

QGIS Tutorial for QStep Students

Reka Solymosi

19/06/2018

The purpose of this tutorial is to give you some experience putting some data on a map. Mapping data is a way of visualising your data in geographical space. The science dealing with spatial information is referred to as Geographical Information Science or GIS for short. GIS is a relatively new field — it started in the 1970's. It used to be that computerised GIS was only available to companies and universities that had expensive computer equipment. These days, anyone with a personal computer or laptop can use GIS software. Over time GIS Applications have also become easier to use — it used to require a lot of training to use a GIS Application, but now it is much easier to get started in GIS even for amateurs and casual users. As we described above, GIS is more than just software, it refers to all aspects of managing and using digital geographical data. In the tutorial that follows we will be focusing on GIS Software. This tutorial is intended as a **crash course** in dealing with spatial data.

However you should be aware that there is a whole field behind this, and that you be motivated to learn more about this area, in order to make sure that any application you create has grounding in this field. Make sure that you are drawing meaningful conclusions from your data, and that you are confident in understanding the meaning behind what you are presenting in your maps.

I strongly recommend that you consult some resources, such as these books:

- Longley, Paul. Geographic information systems and science. John Wiley & Sons, 2005.
(<https://books.google.co.uk/books?hl=en&lr=&id=toobg6OwFPEC>)
- De Smith, Michael John, Michael F. Goodchild, and Paul Longley. Geospatial analysis: a comprehensive guide to principles, techniques and software tools. Troubador Publishing Ltd, 2007.
(https://books.google.co.uk/books?id=SULMdT8qPwEC&dq=geospatial+analysis&source=gbs_navlinks_s)

A topic that this tutorial does not cover is that of **spatial relationships**. The First Law of Geography, according to Waldo Tobler (<http://www.tandfonline.com/doi/pdf/10.1111/j.1467-8306.2004.09402005.x>), is that “everything is related to everything else, but near things are more related than distant things.” This first law is the foundation of the fundamental concepts of spatial dependence and spatial autocorrelation. These concepts tend to account for spatial dependence in data, by using things like spatial weighting. We simply do not have time to cover these here, but it might be worth reading a bit around these concepts if you will be dealing with spatial data at all during your internship.

A quick introduction of terms

Geospatial Perspective - The Basics

Geospatial analysis provides a distinct perspective on the world, a unique lens through which to examine events, patterns, and processes that operate on or near the surface of our planet. Ultimately geospatial analysis concerns what happens where, and makes use of geographic information that links features and phenomena on the Earth's surface to their locations.

We can talk about a few different concepts when it comes to spatial information. These are:

- Place
- Attributes
- Objects

Place

At the center of all spatial analysis is the concept of *place*. People identify with places of various sizes and shapes, from the room with the parcel of land, to the neighbourhood, to the city, the country, the state or the nation state. Places often have names, and people use these to talk about and distinguish places. Names can be official. Places also change continually as people move. The basis of rigorous and precise definition of place is a coordinate system, a set of measurements that allows place to be specified unambiguously and in a way that is meaningful to everyone.

Attributes

Attribute has become the preferred term for any recorded characteristic or property of a place. A place's name is an obvious example of an attribute. But there can be other pieces of information, such as number of crimes in a neighbourhood, or the GDP of a country. Within GIS the term 'attributes' usually refers to records in a data table associated with individual features in a vector map or cells in a grid (raster or image file). These data behave exactly as data you have encountered in your data analysis courses. The rows represent observations, and the columns represent variables. The variables can be numeric or categorical, and depending on what they are, you can apply different methods to making sense of them.

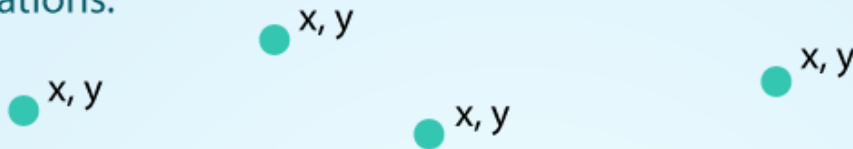
Objects

In spatial analysis it is customary to refer to places as objects. These objects can be a whole country, or a road. In studies of climate change, the objects of interest might be weather stations of minimal extent, and will be represented as points. On the other hand, studies of social or economic patterns may need to consider the two-dimensional extent of places, which will therefore be represented as areas. These representations of the world are part of what is called the vector data model: A representation of the world using points, lines, and polygons. Vector models are useful for storing data that have discrete boundaries, such as country borders, land parcels, and streets. This is made up of points, lines, and areas (polygons):

- Points
 - Points are pairs of coordinates, in latitude/longitude or some other standard system
- Lines
 - Lines are ordered sequences of points connected by straight lines
- Areas (polygons)
 - Areas are ordered rings of points, also connected by straight lines to form polygons. It can contain holes, or be linked with separate islands.

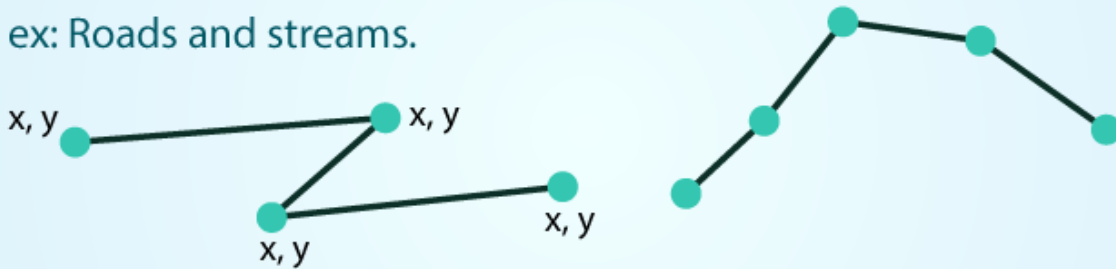
POINTS: Individual **x, y** locations.

ex: Center point of plot locations, tower locations, sampling locations.



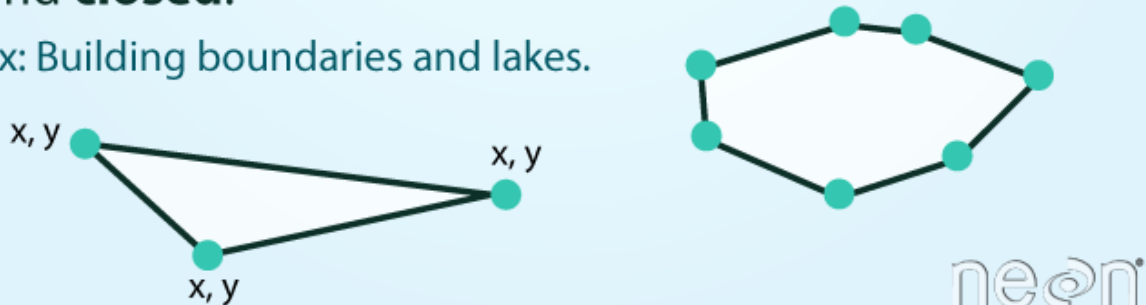
LINES: Composed of many (at least 2) vertices, or points, that are connected.

ex: Roads and streams.



POLYGONS: 3 or more vertices that are connected and **closed**.

ex: Building boundaries and lakes.



neon

Objects can also be Raster data. Raster data is made up of pixels (or cells), and each pixel has an associated value. Simplifying slightly, a digital photograph is an example of a raster dataset where each pixel value corresponds to a particular colour. In GIS, the pixel values may represent elevation above sea level, or chemical concentrations, or rainfall etc. The key point is that all of this data is represented as a grid of (usually square) cells. You will most likely be dealing with vector data in your internships, so we will be focusing on these.

Maps

Historically maps have been the primary means to store and communicate spatial data. Objects and their attributes can be readily depicted, and the human eye can quickly discern patterns and anomalies in a well-designed map.

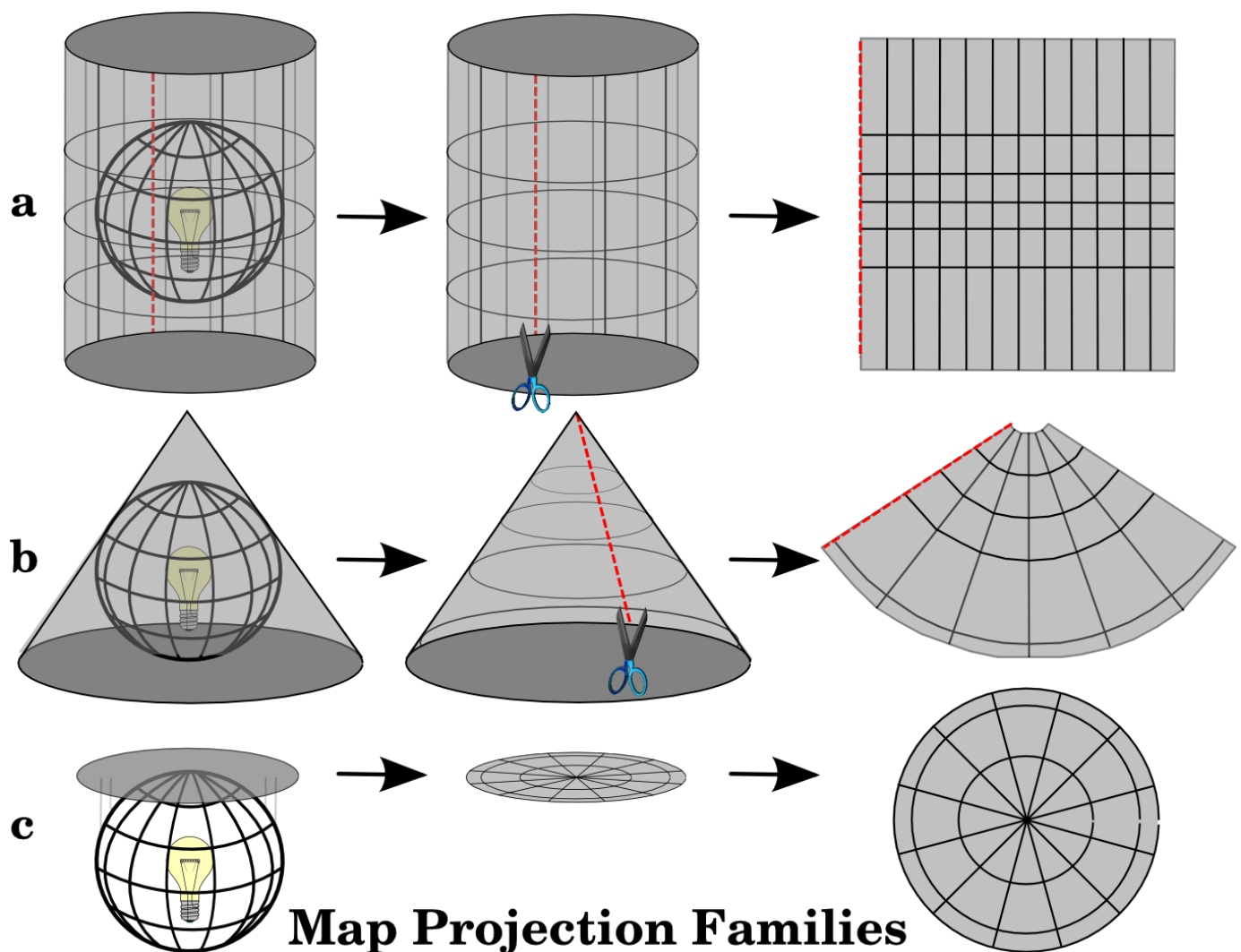
Map projections

Map projections try to portray the surface of the earth or a portion of the earth on a flat piece of paper or computer screen. A coordinate reference system (CRS) then defines, with the help of coordinates, how the two-dimensional, projected map in your GIS is related to real places on the earth. The decision as to which map

projection and coordinate reference system to use, depends on the regional extent of the area you want to work in, on the analysis you want to do and often on the availability of data.

A traditional method of representing the earth's shape is the use of globes. When viewed at close range the earth appears to be relatively flat. However when viewed from space, we can see that the earth is relatively spherical. Maps, are representations of reality. They are designed to not only represent features, but also their shape and spatial arrangement. Each map projection has advantages and disadvantages. The best projection for a map depends on the scale of the map, and on the purposes for which it will be used. For your purposes, you just need to understand that essentially there are different ways to flatten out the earth, in order to get it into a 2-dimensional map.

The process of creating map projections can be visualised by positioning a light source inside a transparent globe on which opaque earth features are placed. Then project the feature outlines onto a two-dimensional flat piece of paper. Different ways of projecting can be produced by surrounding the globe in a cylindrical fashion, as a cone, or even as a flat surface. Each of these methods produces what is called a map projection family. Therefore, there is a family of planar projections, a family of cylindrical projections, and another called conical projections see figure_projection_families (https://docs.qgis.org/2.8/en/_images/projection_families.png)

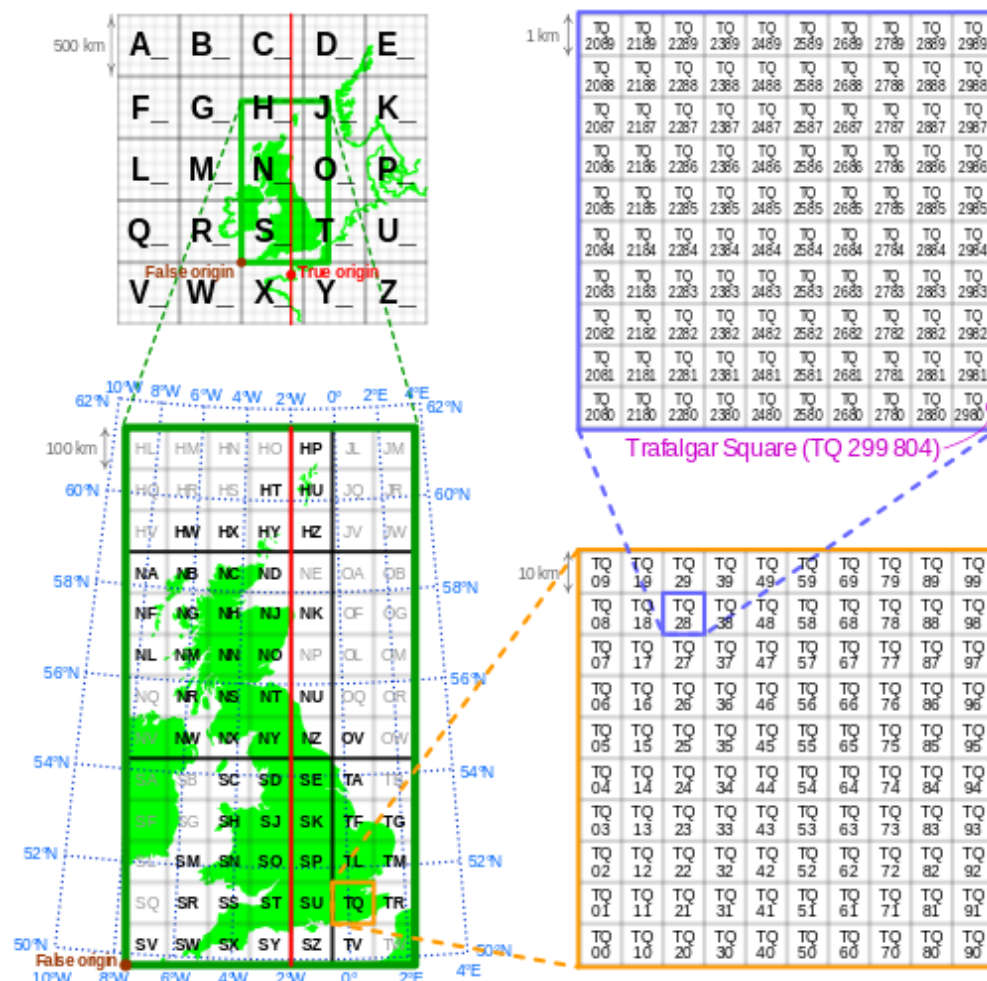


figure_projection_families

Coordinate Reference Systems

With the help of coordinate reference systems (CRS) every place on the earth can be specified by a set of three numbers, called coordinates. In general CRS can be divided into projected coordinate reference systems (also called Cartesian or rectangular coordinate reference systems) and geographic coordinate reference systems.

The use of Geographic Coordinate Reference Systems is very common. They use degrees of latitude and longitude and sometimes also a height value to describe a location on the earth's surface. The most popular is called WSG 84. This is the one you will most likely be using, and if you get your data in latitude and longitude, then this is the CRS you are working in. It is also possible that you will be using a projected CRS. This two-dimensional coordinate reference system is commonly defined by two axes. At right angles to each other, they form a so called XY-plane. The horizontal axis is normally labelled X, and the vertical axis is normally labelled Y. Working with data in the UK, you are most likely to be using British National Grid (BNG). The Ordnance Survey National Grid reference system is a system of geographic grid references used in Great Britain, different from using Latitude and Longitude. In this case, points will be defined by "Easting" and "Northing" rather than "Longitude" and "Latitude". It basically divides the UK into a series of squares, and uses references to these to locate something. The most common usage is the six figure grid reference, employing three digits in each coordinate to determine a 100 m square. For example, the grid reference of the 100 m square containing the summit of Ben Nevis is NN 166 712. Grid references may also be quoted as a pair of numbers: eastings then northings in metres, measured from the southwest corner of the SV square. For example, the grid reference for Sullom Voe oil terminal in the Shetland Islands may be given as HU396753 or 439668,1175316



BNG

This will be important later on when we are linking data from different projections, or when you look at your map and you try to figure out why it might look "squished".

Networks

We already mentioned lines that constitute objects of spatial data, such as streets, roads, railroads, etc. Networks constitute one-dimensional structures embedded in two or three dimensions. Discrete point objects may be distributed on the network, representing phenomena such as landmarks, or observation points. Mathematically, a network forms a graph, and many techniques developed for graphs have application to

networks. These include various ways of measuring a network's connectivity, or of finding the shortest path between pairs of points on a network. You can have a look at the lesson on network analysis in the QGIS documentation (https://docs.qgis.org/2.2/en/docs/training_manual/vector_analysis/network_analysis.html)

Density estimation

One of the more useful concepts in spatial analysis is density - the density of humans in a crowded city, or the density of retail stores in a shopping centre. Mathematically, the density of some kind of object is calculated by counting the number of such objects in an area, and dividing by the size of the area. To read more about this, I recommend Silverman, Bernard W. Density estimation for statistics and data analysis. Vol. 26. CRC press, 1986. (https://books.google.co.uk/books?id=e-xsrjsL7WkC&dq=silverman+density+estimation+for+statistics&lr=&source=gbs_navlinks_s)

Summary

Right so hopefully this gives you a few things to think about. Be sure that you are confident to know about:

- Spatial objects - what they are and how they are represented
- Attributes - the bits of information that belong to your spatial objects
- Maps and projections - especially what WSG84 and BNG mean, and why it's important that you know what CRS your data have

And if you're interested you can read up about density and networks. Again we are not really covering those here, but they are something that you might come accross in your internship, depending on what sort of data you will be working with.

The software

Before we start, let's familiarise ourselves with the software we will be using. Now it's possible that you will enter a workplace where they will be using different systems. Many government agencies might be using something called MapInfo (https://en.wikipedia.org/wiki/MapInfo_Professional). Agencies with a bit more money are likely to be using ESRI Arc GIS (<https://en.wikipedia.org/wiki/ArcGIS>). If you are forced to use these (and if there are people using this and there is support where you are, then you might as well) then don't worry, the concepts are more or less the same. You will be able to search the help function for the same terms, or look through any documentation available and search through that on how to exactly carry out what it is that you want to do. However the issue with these GIS is that they are proprietary, and they cost a lot of money to use. They also have paid-for training and support, which you cannot openly access. QGIS on the other hand is entirely free, and because it's open source, there is documentation and support available everywhere. If you are not sure how to do something in QGIS, all you have to do is GOOGLE IT, and there will be many useful answers for you to browse through and fix your issue. Similarly if you get an error message, just google it, and you will find help.

If you are using your own laptop, or the organisation gives you the option of choosing what you want to use, you can easily make a case for QGIS as it is totally free and therefore gives no cost to the organisation, or to you if you want to use it on your own laptop. Also because it is open source, anyone can contribute, and so you can have these plugins, which people write to help them solve very specific problems. But more about this in the next section, where I tell you all about QGIS.

QGIS

The main tool we will be using is QGIS. QGIS functions as geographic information system (GIS) software, allowing users to analyze and edit spatial information, in addition to composing and exporting graphical maps. If you are interested, you can learn more about QGIS here:

- About QGIS (<http://www.qgis.org/en/site/about/index.html>)

- See some case studies (http://www.qgis.org/en/site/about/case_studies/index.html)
- Browse some tutorials (<http://www.qgistutorials.com/en/>)

Plugins

QGIS has a variety of plugins that you can download, and use for your work. Plugins in QGIS add useful features to the software. Plugins are written by QGIS developers and other independent users who want to extend the core functionality of the software. These plugins are made available in QGIS for all the users.

You can see a tutorial for installing and using plugins here (http://www.qgistutorials.com/en/docs/using_plugins.html)

While we don't use plugins in tutorials, sometimes when you might be googling how to do something in QGIS the answer might be to use the "such and such plugin". In that case you will have to install the plugin first, to use it!

Getting Practical

Right, let's make some maps!

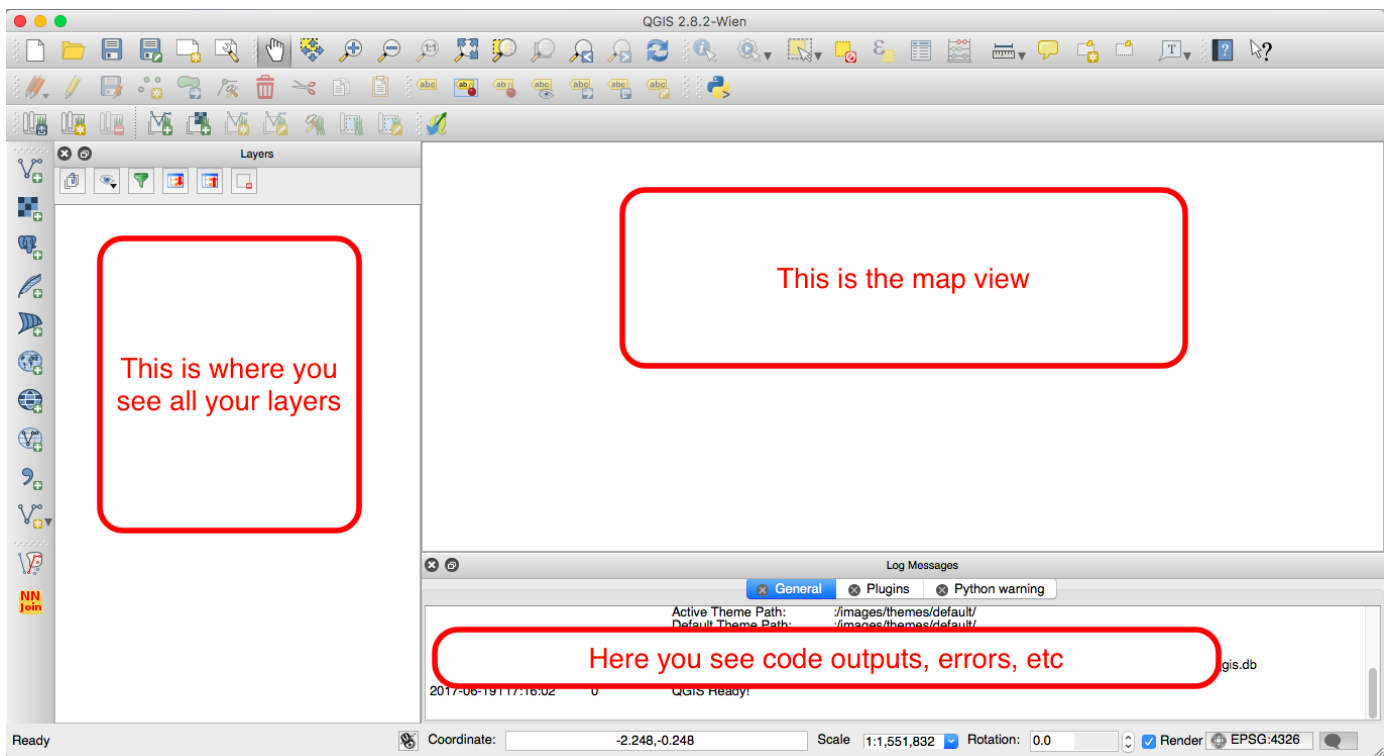
In this tutorial we will import a shapefile, we will consider its projection, and also join some tabular data, to be able to look at such data on a map. Then we will then import some x & y coordinates, we will consider the projection of that as well, and do some reprojection, to align our two spatial objects. We will then use a function called points in polygon to count the number of points in each polygon in the shapefile, and use this to create a thematic map. Hopefully while carrying out these tasks you will learn to:

- Read shapefiles into QGIS
- Understand the difference between spatial and non spatial data
- Know how to link non spatial data to shapefiles, in order to map it
- Read point data into QGIS
- Produce a thematic map for a report

So let's get started.

First things first, you have to create a folder where you will work. It's important to be organised and consistent. All your files should be saved in this one folder, both those you import and those you export from QGIS. So create a folder for your work first.

Then, let's open up the QGIS software.



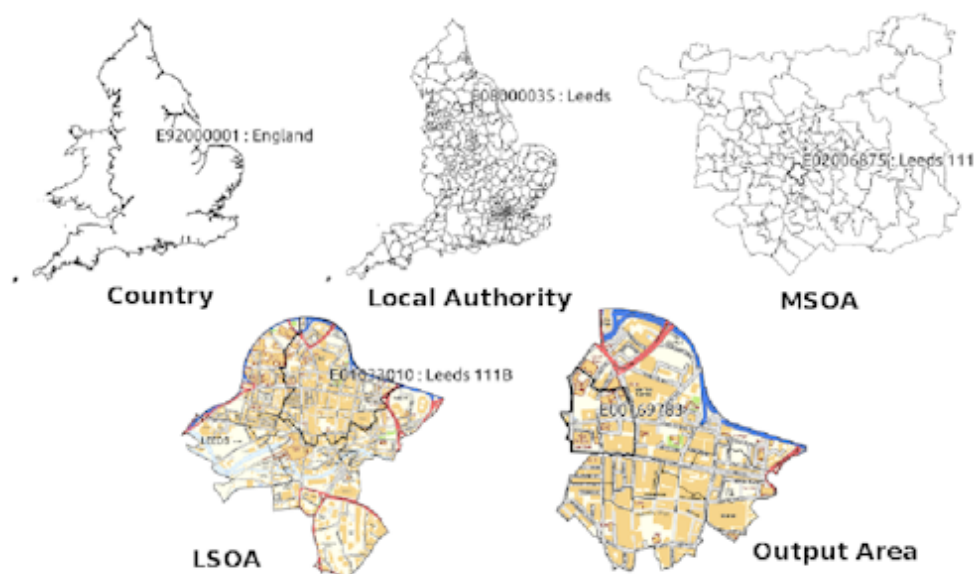
Take a moment to have a look at this, you can see there are quite a lot of buttons there on the side and the top. It might be worth going through this short video (<https://www.youtube.com/watch?v=Z87aAz3M8-0>) that outlines the interface for you quickly (and gives some more tips about why QGIS is a great choice for GIS)

We will return to this shortly. But let's step away for a moment, and talk about how to get some data.

Find a shapefile

You will often need a boundary shapefile for your data analysis. Sometimes you will be given spatial data to begin with, such as a shapefile, or point coordinates with latitudes and longitudes (or eastings and northings). But other times you might have to find this yourself, and join the non-spatial data to these. This latter case is what the first part of this tutorial will demonstrate. In this case, you will have to source the spatial data yourself.

You can acquire spatial data from various sources. An example is Census Boundary Data. You can read more about that here (<https://census.ukdataservice.ac.uk/use-data/guides/boundary-data>). "Boundary data are a digitised representation of the underlying geography of the census". Census Geography is often used in research and spatial analysis because it is divided into units based on population counts, created to form comparable units, rather than administrative boundaries such as wards or police force areas. However depending on your research question and the context for your analysis, you might be using different units. The hierarchy of the census geographies goes from Country to Local Authority to Middle Layer Super Output Area (MSOA) to Lower Layer Super Output Area (LSOA) to Output Area:



Here we will get some boundaries for Manchester. Let's use the Lower Super Output Area (LSOA) level. LSOAs are geographical regions designed to be more stable over time and consistent in size than existing administrative and political boundaries. LSOAs comprise, on average, 600 households that are combined on the basis of spatial proximity and homogeneity of dwelling type and tenure. Neighbourhoods are often operationalised as LSOAs.

So to get some boundary data, you can use the UK Data Service website. There is a simple Boundary Data Selector (<https://borders.ukdataservice.ac.uk/bds.html>)(link text: <https://borders.ukdataservice.ac.uk/bds.html> (<https://borders.ukdataservice.ac.uk/bds.html>))

When you get to the link, you will see on the top there is some notification to help you with the boundary data selector. If you are feeling unsure at any point, feel free to click on that help to guide you.

For now, let's focus on the selector options. Here you can choose the country you want to select shapefiles for. We select "England". You can also choose the type of geography we want to use. Here we select "Statistical Building Block", as discussed above. And finally you can select when you want it for. If you are working with historical data, it makes sense to find boundaries that match the timescale for your data. Here we will be dealing with contemporary data, and therefore we want to be able to use the newest available boundary data.

What and Where?	Format
<p>For help with using the Boundary Data Selector interface, choose one of the following links:</p> <p>How to use Boundary Data Selector link</p>	
<p>Select:</p> <div> <div>England</div> <div>Statistical Building Block</div> <div>2011 and later</div> <div>Find</div> </div>	

Once you have selected these options, click on the "Find" button. That will populate the box below:

Boundaries

English Census Output Areas, 2011
English Lower Layer Super Output Areas, 2011
English Middle Layer Super Output Areas, 2011

Area(s) of Interest: English Districts, UAs and London Boroughs, 2011 (Derived by aggregating 2011 Output Areas using exact-fit OA to higher geography lookup table)

List Areas

Here you can select the boundaries we want. As discussed, we want the census lower super output areas. But again, your choice here will depend on what data you want to be mapping.

Once you've made your choice, click on "List Areas". This will now populate the box below. We are here concerned with Manchester. However you can select more than one if you want boundary for more than one area as well. Just hold down "ctrl" to select multiple areas individually, or the shift key to select everything in between.

Liverpool
Luton
Maidstone
Maldon
Malvern Hills
Manchester
Mansfield
Medway

List Areas

Expand Selection

Previous List

Search Summary:	[Target Geography : English Census Output Areas with OAC, 2011] [Selected Area(s) : Manchester]
Data Format:	SHAPE
Archive Method:	Zip

Extract Boundary Data

Once you've made your decision click on the "Extract Boundary Data" button. You will see the following message:

The geospatial dataset you requested through the UK Data Service is being prepared for you but it may take some time before it is ready to download.

You will be automatically redirected to our data download bookmarking facility if after 3 minutes your data download is still being prepared for you.

To go to the bookmarking facility now click on the Go To Bookmark button below.

[Go To Bookmark](#)

You can bookmark, or just stay on the page and wait. How long you have to wait will depend on how much data you have requested to download.

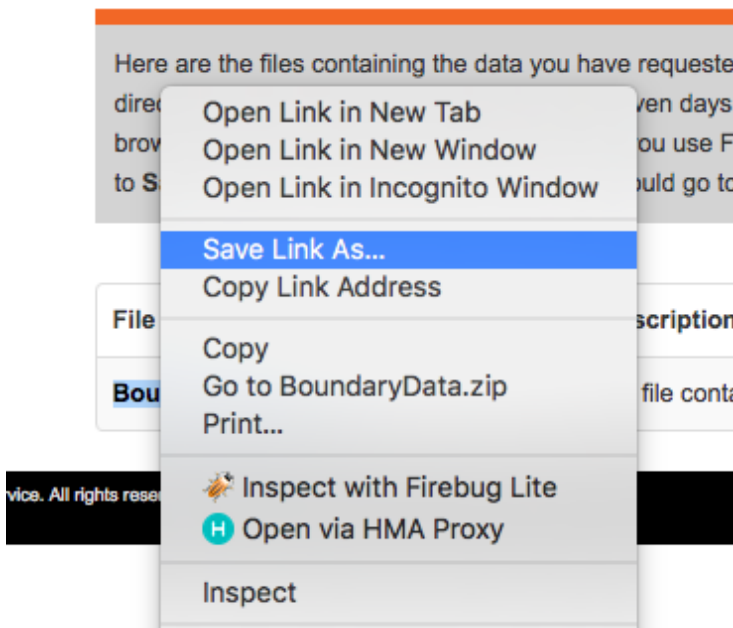
When your data is read, you will see the following message:

FILES EXTRACTED

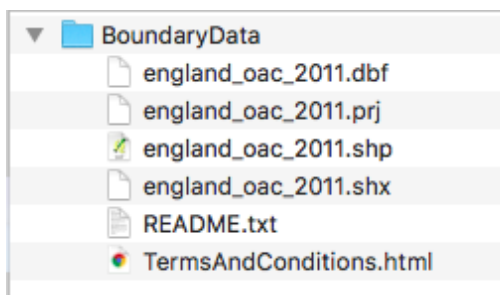
Here are the files containing the data you have requested from Census Support, together with their type and size. The files in this directory will be automatically deleted within seven days of being generated. You can download an individual file using your web browser by clicking on its name in the table. If you use Firefox, 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.

File Name	Description	Size	
BoundaryData.zip	Zip file containing your data	3 MB	

You have to right click on the “BoundaryData.zip”, and hit Save Target as on a PC or Save Link As on a Mac:



Navigate to the folder you have created for this analysis, and save the .zip file there. Extract the file contents using whatever you like to use to unzip compressed files. You should end up with a folder called “BoundaryData”. Have a look at its contents:



So you can see immediately that there are some documentations around the usage of this shapefile, in the readme and the terms and conditions. Have a look at these as they will contain information about how you can use this map. For example, all your maps will have to mention where you got all the data from. So since you got this boundary data from the UKDS, you will have to note the following:

“Contains National Statistics data © Crown copyright and database right [year] Contains OS data © Crown copyright [and database right] (year)”

You can read more about this in the terms and conditions document.

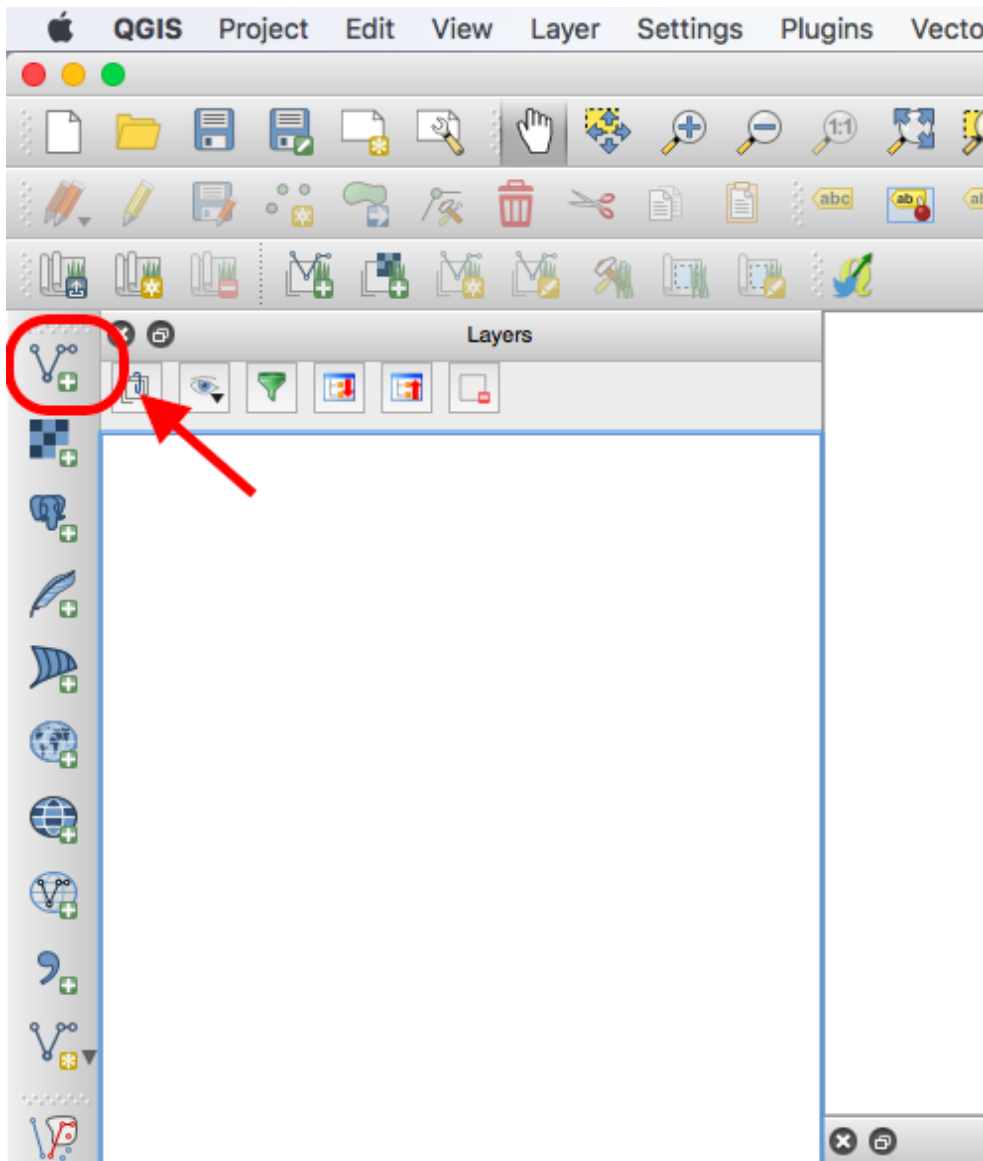
But then you will also notice that there are 4 files with the same name “england_oac_2011”. It is important that you keep all these files in the same location as each other! They all contain different bits of information about your shapefile:

- .shp — shape format; the feature geometry itself - this is what you see on the map
- .shx — shape index format; a positional index of the feature geometry to allow seeking forwards and backwards quickly
- .dbf — attribute format; columnar attributes for each shape, in dBase IV format.
- .prj — projection format; the coordinate system and projection information, a plain text file describing the projection using well-known text format

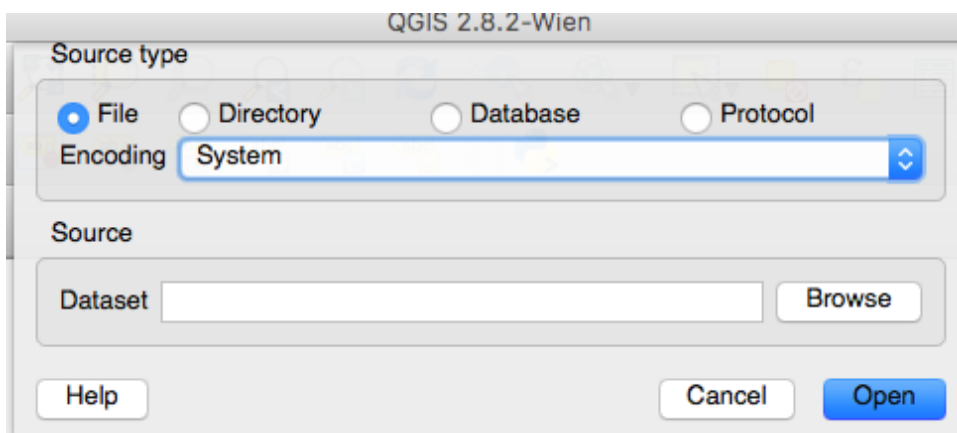
Sometimes there might be more files associated with your shapefile as well, but we will not cover them here.

Getting spatial data into QGIS

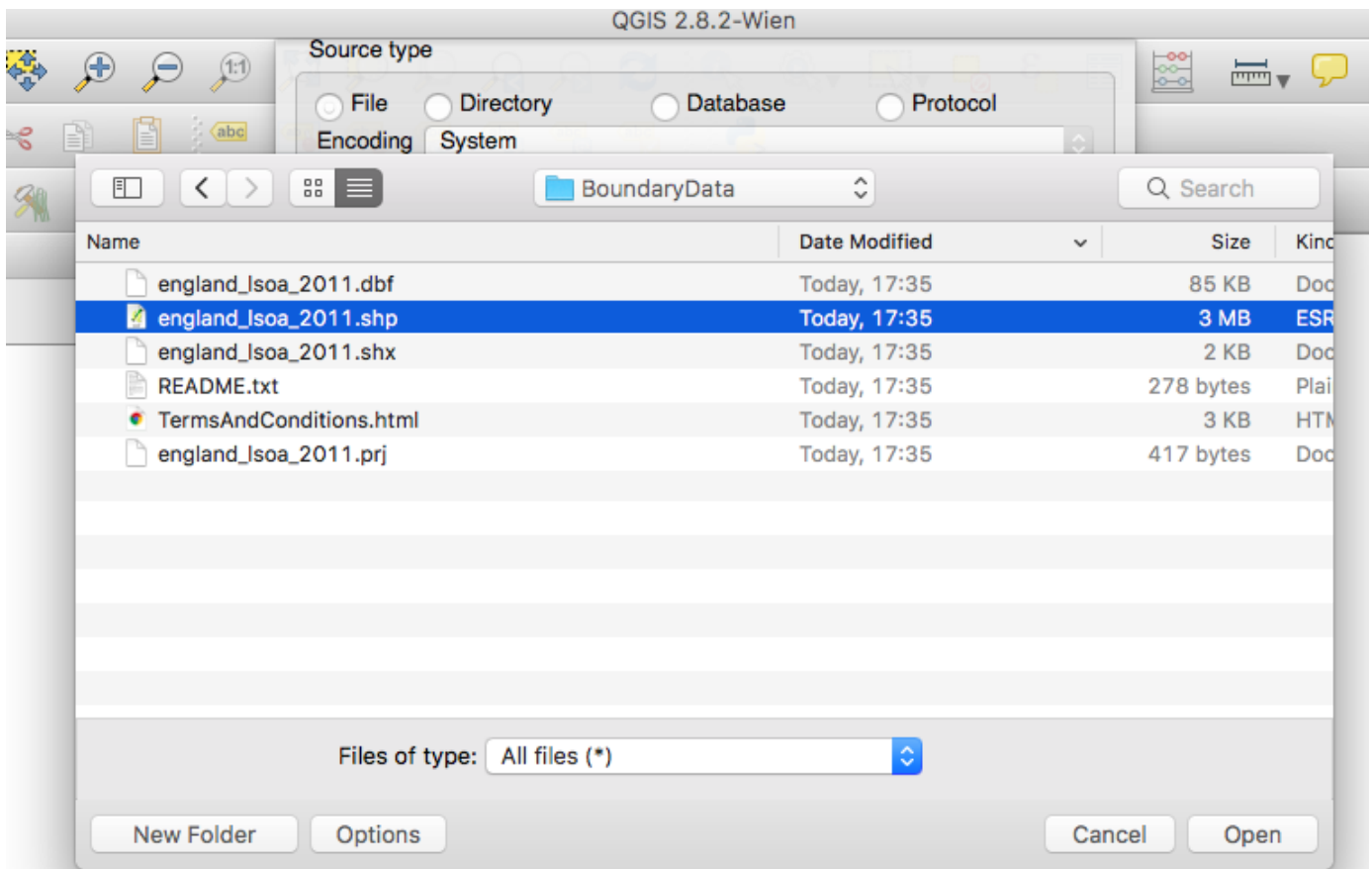
There are two ways to open up a vector shapefile in QGIS. One is to use the “Add Vector Layer” button on the top left hand side:



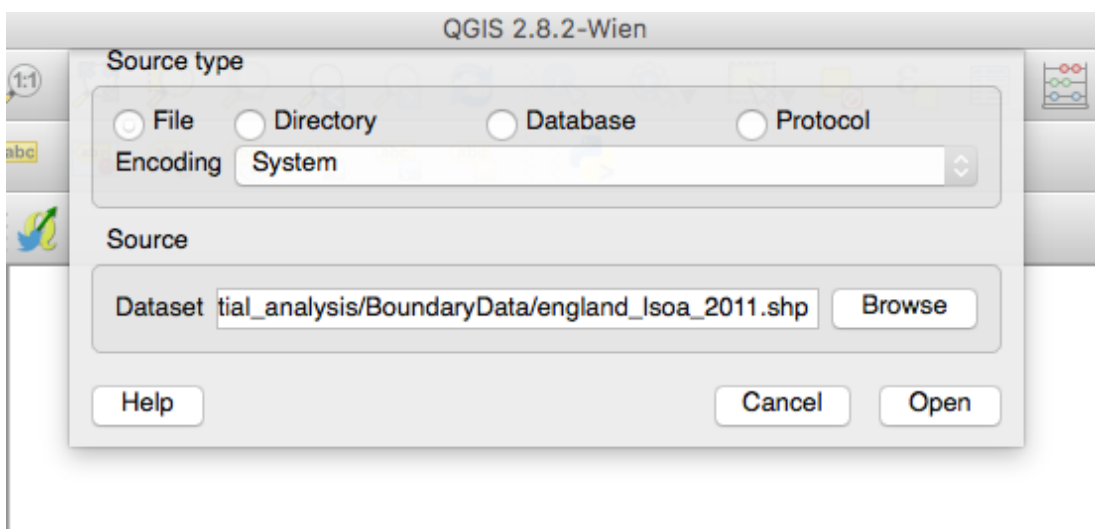
This will open up a dialogue box:



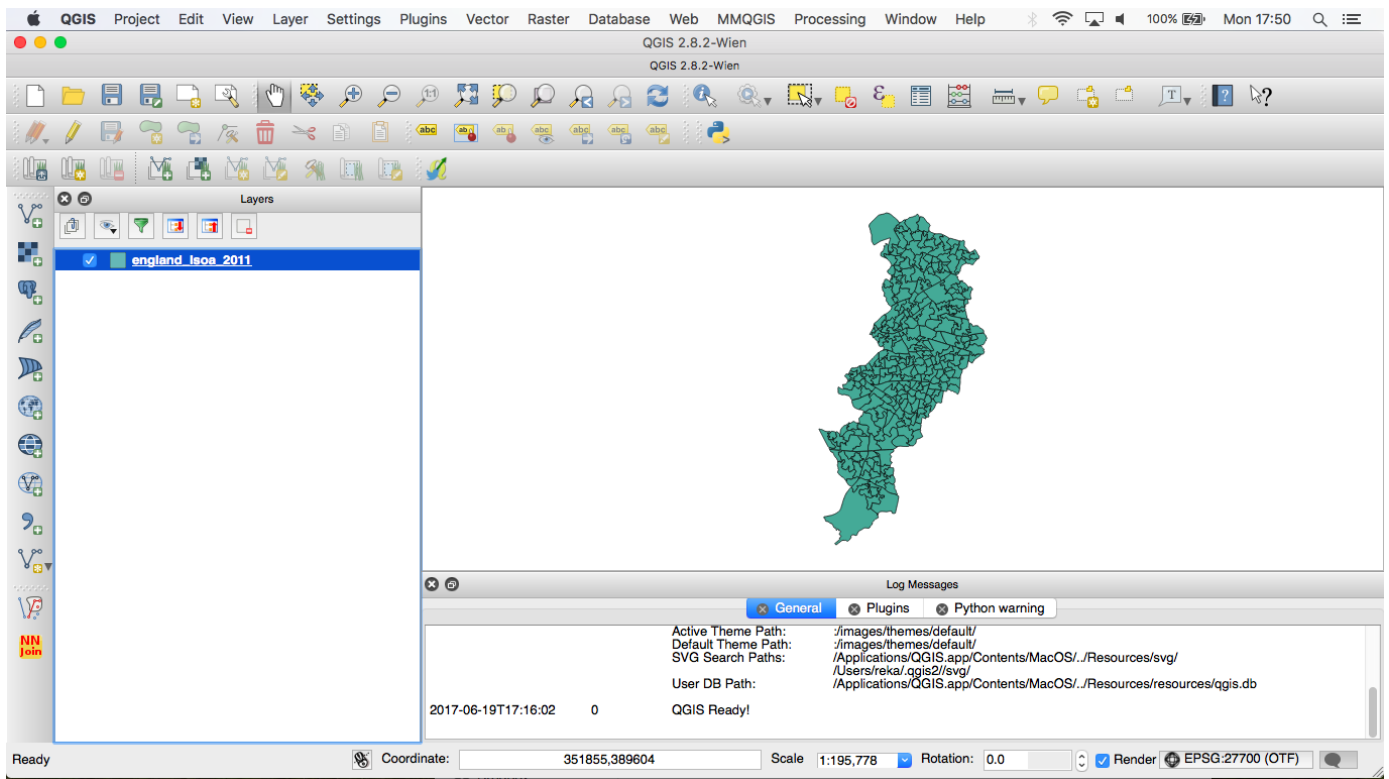
Where you can click on the “Browse” button and navigate to your shapefile, and select it. **Make sure you are choosing the one with the .shp file extension.**



Select the file, click on open, and then you will be taken back to the dialogue box where you click “Open”:



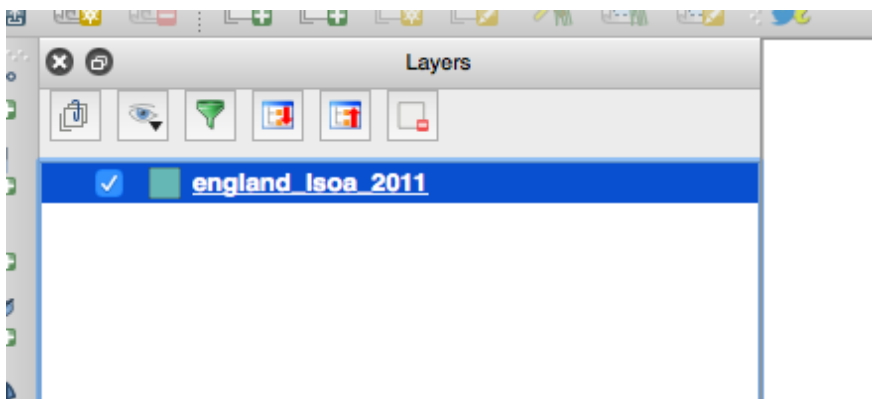
Now you will be able to see your shapefile! Yay!



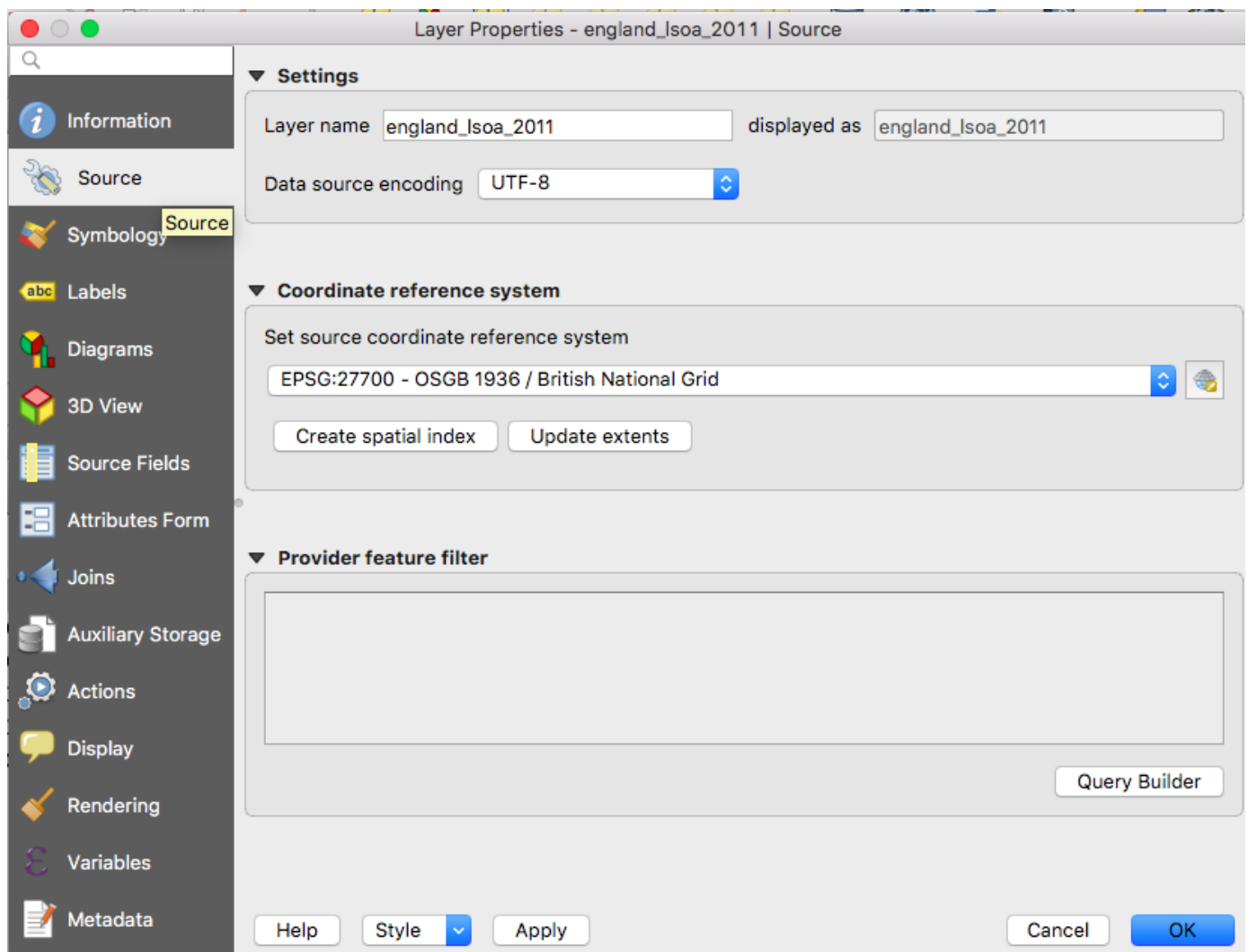
The other way is to very simply drag and drop the shapefile into the QGIS map window. Again, make sure that you are doing this with the file that has the `.shp` extension!

Find out about your shapefile

Once you have your shapefile in the QGIS environment, you can find out some information about it. You can double click its name in your layers pane, to open up a new window with this information.

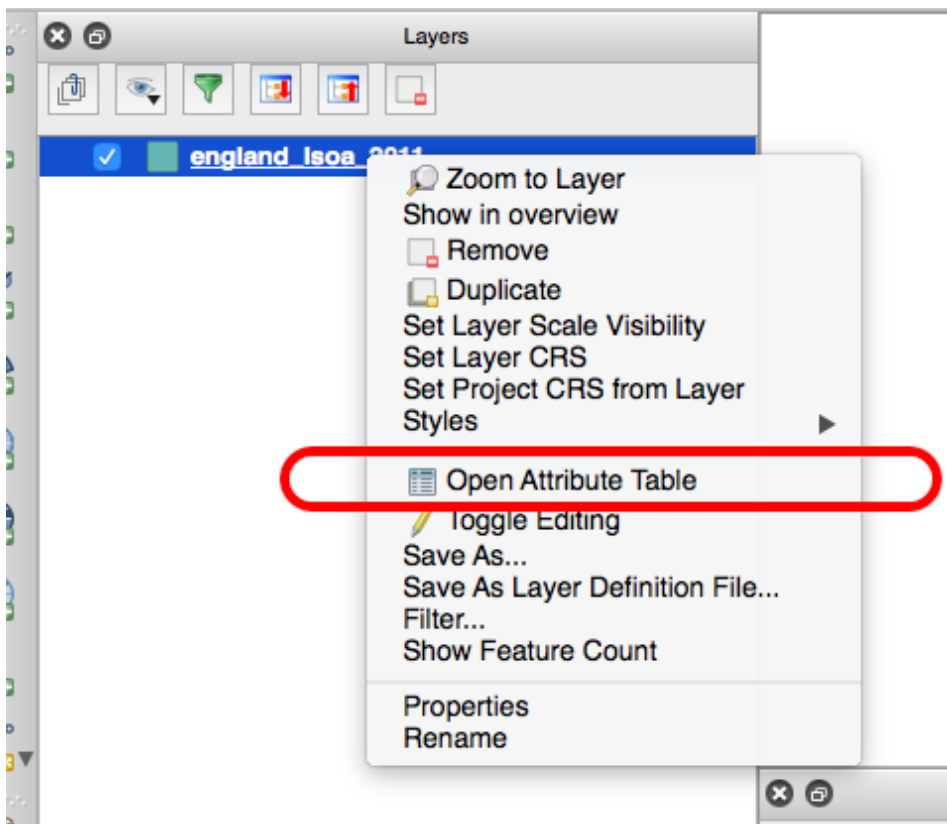


If you click on the Source tab, it tells you some information about your layer, for example it tells you the CRS. You can see here that the CRS is British National Grid. If it weren't we could change it in this window as well, using the drop down menu here.



You can also have a look at other tabs, but we won't get to that just now. Instead we will have a look at the attribute table (if you just wanted to look at the columns in the attribute data you could click the "field" tab here)

So close the window, and this time right click the layer name, and choose "open attribute table"



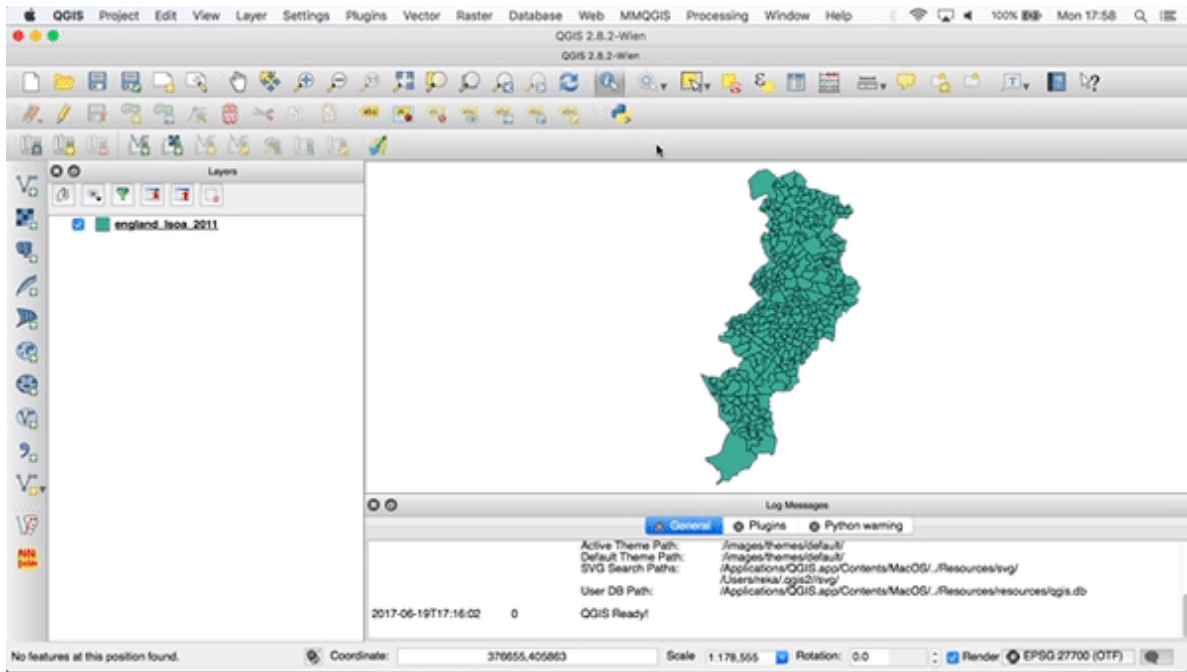
This will open up a table that should look familiar. As discussed, your rows are your observations, and your columns your variables.

The screenshot shows the 'Attribute Table' window in QGIS. The table has 15 rows and 4 columns: 'label', 'name', and 'code'. The first row is highlighted in green. The data is as follows:

	label	name	code
0	E08000003E...	Manchester ...	E01005066
1	E08000003E...	Manchester ...	E01005073
2	E08000003E...	Manchester ...	E01005061
3	E08000003E...	Manchester ...	E01005062
4	E08000003E...	Manchester ...	E01005063
5	E08000003E...	Manchester ...	E01005065
6	E08000003E...	Manchester ...	E01005067
7	E08000003E...	Manchester ...	E01005078
8	E08000003E...	Manchester ...	E01005068
9	E08000003E...	Manchester ...	E01005097
10	E08000003E...	Manchester ...	E01005069
11	E08000003E...	Manchester ...	E01005070
12	E08000003E...	Manchester ...	E01005071
13	E08000003E...	Manchester ...	E01005072
14	E08000003E...	Manchester ...	E01005074

Not too many variables in there at the moment.

You can also get information from the shapefile. By clicking on the little information arrow, and then with that selecting a LSOA, you get the information from that LSOA:



Right now let's consider that we want to add some data to this map.

Getting non spatial data into QGIS

So it's quite easy and intuitive to get spatial data in here. But non spatial data can only be mapped if it is linked with spatial information. The main way that this will happen, is that your non-spatial data will have some sort of spatial information still included with it. Remember in the slides the country names column? Those names can be matched to the **polygons** for those countries. I will demonstrate here with some police recorded crime data, which can be downloaded from the police.uk (<https://data.police.uk/data/>) website.

Let's stick local and download some data for crime in Manchester.

To do this, open the data.police.uk/data (<https://data.police.uk/data/>) website and follow instructions to download the data.

- In Date range just select June 2016 - June 2016
- In Force find Greater Manchester Police , and tick the box next to it.
- In Data sets tick Include crime data .
- Finally click on Generate File button.

This will take you to a download page, where you have to click the Download now button. This will open a dialogue to save a .zip file. Navigate to the working directory folder you've created and save it there.

Unzip the file.

If you want you can have a look at this data in Excel:

The screenshot shows an Excel spreadsheet with the following columns: Crime ID, Month, Reported by, Falls within, Longitude, Latitude, Location, LSOA code, LSOA name, Crime type, Last outcome, and Context. The data rows show various crime incidents from 2016-06, mostly in the Greater Manchester area, with LSOA codes like Bolton 001A and Bolton 001B.

Crime ID	Month	Reported by	Falls within	Longitude	Latitude	Location	LSOA code	LSOA name	Crime type	Last outcome	Context
1	2016-06	Greater Manch	Greater Manch	-2.464422	53.612495	On or near Par	E01004768	Bolton 001A	Anti-social behaviour		
2	2016-06	Greater Manch	Greater Manch	-2.441166	53.616044	On or near Piti	E01004768	Bolton 001A	Anti-social behaviour		
3	2016-06	Greater Manch	Greater Manch	-2.440584	53.610958	On or near Wh	E01004768	Bolton 001A	Anti-social behaviour		
4	2016-06	Greater Manch	Greater Manch	-2.441166	53.616044	On or near Piti	E01004768	Bolton 001A	Violence and s	Local resolution	
5	2016-06	Greater Manch	Greater Manch	-2.441166	53.616044	On or near Piti	E01004768	Bolton 001A	Other crime	Local resolution	
6	2016-06	Greater Manch	Greater Manch	-2.442588	53.62624	On or near De	E01004803	Bolton 001B	Anti-social behaviour		
7	2016-06	Greater Manch	Greater Manch	-2.442588	53.62624	On or near De	E01004803	Bolton 001B	Anti-social behaviour		
8	2016-06	Greater Manch	Greater Manch	-2.442588	53.62624	On or near De	E01004803	Bolton 001B	Anti-social behaviour		
9	2016-06	Greater Manch	Greater Manch	-2.445531	53.630013	On or near Lor	E01004803	Bolton 001B	Violence and s	Local resolution	
10	2016-06	Greater Manch	Greater Manch	-2.438577	53.627297	On or near Bri	E01004804	Bolton 001C	Anti-social behaviour		
11	2016-06	Greater Manch	Greater Manch	-2.438577	53.627297	On or near Bri	E01004804	Bolton 001C	Anti-social behaviour		
12	2016-06	Greater Manch	Greater Manch	-2.435403	53.628909	On or near Bar	E01004804	Bolton 001C	Anti-social behaviour		
13	2016-06	Greater Manch	Greater Manch	-2.438458	53.626031	On or near Ma	E01004804	Bolton 001C	Criminal dama	Investigation complete; no suspect identified	
14	2016-06	Greater Manch	Greater Manch	-2.438577	53.627297	On or near Bri	E01004804	Bolton 001C	Criminal dama	Investigation complete; no suspect identified	
15	2016-06	Greater Manch	Greater Manch	-2.428078	53.622221	On or near Upj	E01004807	Bolton 001D	Other theft	Investigation complete; no suspect identified	
16	2016-06	Greater Manch	Greater Manch	-2.428078	53.622221	On or near Upj	E01004807	Bolton 001D	Other theft	Investigation complete; no suspect identified	
17	2016-06	Greater Manch	Greater Manch	-2.427535	53.619401	On or near Thx	E01004807	Bolton 001D	Vehicle crime	Investigation complete; no suspect identified	
18	2016-06	Greater Manch	Greater Manch	-2.423752	53.619073	On or near Elr	E01004807	Bolton 001D	Violence and s	Investigation complete; no suspect identified	
19	2016-06	Greater Manch	Greater Manch	-2.430268	53.617584	On or near Hill	E01004808	Bolton 001E	Anti-social behaviour		
20	2016-06	Greater Manch	Greater Manch	-2.430224	53.619211	On or near Lo	E01004808	Bolton 001E	Anti-social behaviour		
21	2016-06	Greater Manch	Greater Manch	-2.429102	53.614425	On or near Co	E01004808	Bolton 001E	Vehicle crime	Investigation complete; no suspect identified	
22	2016-06	Greater Manch	Greater Manch	-2.428361	53.623353	On or near Ro	E01004808	Bolton 001E	Violence and s	Investigation complete; no suspect identified	
23	2016-06	Greater Manch	Greater Manch	-2.432717	53.623535	On or near Lo	E01004808	Bolton 001E	Other crime	Investigation complete; no suspect identified	
24	2016-06	Greater Manch	Greater Manch	-2.400062	53.60773	On or near Thx	E01004788	Bolton 002A	Burglary	Status update unavailable	
25	2016-06	Greater Manch	Greater Manch	-2.395748	53.607043	On or near Mn	E01004788	Bolton 002A	Burglary	Defendant found not guilty	

You will see that there are actually coordinates associated with this data. Let's ignore that for a second, and pretend that there is no spatial information here. Instead we have another column, which has another type of spatial information, similar to that of the country name in the slides. This column that you might notice is the one called **LSOA code**. If you recall our variables in the LSOA boundary data file, one of these was **code**. These match in the two data sets. So by linking the non-spatial data to the spatial data using these codes, you can put the data on a map!

OK let's say we want to map the number of crimes in each neighbourhood (neighbourhood measured as LSOA). For this we first want to create a frequency table. You may have encountered this in your data analysis courses or in Excel Q-step training, but just real quick, you should probably familiarise yourselves with pivot tables in excel.

Creating a frequency table in Excel

Making a frequency table in Excel is quite simple, and it is achieved by using something called a **pivot table**. As far as I know this name is specific to Excel. If you apply to public sector jobs, especially where excel is a requirement, the word pivot table is likely to come up in interview. It's a handy tool for summarising categorical data. A pivot table is a tool that lets you build different types of summary tables from your data. One of these is a frequency table.

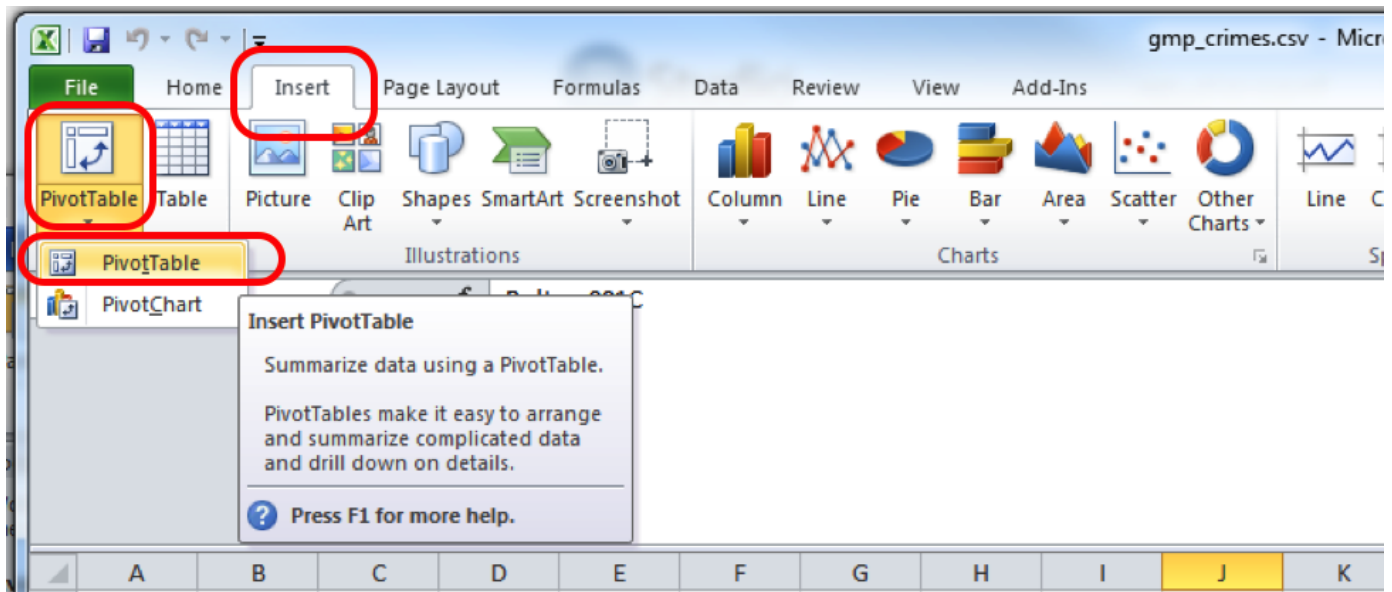
PivotTables are a great way to summarize, analyze, explore, and present your data, and you can create them with just a few clicks. PivotTables are highly flexible and can be quickly adjusted depending on how you need to display your results.

- The microsoft excel sales pitch (<https://support.office.com/en-gb/article/Create-a-PivotTable-to-analyze-worksheet-data-a9a84538-bfe9-40a9-a8e9-f99134456576>)

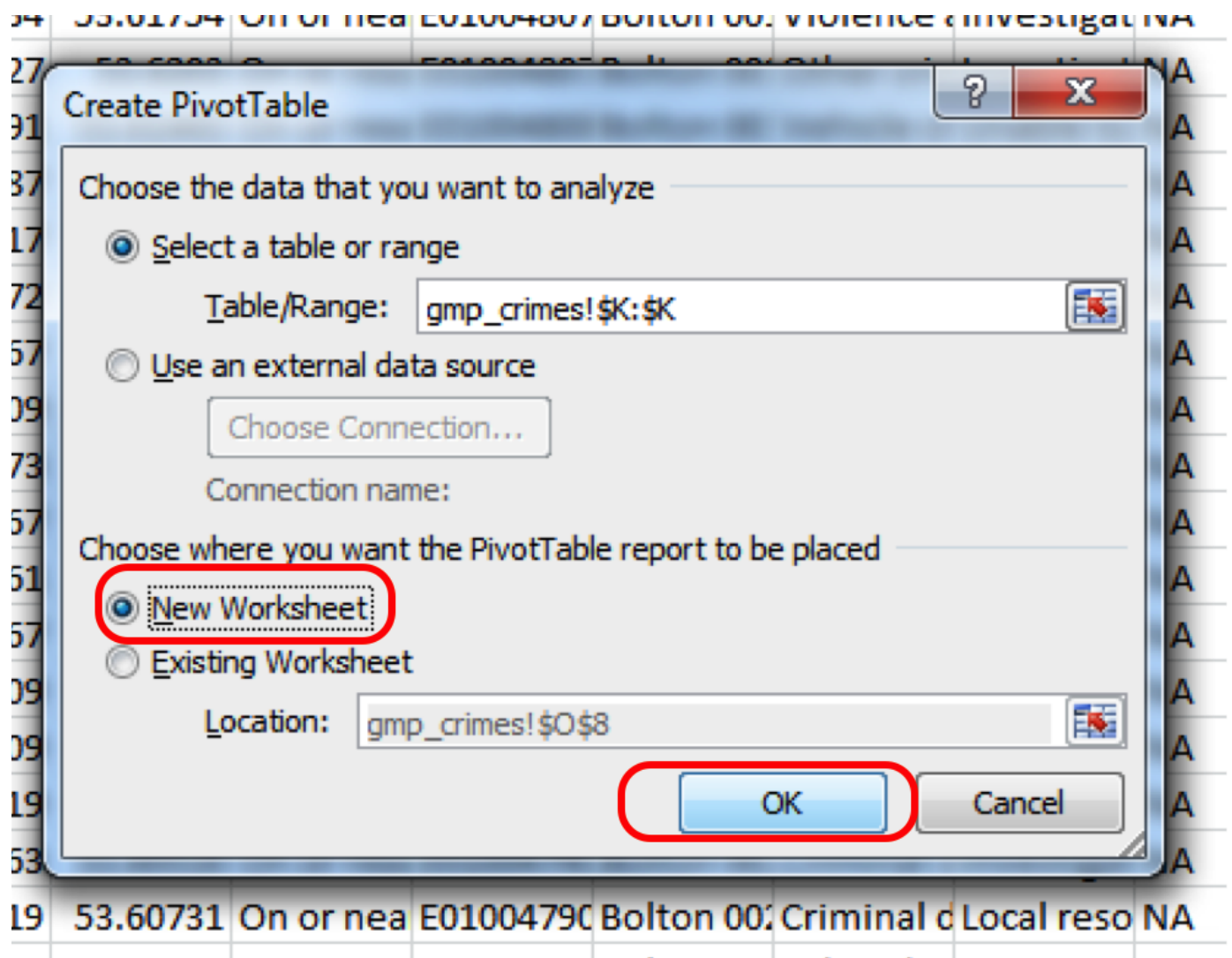
If you want to go a bit further in the pivot table knowledge, here's a handy list of 23 things you should know about pivot tables (<https://exceljet.net/things-to-know-about-excel-pivot-tables>). I like it because it's a list, and BuzzFeed has taught me that all information is best presented in list format, preferably with a random number of items in the list, like 23. We'll cover most of these items during the upcoming weeks.

So for now, we will now use a pivot table to create a frequency table of the crime type variable in the GMP crimes data. To do this, go to your gmp_crimes data set, opened up in Excel. Download the data from BB, as we did last week. If it's not downloading for any reason stick up your hand, we can come around and trouble shoot this for you! Now once you have the data open in Excel, you can easily create a frequency table following the below steps:

First you will have to select the pivot table option. Click into the **Insert** tab, click on **pivot table** and then again on **pivot table**:



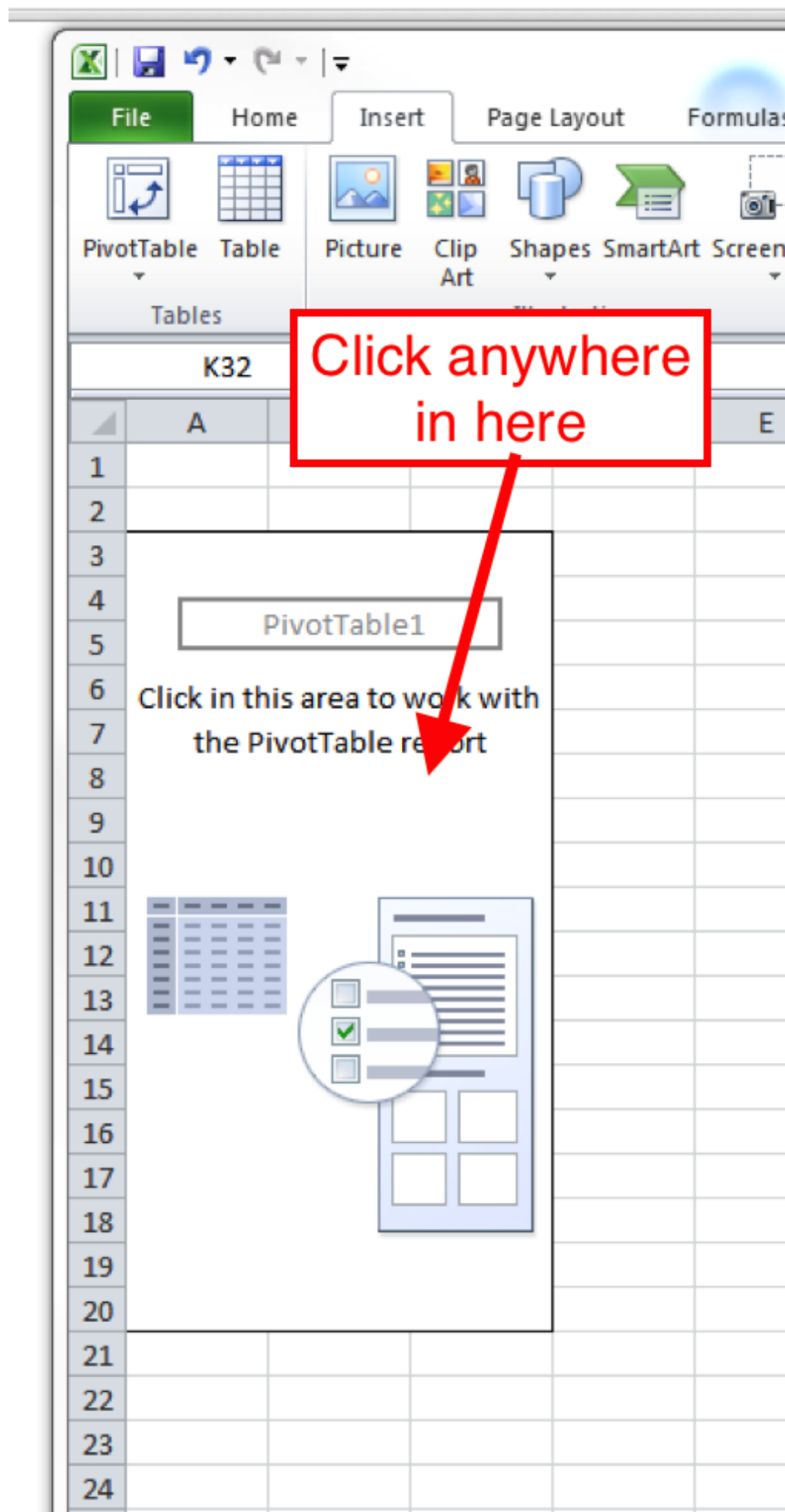
This will open a popup window, where you want to make sure that you select 'New worksheet' where it asks where your pivot table should be placed, and then click OK:



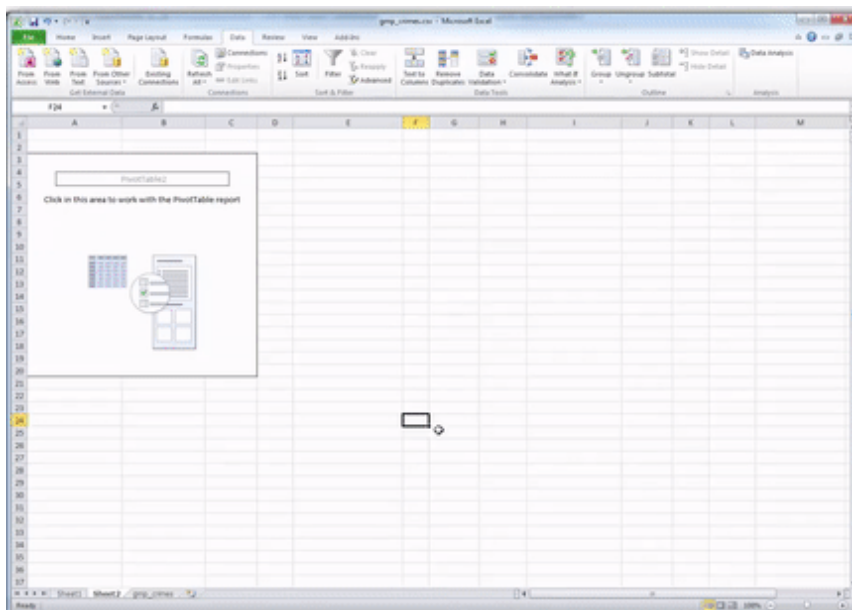
Don't worry too much about the top option where you select your data, because the pivot table will let you select your variables retrospectively. But just make sure the 'Select table or range' option is selected, and not the 'use external data source one'.

Now when you click OK, excel should take you to the new worksheet where it has set up a pivot table for you, ready to get into your data.

It might also open a toolbar on the side, but it might not do this automatically. In any case, if the toolbar ever disappears, to summon it you have to do one simple step, which is to click *anywhere* inside the pivot table area:



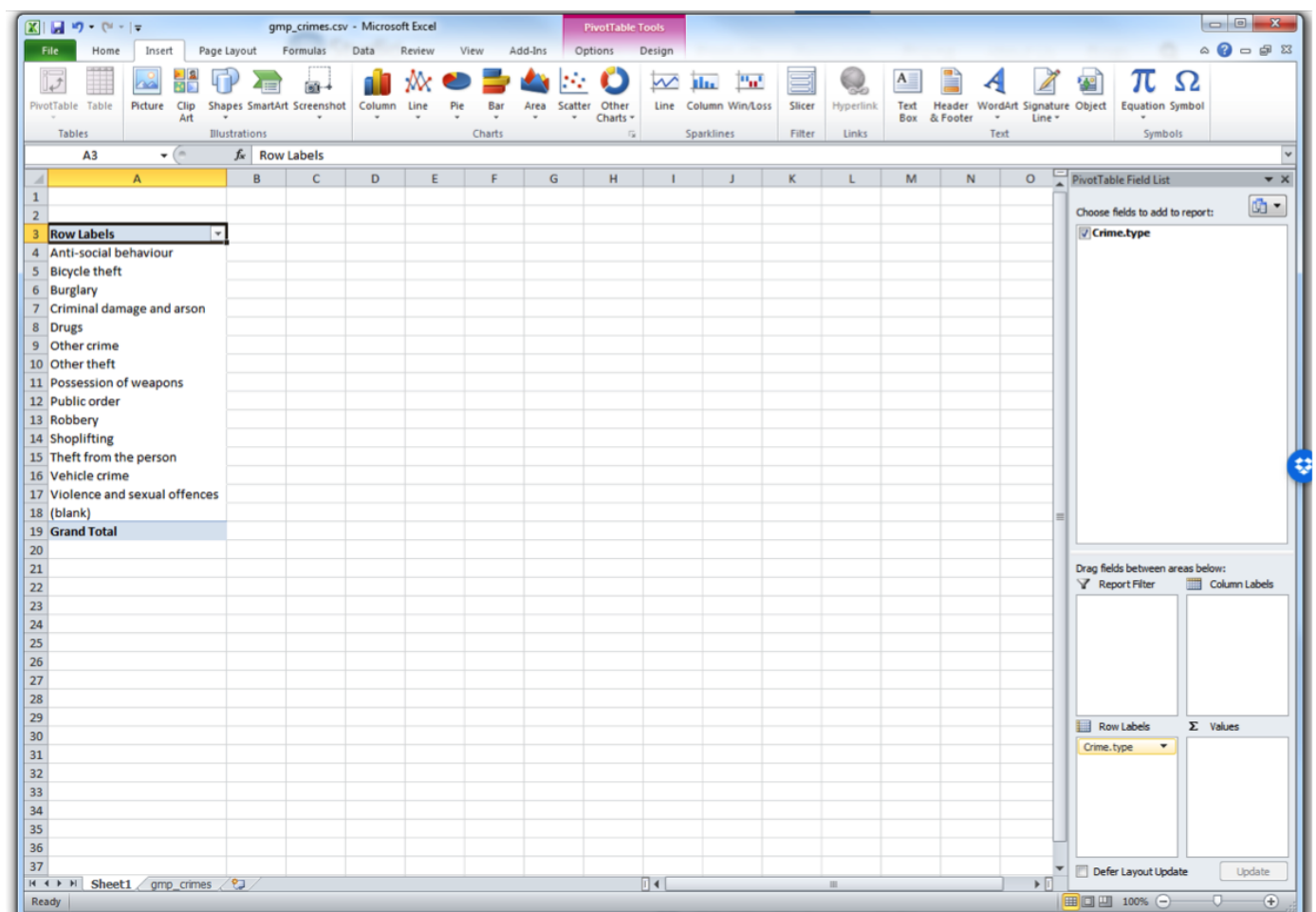
Once you do that, a navigation pane should appear. Just like this:



Now you should see all your variables on the side there as well, in this little panel that has just appeared.

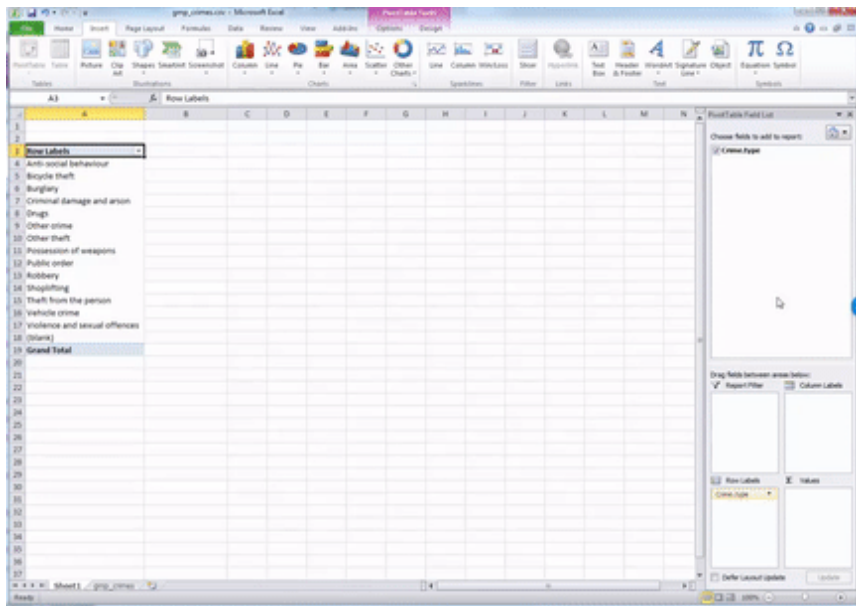
You can scroll through and find the variable “LSOA code”. This is the variable we want to look at in this case.

You can see four windows within the pivot table panel. You’ve got **Filters**, **Columns**, **Rows**, and **Values**. You can drag your variables into these boxes in order to create a table. Whatever you drag into the Columns box becomes the columns, and whatever you drag into the Rows box becomes the Rows. Try it out, drag “LSOA code” into the Rows box. You should see a list of all the possible values that the crime type variable can take in the rows. Now drag it over to columns box, and you’ll see it across there. Drag it back to rows and leave it there:

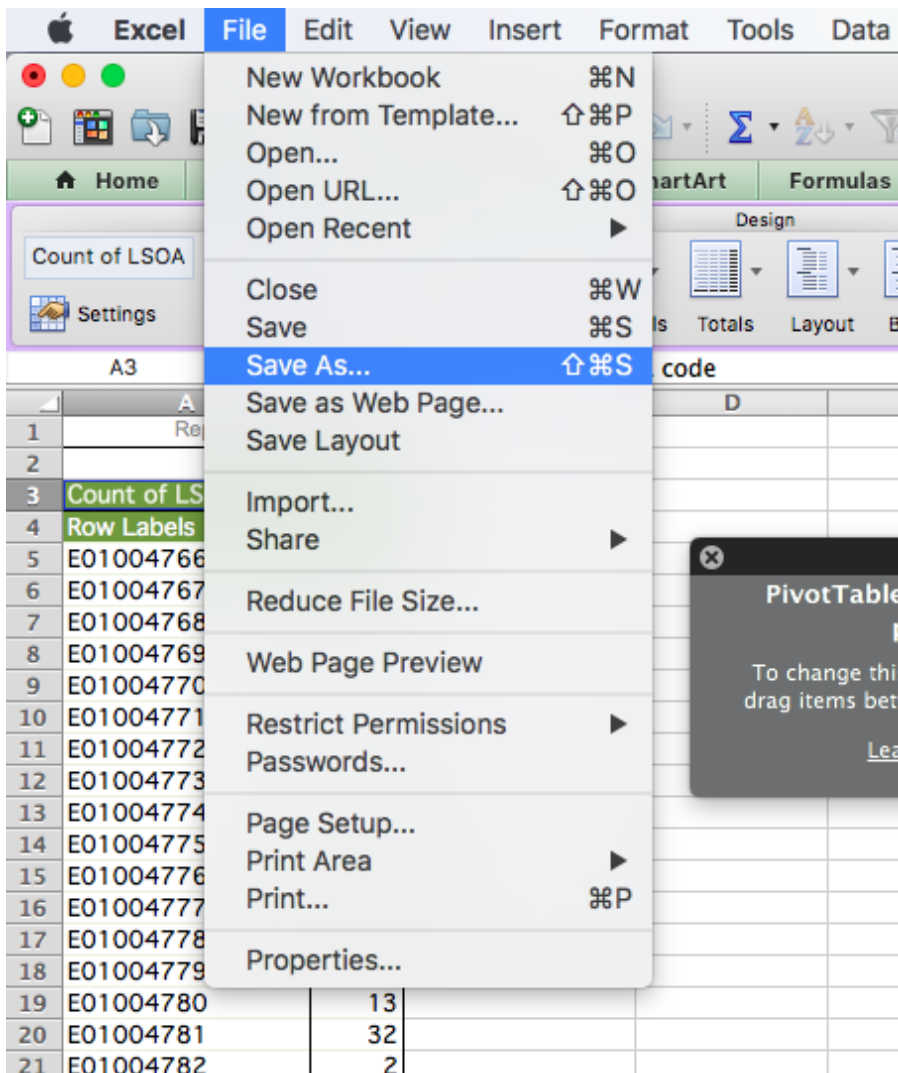


While you see the list of possible crime types, there is no value next to it - it is not yet a frequency table. This is where you need the **Values** box on the pivot table toolbar. What you drag into there determines what values will be displayed. So now grab the “LSOA code” label from the top again, and drag down, this time to the values

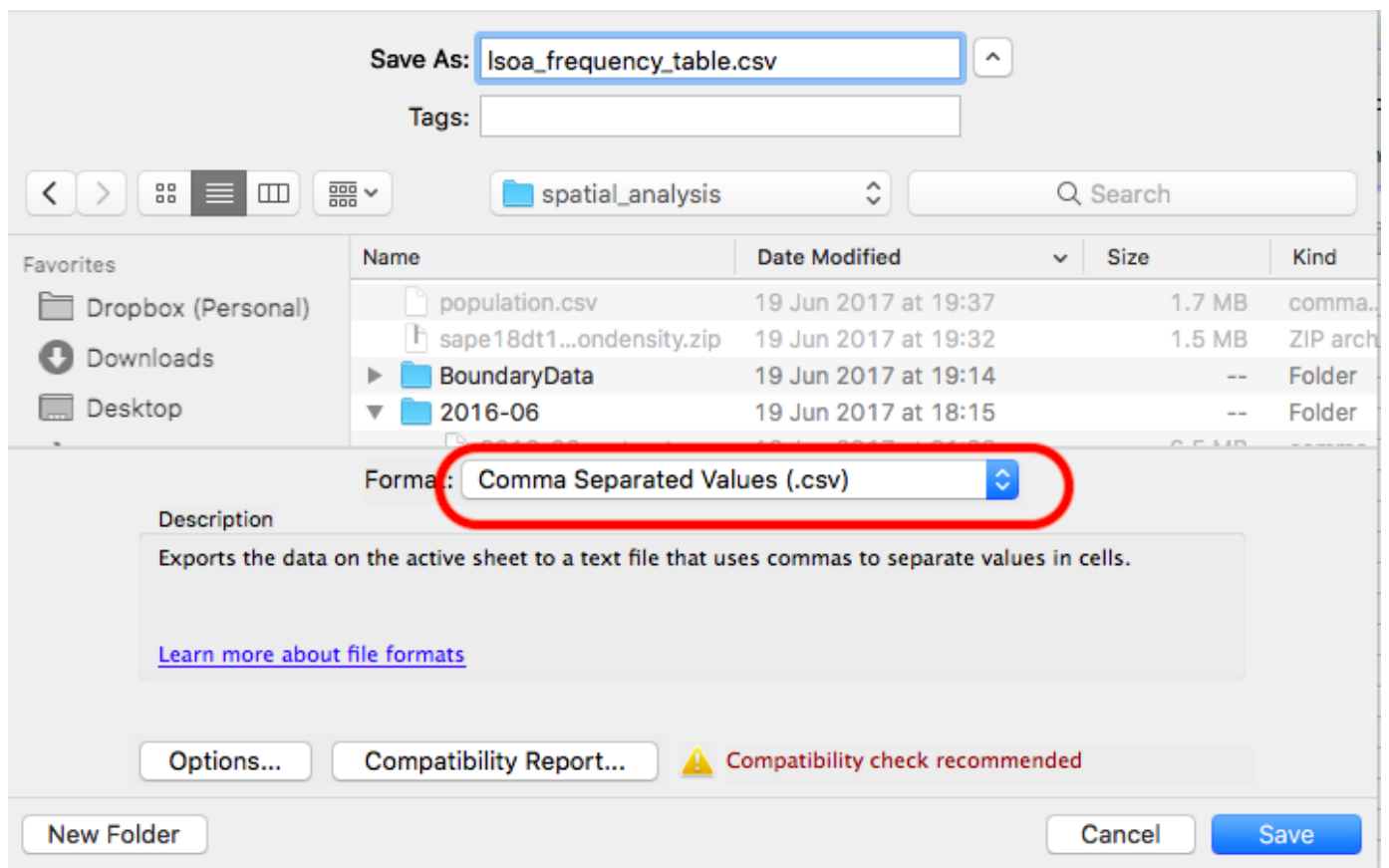
box, like this:



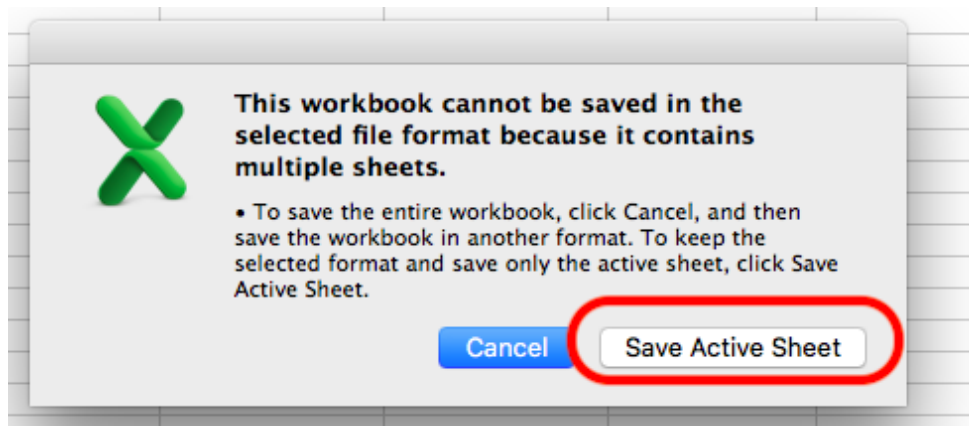
You now have a table where each LSOA code has the number of crimes in that LSOA (neighbourhood) next to it. Save this as a csv by selecting “Save as...”



and choosing .csv as the option:



It might come up with a popup message, just select "save active sheet":



(there might be another popup just click "Continue").

Make sure you know where you saved this data, and navigate there, and just open it up in excel to make sure everything looks OK. You want 2 column, one with the LSOA code, and one with the number of crimes. You might have an extra line at the top, like I do here:

The screenshot shows the Microsoft Excel interface. The spreadsheet has the following data:

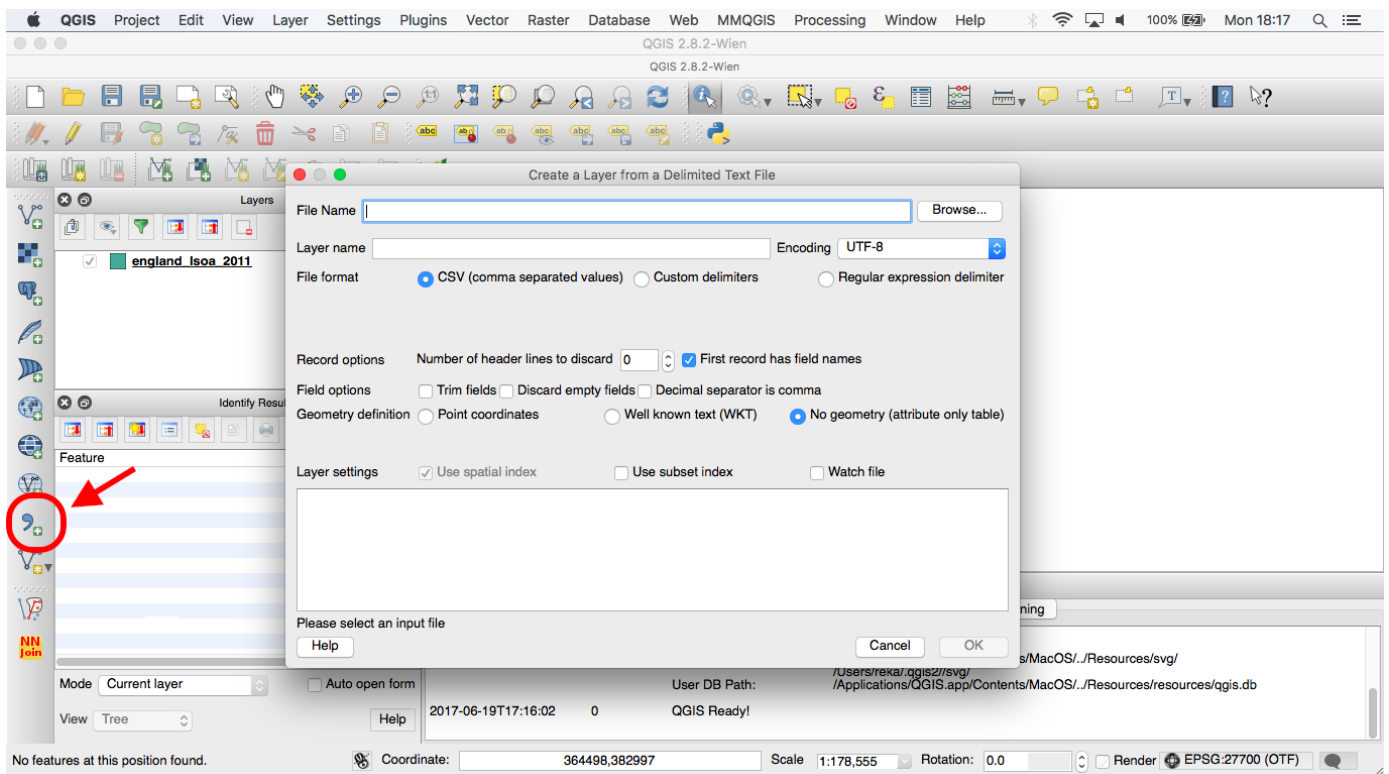
Count of LSOA code	
Row Labels	Total
E01004766	20
E01004767	25
E01004768	5
E01004769	4
E01004770	7
E01004771	12
E01004772	10
E01004773	3
E01004774	13
E01004775	20
E01004776	4
E01004777	18

A red box highlights the first row (row 1), and a red arrow points to it with the text "delete this".

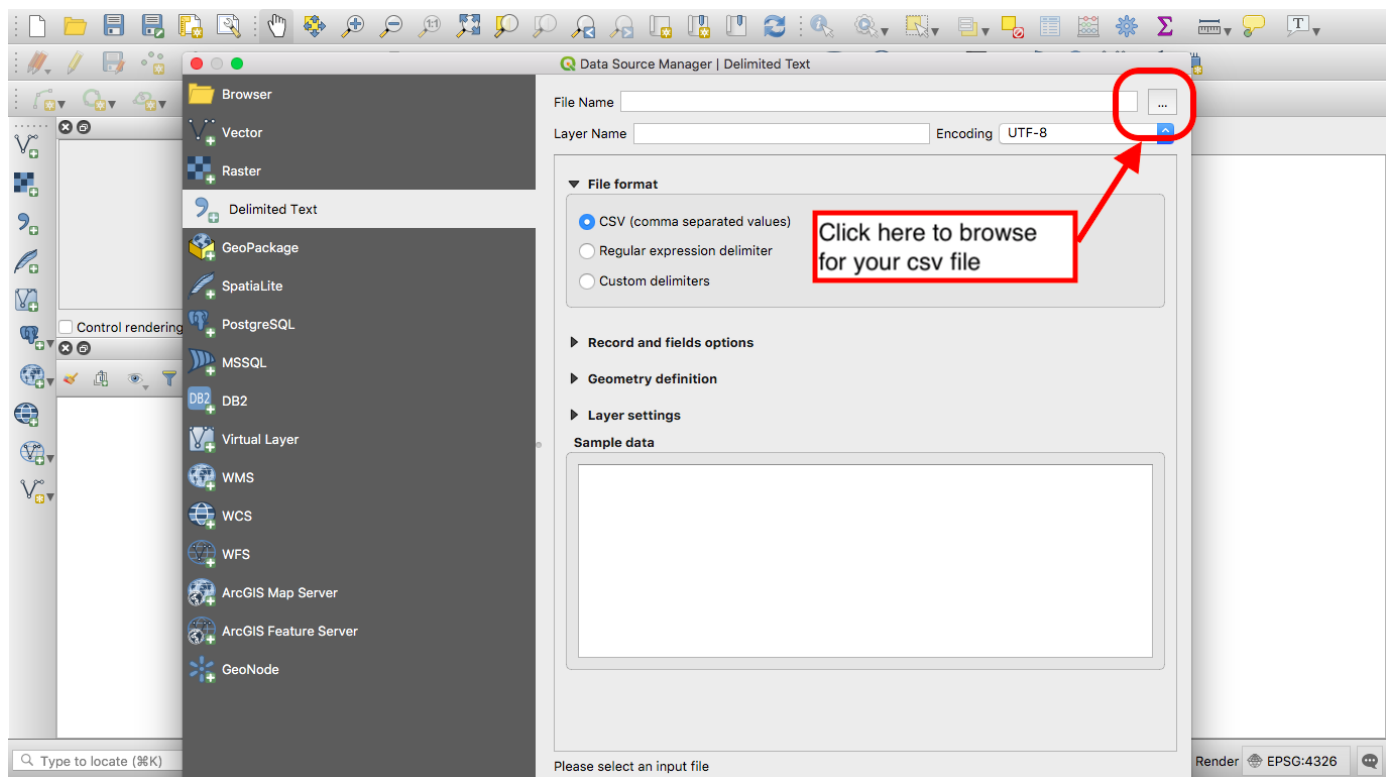
If you do, just delete this row, and save the csv again. This is because the first row has to be the column names to be able to import to QGIS. And we want to do this, as we will now import it into the QGIS environment to join it up with our spatial data.

Adding the non-spatial data (csv) to a map

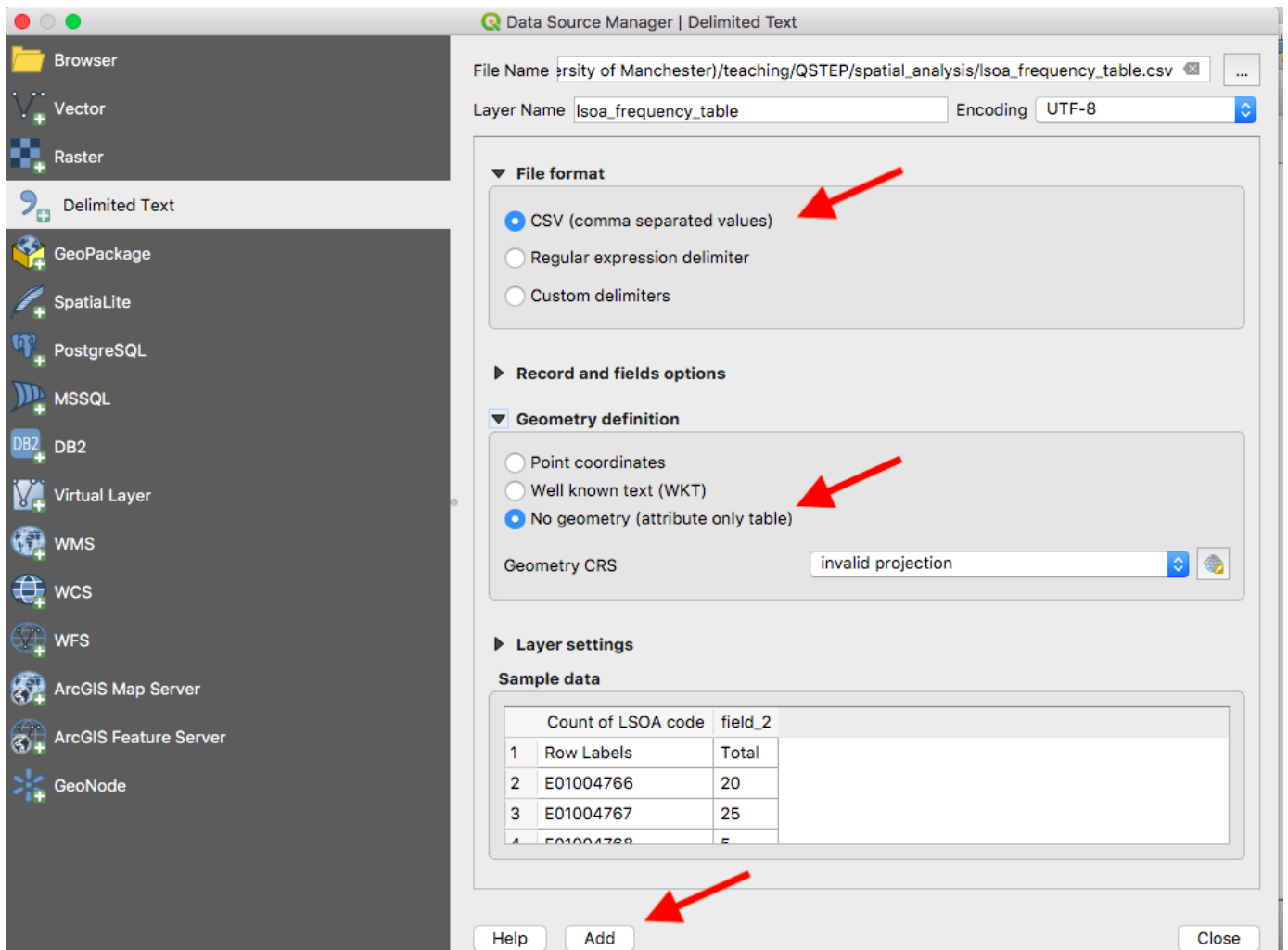
Get back into QGIS, and import the crime data. To do this we use the **Add Delimited Text Layer** button which will make a new window popup:



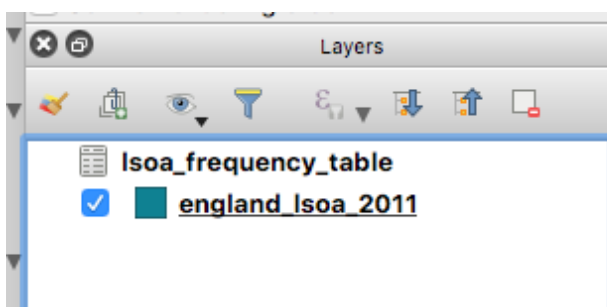
Click on “Browse” button next to the “File Name” bar, and navigate to your .csv file.



Select it, and then when you get back to this dialogue box, select “CSV” for “File format”, and then on the “Geometry Definition” option, select “No geometry (attribute only table)”. Then click on “Add”. Like so:



Then click on “Close” and you will see your new attribute table appear in your layers window:



You can right click this, and select “Open Attribute Table” to see all the data that is there.

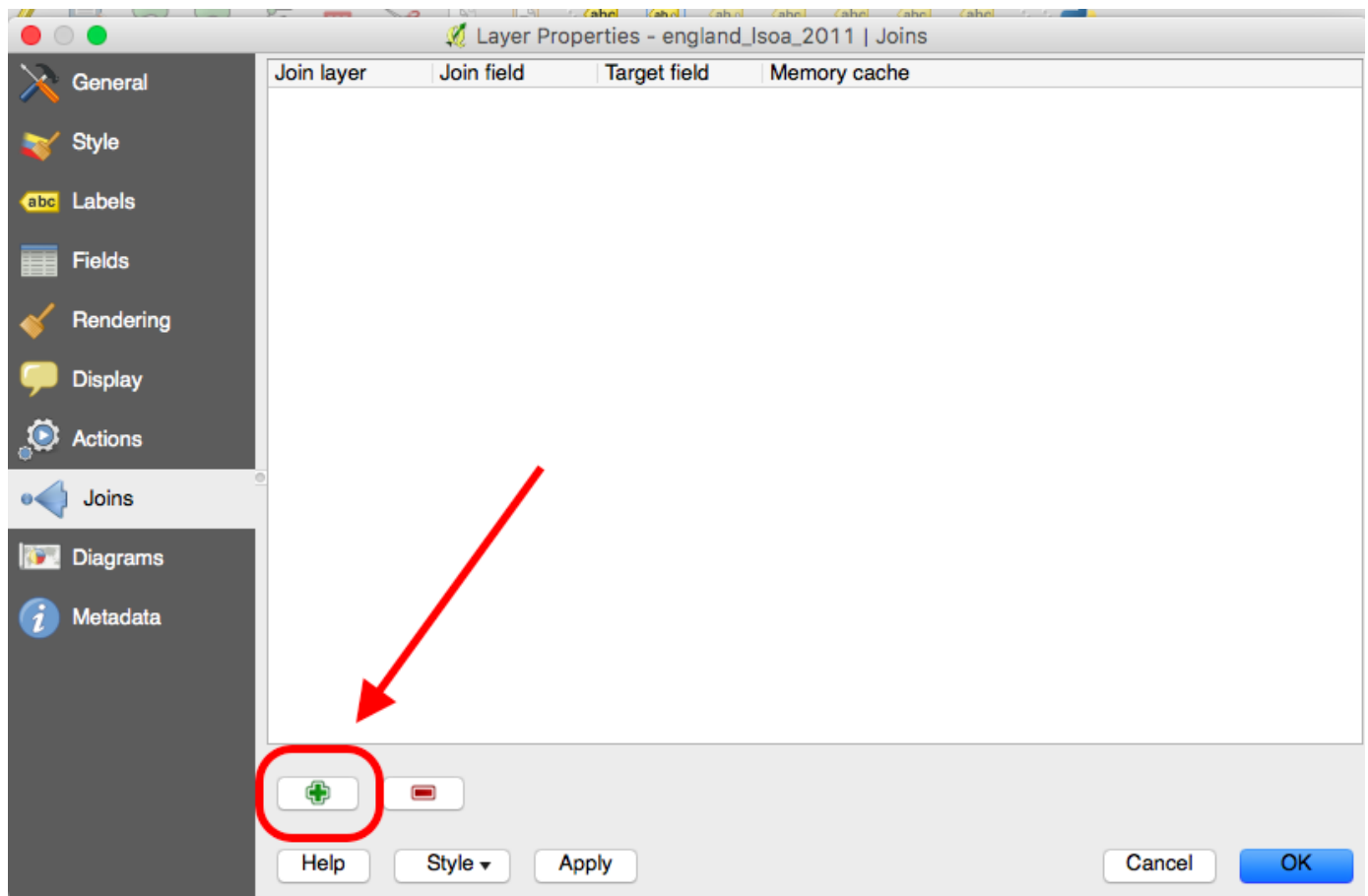
Now this is not spatial data. Unlike the shapefile, it does not appear on the map. It is only a table of data. For it to appear on the map, it needs to be joined to a spatial layer. We can do this because it has a matching column to an existing spatial layer that we have.

You see there are two columns, one with the LSOA codes called

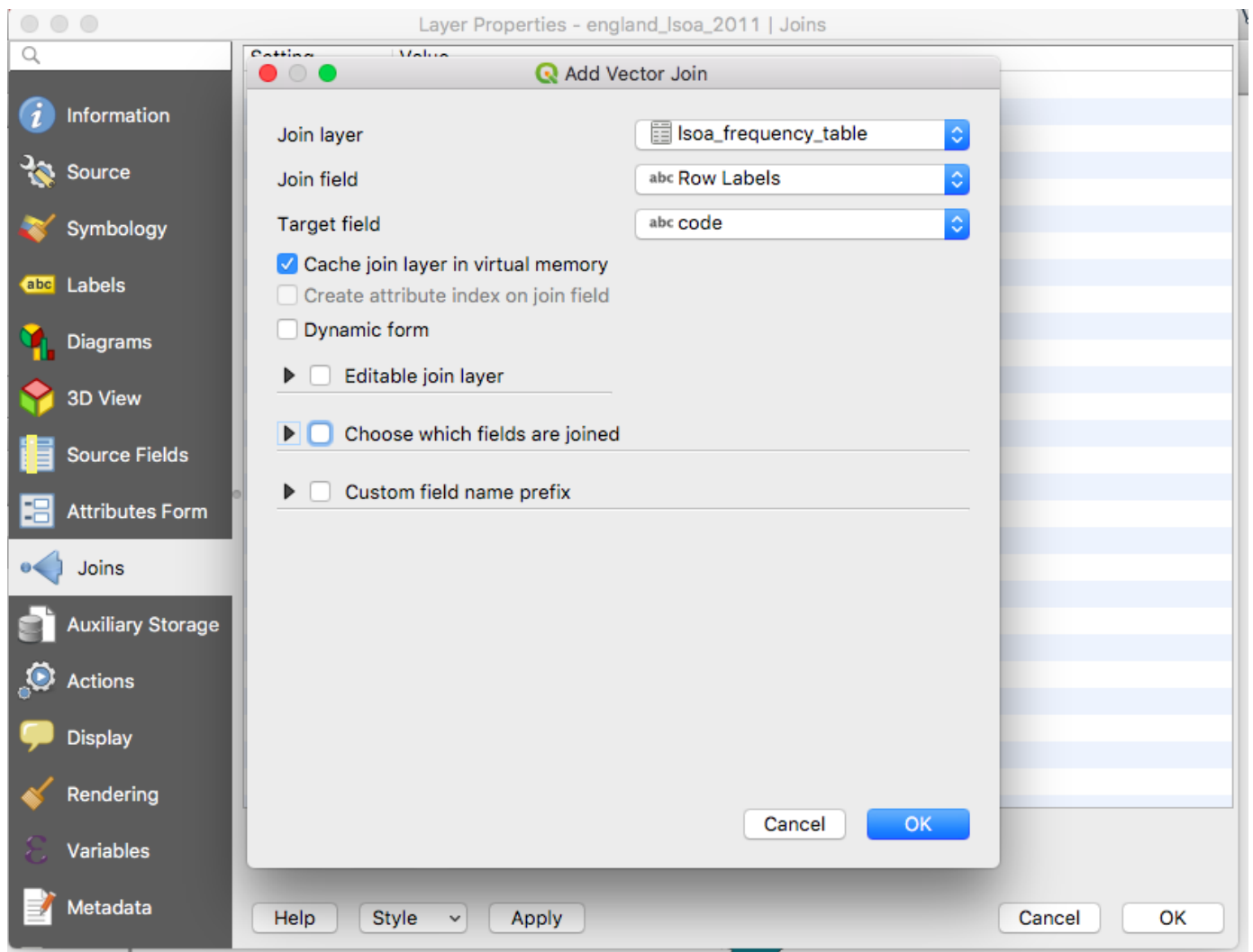
Link non spatial data to shapefiles

So if we want to map our data, we need to join it to a shapefile. We can do this, as they have a matching column.

To do this, double click the spatial layer `england_lsoa_2011` , and this time select the “Joins” tab. Then click on the little green plus sign on the bottom left:



This will open up a dialogue box where you have to select what you want to join to this spatial layer, as well as designate which column in its attribute table matched which column in the attribute table of the shapefile:



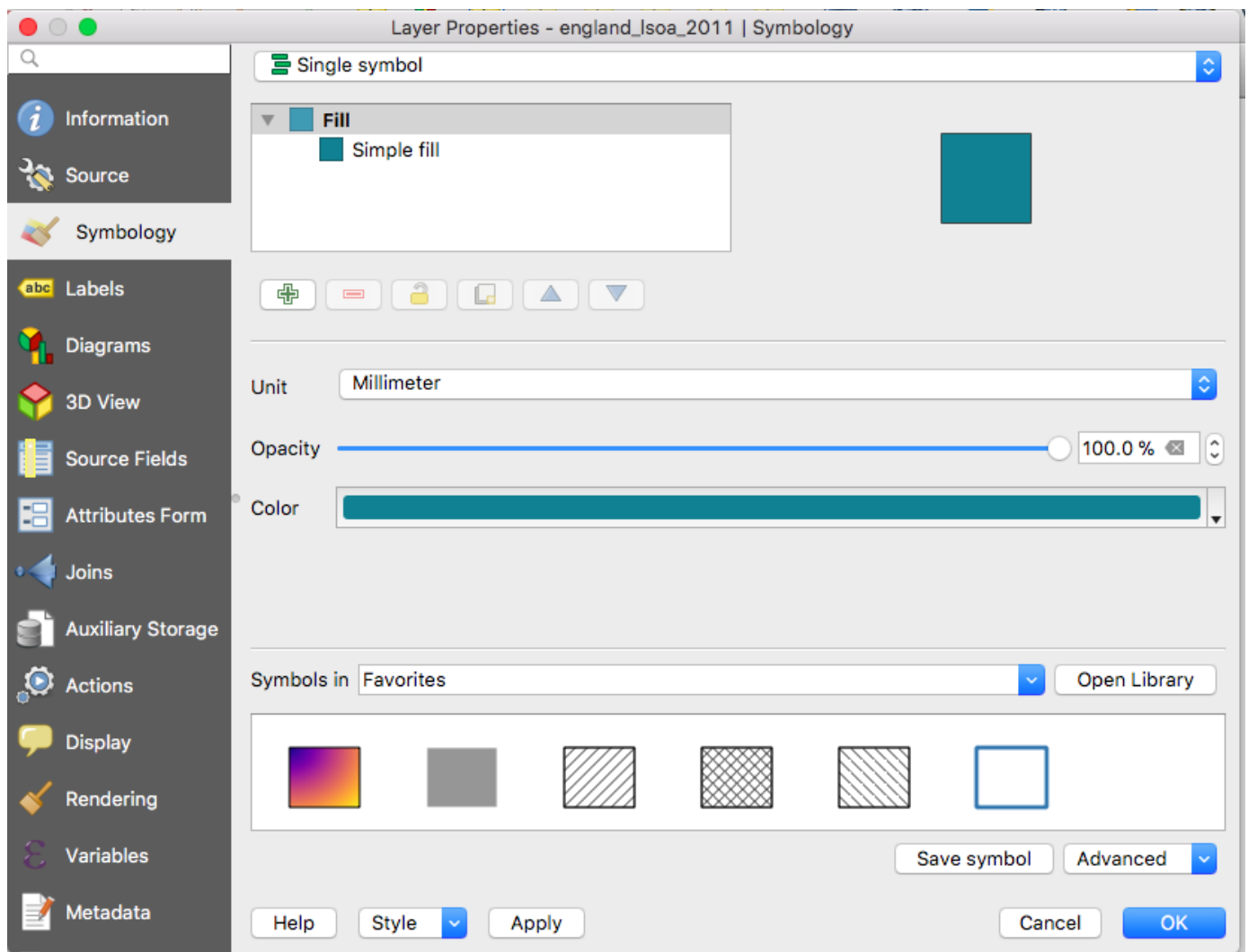
Select the appropriate fields and click “OK”. Then you will see the name of the attribute table appear in the list of joined tables. You can now close this window by clicking “OK”.

To check whether your join was successful, right click on the spatial layer and open the attribute table. You should see a whole load of new columns have appeared, all to do with crime. YAY!

You can now use this to create a thematic map.

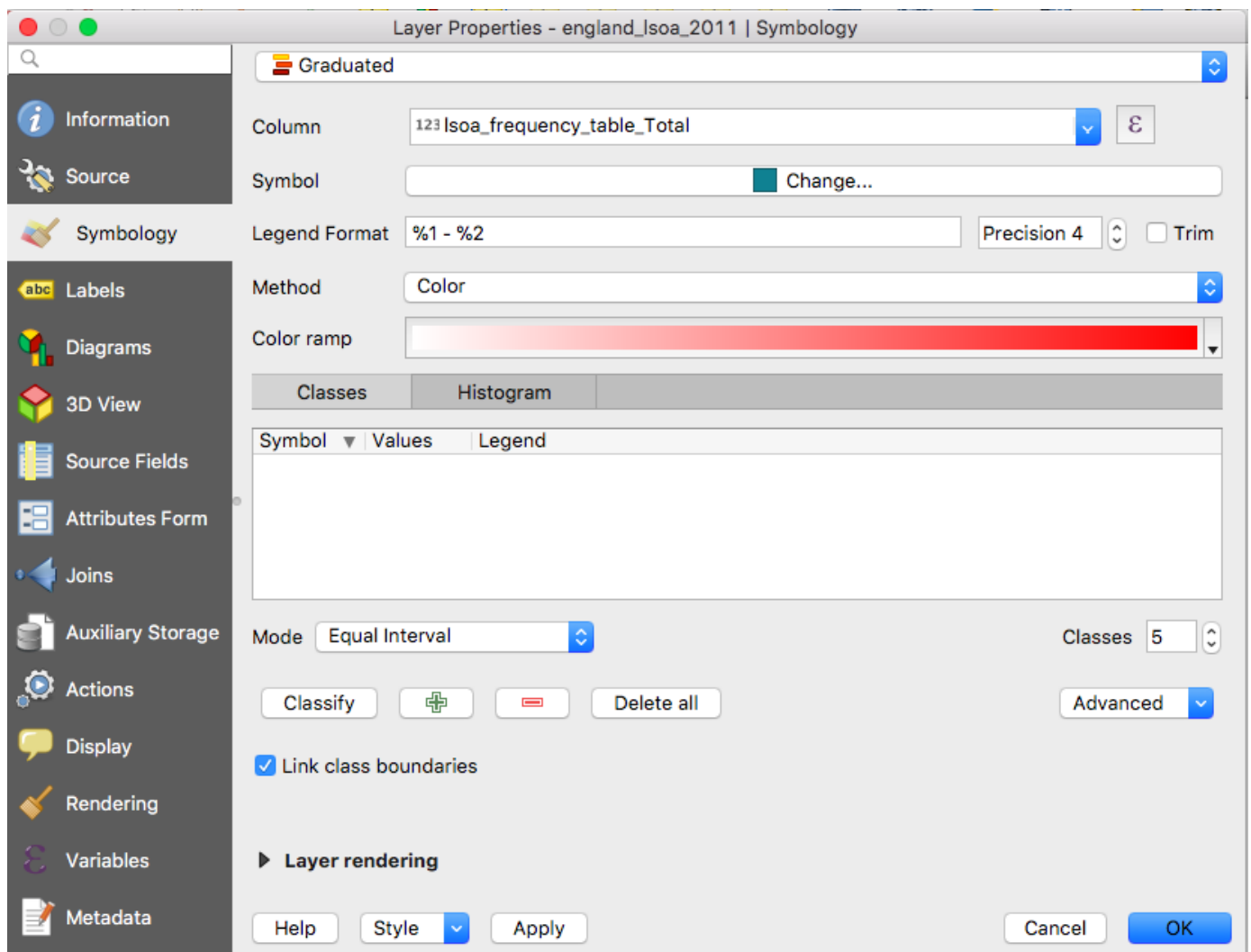
Create a thematic map

Now we want to visualise which neighbourhoods have more crimes and which have less. To do this navigate to the **Symbology** tab.



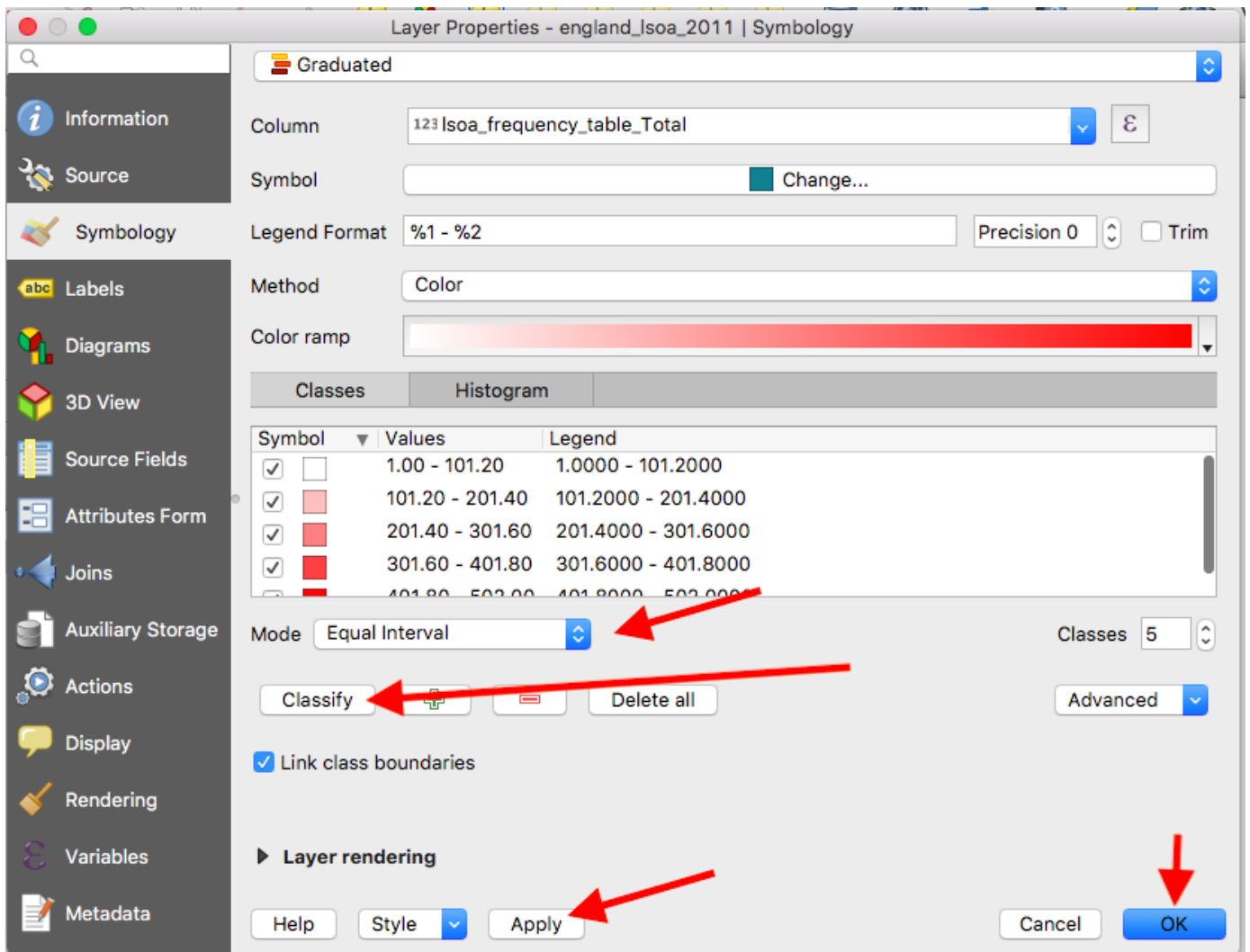
On the top dropdown menu, where it says “Single Symbol”, select “Graduated”. This means that you are choosing a continuous variable (numeric) to shade the neighbourhoods by. If you had a qualitative variable (categorical) then you would select the “Categorized” option.

Then, under “Column” select the variable you want to shade by. Here we select the “Total” column (might be called your table name + Total, so for me it’s “Isoa_frequency_table_Total”):

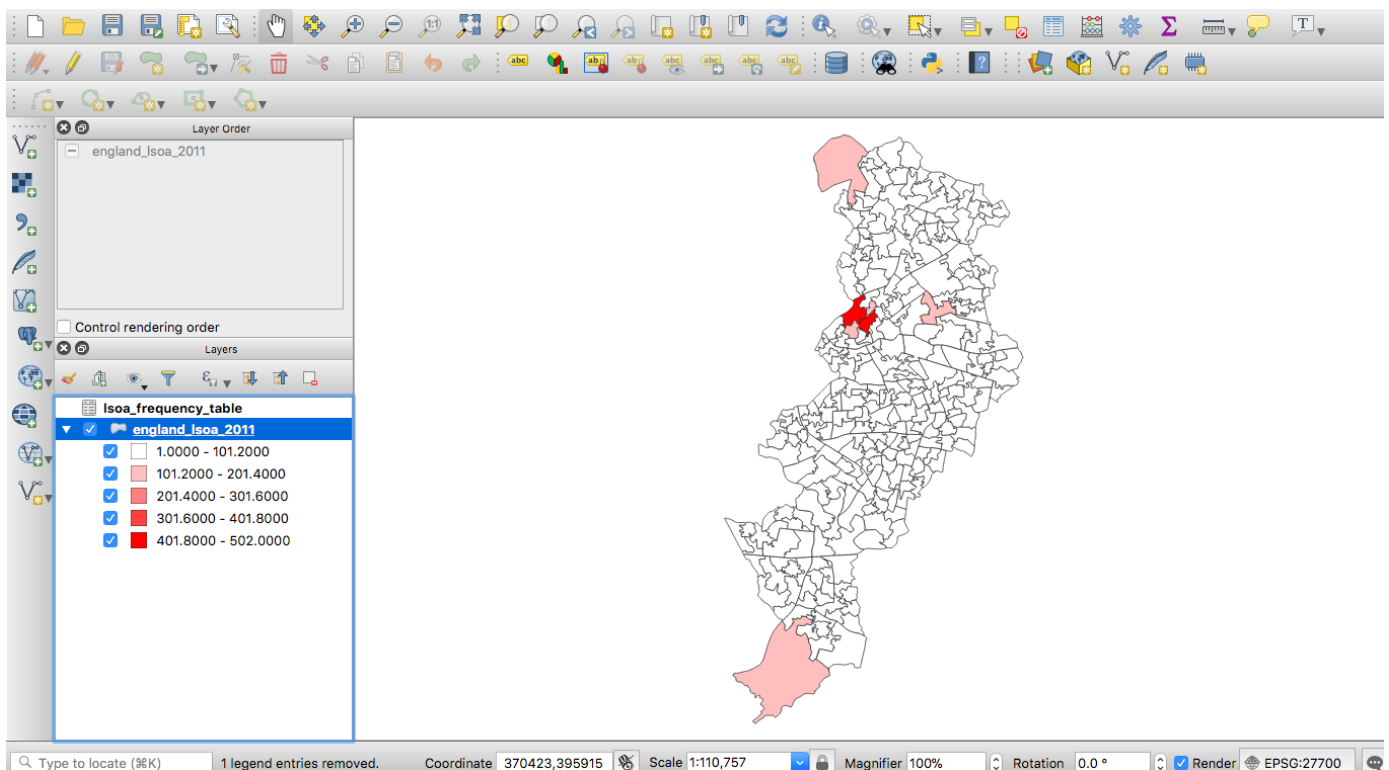


Now at this point I want to draw your attention to the “Classes” and “Mode” options. These determine what your map will look like. Classes determined the number of groups you will have. Here we split the neighbourhoods into 5 groups. But if you are looking to rate something on a red amber green scale for example for police, you would want this to only have 3 categories. Similarly the Mode you choose will depend on what you want to say with your data. There is no right or wrong option, but it depends very much on the question you’re asking.

So let’s try using the 5 classes with the equal interval mode. To do this, select these options, and click on the “Classify” button on the left just under that (currently) blank window. You will now see the classes appear. Once they appeared, click “Apply” then “OK”:



So now you should have a map. Does it look something like this:



What story does this map tell? You can see that we have equal interval, the groups are broken up by 100 crimes. You can see that most neighbourhoods in Manchester have between 1-100 crimes. There are a few (you can see actually that they are the larger ones) that have between 100-199. And then you have city centre, where you have the majority of the crimes recorded.

Have a go at selecting the different classifications and see how your map changes!

Read in spatial points

Now we mentioned earlier that our police data actually does have a latitude and longitude column. Because of this, we can actually read it in as spatial data. Let's give this a go.

To do so, again click on the **Add Delimited Text Layer** button. This time navigate to the downloaded police data, and this time, click the "Point coordinates" option under "Geometry definition". You can then select your X and Y coordinates. **X should be the column for longitude, and Y the column for latitude.**

Data Source Manager | Delimited Text

File Name: `iching/QSTEP/spatial_analysis/2016-06/2016-06-greater-manchester-street.csv`

Layer Name: `2016-06-greater-manchester-street` Encoding: `UTF-8`

File format

- ☒ CSV (comma separated values)
- ☐ Regular expression delimiter
- ☐ Custom delimiters

Record and fields options

Number of header lines to discard: `0`

- ☒ First record has field names
- ☐ Decimal separator is comma
- ☐ Trim fields
- ☐ Discard empty fields

Geometry definition

- ☒ Point coordinates
- ☐ Well known text (WKT)
- ☐ No geometry (attribute only table)

X field: `Longitude` Y field: `Latitude`

☐ DMS coordinates

Geometry CRS: `Project CRS: EPSG:4326 - WGS 84`

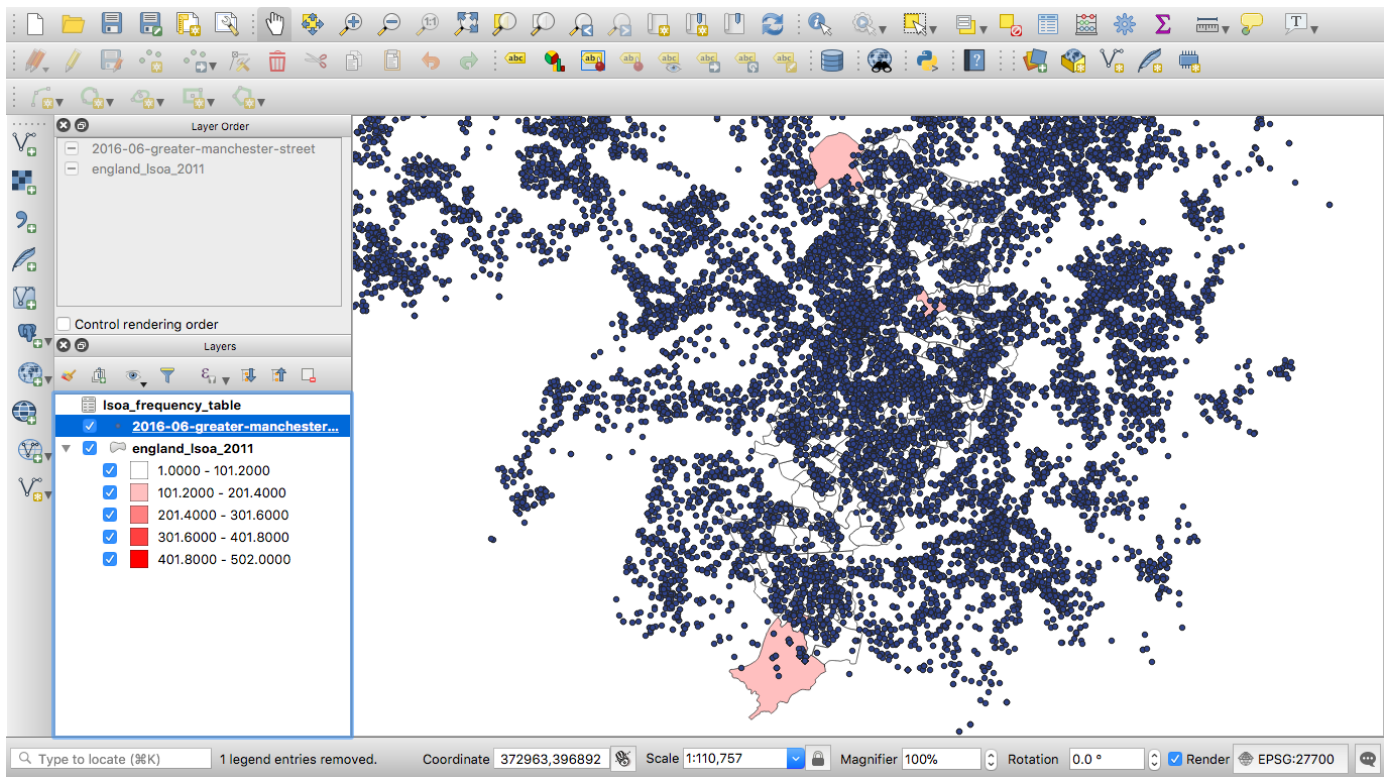
Layer settings

Sample data

Crime ID	Month
1	2016-C

Help Add Close

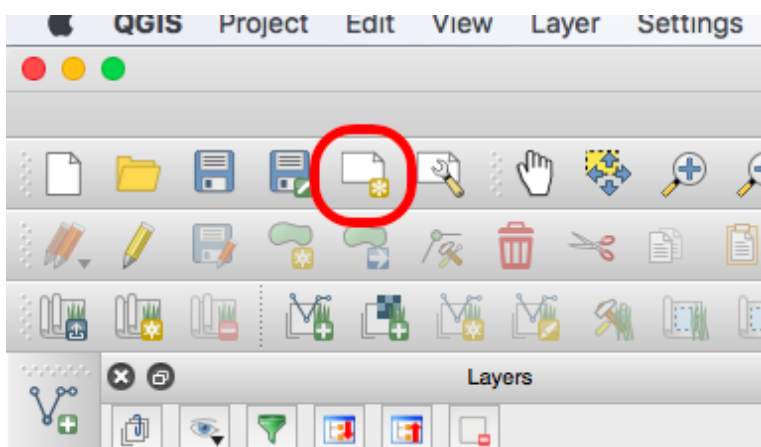
When you've specified all this, click "Add" and watch your points get added to the map!



You can see that the points cover a lot more area than what our polygon shapefile for Manchester covers. This is because Greater Manchester Police cover... well Greater Manchester, which also includes for example Bolton and Bury as well. But let's say here we're only interested in those in the Borough of Manchester. In this case, our shapefile is just fine. If you were interested in all, you can go back, and when downloading the boundary data select all the boundaries that would be relevant.

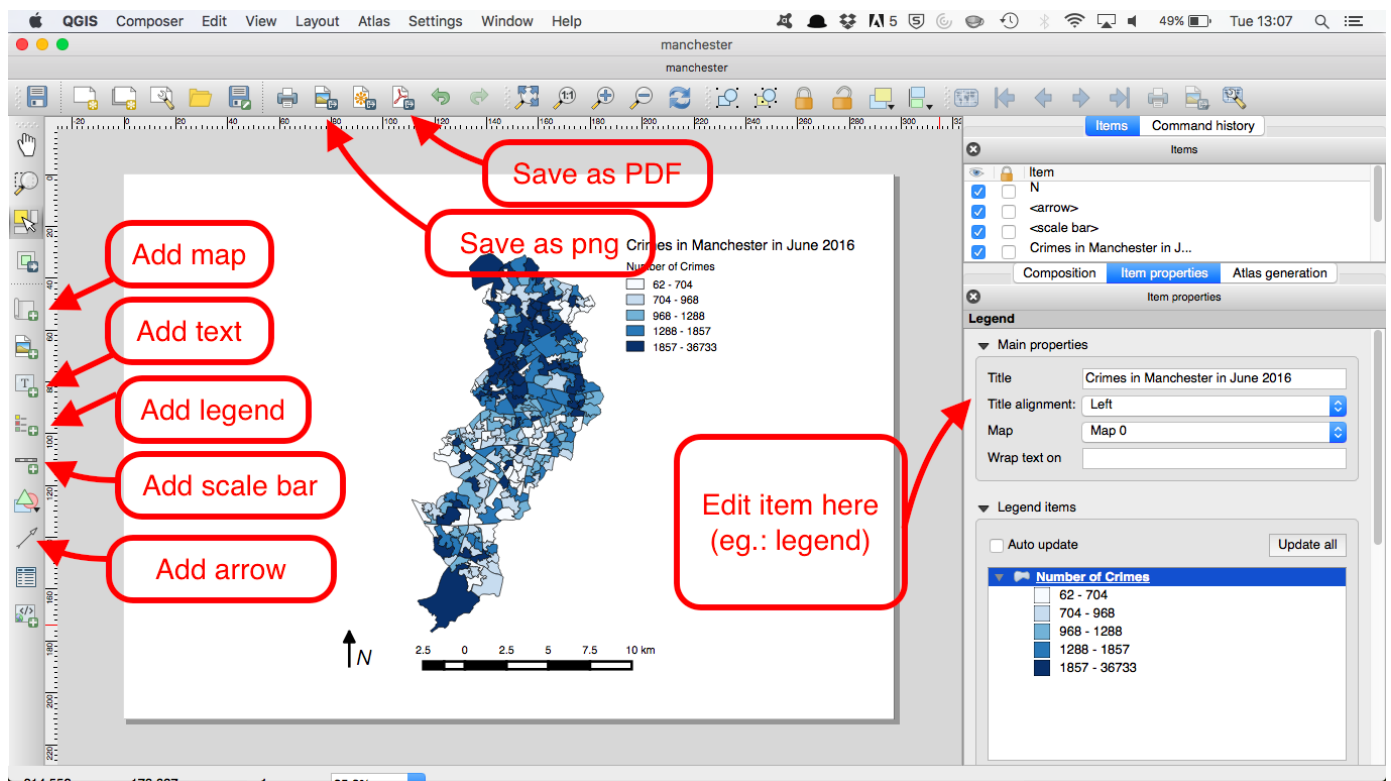
Saving your map

Anyway that is all I have for you now. You will of course also have to save your map. For this you can use the print composer. You open that by clicking this icon:



Here is an excellent video (<https://www.youtube.com/watch?v=qwmPU-LWpGU>) you can watch that will walk you through how to export your map using this print composer.

You should probably watch that, but just as a note here are some useful buttons:



Parting words

As I said at the beginning there is loads of help online for QGIS. If you want to know how to do something, just google it. For example, if you want to know how to make a heatmap in QGIS, just google "HOW TO MAKE A HEATMAP IN QGIS" and I guarantee that you will find a nice step-by-step tutorial to follow online.

If you are stuck though, and need some help, don't hesitate to get in touch and email me at reka.solymosi@manchester.ac.uk