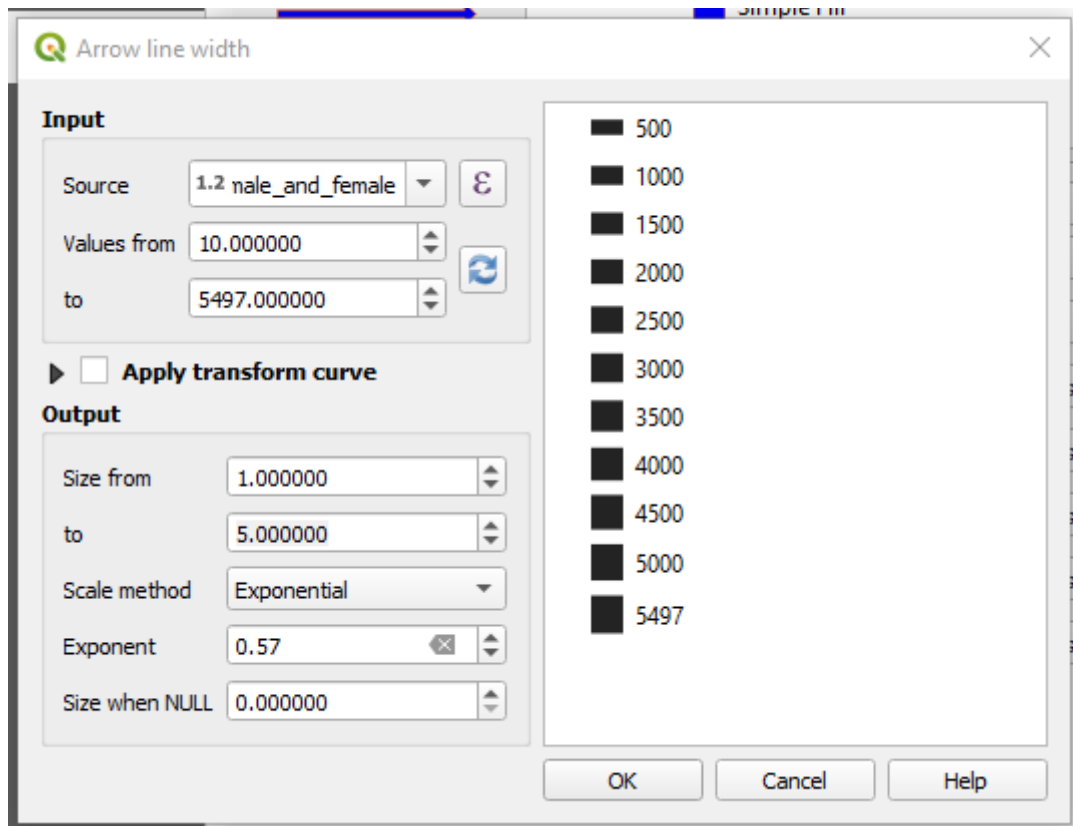
Doing so will populate the Values from / to boxes with the lower and upper limits of the data values present in the male\_and\_female column.

At the right of the dialogue QGIS will give a preview of the sort of arrow widths that this will produce:



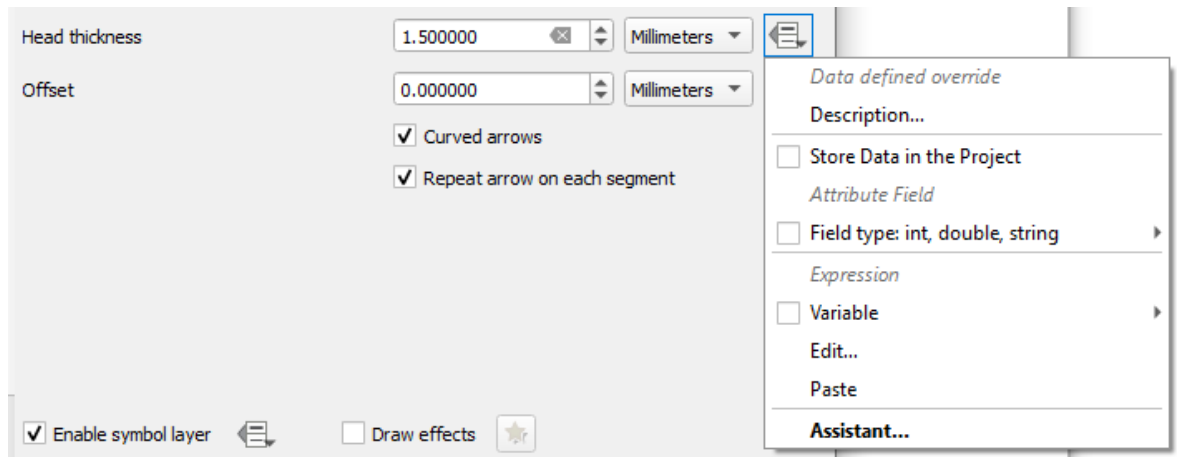
So we can see that small values will be drawn substantially smaller than large values.

By default QGIS will draw all flow arrows between a width of 1 and 10 but this will produce very thick lines so lets reduce the maximum arrow width to 5, after which you should have something like this:

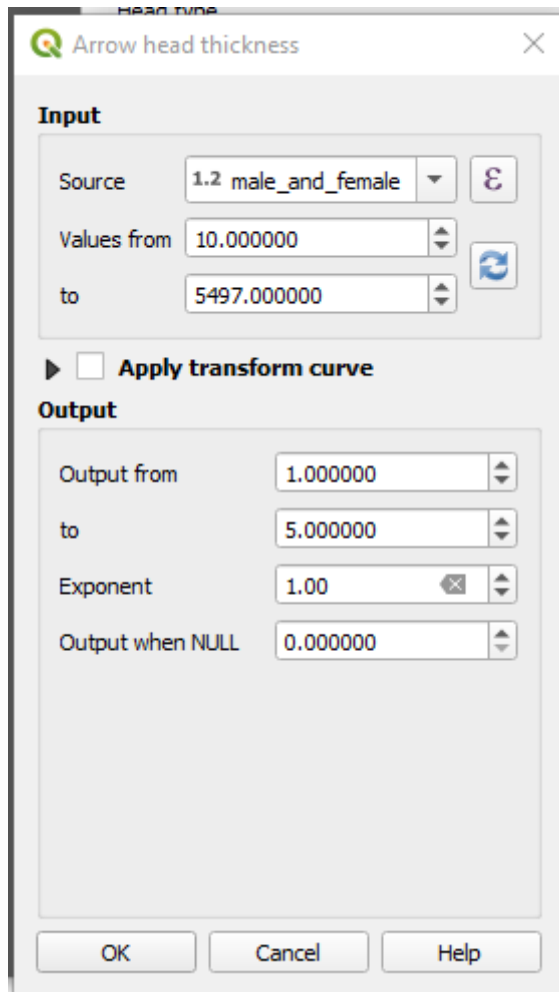


After doing so click OK to return to the main Layer Properties window.

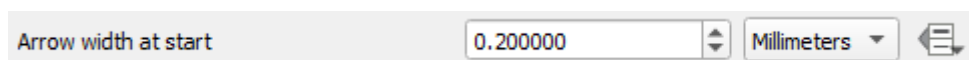
To instruct QGIS to similarly alter the width of the arrowheads we need to apply the same range of input values and output widths to the arrowhead (which QGIS terms Head thickness). So to the right of the Head thickness section, click the dropdown and choose Assistant...



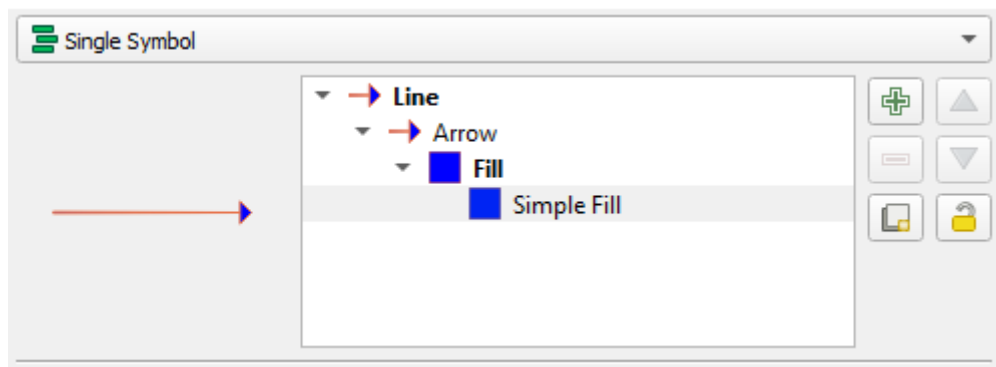
This will open an Arrow head thickness dialogue. Set the values of this as shown below by setting the Source dropdown to male\_and\_female, clicking the Fetch value range from layer button and changing the upper Output value to 5:



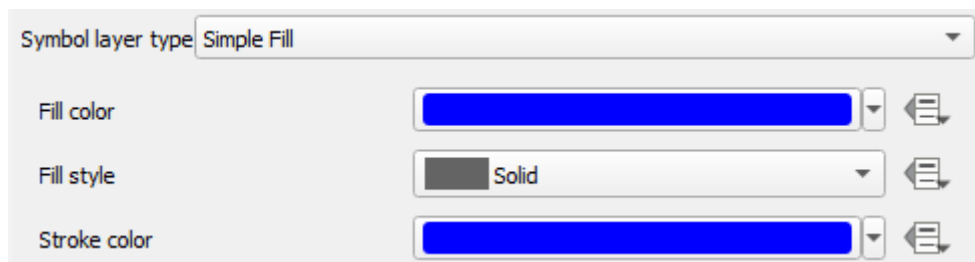
Then click OK to return to the Layer Properties dialogue. Change Arrow width at start property to 0.2 like so:



Finally at the top list box, click on Simple Fill

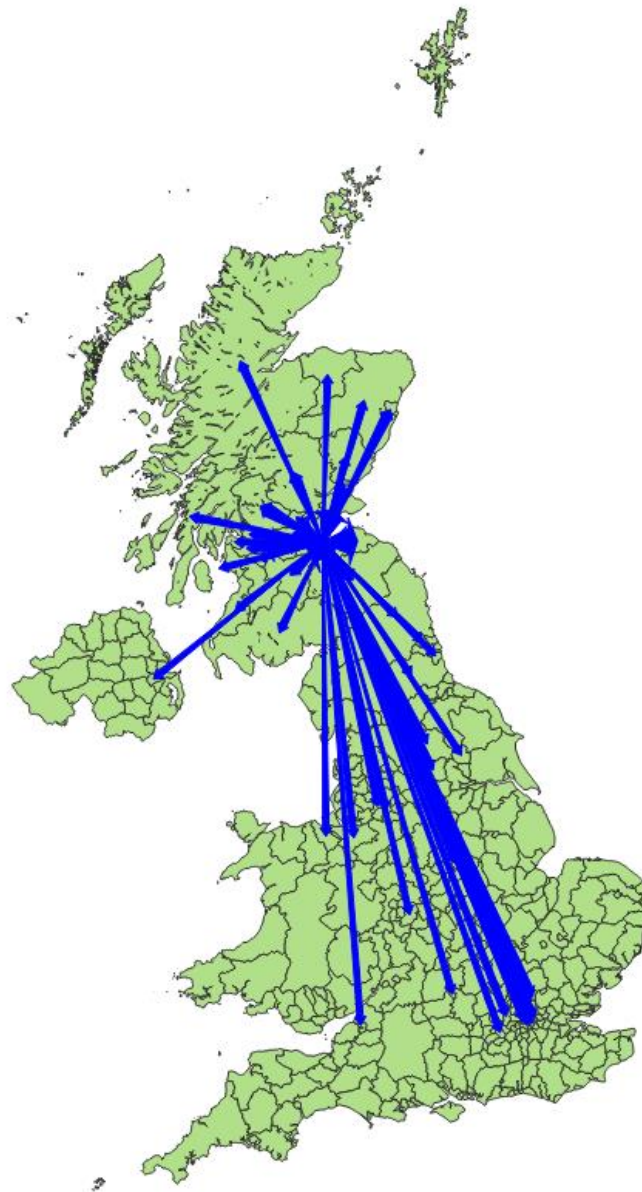


Ensure that as here that Fill color and Stroke color are set to the same colour:



Finally click the Apply button, followed by OK to close the Layer Properties dialogue and return to the main QGIS map window.

You should see that the flow arrows are now drawn as arrows, the width of which varies depending according to the volume of that flow:



So in the case of Edinburgh it is clear from the flow map that there as might be expected the highest travel to work flows between the City of Edinburgh local authority and the local authorities such as East; West and Midlothian that surround the city. There are however also significant travel to work flows to the north of England and London.

## 5. Appendix 1 - Accessing Census Flow data at the UK Data Service

Although you are provided with pre-processed prepared data for this exercise, it's worth spending a moment to consider how to access the raw source 2011 Census Flow data from the UK Data Service.

Go to <https://census.ukdataservice.ac.uk/>

From Get census data at the top go to Flow data

As described census flow data (also known as interaction data) are data that relate to flows of people between places.

To download some census flow data from the 2011 census click on the link to WICID.

This will take you to <http://wicid.ukdataservice.ac.uk/>

The WICID application provides different ways to access census flow data. Click on WICID Downloads, then:

### Standard login (public access datasets)



[Start a new session - no username or password required](#)

which will give you access to public access census flow datasets that require no authentication to access. Click on the 2011 Census United Kingdom – Open link to see a list of datasets that are available.

If you scroll down towards the bottom of the list there is a dataset called WU01UK. This is the 2011 Census Flow dataset that we create a flow map from in this exercise. From the description we can see that this dataset provides the location of usual residence and place of work by sex for all usual residents aged 16 and over in employment the week before the census. Data coverage is UK and level of output geography is by UK Local Authority.

## 6. Appendix 2 - Processing UK Data Service Census Flow Data in PostgreSQL and PostGIS to create flow line geometries.

From WICID downloaded the WU01UK table as wu01uk\_v2.csv. From ONS obtained a Shapefile of UK Census Merged Local Authority District 2011 boundaries. In a PostgreSQL / PostGIS database loaded the wu01uk\_v2.csv OD data as wu01uk\_v2 table and loaded the LAD shapefile as uk\_census\_merged\_local\_authority\_districts\_2011 table.

Created a new empty table to hold flowlines:

```
CREATE TABLE wu01uk_v2_no_quasi_workplaces_w_geom(
gid serial,
area_of_residence text,
area_of_workplace text,
male_and_female numeric,
male numeric,
female numeric,
geom geometry(LINESTRING, 27700)
);
```

Then populate this new table by joining the geometries from the LAD boundaries to the OD data using LAD geographic identifiers, computing the centroid for each LAD and then generating a line linking the origin to the destination. Excluded are records where workplace = residence and where workplace is a quasi-workplace (workplace code is something like 'OD%'):

```
INSERT INTO wu01uk_v2_no_quasi_workplaces_w_geom(
area_of_residence,
area_of_workplace,
male_and_female,
male,
female,
geom
)
SELECT
a.area_of_residence,
a.area_of_workplace,
a.male_and_female,
a.male,
a.female,
st_makeline(st_centroid(b.geom), st_centroid(c.geom)) as geom
FROM wu01uk_v2 a
LEFT JOIN uk_census_merged_local_authority_districts_2011 b ON (a.area_of_residence =
b.geocode)
LEFT JOIN uk_census_merged_local_authority_districts_2011 c ON (a.area_of_workplace =
c.geocode)
WHERE a.area_of_workplace != a.area_of_residence
```



AND a.area\_of\_workplace not like 'OD%';

Finally export the table from PostGIS to an output GeoPackage using GDAL ogr2ogr command to enable the data to be easily shared and opened in QGIS.



T +44 (0) 1206 872143  
E [help@ukdataservice.ac.uk](mailto:help@ukdataservice.ac.uk)  
W [ukdataservice.ac.uk](http://ukdataservice.ac.uk)

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

© Copyright 2017  
University of Essex and  
University of Manchester

