Making maps from public health data using free and open-source software.

A one day crash course

Tephinet Learning Session October 2019

The Global FETP Enterprise:
Applied Epidemiology in the 21st Century
10th TEPHINET Global Scientific Conference

Trainers. Andy South, Sophie Dunkley & Tiago Canelas

https://etch.lstmed.ac.uk/gives

LSTM
LIVERPOOL SCHOOL
OF TROPICAL MEDICINE

GIVeS

TEPHINET
Training Programs in Epidemiology and
Public Health Interventions Network

This work is licensed under a Creative Commons Attribution 4.0 International License.

Introduction and course philosophy

Welcome to this one day crash course on making maps from public health data with QGIS.

Maps can be a useful means of displaying data. Maps can help with communicating where things occur, detecting spatial patterns and directing management responses. They are popular with the public and politicians. Scientists and operational staff in a range of fields can benefit from making them.

The course is predominantly practical to provide you with hands on experience of map making using QGIS in a supported environment.

By the end of the day you should be able to make maps from your own data.

Learning Objectives

1. Be able to use QGIS to manipulate geographic information and make maps.
2. Be able to describe and use the most important types of spatial data.
3. Be able to modify software settings to make maps useful to inform decision making.
4. Be able to get public health data into QGIS and make maps from it.
5. Know how to attempt to solve inevitable technical difficulties with map data and software.

We will be focusing on practical use and skimming over some areas of theory due to limitations of time. There will probably be aspects that you don't understand. Please do feel free to ask the instructors and your neighbours. In some cases not knowing exactly how something works won't matter, in some cases it will and we will try to explain (if indeed we know).

Also remember that in this aspect the internet is your friend. There are excellent (although varied) help resources available on the internet to support learners (and instructors) to make maps with free and open-source software.

The beauty of free software is that learners can install it on their own or work computers, can share with others and can seek help on the internet from diverse user communities who are not constrained by cost. This is especially the case in low resource settings. Yes, sometimes software vendors can offer deals for educational or low income country users but this doesn't accommodate students after graduation, or the member of staff wants to install on their home computer, or the learner who needs to ask a question from someone who doesn't have a licence.

Manual conventions

In this manual commands for you to carry out yourself are indicated with small arrows :

- ➔ this small arrow indicates something for you to do

GIS & QGIS - a very brief intro

A *Geographical Information System (GIS*) is a computer-based system for capturing, storing, visualising and analysing geographic data.

In this course we will be using QGIS, an open-source GIS software that is freely available to download from www.qgis.com.

QGIS is a professional GIS application that is built on top of and proud to be itself Free and Open Source Software (FOSS). QGIS development is driven by a community of volunteers. We are not a part of the developer community but we have made a financial contribution to support continued QGIS development.

For this course, we will be using either versions 3.8.* or 3.4.*. v3.4 is the long term release (LTR) that is the most stable and least likely to have bugs. v3.8.* is the most recent release that has more additions but may have bugs because it has not yet been tested so much. The QGIS maintainers plan for v3.4.* to be the 'Long Term Version' or LTR until the end of 2019. Note that minor versions e.g. 3.8.1 or 3.8.2 indicate bug fixes and will look very much the same as any version of 3.8.

Installing QGIS

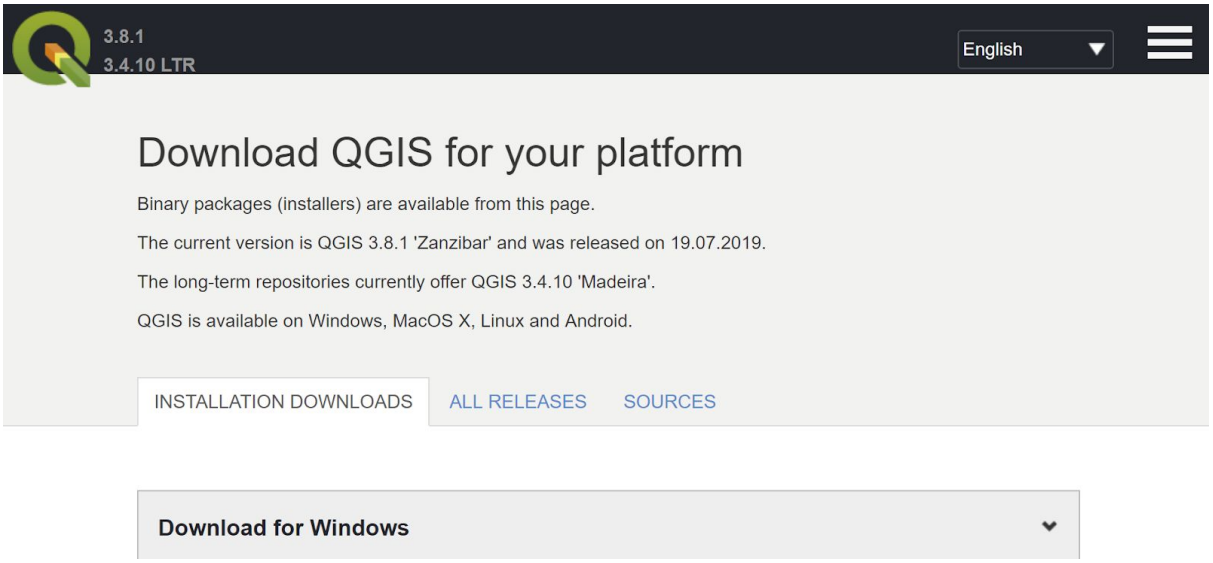
We recommended that you install QGIS before the course. If you have done that you can ignore this section :-).

We have provided the installation files for windows on the course USB drive. Latest versions of QGIS can be downloaded from the internet. Here we step through installation on Windows, installers and instructions for Linux and Mac can also be found at the link.

➔ To download the latest version of QGIS go to this link :

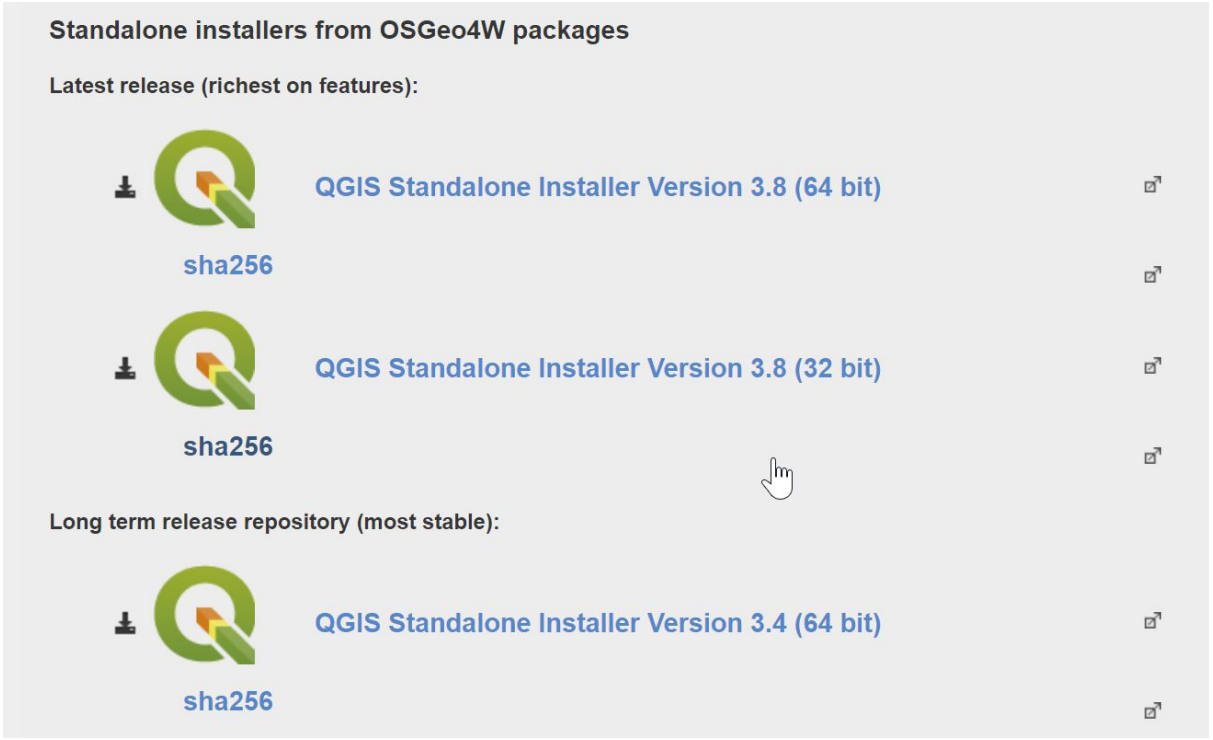
https://download.qgis.org

or google search for "qgis download" to find it. The top of the download page should look like this :



➔ Scroll down to find the standalone installers
➔ Download the file for 3.4 Long term release (64 bit if your computer is less than 5 years old)

If your computer is older than 5 years, you can look in system settings to see if you have a 32 or 64-bit version of windows, type 32 into the windows search dialog and it should bring up an option to find which version you have.



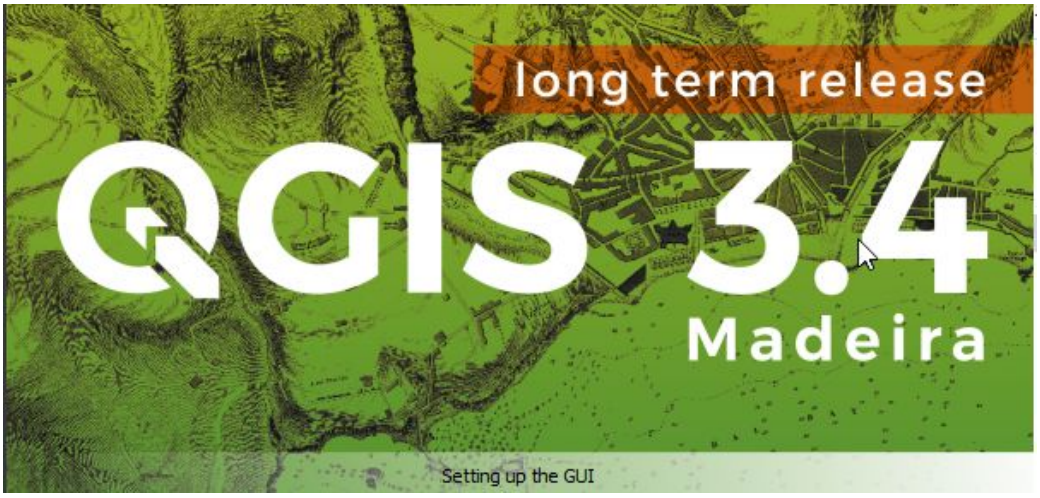
➔ Double click on the file to run the installer.

If you already have an earlier version it will ask you whether you wish to replace with the newer version. (Click yes, you do).

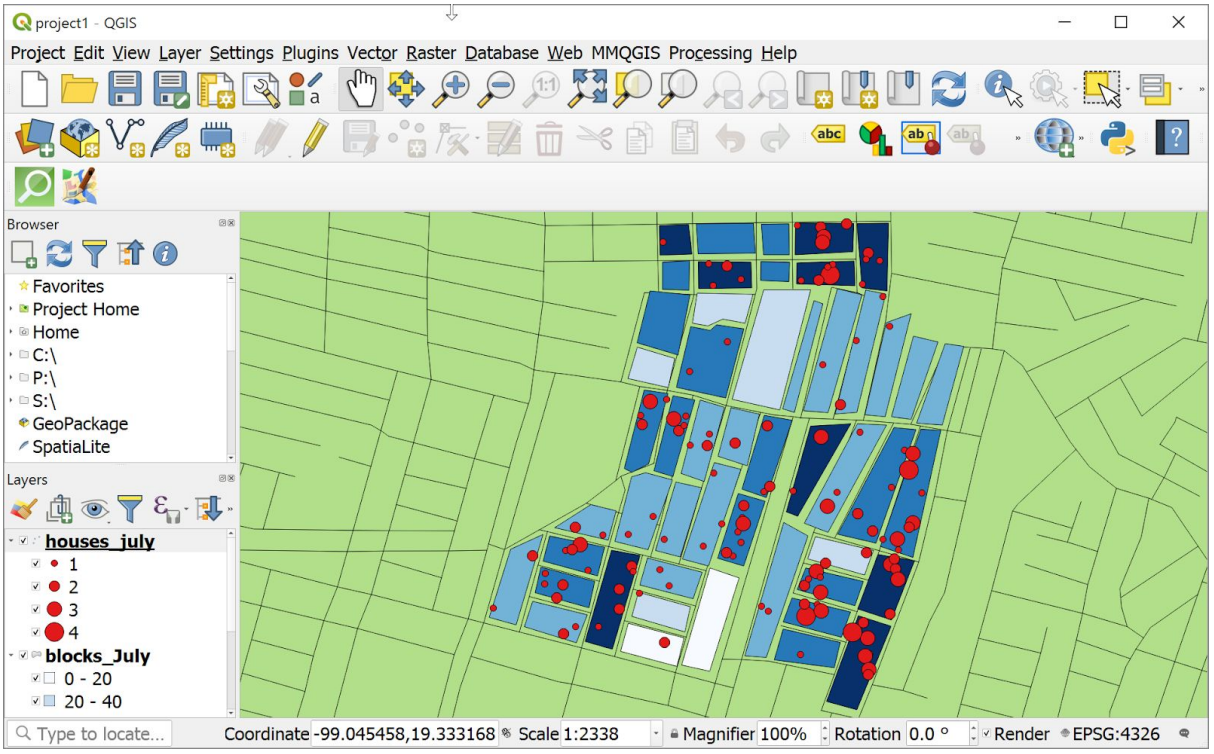
Getting started with an existing QGIS project

- ➔ Find the file on the course USB : **training-QGIS-201910/project1.qgs**
- ➔ Double click on it

This should open QGIS, which will take a few seconds. You should initially see something like this :



Then you should see something like this :

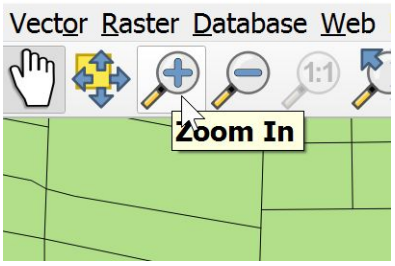


This is great we can see a map already.

The map shows some sample data that we have generated solely for demonstration purposes for an area of a few kilometres in Mexico City. Initially you can see points at the locations of individual houses, blocks and a road network. We will explain these more later.

There are also a slightly frightening number of icons and text labels around the edges.

To get some indication of what the different icons along the top do you can pause the cursor ('hover') over and a little message will appear e.g. as seen for 'Zoom In' :



➔ Try 'hovering' over some of the buttons yourself. The hover message disappears after a while.

QGIS is a very powerful application with lots and lots of options. This can be intimidating for new users. However, it can be useful to realise that you don't have to understand everything that QGIS does. We don't understand everything that QGIS does !

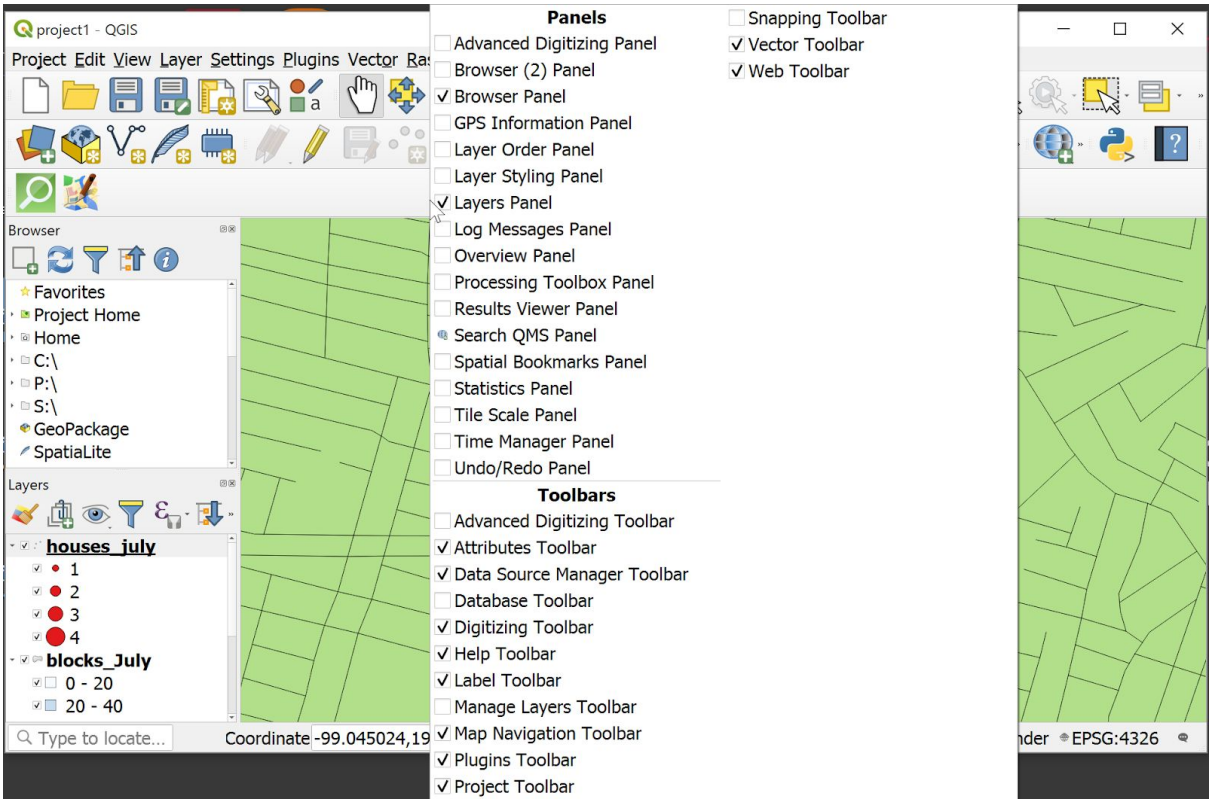
Modifying the QGIS user-interface to focus on key tasks

QGIS is also very flexible and we are going to take advantage of that to make it easier for us in our early steps. There are options to make the different user interface components visible and invisible. We are going to make some of the default components invisible so that we can focus on the most important ones for new users.

Components currently on the left of the user interface, titled 'Browser' and 'Layers' are called 'Panels'. The icons along the top are contained within components called 'Toolbars'. By right-clicking in some of the empty grey areas we can bring up a list of tick-boxes which allow us to turn panels and toolbars on and off.

➔ Right click in one of the empty grey areas above the map.

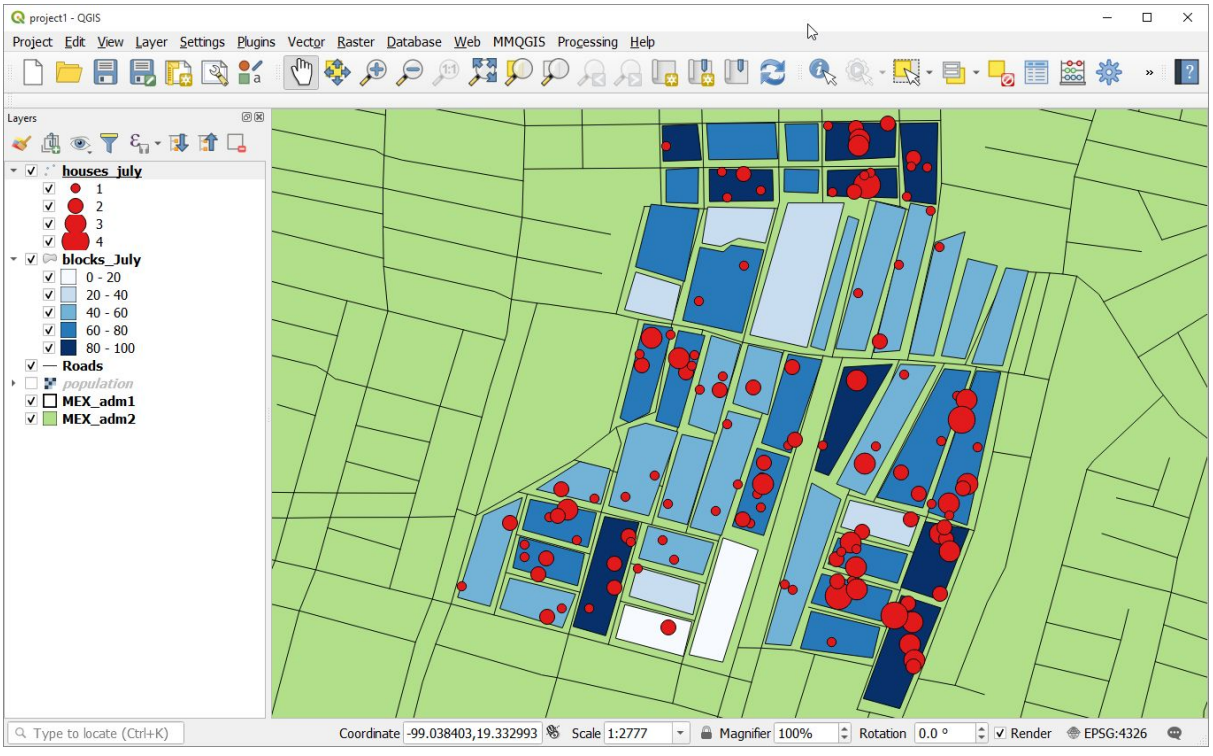
This should bring up a list of Panels and Toolbars



➔ Untick some of the boxes so that you just have :

- Panels : Layers Panel
- Toolbars : Attributes, Project, Map Navigation

Now you should have something that looks like this :



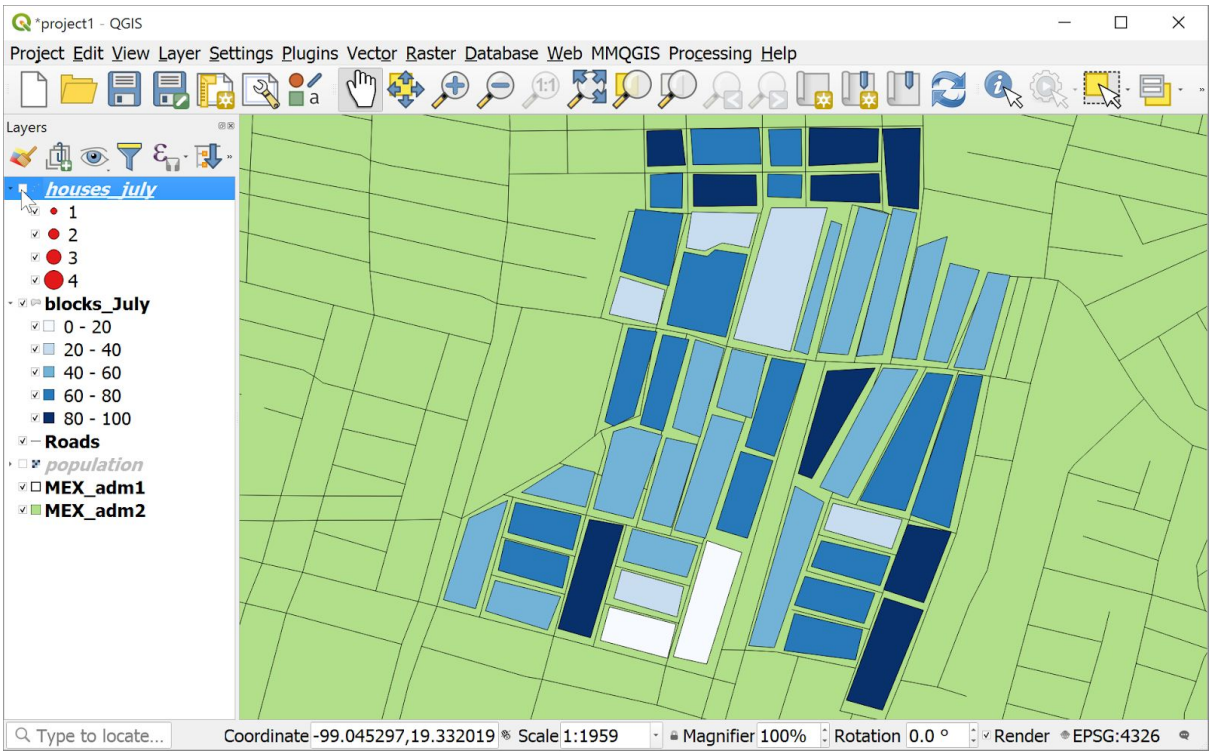
Layers

Data are displayed in the map as 'layers' that appear in the Layers Panel on the left. We will be using the Layers Panel a lot.

The checkboxes to the left of each layer name within the Layers Panel allow you to change what is displayed in the map view.

➔ Experiment with unchecking boxes and see how features no longer appear in the map view.

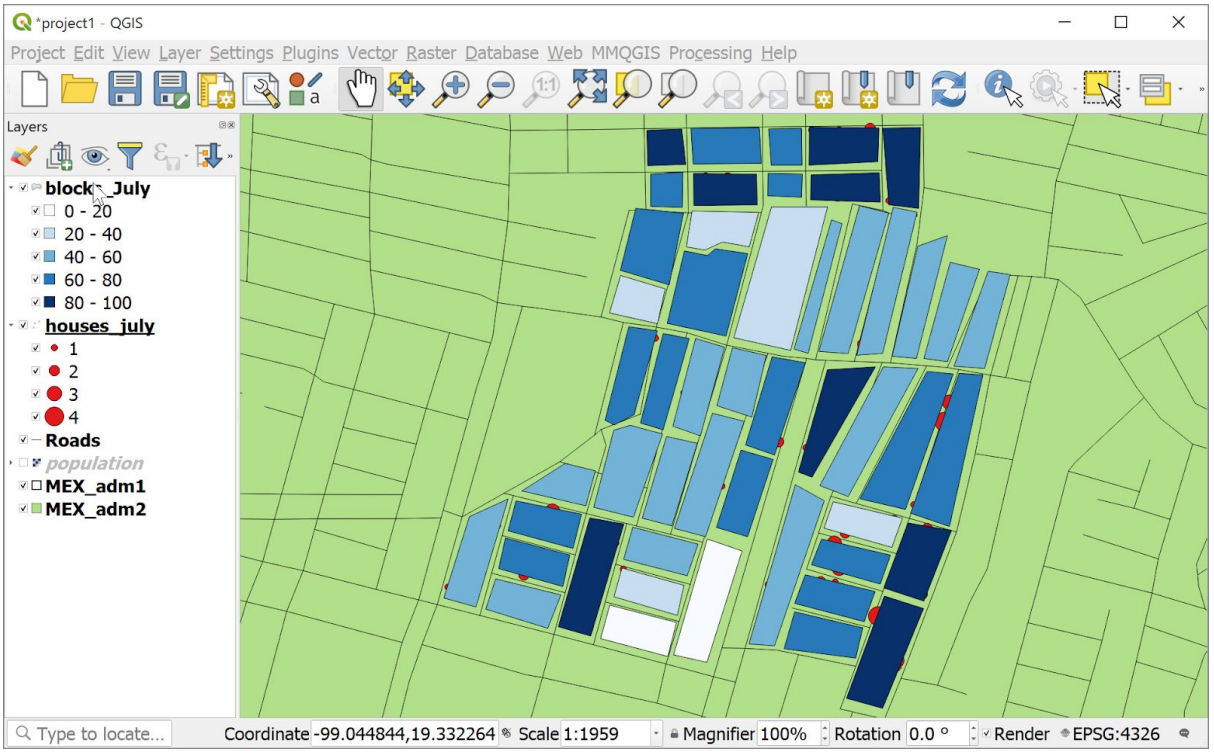
Checking the box again makes them come back (they have not been deleted they are just not shown when they are unchecked). You may also notice that the layer name in the layers panel becomes greyed out and italicized when it is unticked.



The order of layers in the Layer Panel determines how they appear in the map view. The layers at the top are displayed last and those at the bottom are displayed first. Thus it is like the higher up layers are painted on top of the lower layers. This can mean that some layers get hidden by others. Later we will see that layers can be made partially transparent to allow them to be seen underneath others. You can change the order of layers by clicking and dragging them within the Layers Panel.

➔ In the Layers Panel Left click, hold and drag the blocks layer so that it is on top of the points layer

This is a little tricky, you need to let go of the mouse button in a certain place and if you don't the layer will not move. Try experimenting.

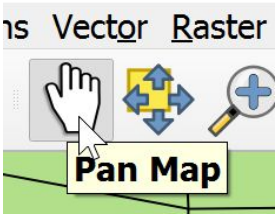


Moving and zooming a map

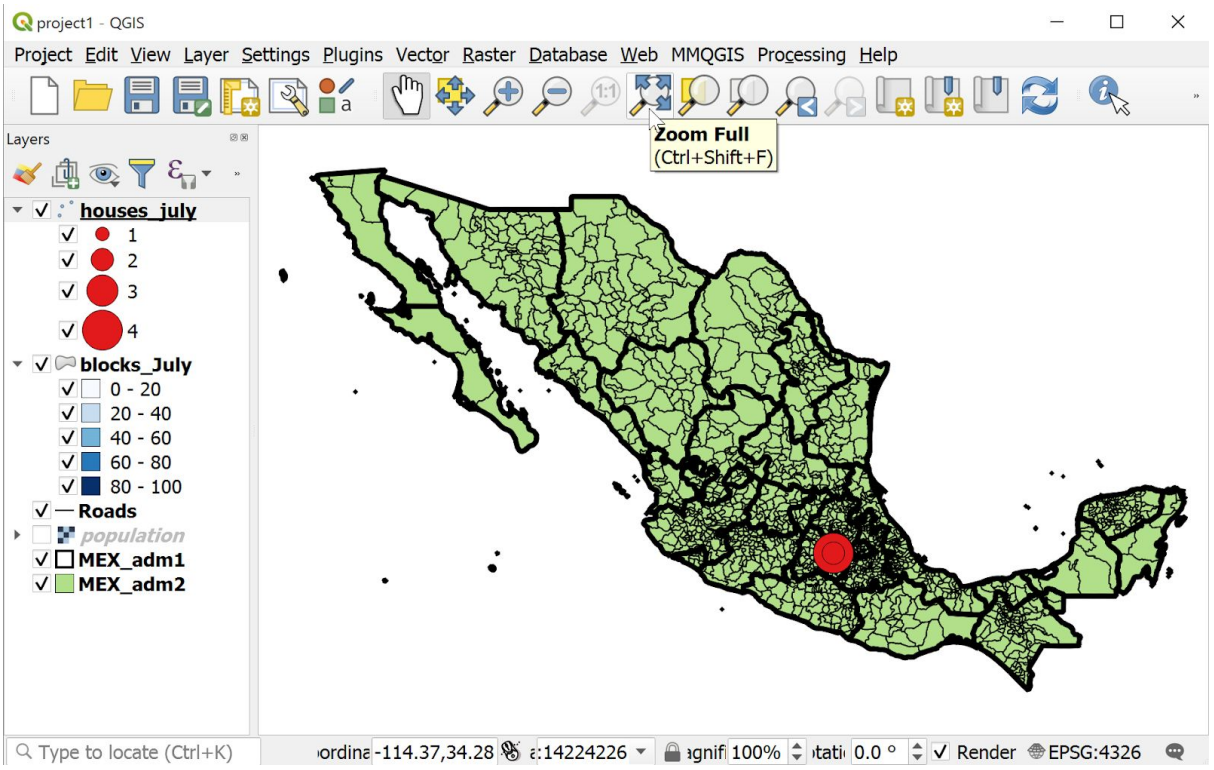
There are many ways to change the position and size of the area that is shown within the map view.

Experiment with these :

- ➔ zoom in and out with the mouse wheel (if you have one)
- ➔ left click and drag to move the map when the Pan map hand icon is selected



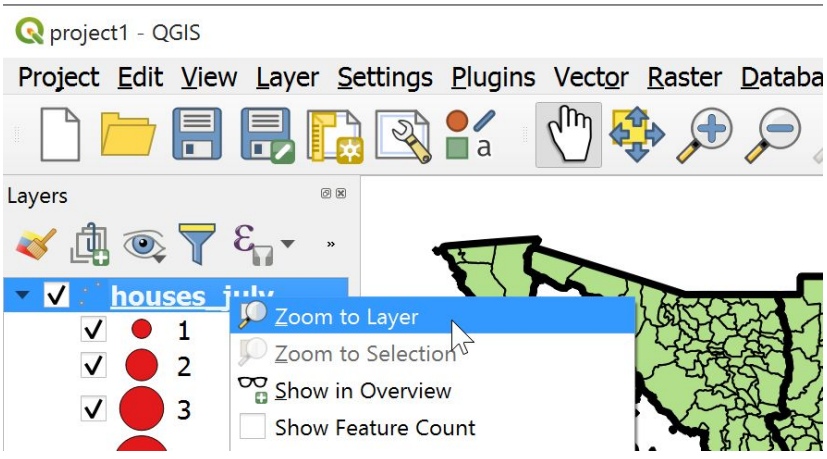
- ➔ use zoom icons from the toolbar (those with a magnifying glass)
- ➔ Zoom full takes you to the full extent of all of the map layers.



QGIS works easily with layers that have different geographic extents. In this case we have the layers for houses and blocks that are just a few metres across, and layers for the whole of Mexico that are hundreds of kilometres across. Layers do not have to be the same size.

A really useful command allows us to zoom in or out to see the whole of particular layer :

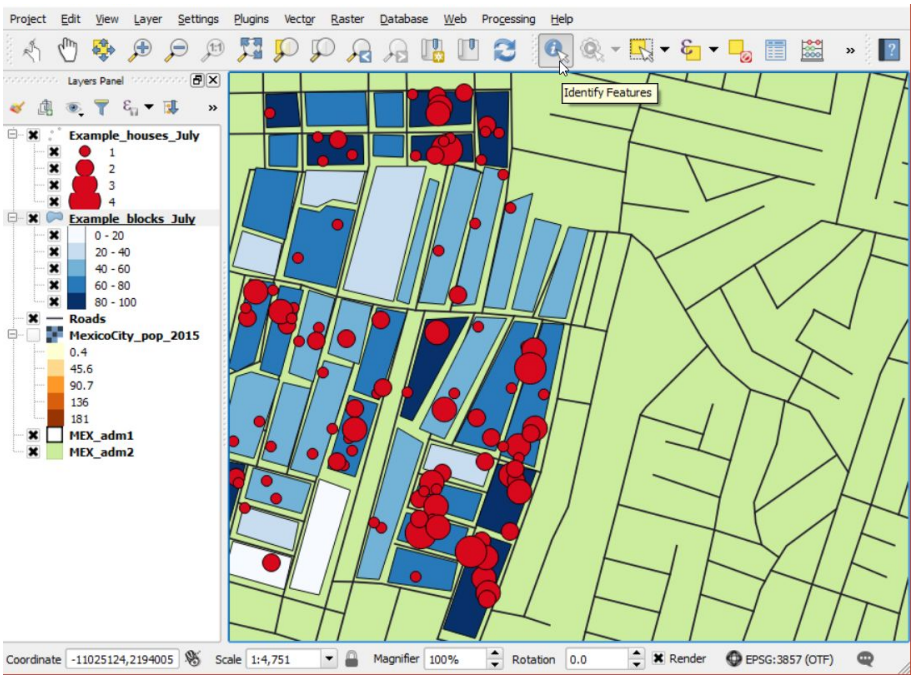
- ➔ Right click a layer in the Layers Panel on the left
- ➔ Select 'Zoom to layer'



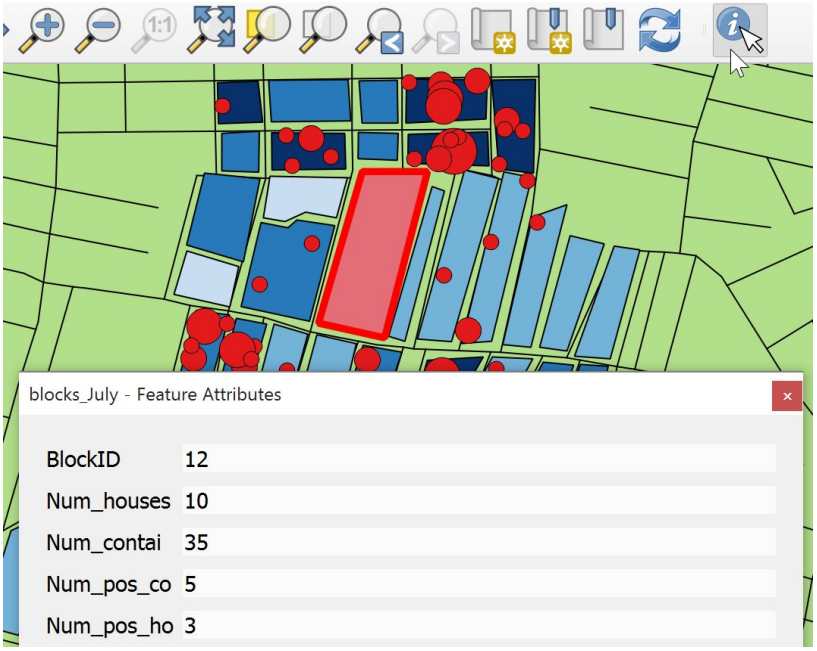
This should zoom us back in to the houses, blocks and roads that we saw at the start.

There are many ways to find out information about the map. One way is to use the 'Identify features' button to bring up information about map features stored in the currently selected layer :

- ➔ Left click the blocks layer in the Layers Panel to select it
- ➔ Left click the 'Identify Features' button in the toolbar
- ➔ Left click in one of the blue polygons on the map to select it



This should bring up an information panel on the where you should be able to see data that we have stored for each block, e.g. the number of houses ('Num_houses') within each block.

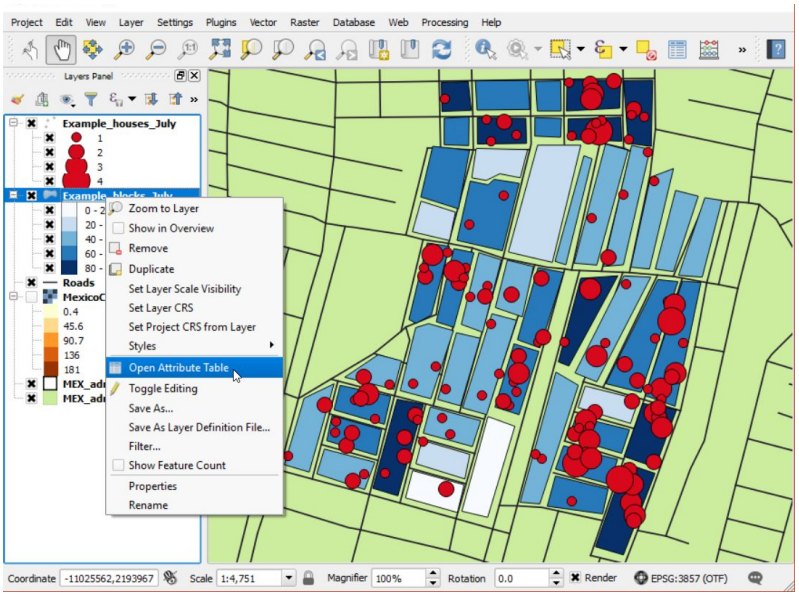


➔ Try selecting the 'Example_houses' layer and use the identify button to find out information about individual houses (the red circles).

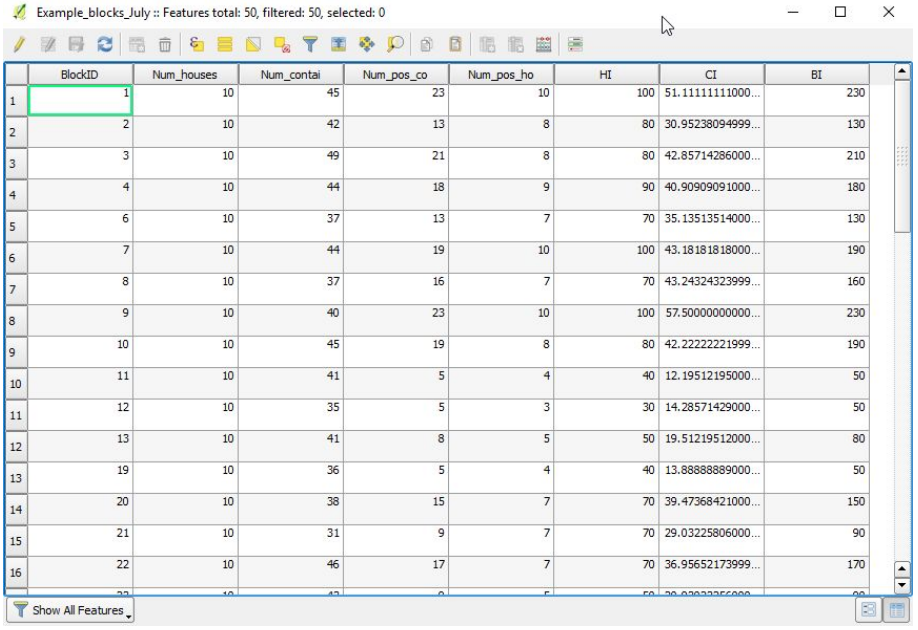
Attribute tables

This information about the layers is stored in 'Attribute tables' which are like spreadsheet tables. Attribute Tables are very useful. This is a key command to view attribute tables :

➔ Right click on a layer in the Layers Panel
➔ select 'Open Attribute Table'



The attribute table for the blocks layer should look like below. Each row stores data about one feature. In this example, a feature is a block so there is one row per block. Have a look at the columns and values.



	BlockID	Num_houses	Num_contai	Num_pos_co	Num_pos_ho	HI	CI	BI
1	1	10	45	23	10	100	51.11111111000...	230
2	2	10	42	13	8	80	30.95238094999...	130
3	3	10	49	21	8	80	42.85714286000...	210
4	4	10	44	18	9	90	40.90909091000...	180
5	6	10	37	13	7	70	35.13513514000...	130
6	7	10	44	19	10	100	43.18181818000...	190
7	8	10	37	16	7	70	43.24324323999...	160
8	9	10	40	23	10	100	57.50000000000...	230
9	10	10	45	19	8	80	42.22222221999...	190
10	11	10	41	5	4	40	12.19512195000...	50
11	12	10	35	5	3	30	14.28571429000...	50
12	13	10	41	8	5	50	19.51219512000...	80
13	19	10	36	5	4	40	13.88888889000...	50
14	20	10	38	15	7	70	39.47368421000...	150
15	21	10	31	9	7	70	29.03225806000...	90
16	22	10	46	17	7	70	36.95652173999...	170

The column names give some indication to the data that are held in them. These made-up data are for *Aedes* mosquito surveillance. They show the numbers of houses and the number of water containers found to have breeding *Aedes* per block. We don't need to spend much time here on exactly what they show - it could be any other variables of public-health interest, whatever data you have.

blockID	a unique identifier for each block
num_houses	number of houses within the block
num_contai	number of containers
num_pos_co	number of positive containers (positive=containing larvae or pupae)
num_pos_ho	number of positive houses
HI	House Index, percentage of positive houses
CI	Container Index, percentage of positive houses
BI	Breteau Index, number of positive containers per 100 houses

- ➔ Close the attribute table by clicking the x in the top right corner.
- ➔ Open the attribute table for the houses (points) layer.
- ➔ (Remember, right click on the layer name in the Layer Panel, then select 'Open Attribute Table'.)

The houses data has more columns and you can use the lower scroll bar to view the columns to the right. In this table there is one row for each point (house).

Spatial data types : Vector (points, lines or polygons) and Raster (gridded)

There are two main types of spatial data, namely **vector** data and **raster** data. The differences between these two types are described below. Note that the word 'vector' in this context does not refer to disease vectors such as mosquitoes.

Vector data

Vector data represent *points*, *lines* or *polygons*.

Points: For example, point data may represent the exact location of different houses included in an entomological survey or general points of interest, or the points may represent larger geographical areas such the centre of a community.

Lines: For example, line data may represent physical features such as rivers or roads, or routes travelled such as the route taken by a spraying truck, or the route walked by an individual. A single line can be stored as one object, or it may be split into smaller segments.

Polygons: Polygons are closed shapes. For example, a polygon may represent a physical feature such as a pond, lake or park, or an administrative boundary such as a region or district. You can also create custom polygons relevant to your project, such as survey blocks in our example.

Vector (point, line, polygon) data are commonly stored as a **shapefile**. A shapefile, somewhat confusingly, actually consists of multiple files with the same initial name and different extensions after the '.' (e.g. a.shp, a.shx & a.dxf). These files contain different elements of the data, including the locations and variables (known as attributes) associated with each. All of the individual files which make up the shapefile must be saved in the same folder on your computer. Minimally, a shapefile consists of three files. These three files end in .shp, .shx, and .dbf . There may be other files that comprise your shapefile. You don't need to know what is in each file to be able to use them.

Raster data

Raster data represent something that is geographically continuous. The data are stored as a continuous grid of cells (or pixels), and each cell is assigned a value (single band raster), or possibly multiple values (multiband raster). The dimensions of the cell vary depending on the data that you are using, and is referred to as the 'spatial resolution' of the raster data. For example, we will be looking at gridded population estimate data that has a spatial resolution of 100m, meaning that each cell is 100m by 100m. It is a single band raster, the value associated with the cell represents the estimated number of people living in the area covered by the cell.

Raster data are frequently used to represent information measured by remote sensing (usually satellites). For example, the images used by Google in their mapping products such as Google Earth are obtained from satellites. Sensors on satellites can also be used to derive other physical and environmental measures such as elevation, vegetation, land use, plus climate measure such as temperature and rainfall. Data from numerical models can also be represented by a raster. For example, heat maps are often used to represent disease transmission risk.

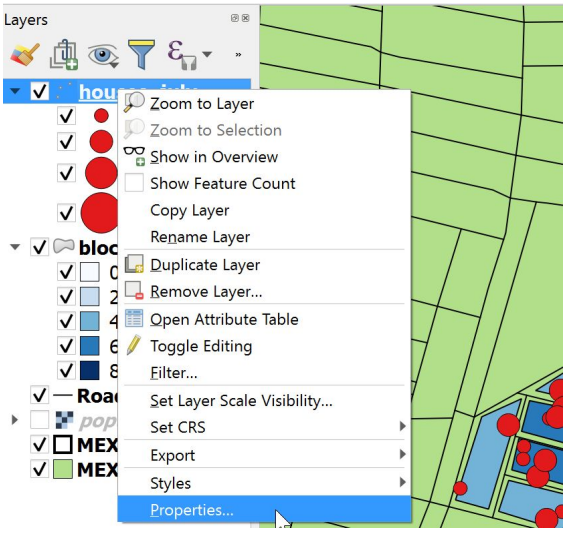
Raster data can be stored in the form of an image file with associated geographical information. For example, the population data we're using in the exercises consists of two files, namely the TIFF image file (.tif) plus an additional file with the same file name within which the geographical information is stored (.tif.aux).

We will look at how to add these different data types and files to a project and map later. Before that we are going to show you how to modify the appearance of a map.

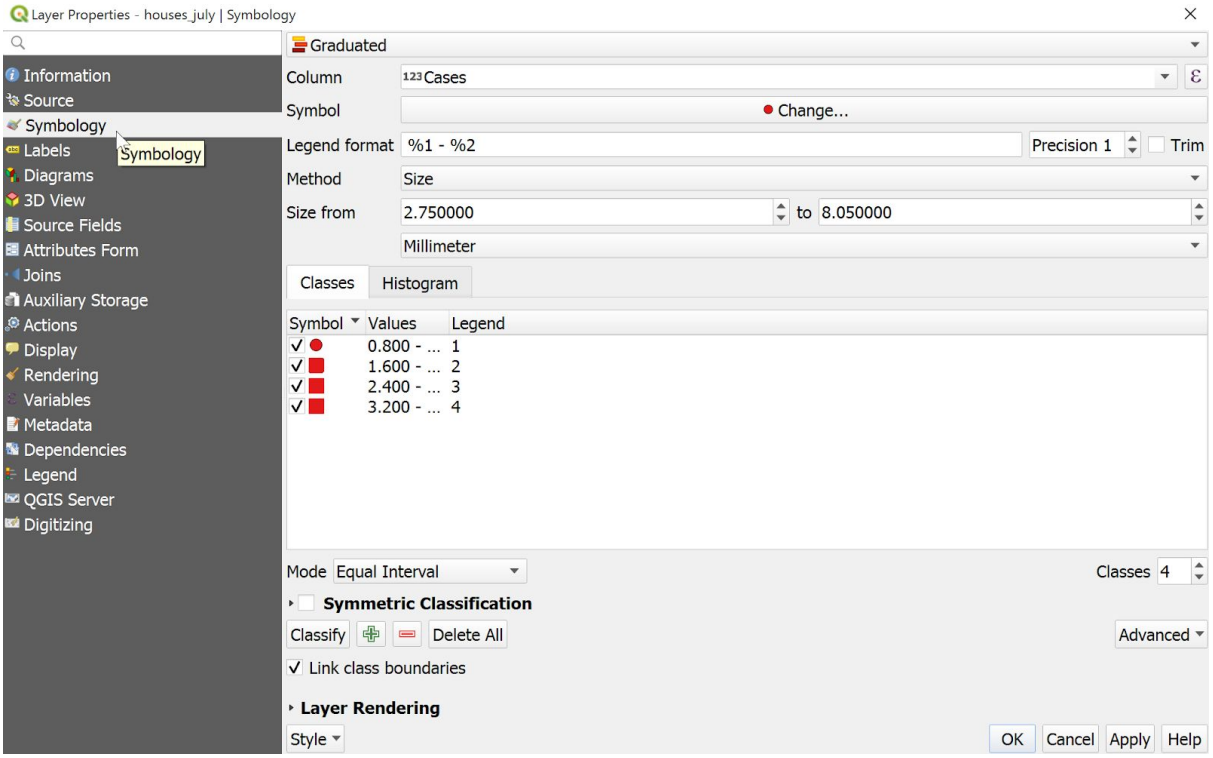
Symbology - changing appearances

Symbology determines how data are displayed on the map. For example, the current symbology we are using results in the locations of cases being presented on the map by a red circle, such that the size of the circle represents the number of cases identified. Symbology can be modified using Layer properties. Symbology options are different for point, line, polygon and raster data. We will start by looking at point data.

- ➔ Right click the houses layer in the Layers panel
- ➔ Select Properties



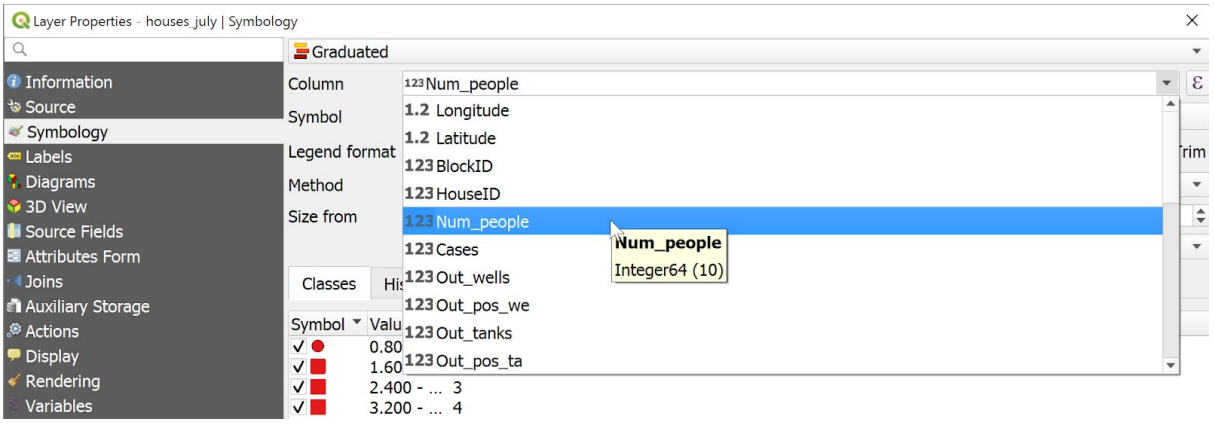
➔ Select the Symbology option on the left in the Layer properties window.



There are lots of Symbology options and we don't need to understand them all at first. Currently the house layer is set to present the variable ‘cases’ using the 'Graduated' option, as seen at the top of the Symbology screen. The graduated option enables the size of the points to be determined by the values in one of the columns of the Attribute table, which initially is *cases*. Usually, we use increasing point sizes to represent increasing values.

➔ in the ‘Column’ choice box choose ‘Num_people’ to change the column that sets the appearance of the points

➔ click Apply or OK at the bottom of the window

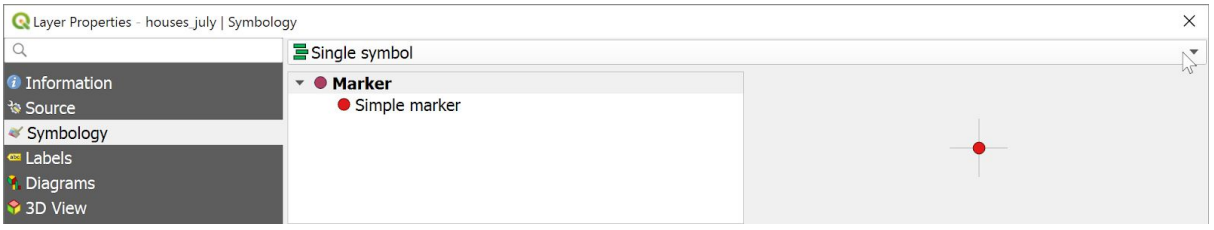


Changing Column to num_people means that the size of the points is now determined by the number of people in each house and the map will look different.

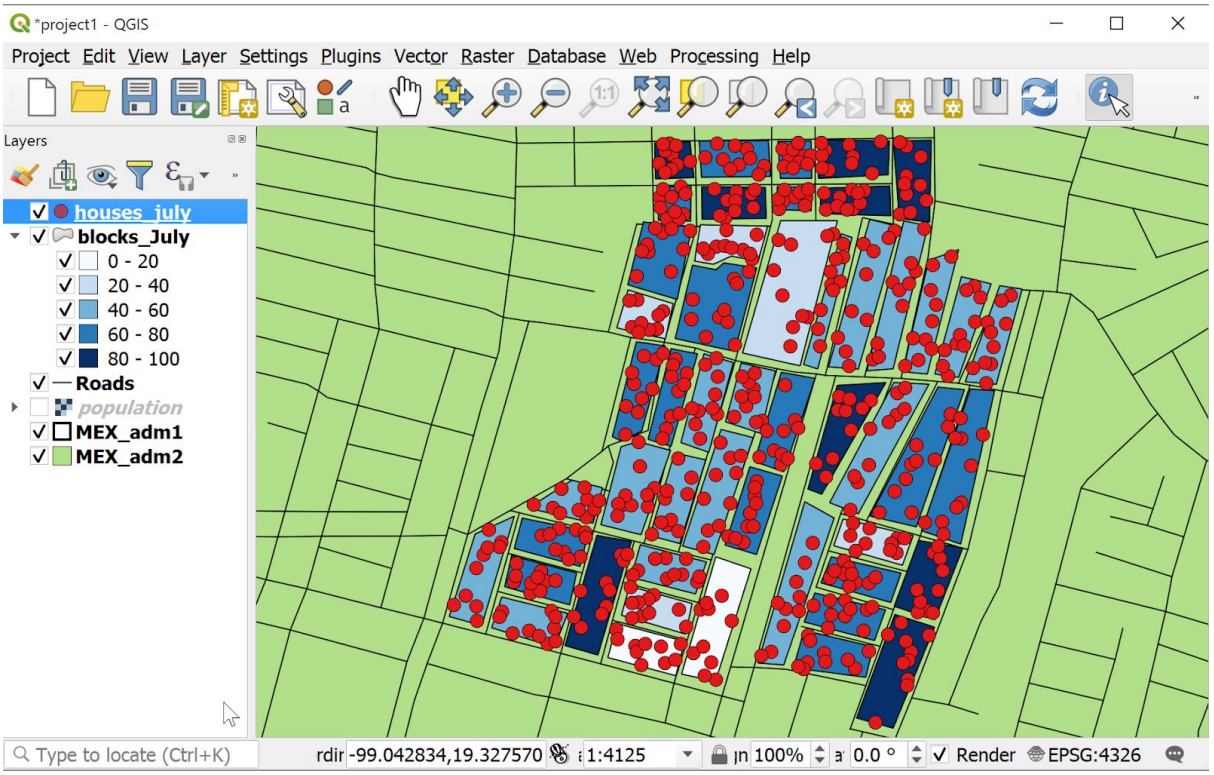
Note that you can only graduate your data by numeric data. Points and lines can be graduated by size. Points, lines, polygons and rasters can be graduated by colour. (The size of the polygons and raster cells is predetermined).

We can also have points of a constant size :

- ➔ change from the current 'Graduated' option at the top to 'Single Symbol'
- ➔ click Apply or OK at the bottom of the window



Now the map should like the below with both the point symbols and annotation in the layers panel changed :

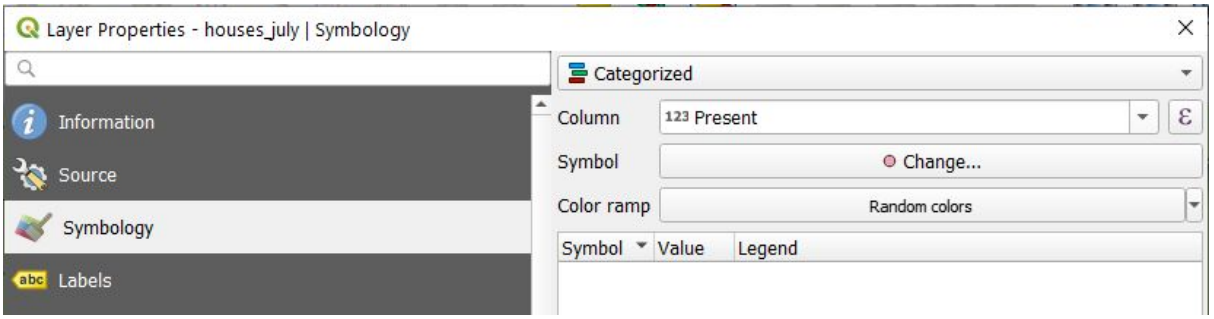


Symbology also has the option to represent categorical data. Colour can be used to visualise the different categories of your attributes, for example positive/negative, resistant/susceptible, study groups A-Z. The option categorized will split your data wherever it finds even the smallest of differences and create categories.

- ➔ Right click on the houses layer again to bring up the Properties, Symbology window
- ➔ Set the top option to 'Categorized'
- ➔ Set 'Column' to 'Present'

The variable 'present' in this data represents whether mosquitoes were present (1) or absent (0) at the surveyed household.

- ➔ The Color ramp can remain 'Random colors'.

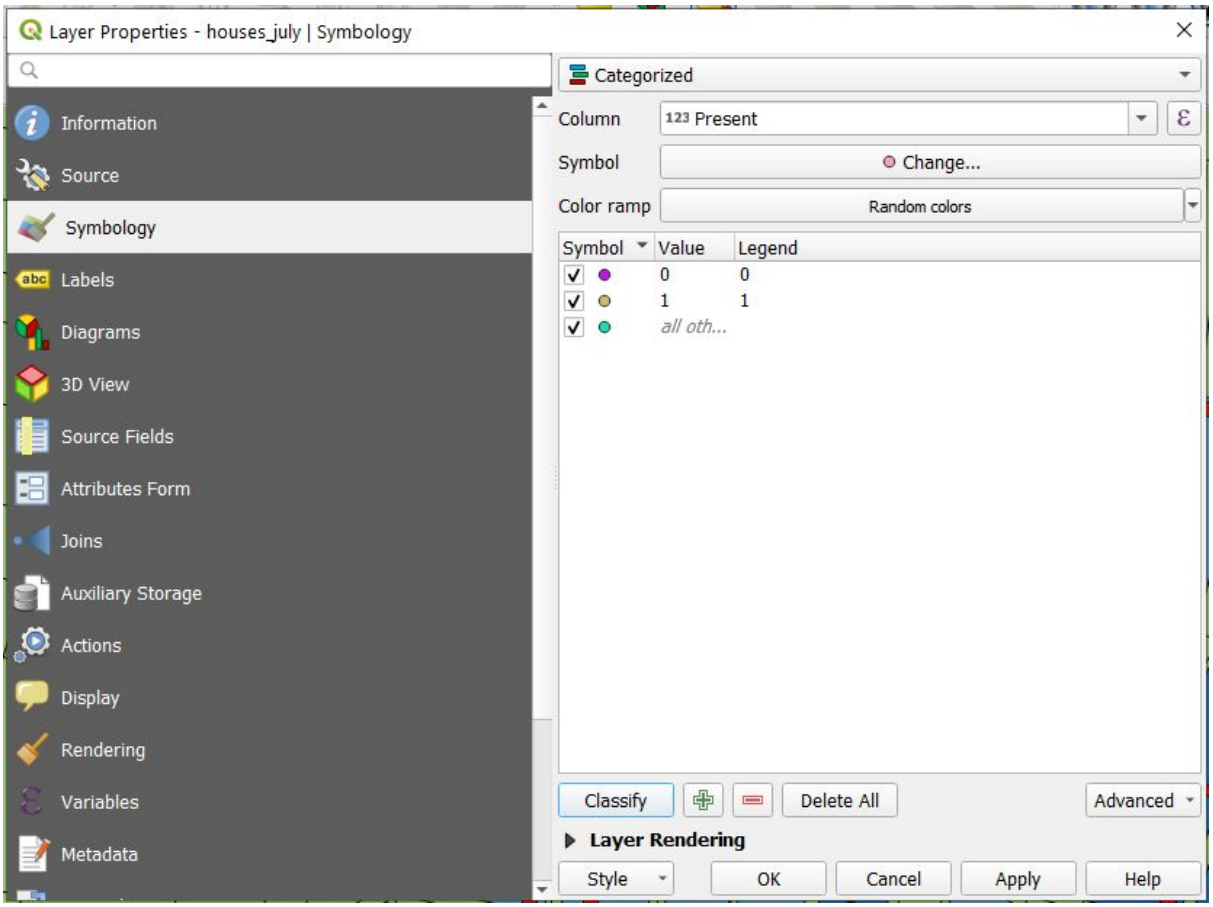


When we have made any change to how features are displayed these changes don't register until the 'Classify' button is pressed. This frequently confuses new users (and us too).

➔ Click ‘Classify’

The ‘Classify’ button assigns colours to data, according to your chosen settings. Note that these will then appear as a preview in the central section.

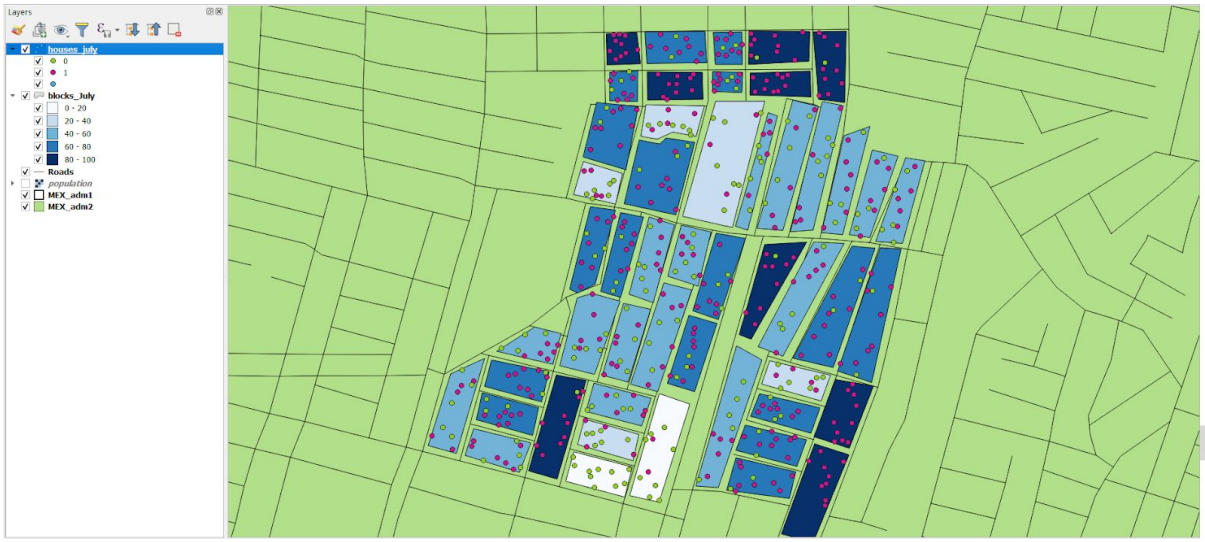
If you do not click on ‘Classify’, it may seem that your points have disappeared from your map. This can be a bit disconcerting, but you simply need to return to Symbology (right click on the layer, properties, symbology) and complete the process by clicking on classify to assign colours to categories.



➔ Try using the ‘Shuffle Random Colors’ option to change the colours.
➔ Click OK to make the changes.

Now the map should appear as below with the points displayed in two colours, representing the presence or absence of mosquitoes.

You may have noticed a third category, what does this represent?



Similar options can be set for the blocks (polygon) layer. You can adjust the colour of polygons to be all the same, graduated, or categorized as appropriate for your data. You cannot change their size or shape.

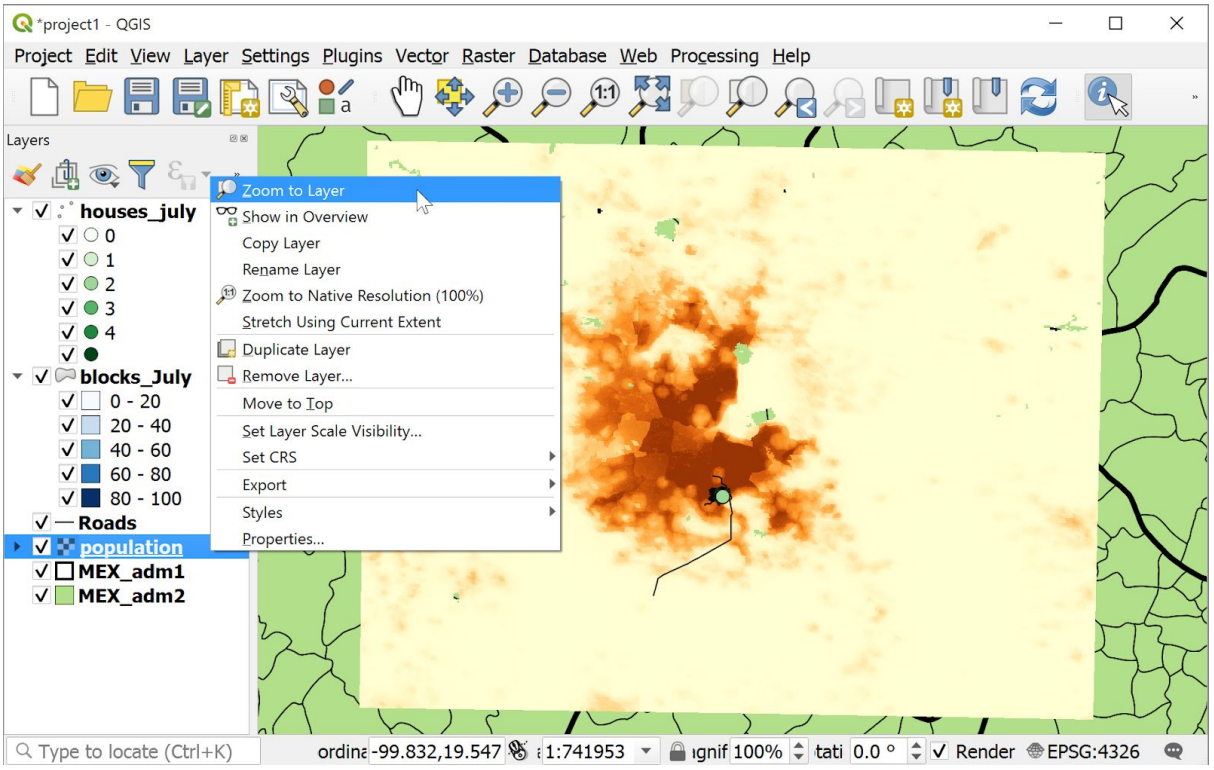
Because we have made changes within our project it is a good idea to save it. The project file just contains 1) settings that determine the appearance of layers and 2) links to the data. Saving the project will not change the data. To save the project :

- ➔ from the menu bar, select Project, Save As
- ➔ Browse to **training-QGIS-201910/my_work/projects_QGIS**
- ➔ name the new project : **project1_modified.qgz**

When you make any later changes you can just click the Save icon to save the project again.

You can control the Symbology of raster (gridded) data too. We have added raster data of estimated human population for Mexico city in the layer 'population'. This layer is derived from data downloaded from the WorldPop project at : http://www.worldpop.org.uk

- ➔ Click the check box for ‘population’ in your layers panel.
- ➔ To zoom to the extent of this layer, right click on ‘population’, select Zoom to Layer

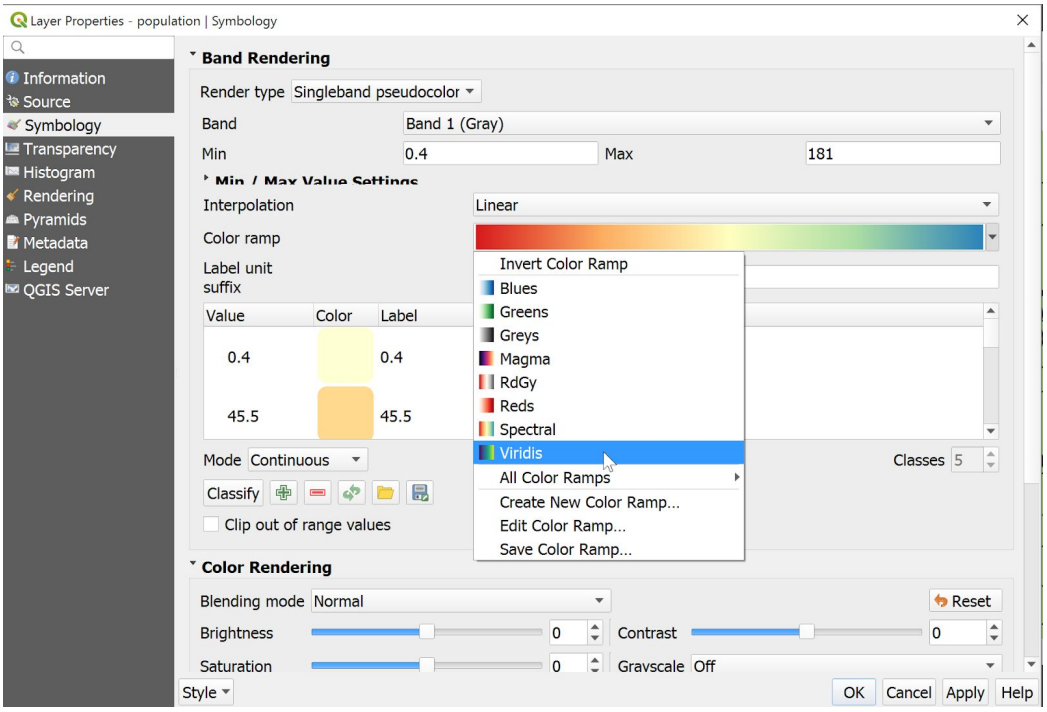


➔ In the layers panel Right click again on population, select Properties

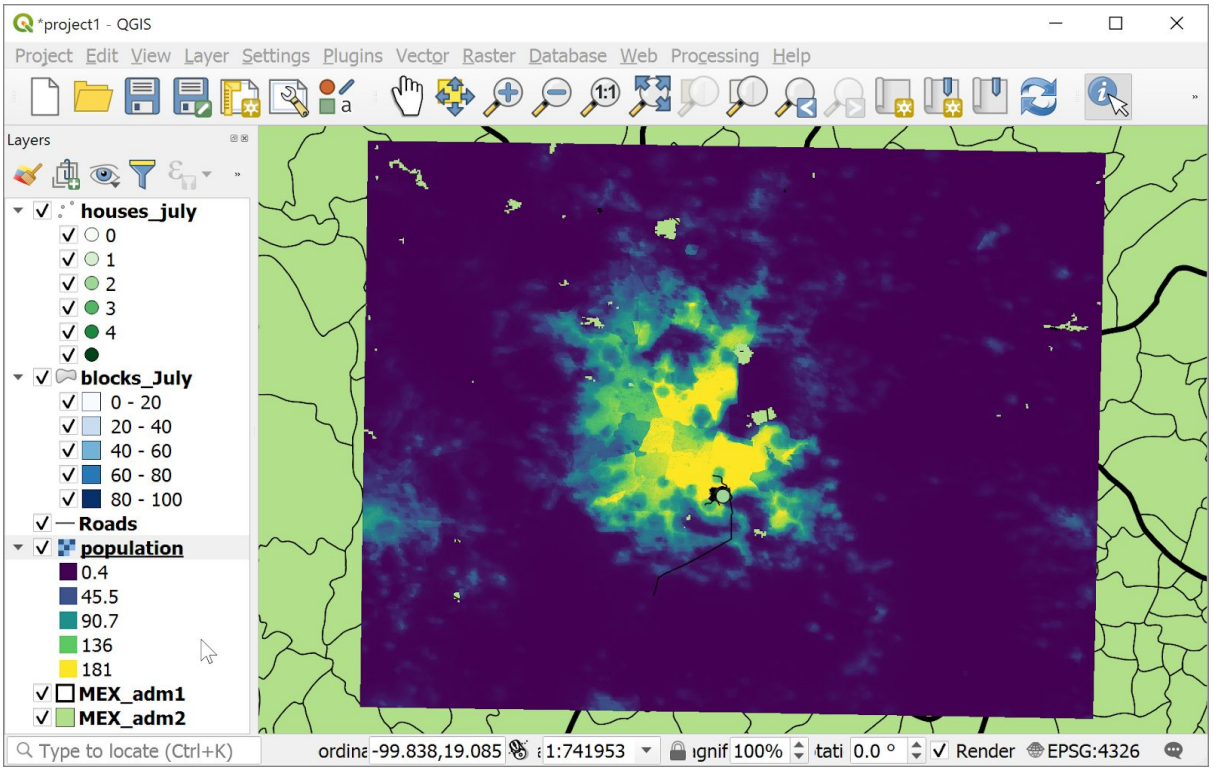
➔ Select Symbology on the left

➔ Change 'Color ramp' to *Viridis* and press OK.

Viridis is a modern colour scale that works well for colour blind people and if it is viewed in black and white. Rainbow colour scales have various problems and should usually be avoided.



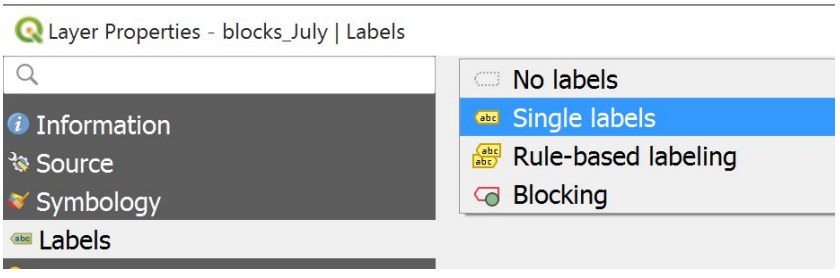
Now the map should look like this, the yellow areas represent high population :



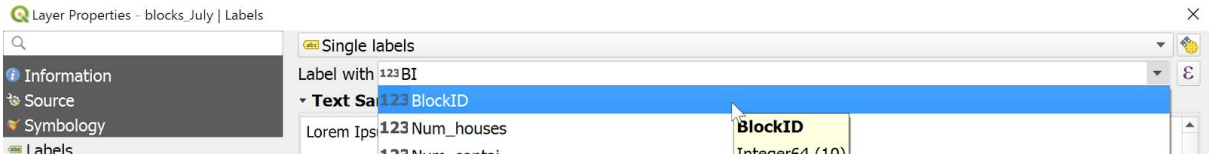
Add labels to features

Labels allow us to put text on the map associated with points, lines or polygons. The text for these labels can be anything in the attribute table, i.e. the names or values of your data.

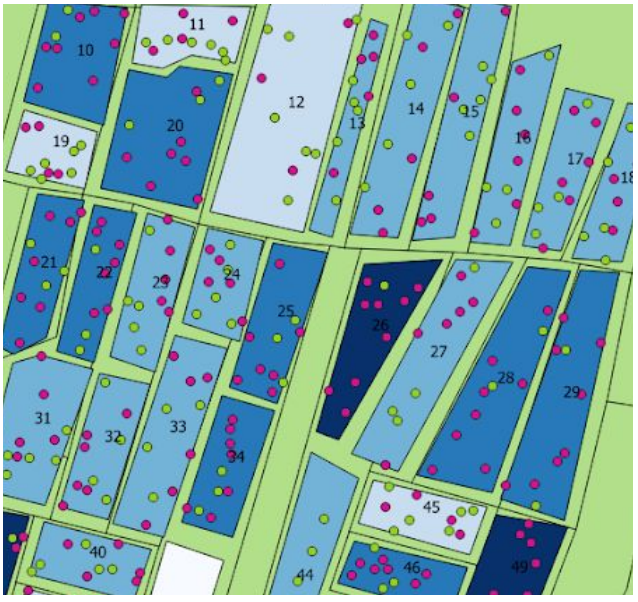
- ➔ In the layers panel uncheck the population raster layer
- ➔ Zoom back to the extent of the blocks layer (right click, zoom to layer)
- ➔ Right click on the blocks layer and select Properties.
- ➔ Select 'Labels' on the left
- ➔ At the top change from ‘No labels’ to ‘Single labels’



- ➔ Set 'Label with' to 'blockID' and press OK at the bottom.



The blockID column in the Attribute table is used to label the blocks :



Note that there are options available for formatting label text within the Labels window (right click, properties, labels) . These can be to useful make labels more visible to the map reader.

'Basemaps' & 'plugins'

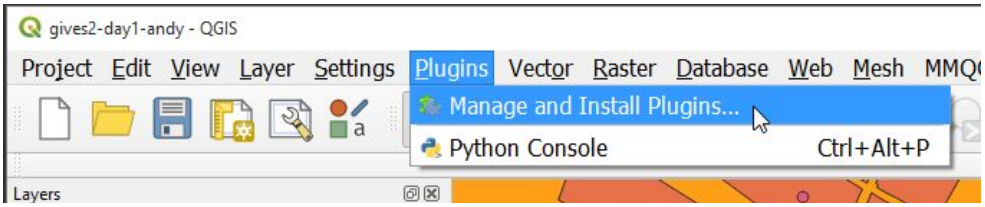
Two new QGIS words for us to learn, basemap and plugin.

Adding a background 'base map' quickly adds context to our data, allows us to visualise the location of our data without the need for other layers.

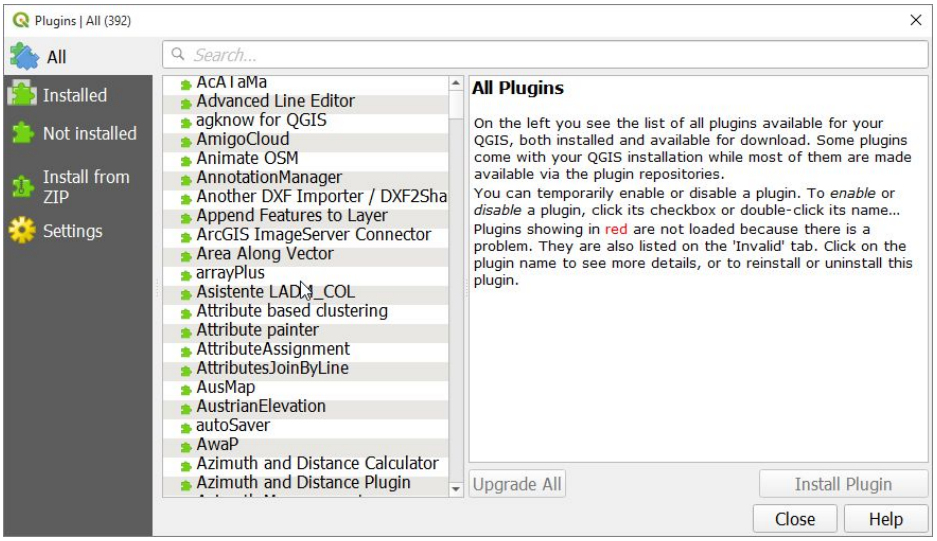
However, the base maps we are going to use are downloaded live from the internet, so do require an internet connection. The base map will automatically update as you adjust the zoom, or pan across the map window.

To add a basemap we first need to add a 'plugin'. Plugins add extra options to QGIS. We are going to use the **QuickMapServices** plugin.

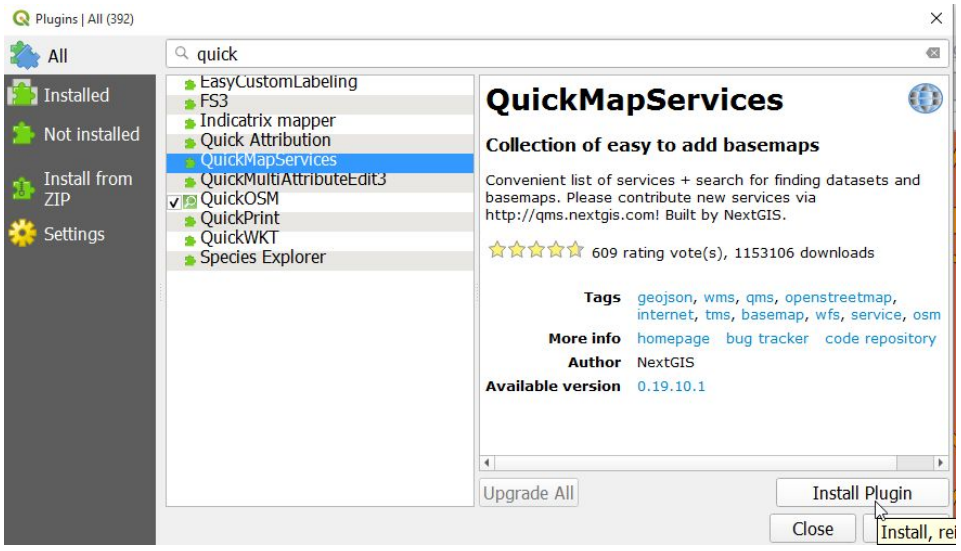
➔ Click on Plugins, Manage and Install Plugins



That should display this window. It should show > 300 available plugins. **If it only shows a low number ~ 10 that can indicate that you are not connected to the internet.** Plugins are stored on the internet. Check your internet connection and repeat.

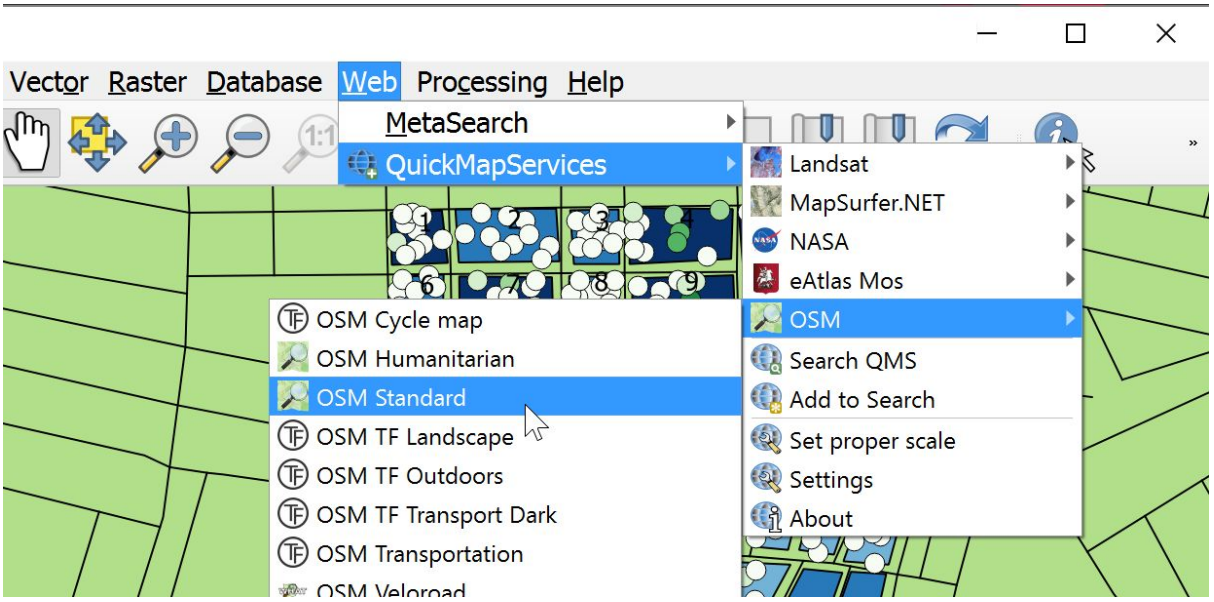


➔ Type **quick** into the search box
➔ Click on **QuickMapServices**
➔ Click the 'Install Plugin' button in the lower right hand corner.



➔ After it has installed, click the 'close' button to shut the plugins window.

➔ Select **Web, QuickMapServices, OSM, OSM Standard** from the main QGIS menus



OSM stands for OpenStreetMap, an excellent, freely available map data created by volunteers. In remote areas OpenStreetMap often has better data than Google or Bing Maps. Also OpenStreetMap has the advantage that you are allowed to use images from it in your own publications. Google and Bing have licenses that restrict you using them in your publications.

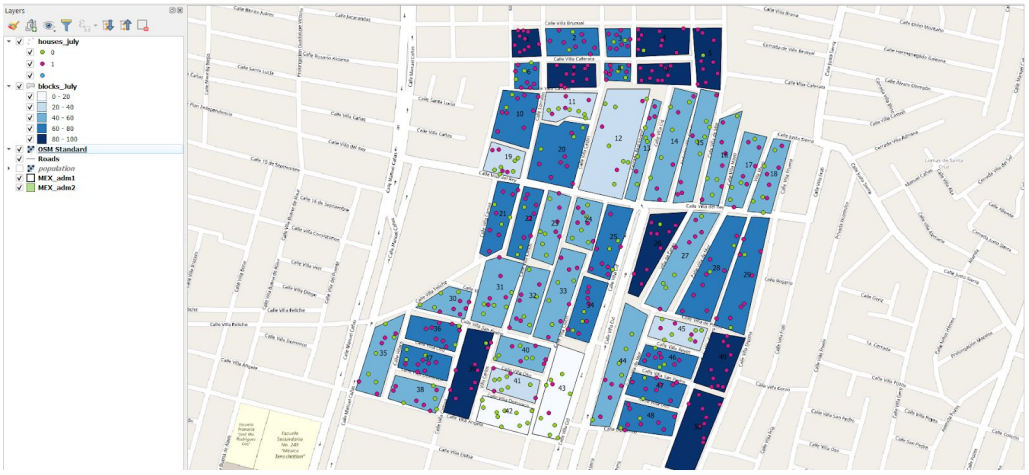
You should see OSM Standard added to the bottom of the layers list in the Layers panel. But nothing seems different about the map display.

Can you think why we cannot see the new layer ?

It is because it is covered by the layers above it in the list.

➔ Drag OSM Standard above Roads in the Layers list

Now you should be able to see the basemap including street names, buildings etc.



➔ Try zooming and moving the map to see how the basemap changes

You should see that new map sections (they are sometimes called map tiles) appear if you move the map to new areas. If you zoom in you should see that the map gets more detailed.

Note that these operations can require significant data transfer over the internet. It can lead to your map freezing or updating slowly If you have a poor internet connection (or if, for example, there are lots of people in a training room doing this at the same time!). If you have this problem untick the OSM Standard layer.

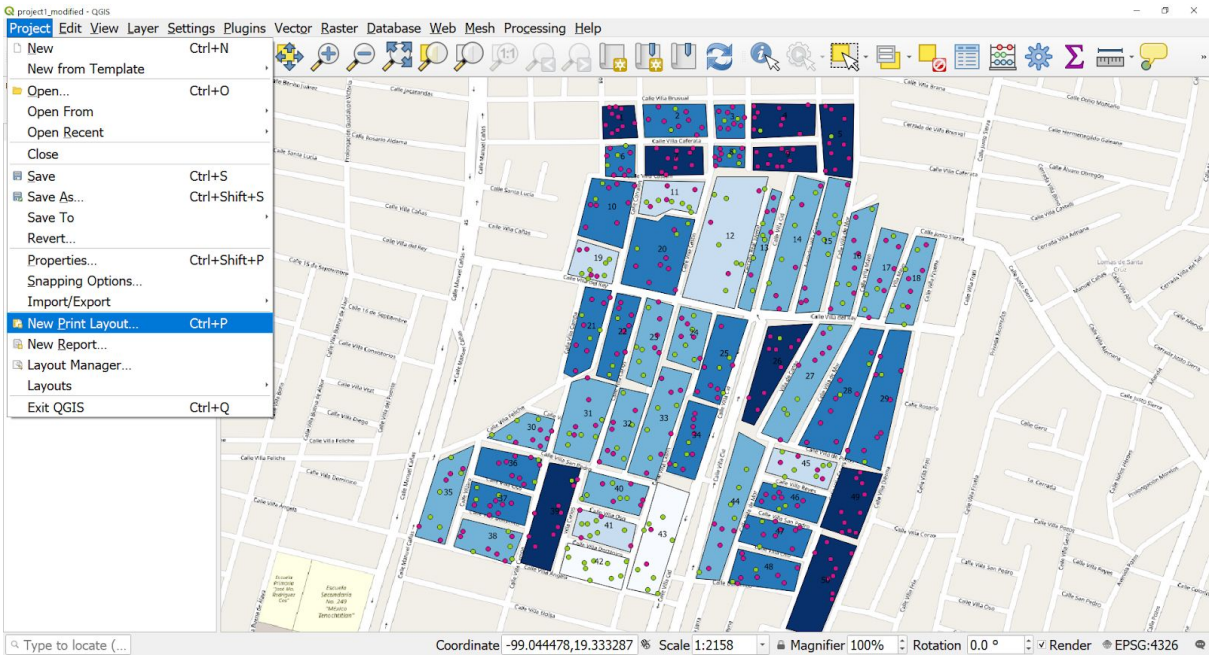
Print Layouts : outputting maps for others

Now we are going to look at how we can make maps to show to other people, for example in documents, presentations or web pages.

We will see how to create maps as images or pdfs that can be sent to others to view without them needing GIS software and your files of data.

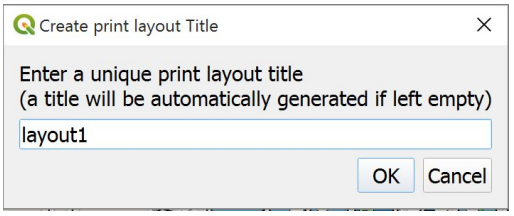
In QGIS, Print Layouts allow us to compose a map with legends, titles, labels etc., to set their size and arrangement on the page and then to output that map in a variety of formats including images or pdfs.

➔ Click Project, New Print Layout



➔ type a title for the layout, e.g. layout1

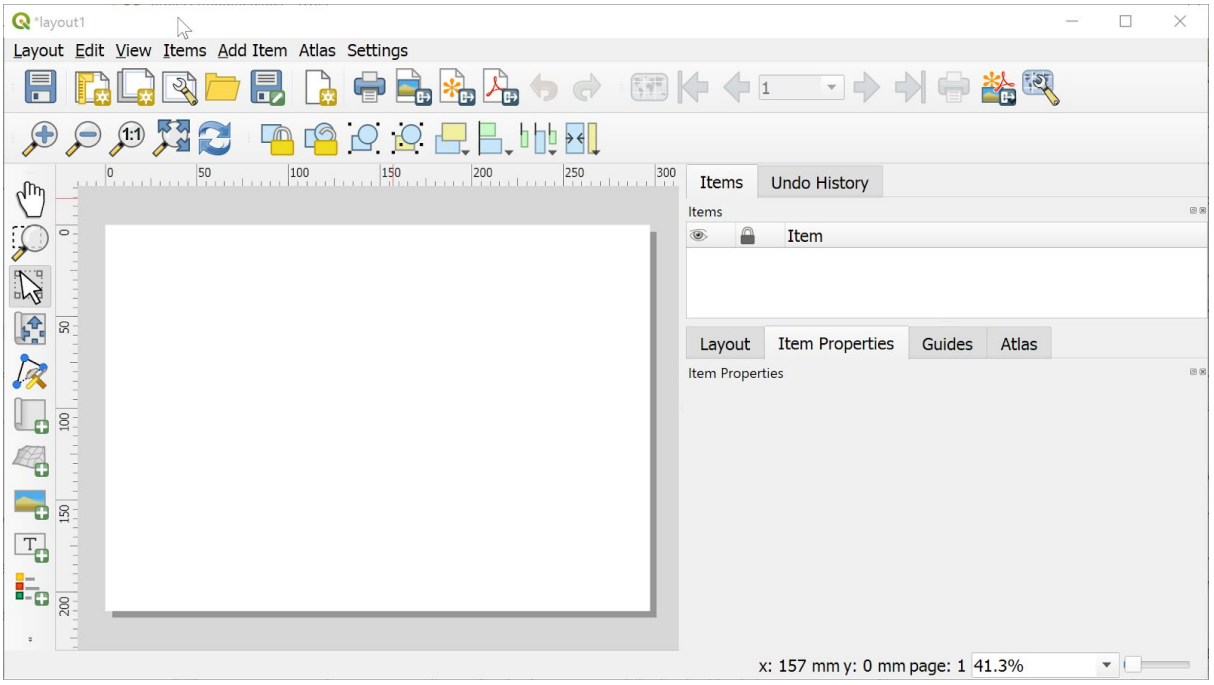
Layouts are saved as a part of your project and you can have more than 1 to be able to present maps in different ways.



The layout should appear as a separate window as shown below. Your main project window remains open.

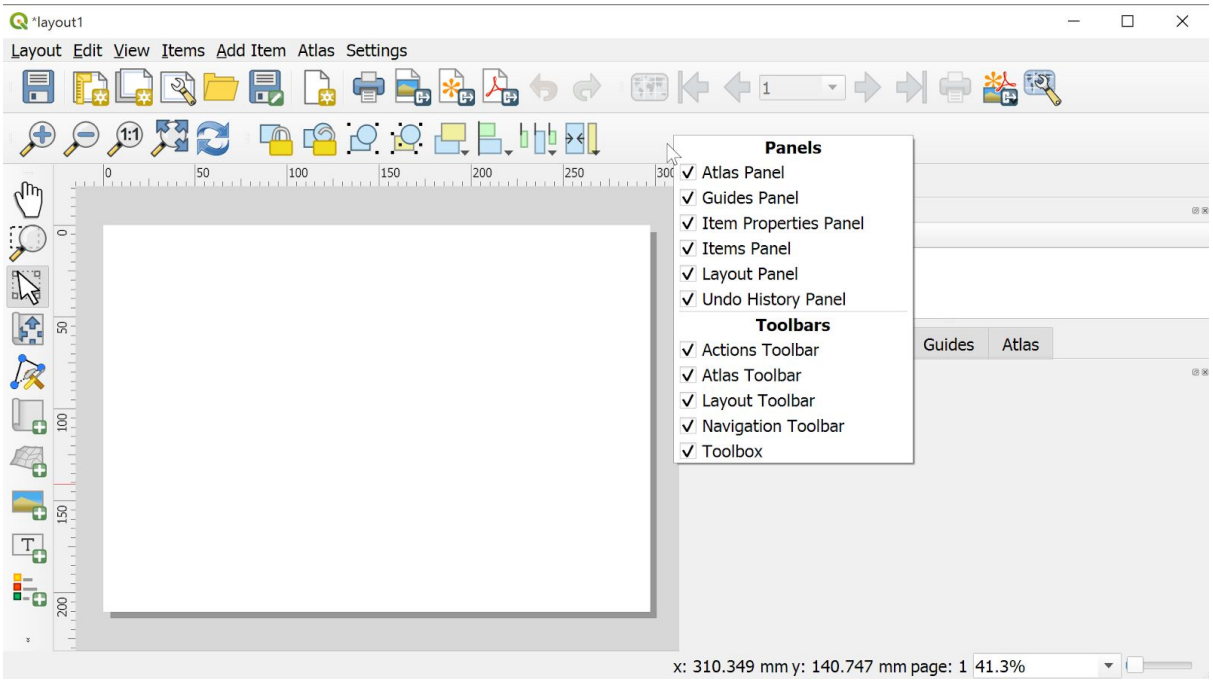
There are 3 main areas in the Print Layout :

1. A canvas on the left - effectively a blank sheet of paper that you can add things to.
2. A list of Items on the top right, initially blank. Will show one row per item you add.
3. Item properties on the lower right. Allows you to edit properties of selected item.

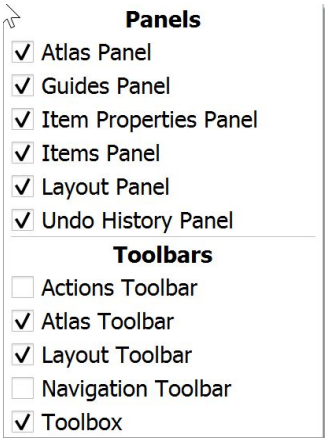


Similar to the main QGIS window the print layout window can be modified to show more or less panels and toolbars.

➔ Right click in the empty grey area to the right of the menu icons.



➔ For now turn off the Actions and Navigation toolbars to give us more space

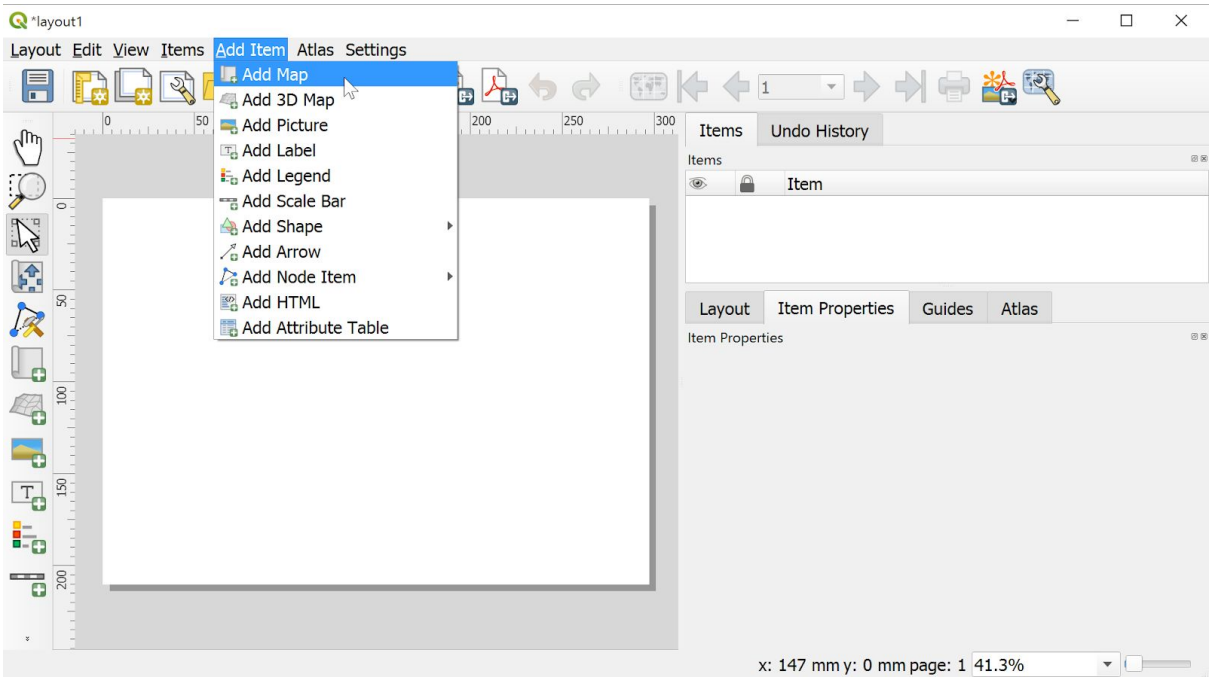


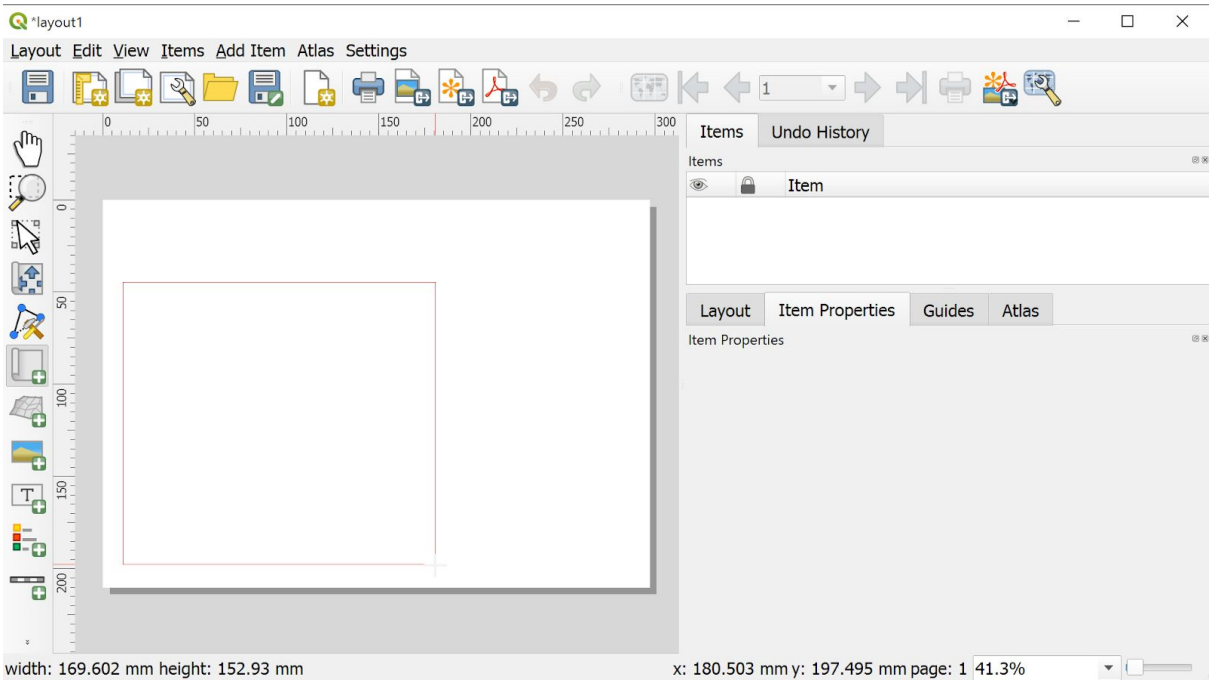
Sometimes the Items and Item properties panels can get turned off by mistake. If that happens you can just right click in an empty grey area to turn them back on again.

Adding items to Print Layout

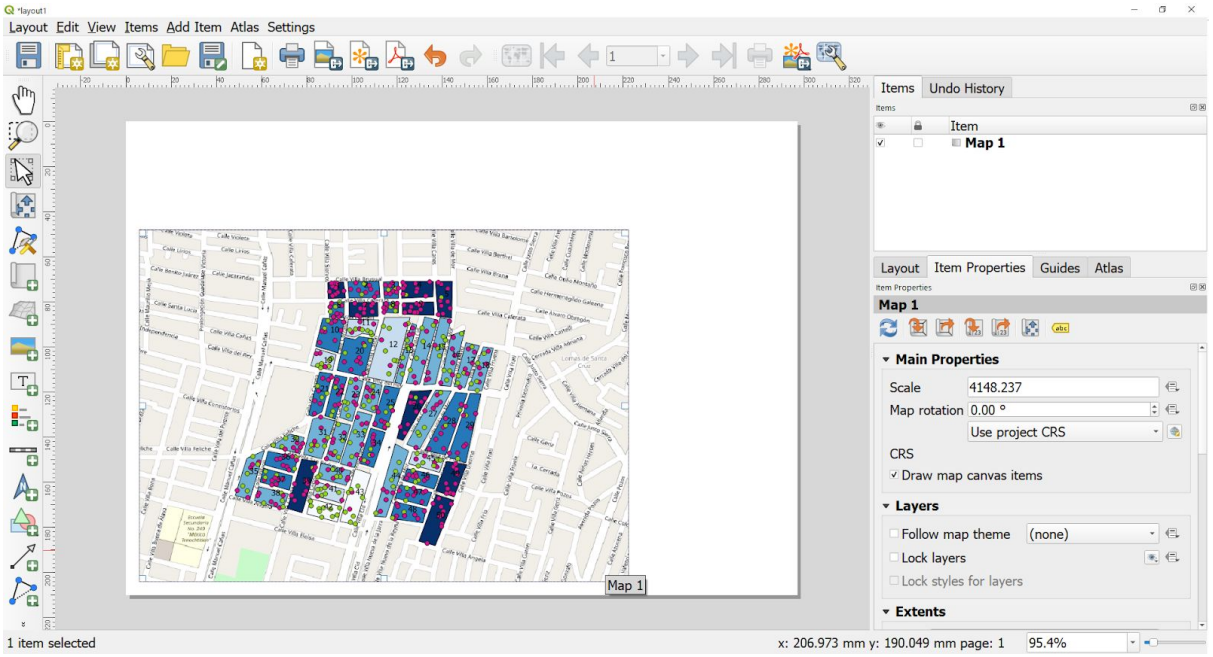
First we will add a map to our Print Layout.

➔ From the Print Layout Menu select Add Item, Add Map
➔ Click and drag to draw a box within the blank canvas on the left





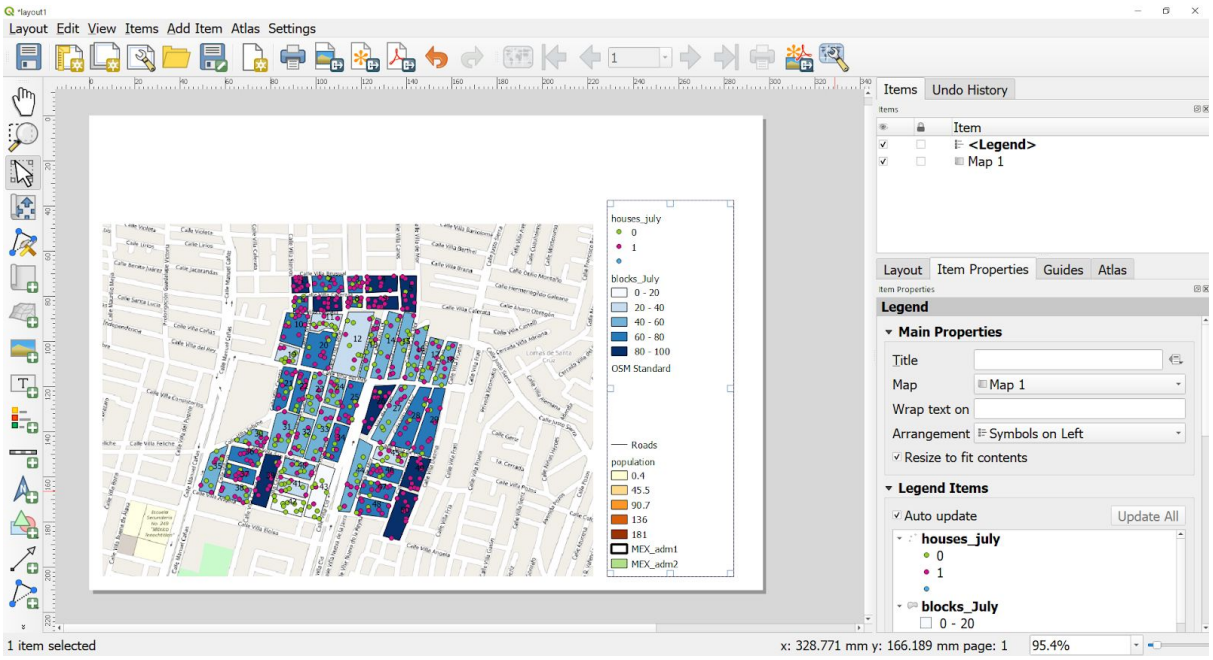
You should see a map appear and that an item called Map 1 has been added to the Items list at the top right. The map shows the current view from your main QGIS project.



Similarly we can add a legend.

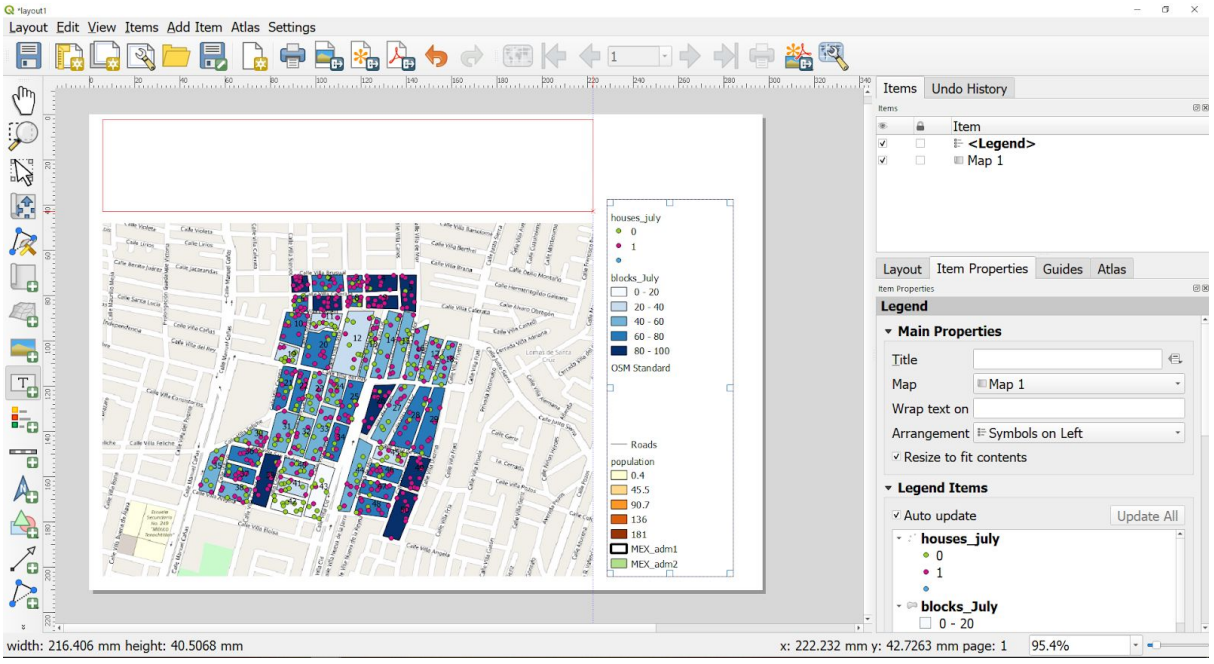
- ➔ From the Print Layout Menu select Add Item, Add Legend
- ➔ Click and drag to draw a box within the blank canvas to the right of the map

QGIS will add a legend and automatically modify the size to fit in all of the components. You should also see that an Item called <Legend> has been added to the list of items at the top right.

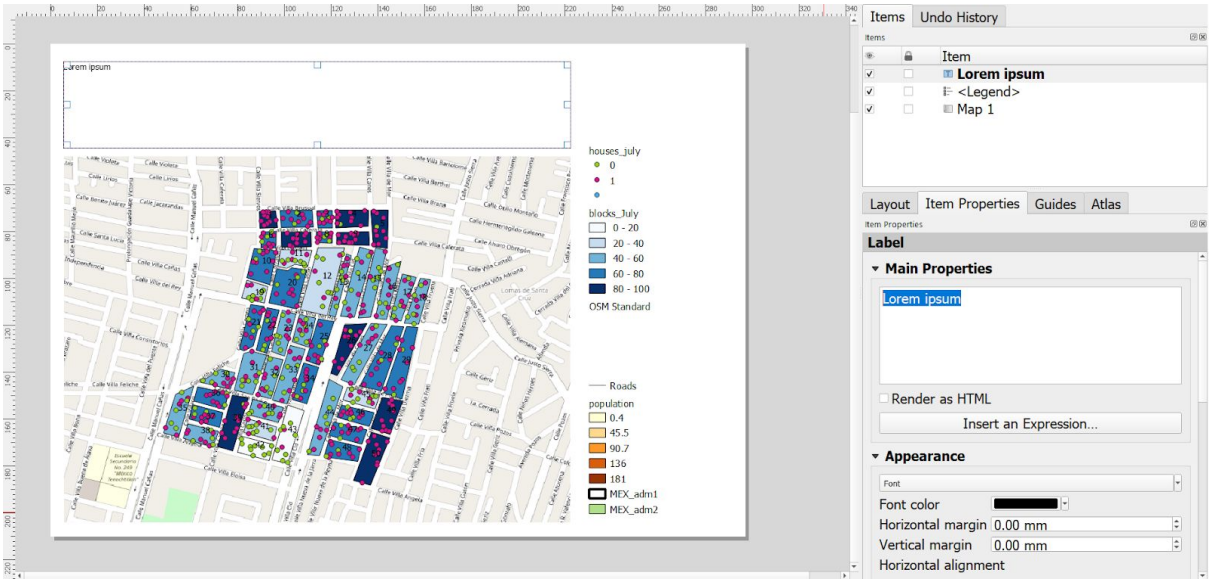


Next we can add a Title by using the Add Label option.

- ➔ From the Print Layout Menu select Add Item, Add Label
- ➔ Draw a box above the map.



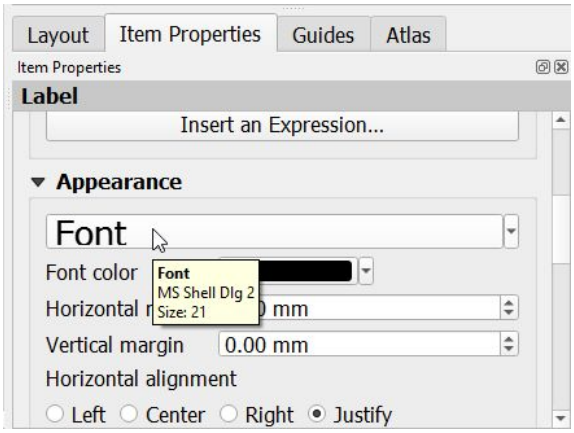
- ➔ In the Item Properties box at the lower right, select the 'Lorem ipsum' and replace with your own Title.



Here is a nice trick that I found by accident to change the size of the title text (which is too small to start with) :

- ➔ scroll down in item properties until you see Font
- ➔ place the cursor over Font

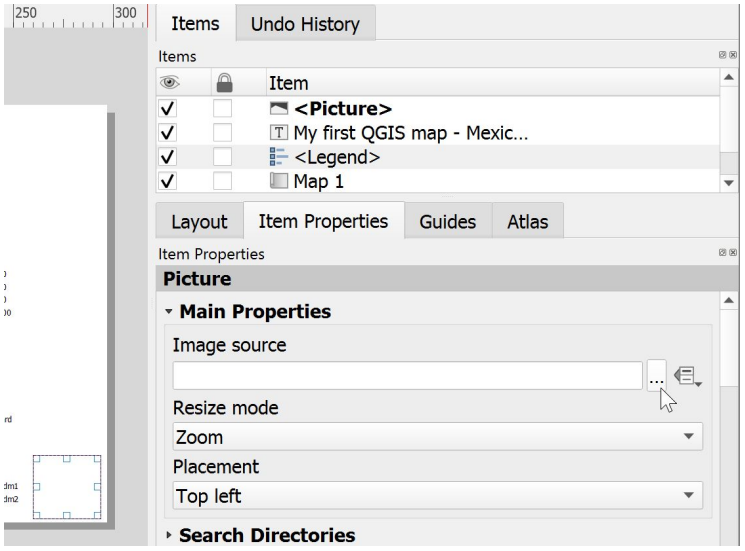
➔ using the mouse scroll wheel you can increase and decrease the size of the text.



➔ Alternatively click on Font and it will bring up a window allowing you to select font size

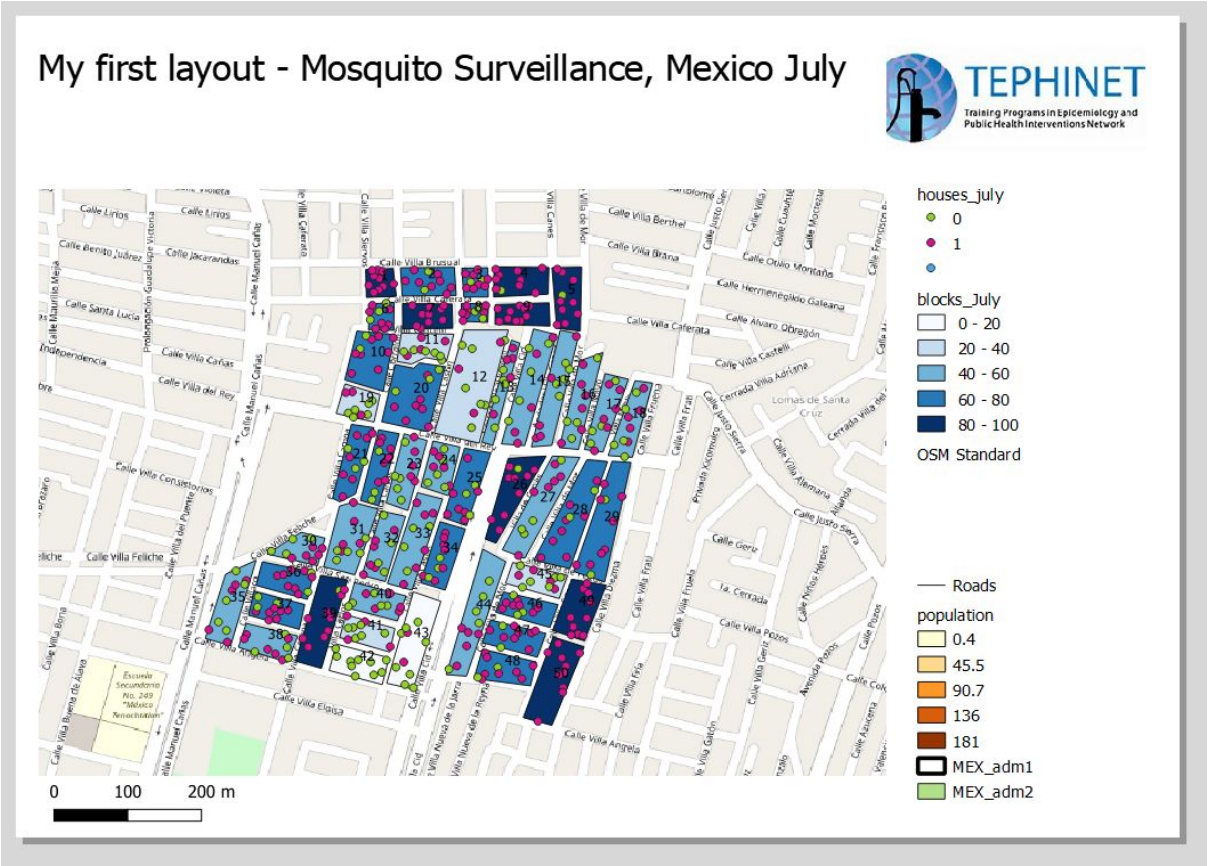
To add a logo.

➔ Add Item, Add Picture.
➔ Click and drag to add a rectangle to the canvas.
➔ In the Item properties click the three dots (...) to browse to find a file
➔ Tephinet logos are in the course folder, or you can use your own.



To add a scalebar to the Print Layout.

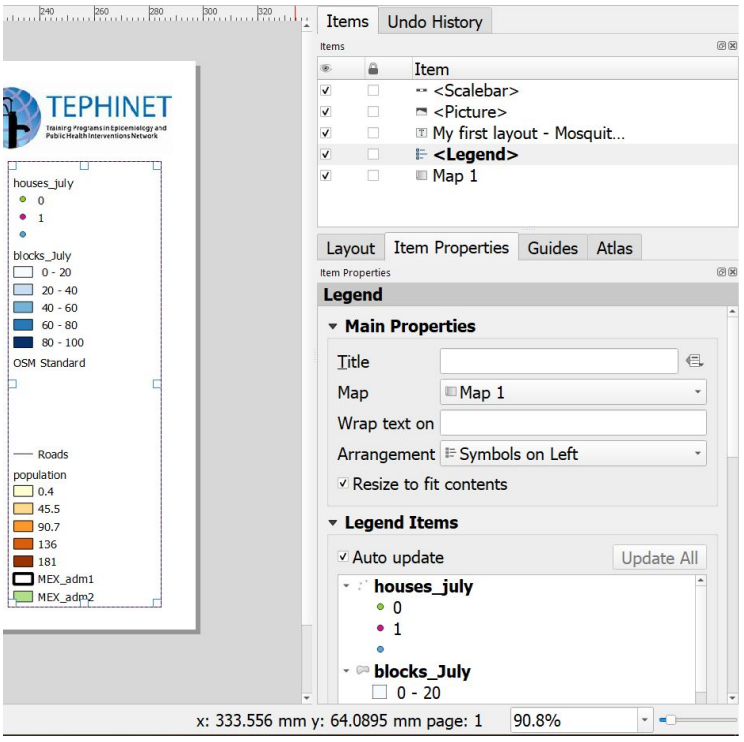
➔ Add Item, Add Scale Bar.
➔ Click and drag to add a rectangle to the canvas.



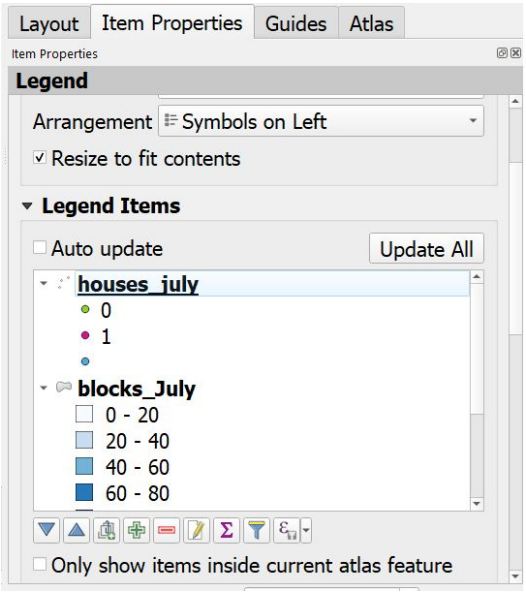
You may notice that the legend contains elements that don't actually appear in the map.

We can edit the legend to remove the elements we don't want.

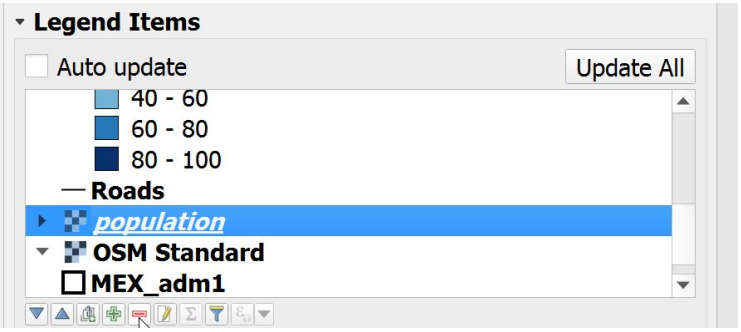
- ➔ select the legend either by clicking on it in the canvas or the Items list
- ➔ In Item properties, scroll down until you see Legend Items
- ➔ Untick the box Auto Update



This allows us to edit the legend from within Print Layout (otherwise you can only edit from within the main QGIS project, note that if you tick the box again after making any edits those edits will be lost). It should make a line of small icons, including a green plus and a red minus, visible below the Legend Items.

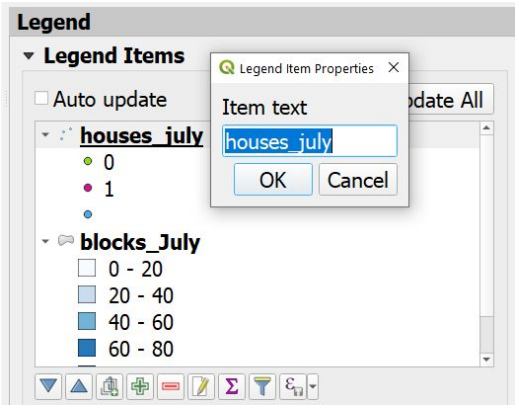


- ➔ Click population (one of the layers we can't see in the map)
- ➔ click the red minus icon to remove it from the legend
- ➔ Repeat this to remove the other items below that are not useful in the legend



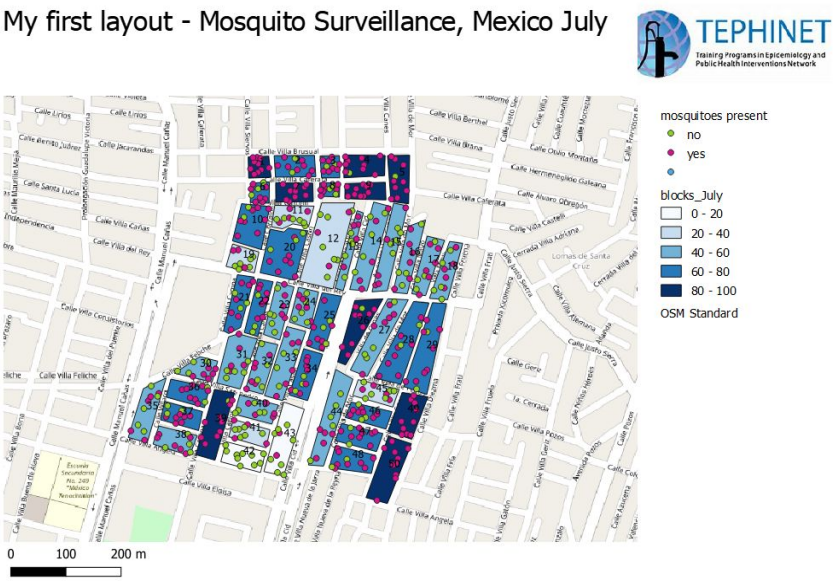
Here we can also modify the text in the legend to make it more useful.

- ➔ Double click on houses_july
- ➔ In the item text window change it to 'mosquitoes present'



Remember to press the save icon at the top left after you've made changes to save your Print Layout.

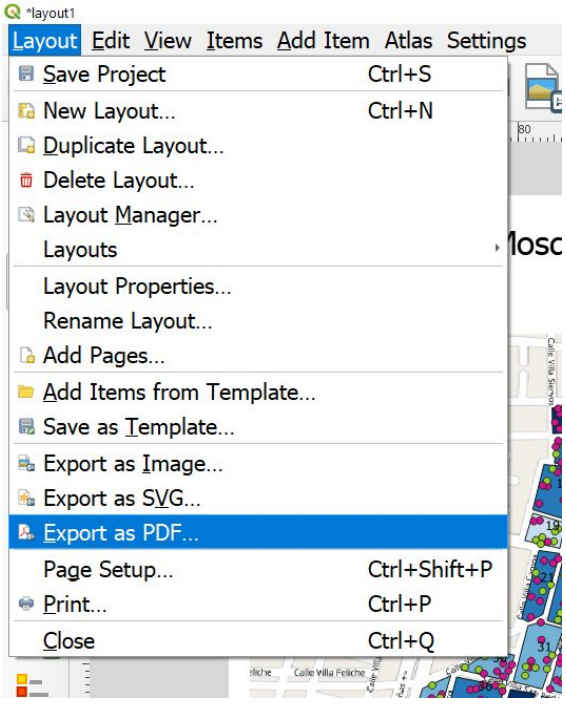
Now your map in the canvas window should look something like this :



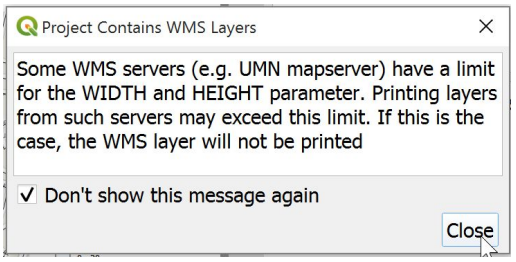
Now we can output the map by exporting to various formats including an image or pdf.

A pdf preserves lines and text in better quality, an image may be easier to insert into a document or presentation. If you are exporting a map with lots of features, be aware that a pdf can become quite large and that then an image may be better.

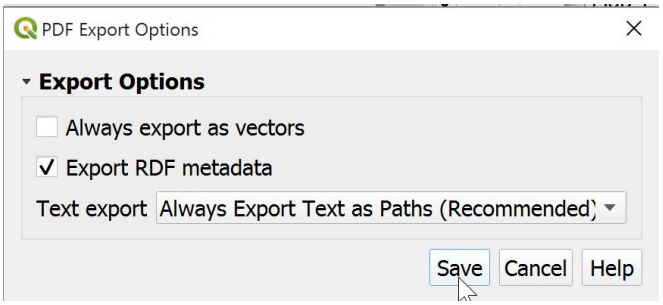
➔ Select Layout, Export as pdf



You may get this message window, don't worry about it just press close.



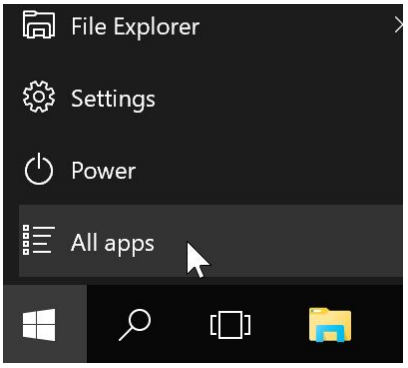
➔ Save the file to **my_work/my_first_qgis_map.pdf**
➔ Just press save if you get this window too.



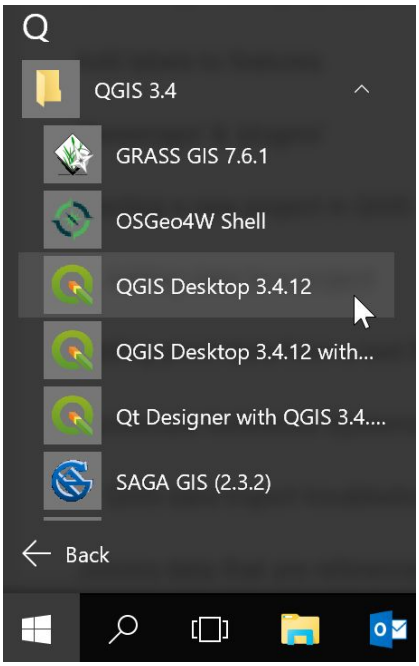
Starting a new project in QGIS

When we opened QGIS for the first time we did it by double clicking on a project file. Now we will see how we can open QGIS without any files. We need to find **QGIS Desktop** on our computers, there are a few similar options. We can find the link to QGIS Desktop under Q in the Windows start menu at the lower left.

➔ Select 'All Apps' from the Windows start menu at the lower left.



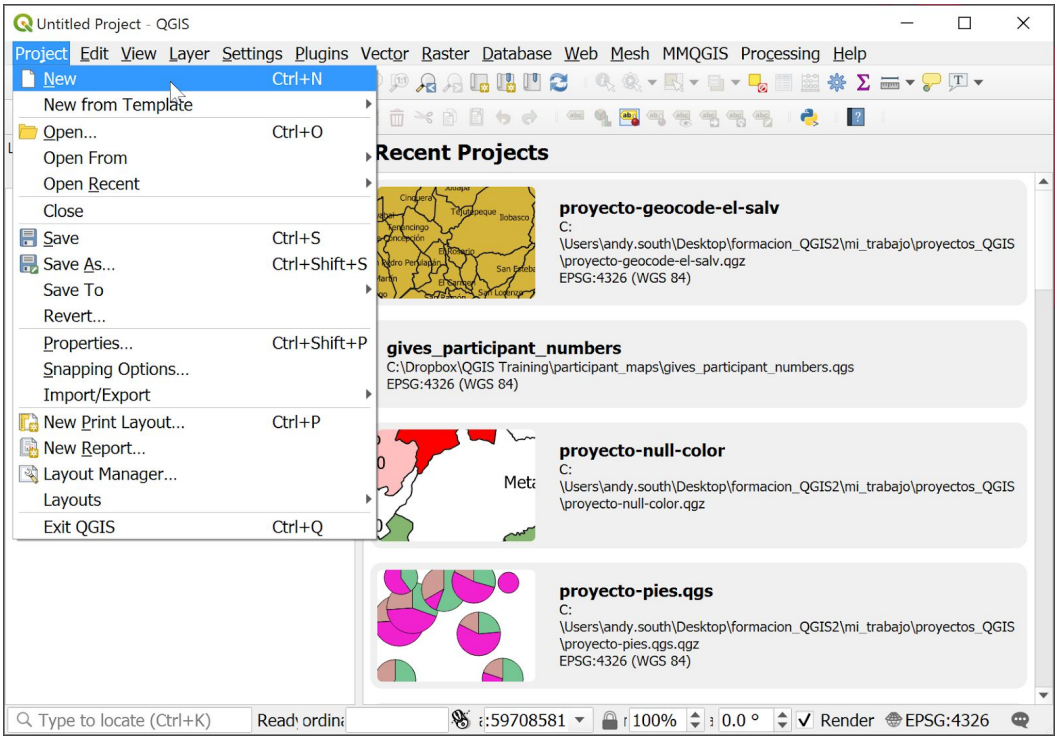
➔ Under Q find the link to QGIS Desktop (you may need to click on the folder to open it first)
➔ Left click and drag QGIS Desktop to your own desktop to create a shortcut that is easier to find



➔ Double click on either icon to start QGIS Desktop, it may take a minute or so to get started.

When QGIS opens it usually gives you the option to open recent projects. Instead we will create a new blank project.

➔ Click Project, New



In QGIS **project files only contain links to data layers and settings that determine how the data are displayed.** Project files do not contain the data and if you move or delete the datafiles the project will no longer display what it did before.

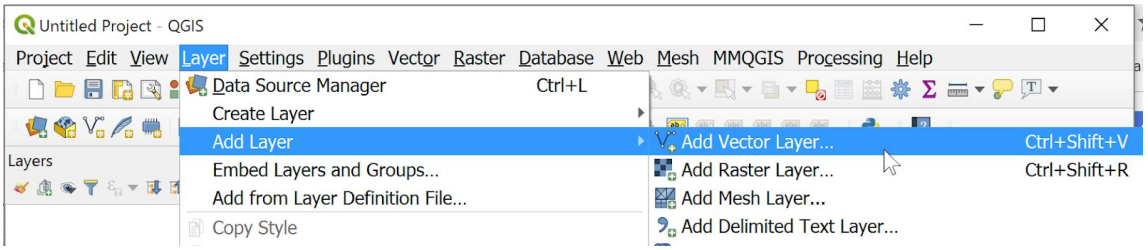
If we move a QGIS project file or a data file that can break the link between them and cause the data not to display. This is a good reason to decide on your file organisation system early in a project. Note that QGIS project files store relative addresses so if you have the following folder structure and you copy the whole study1 folder to a new location then links to the data files should still be OK.

study1/QGIS_projects/
study1/data/

Adding data to a project

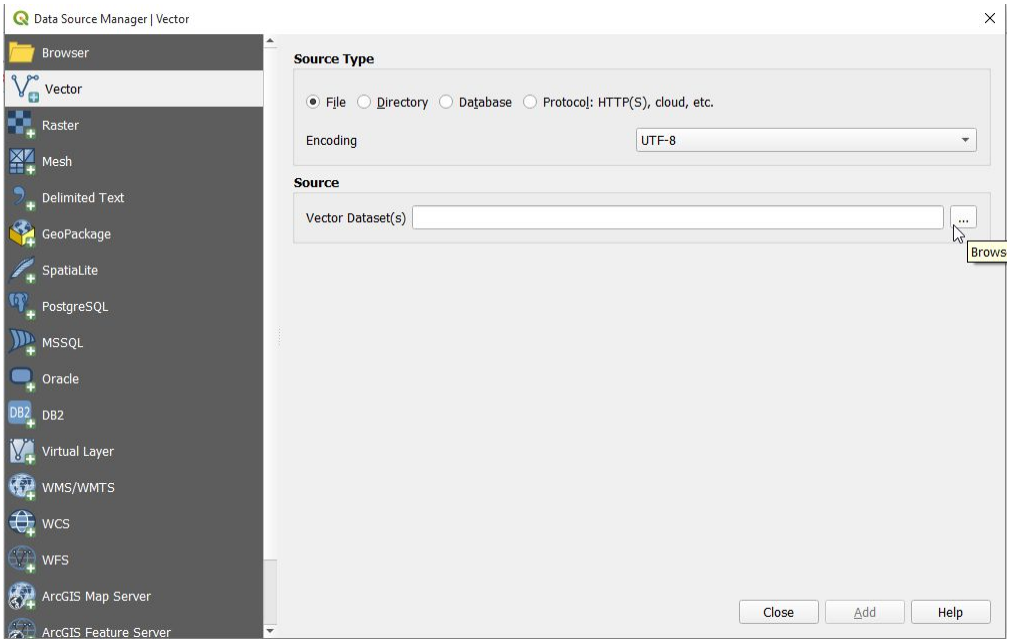
Remember that data are displayed as layers in the map. One of the most frequent actions you will need is to add data layers. Also remember that there are different types of data and that vector data are one of the most common. Thus the command to add vector layers to a project is one of the most common that you will need.

➔ From the menus at the top select Layer, Add layer, Add vector layer



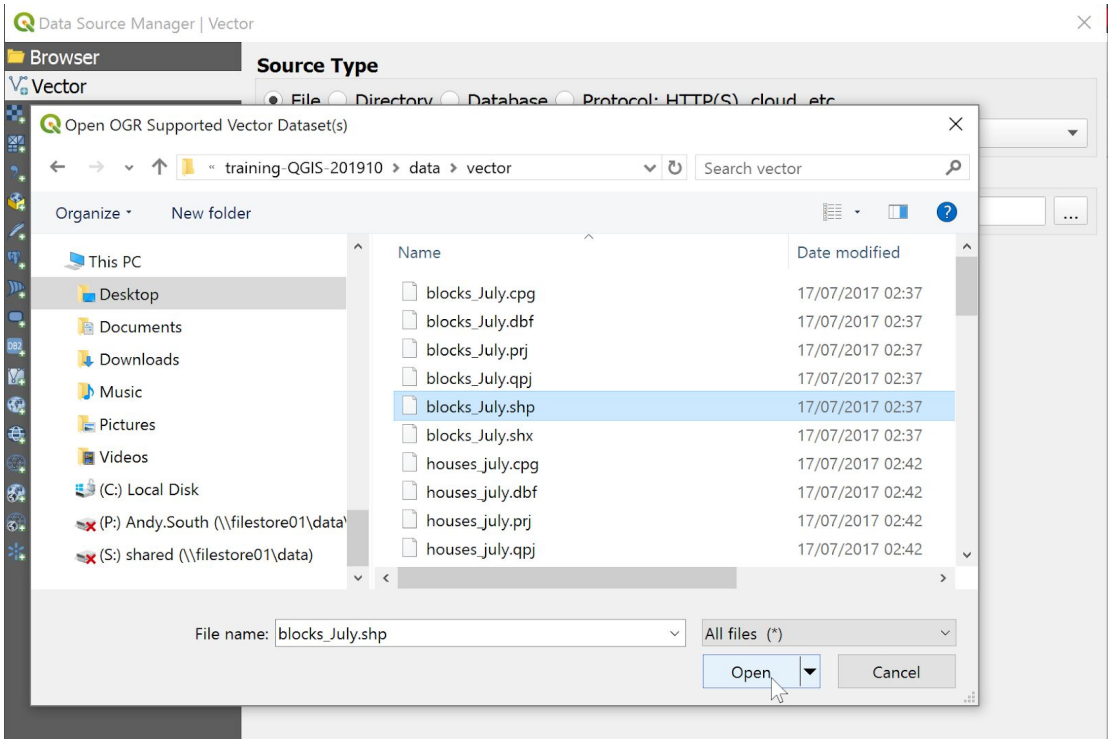
This will bring up a window called Data Source Manager.

➔ Click the three dots ... to browse your files



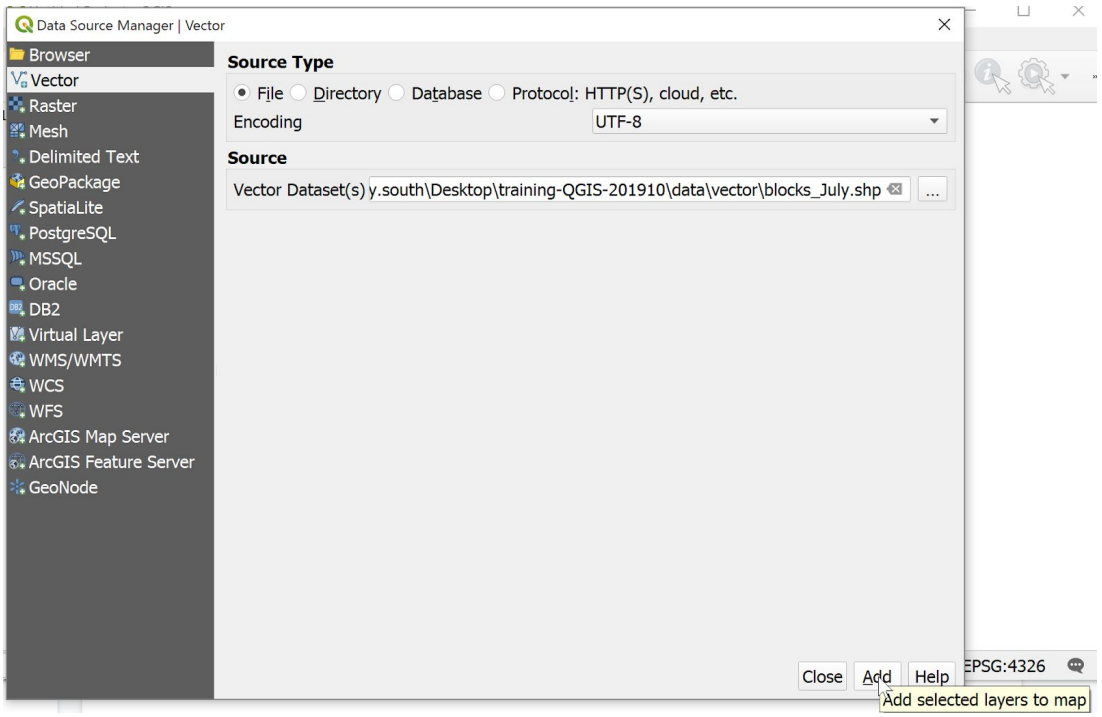
➔ find the file in the training folder : **training-QGIS-201910/data/vector/blocks_July.shp**

➔ click Open



You then still need to click 'Add' to add the layer to the map.

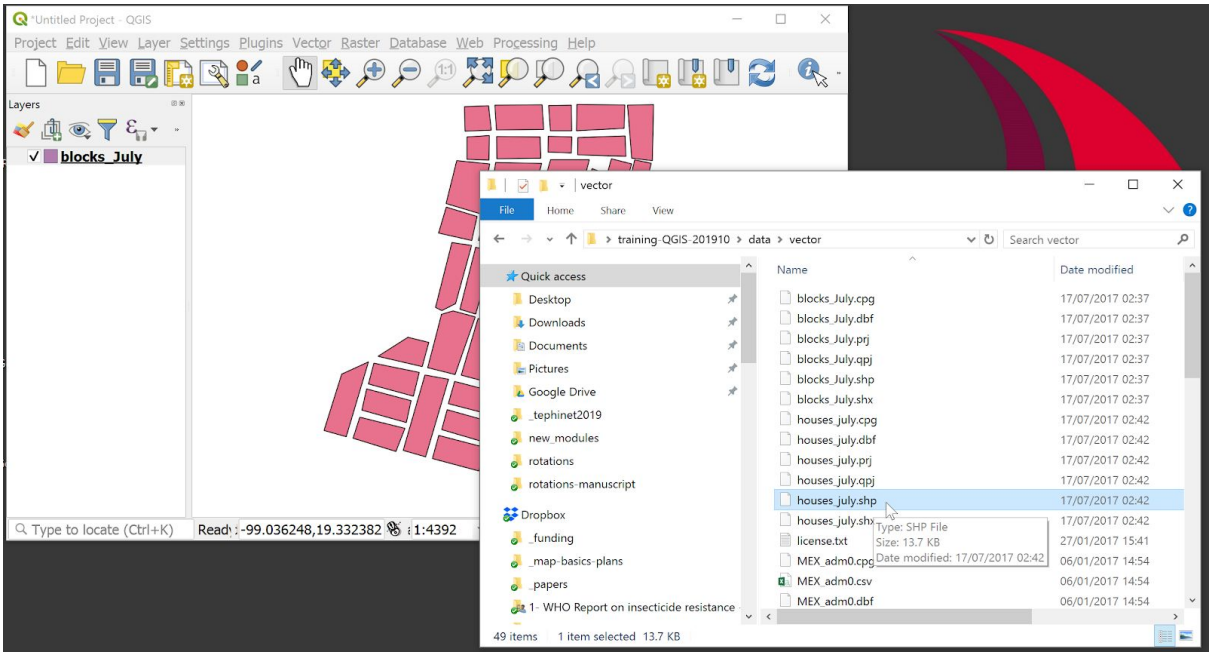
➔ click Add and then Close

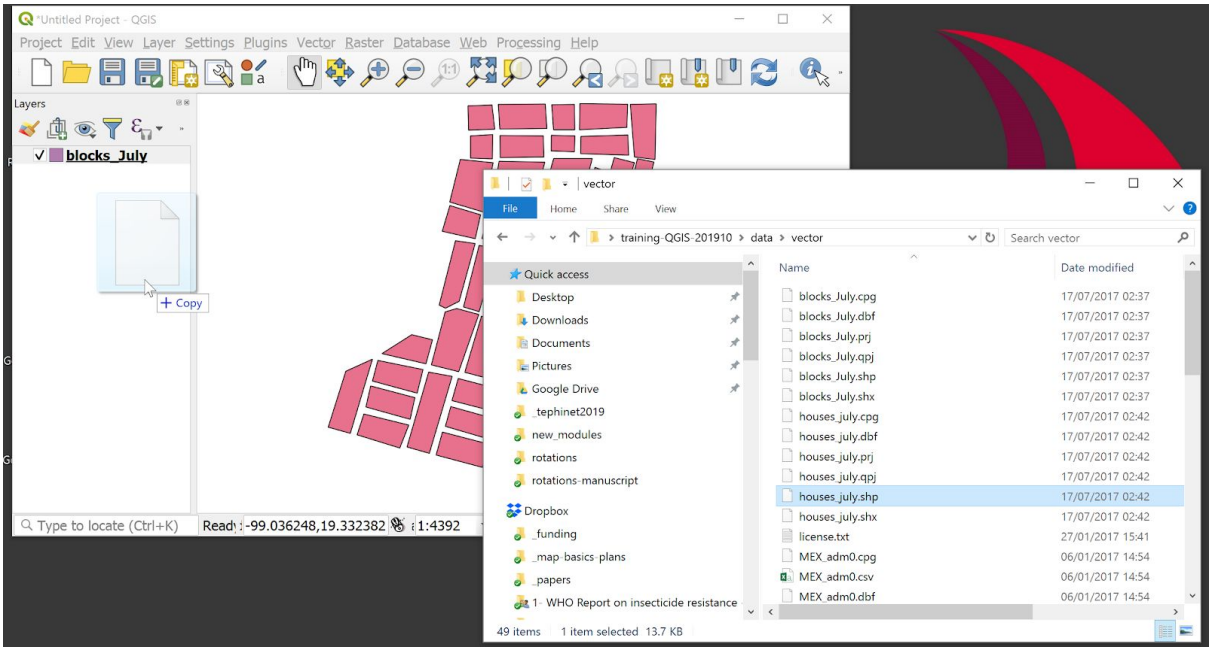


Remember that vector shapefiles are made up of a number of separate files with the same name and different extension and that we need to select the file with the extension **.shp**.

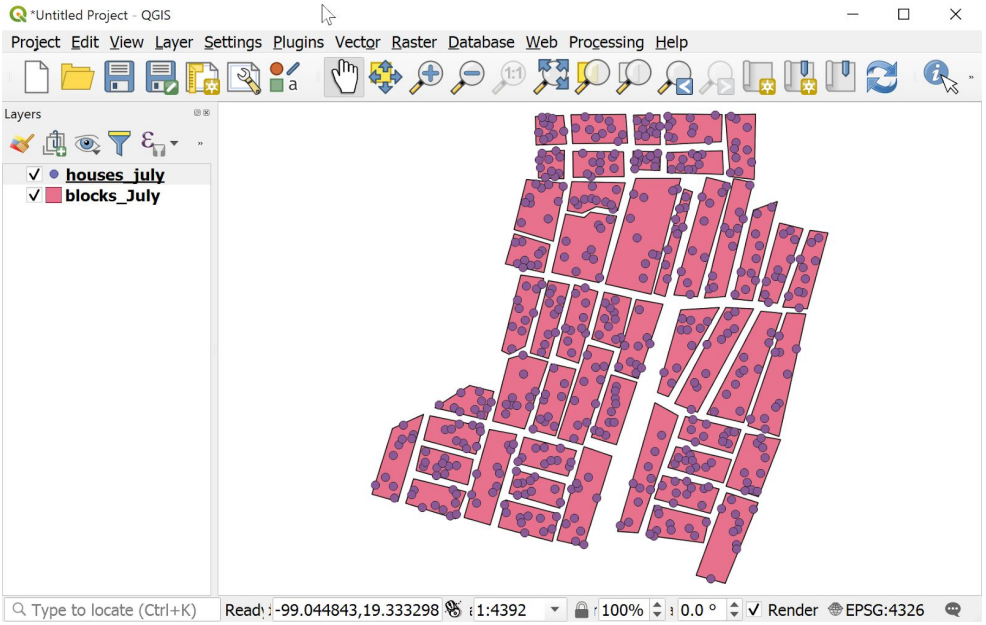
There is also an option to add layers by left clicking and dragging files into the layers panel. We will try that out, it can be quicker.

- ➔ In your windows file manager find the file : **training-QGIS-201910/data/vector/houses_July.shp**
- ➔ Left click on the .shp file, keep the mouse button pressed
- ➔ drag the file into the Layers panel (you should see a copy icon)
- ➔ release the mouse button





The points and polygons layers we recognise from before should now be visible in the map window.

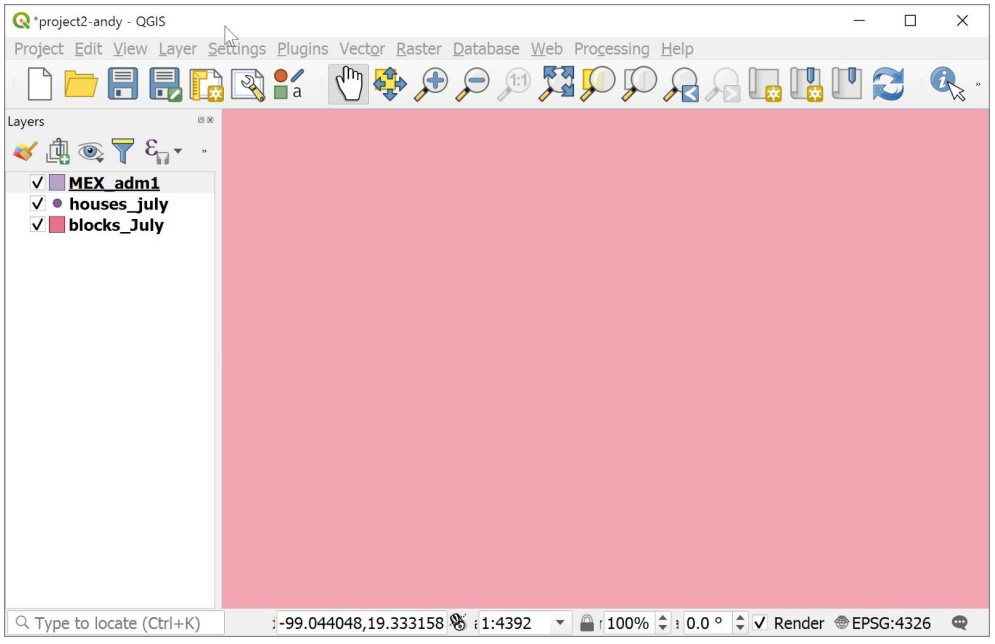


It is good to save your projects frequently so that you don't lose work if a crash occurs.

- ➔ From the menus select, Project, Save As
- ➔ Browse to **training-QGIS-201910/my_work/projects_QGIS**
- ➔ name the new project : **project_201910_mexico.qgz**
- ➔ Repeat the layer adding process for **training-QGIS-201910/data/vector/MEX_adm1.shp** (the level 1 administrative boundaries for Mexico).

The screen may now look something like this, it doesn't matter that the colours will be different.

This is not very informative ! Where have our points and blocks gone ?

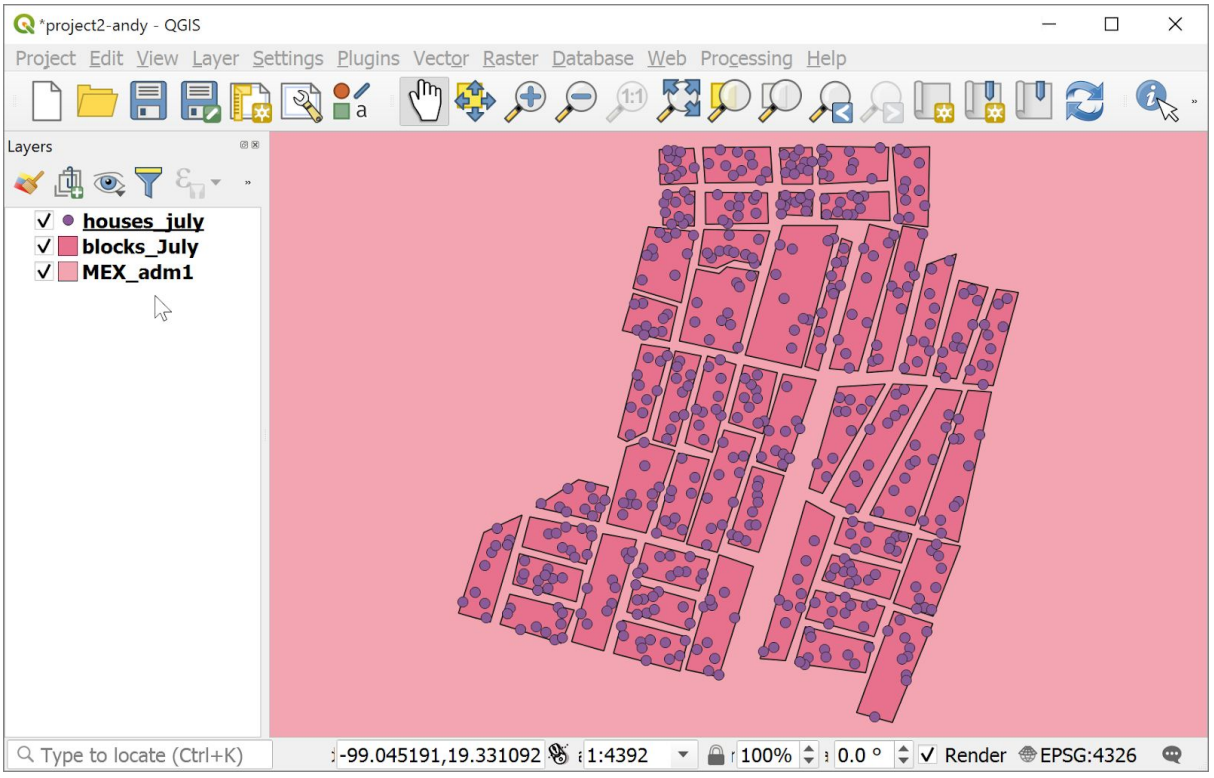


You may remember from earlier that layers can cover up others beneath them in the Layers panel. The default in the newer versions of QGIS is that new layers are added to the top of the layers panel and thus they can hide earlier layers.

Can you remember how to reorder layers and display hidden ones ?

- ➔ Left click and drag the MEX_adm1 layer so it is bottom in the Layers panel
- ➔ (sometimes it is tricky to get the layers in the order you want, keep trying)

Now the map should look something like this :



We have added points and polygons layers, these are 2 out of the 3 types of vector data. Can you remember what the third vector data type is ?

The 3 types of vector data are points, lines and polygons.

Save the project again.

Remember project files only contain links to data layers and properties settings that determine how the data are displayed. Project files do not contain the data and if you move or delete the datafiles the project will no longer display what it did before.

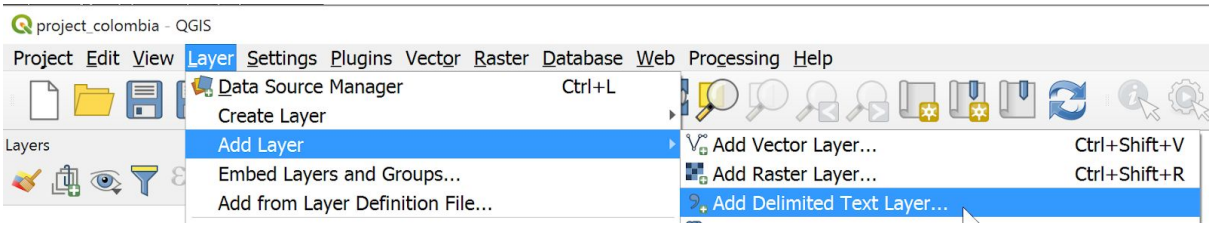
Adding point data from a text file

Earlier we added data that were already saved as geographic shapefiles. Now we are going to try adding some test points that are stored in a delimited text (csv) file that could come directly from a spreadsheet.

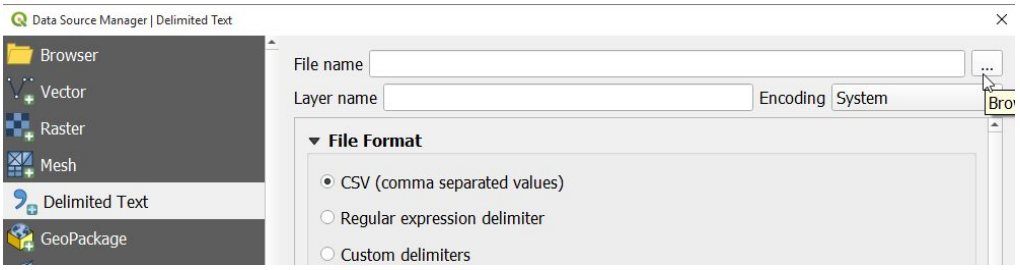
First we are going to create a new blank project to work with.

- ➔ In QGIS, Click Project, New
- ➔ Save the project as :
  - ◆ **training-QGIS-201910/my_work/projects_QGIS/project_colombia.qgz**

- ➔ click Layer, Add layer, Add delimited text layer



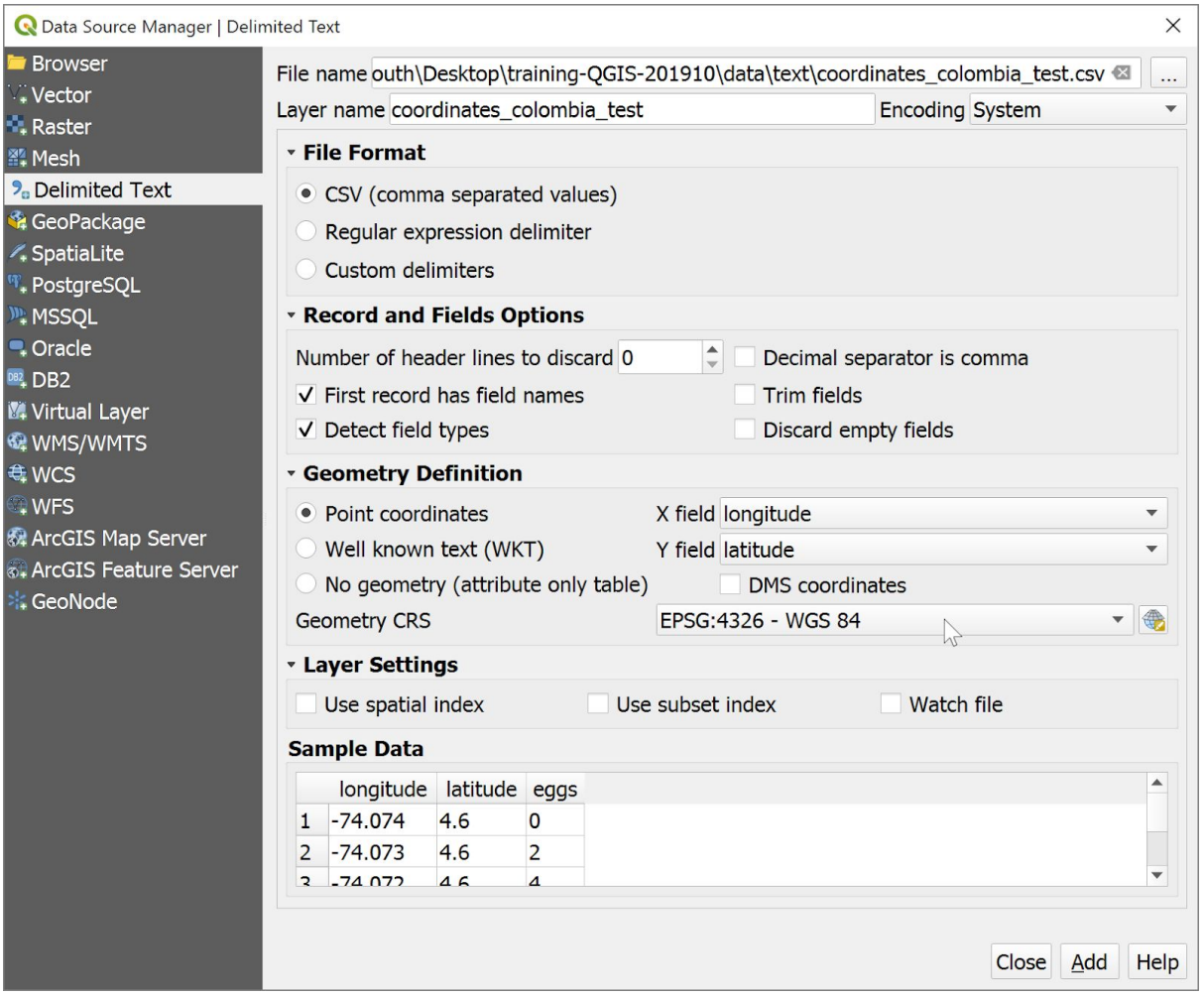
- ➔ Click the three points (...) to browse and find the file:
  - ◆ **training-QGIS-201910/data/text/coordinates_colombia_test.csv**



The window that appears should look like this. It has automatically detected that there are columns named longitude and latitude and it is using these for the X & Y coordinates.

- ➔ Check that the Sample Data section lower down has 3 columns

Sometimes problems with delimiters (commas or semicolons) can cause all the data to be put in one column and then the import will fail.



Coordinate Reference Systems (CRS)

Sometimes you may get a problem if the Coordinate Reference System (CRS) is not defined. In that case you may get a message saying : “You must select the CRS”.

This is an opportunity for us to talk very briefly about Coordinate Reference Systems which are important for GIS but we only have time to touch on briefly. For the entry-level things we are doing this will be enough for now.

The Coordinate Reference System specifies where the coordinates will appear on a map of the world. The CRS is made up of two parts :

1. datum – a coordinate grid for the globe
2. projection - something to convert globe coordinates to a flat map

QGIS is able to display layers that are in different CRS. Problems occur (i.e. points appear in the wrong place) when a layer has been assigned the wrong CRS. This can happen and still happens to us.

By default when you add a delimited text layer QGIS doesn't ask you for the CRS it assumes that it will be unprojected latitude, longitude using a WGS 84 datum (this is the most common CRS used, e.g. by GPS).

We also see the letters EPSG in QGIS. The EPSG is just another code that is a different way of naming the CRS. It can be useful to remember that most often you need **EPSG 4326 : WGS84**. Some common CRS are shown in the table below :

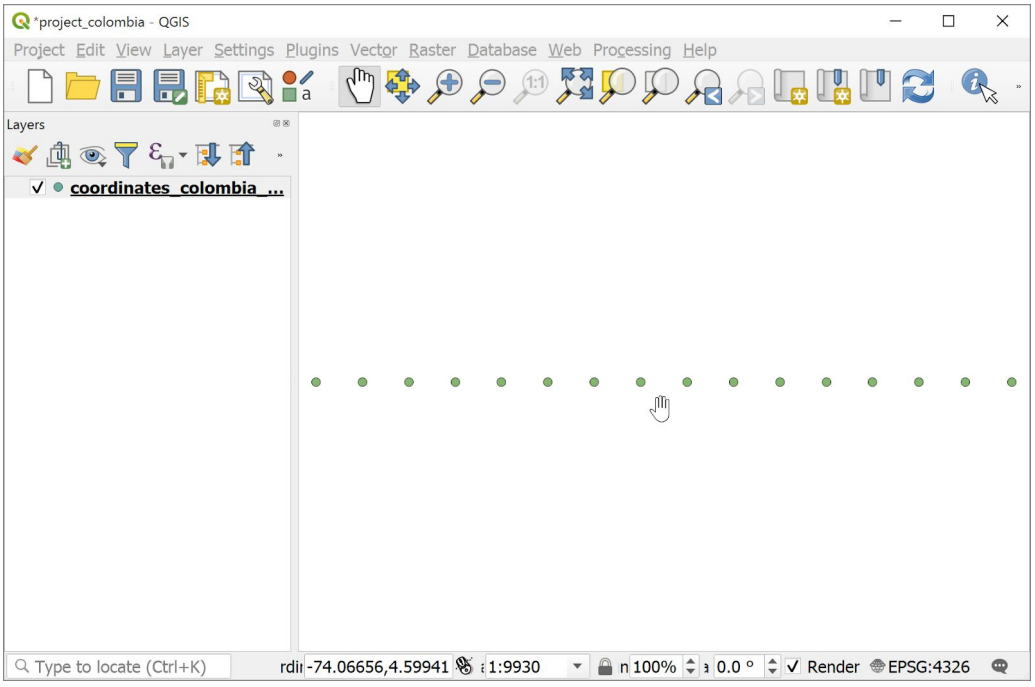
datum	projection	EPSG code	notes
WGS 84	none	EPSG 4326	Probably the most common CRS you will see. Used for GPS longitude latitude.
WGS 84	Pseudo Mercator	EPSG 3857	Used for much web-mapping. e.g. if you add an OpenStreetMap basemap to a blank QGIS project then it may adopt this CRS.
WGS 84	UTM 18N	EPSG 32618	The UTM zone for Colombia. UTM projections are useful when you want to measure distances in metres. The world is divided into 60 zones each denoted by a number, and into the north or south hemisphere indicated by an N or S

UTM zone maps are available online if you want to identify which zone your geographical area of interest is in. There is also this handy site you can use:
https://mangomap.com/robertyoung/maps/69585/what-utm-zone-am-i-in-#

Back to our example, we are happy to accept the default CRS of WGS 84 (EPSG 4326) for our layer.

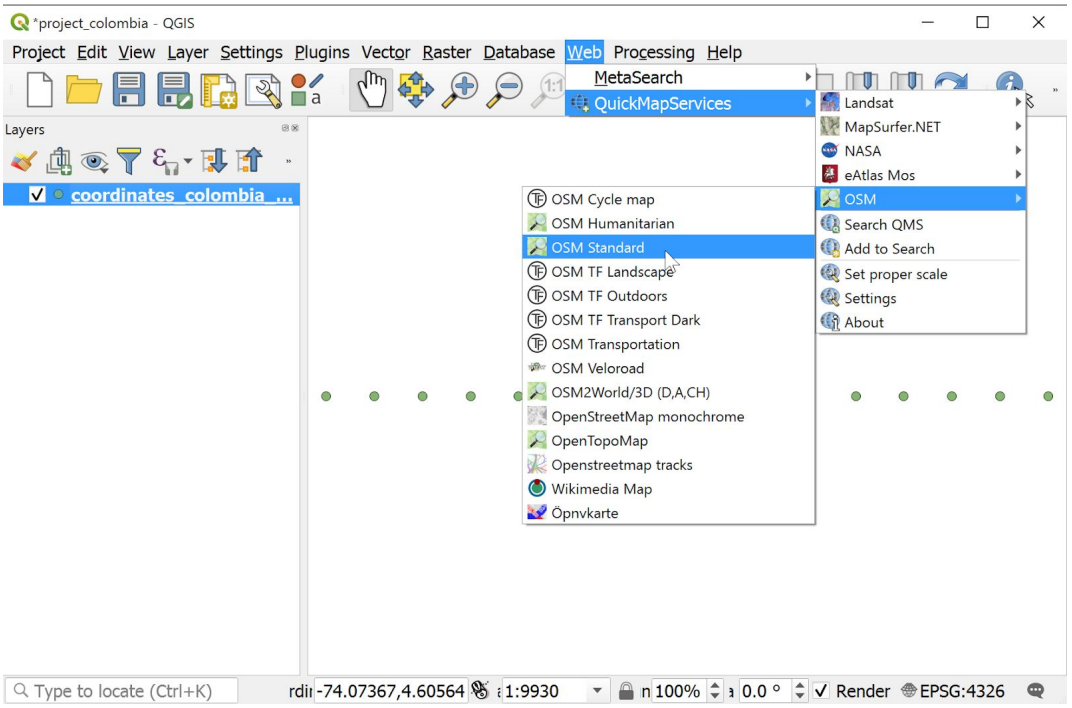
➔ Click Add and then Close

You should see that a horizontal line of test points have appeared in the map.

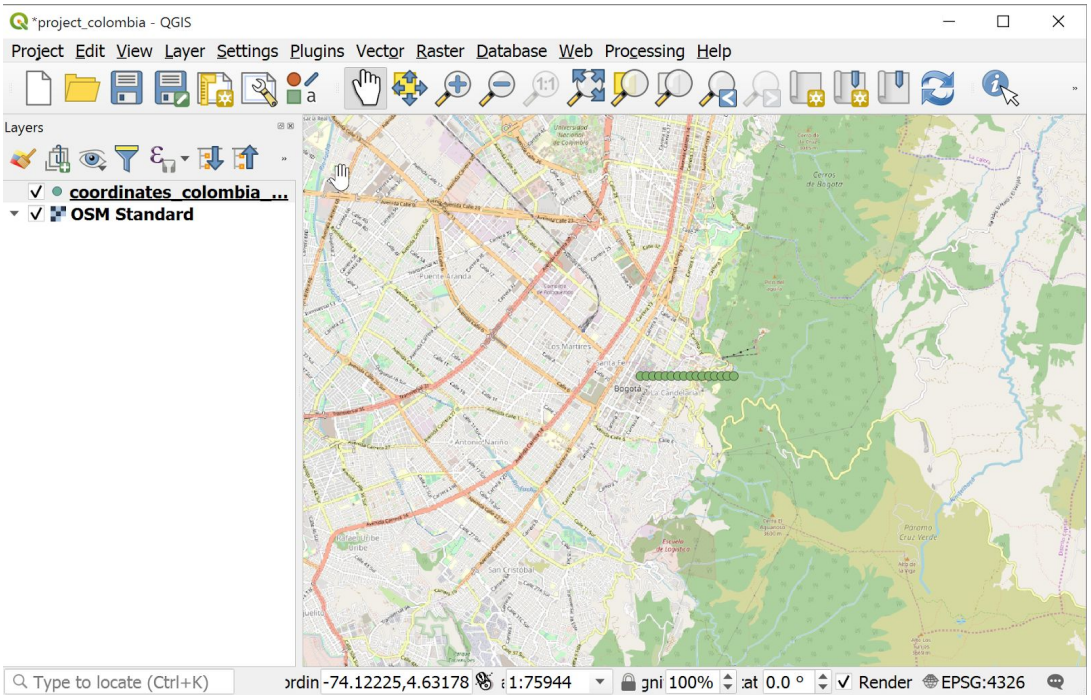


In order to check that these points have imported successfully in the area in Colombia that we expect we can add a basemap from OpenStreetMap.

➔ select **Web, QuickMapServices, OSM, OSM Standard** from the main QGIS menus



By zooming out you should see that the points appear in Bogota, Colombia.



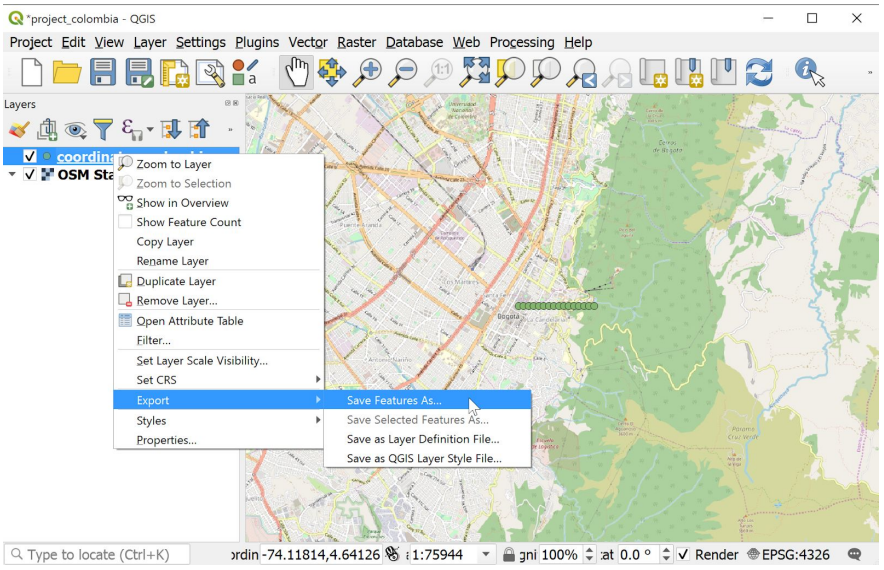
Exporting text files to geographic format files

To be able to use these data in QGIS in future without having to import them we can export them in a geographic file format.

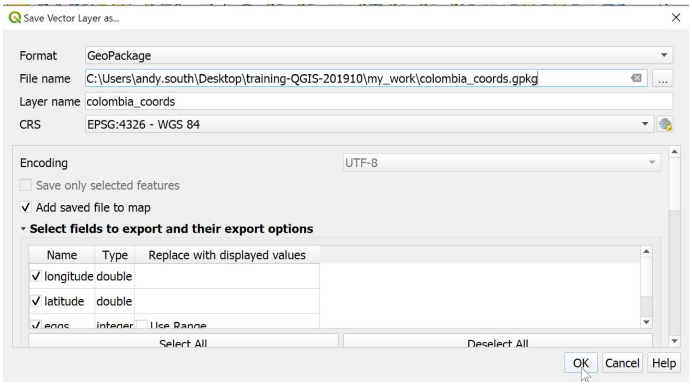
The geographic files that we have used before in this course have been in **ESRI shapefile (shp)** format. This is probably the commonest vector file format that you will see. However shapefiles are an ageing file format and a few things about them are not ideal. Shapefiles consist of more than one file with the same name and different extension (e.g. *.shp, *.shx, *.dbf) and if you lose one of the files they may not work. Also the column names in shapefiles can only contain a maximum of 10 characters. A more modern geographic file format is **geopackage**. A geopackage format file has the advantage that it consists of just a single file and can have longer, more informative, column names.

Just as a demonstration we will show how we can use QGIS to export our text coordinates as a geopackage.

- ➔ right click on the coordinates_colombia layer within the layers panel
- ➔ select Export, Save Features As ...



- ➔ select Format : Geopackage
- ➔ File name : **my_work\colombia_coords.gpkg**
- ➔ OK

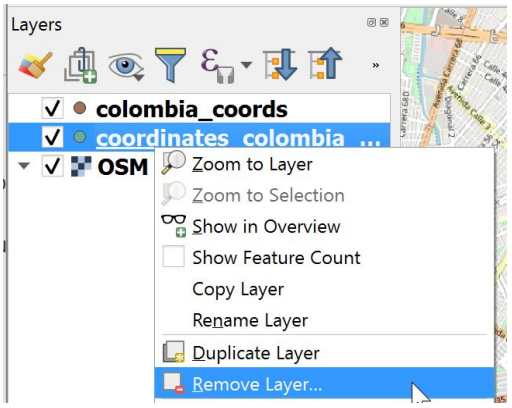


You should see that the new layer gets added to the list in the Layers Panel.

We can now remove the text file.

- ➔ right click on the coordinates_colombia layer within the layers panel
- ➔ select Remove Layer

This removes the layer from the project, but does not delete the data file.



QGIS data import troubleshooting advice

There are various solvable issues that we have seen cause difficulties with text files when we have been giving previous QGIS courses. For reference, we summarise them briefly and their solutions in the following table. Also there are some troubleshooting exercises at the end of the manual that you can try out if you have time now or after the course.

Symptom	Checking	Cause	Solution
Layer points do not appear at all. In the 'add layer' data preview window the data do not appear in columns.	Open the coordinates text file in MS Notepad or Word, are there lots of ';' ?	The file may use ';' as the separator instead of ','. This is common in Spanish where the ',' is used to indicate decimals instead.	In the add layer window choose 'File format, Custom delimiters' and select 'Semicolon', also tick the box : 'Decimal separator is comma'.
	The coordinates in the file look something like this : 74°04′ 00″ W 4°36′ 07″ N	The longitude, latitude coordinates are saved in the format degrees, minutes, seconds instead of decimal degrees.	When adding the layer to QGIS, you need to tick the box : 'DMS coordinates'

Symptom	Checking	Cause	Solution
Layer points do appear but are in the wrong place.	The x and y coordinates are reversed.	The columns may have been named incorrectly, or specified in the wrong order in the 'add layer' window.	Rename the columns in the text file or change the order in the 'add layer' window.
	The x or y coordinates are positive where the map coordinates are negative.	A '-' symbol has been lost from the coordinates.	Add a '-' back in to values in the column using Excel.

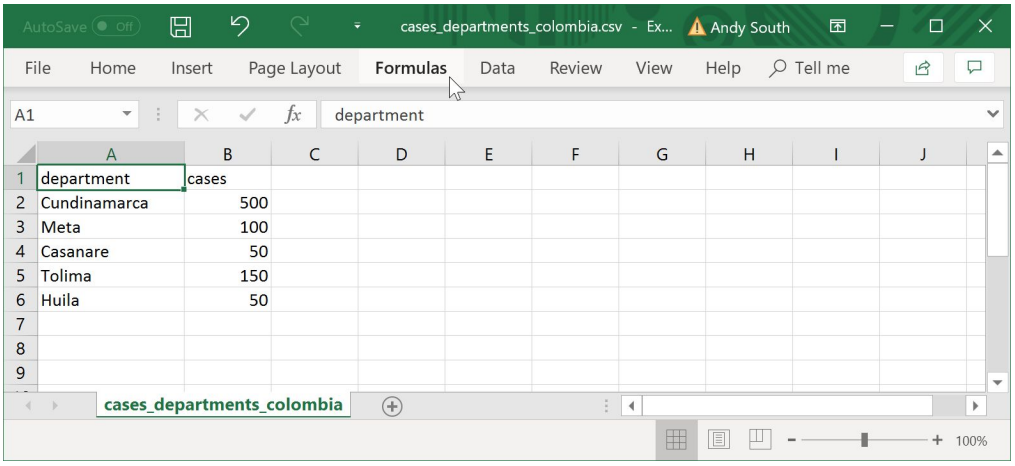
	Your points appear in the north rather than the south or east rather than the west, similar to the + / - situation.	The S or W is missing from the coordinates, causing a similar effect to the missing "-" symbol.	Add the required letter back into the coordinates in Excel.
	The coordinate values for the new layer are very different from the longitude latitude values in the area you expect and may be outside of the usual longitude latitude range of +-90, +-180.	A different CRS has been used e.g. it could be UTM. Either ask the person you got the file from or the website www.projfinder.org can be useful.	Specify the different CRS in QGIS in the properties of the added layer. (If you are unsure you can experiment with different potential ones, e.g. try the UTM zone for your region).

Joining data that are referenced by names rather than coordinates

We can use QGIS to add data that are defined by the names of areas rather than coordinates by creating a 'join' between those data and an existing polygons layer.

Here we will imagine that a colleague has sent you a file with the number of disease cases in Colombian departments. That file is saved as '**cases_departments_colombia.csv**'.

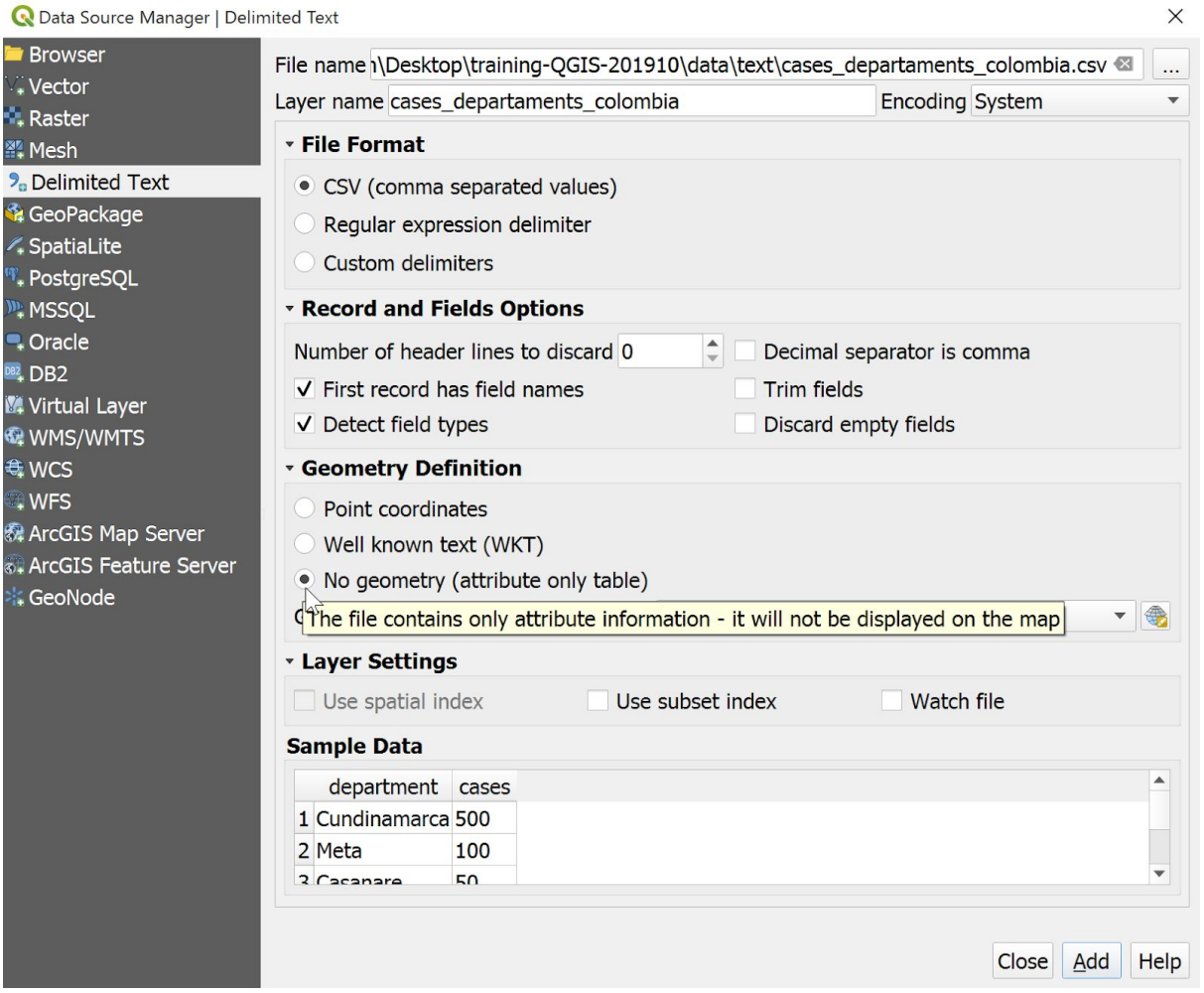
It is a small simple file for demonstration.



	A	B
1	department	cases
2	Cundinamarca	500
3	Meta	100
4	Casanare	50
5	Tolima	150
6	Huila	50

What we need to do is first add it as a layer (with no geometry) and then join that layer to COL_adm1 which has the department polygon boundaries. We will run through the steps here.

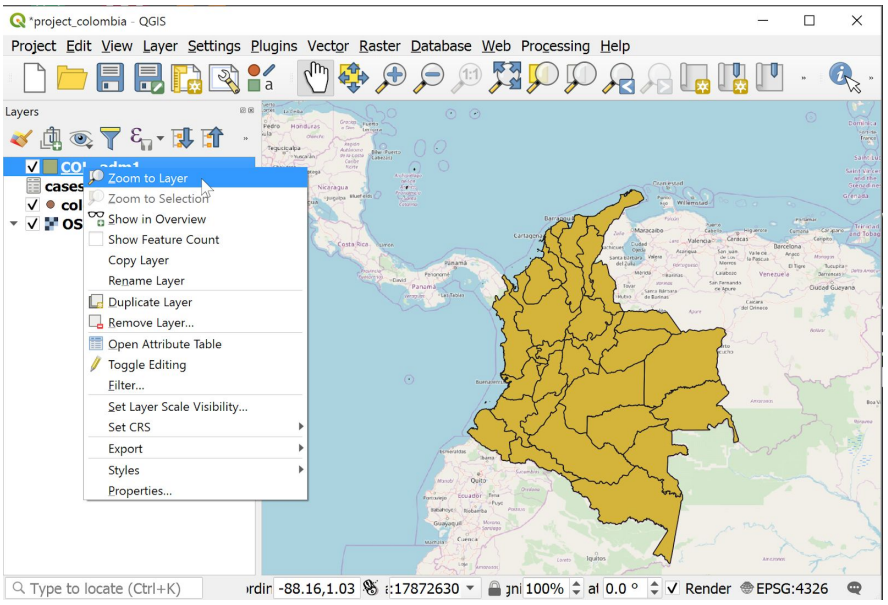
- ➔ Layer, add layer, add delimited text layer and press browse to find the file : **data\text\cases_departments_colombia.csv**
- ➔ Under Geometry definition : Tick '**No geometry**' because this file has no coordinates
- ➔ Click Add and then Close



This should add 'cases_departments_colombia' to the layer list with a little table symbol indicating that it is only a data table and does not have any spatial information.

Next we need to add the departments polygons layer that we will join the text data to.

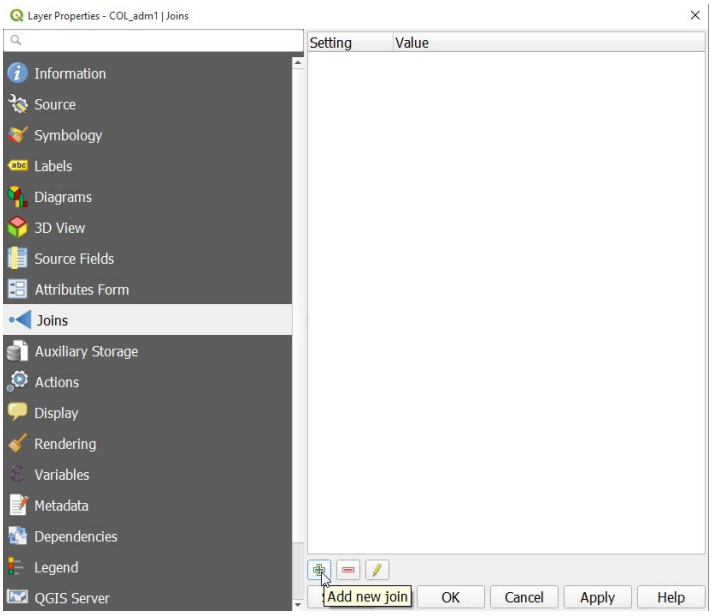
- ➔ drag **COL_adm1.shp** into the Layers panel or click Layer, Add layer, Add vector layer
- ➔ Right click on COL_adm1, Zoom to Layer



Now we can join the text data we added to the polygons in COL_adm1. It is important that we do the join from the layer that has the spatial information, doing it from the new table will not work.

➔ Right click on COL_adm1 in the layers panel and select 'Properties'

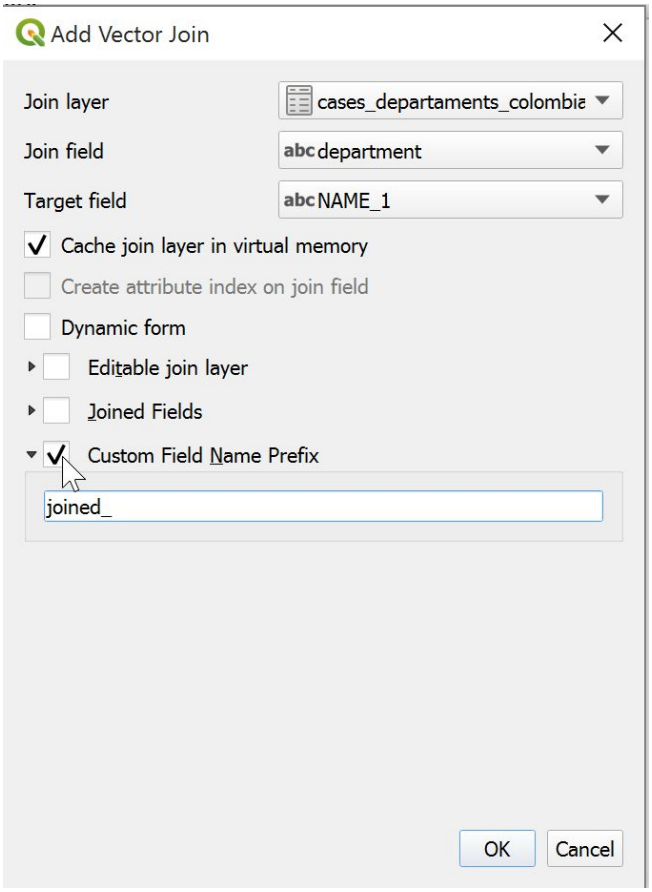
➔ Select 'Joins' from the options on the left, then click the green plus symbol lower left to ‘Add new join’.



➔ Set the following options :

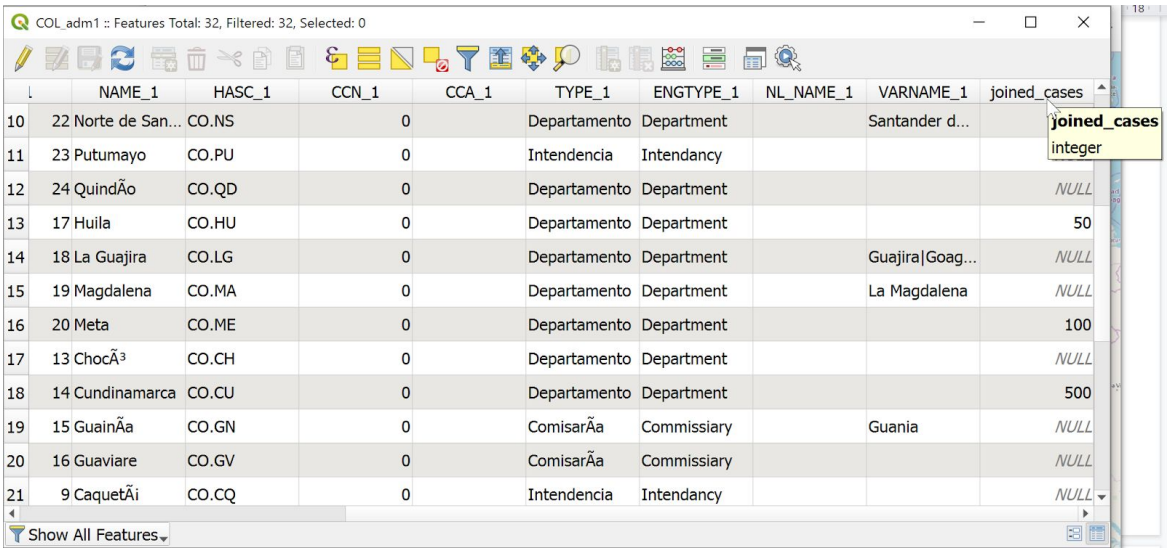
Join layer : cases_departments_colombia
Join field : department
Target field : NAME_1

➔ Tick ‘Custom field name prefix’ (this is text that will be added to the start of the joined columns, by default it uses the whole filename which creates very long column names)
➔ Type joined_ in the text box
➔ click OK, and then OK again to close the properties window



➔ right click on COL_adm1, Open Attributes Table

You should see that the new cases column has been joined on and named ‘joined_cases’. There were only four rows in the csv file, departments that were not in that file will just get a NULL entry in the new column in the new file.



		NAME_1	HASC_1	CCN_1	CCA_1	TYPE_1	ENGTYPE_1	NL_NAME_1	VARNAME_1	joined_cases
10	22	Norte de San...	CO.NS	0		Departamento	Department		Santander d...	
11	23	Putumayo	CO.PU	0		Intendencia	Intendancy			
12	24	QuindÃo	CO.QD	0		Departamento	Department			NULL
13	17	Huila	CO.HU	0		Departamento	Department			50
14	18	La Guajira	CO.LG	0		Departamento	Department		Guajira|Goag...	NULL
15	19	Magdalena	CO.MA	0		Departamento	Department		La Magdalena	NULL
16	20	Meta	CO.ME	0		Departamento	Department			100
17	13	ChocÃ³	CO.CH	0		Departamento	Department			NULL
18	14	Cundinamarca	CO.CU	0		Departamento	Department			500
19	15	GuainÃa	CO.GN	0		ComisarÃa	Commissiary		Guania	NULL
20	16	Guaviare	CO.GV	0		ComisarÃa	Commissiary			NULL
21	9	CaquetÃi	CO.CQ	0		Intendencia	Intendancy			NULL

As a quick aside while we are here you may notice that some of the department names in the NAME_1 column have strange looking symbols in them, e.g. capital letters with accents. This can happen when accents are not interpreted correctly due to text format differences. Accents are a bit like mosquitoes, wherever they occur they often cause problems !

Googling the problem offered this solution. If accents don't look right on your maps try this :

- ➔ Right click, Properties, Source
- ➔ Change Data Source encoding to either System or UTF-8

Look at the attribute table again to see if the names in the NAME_1 column now look better :

NAME_1
Meta
Chocó
Cundinamarca
Guainía
Guaviare
Caquetá
Casanare
Cauca
Cesar
Bolívar
Boyacá
Córdoba

In our experience sometimes accents work with ‘UTF-8’ and sometimes they work with ‘System’.

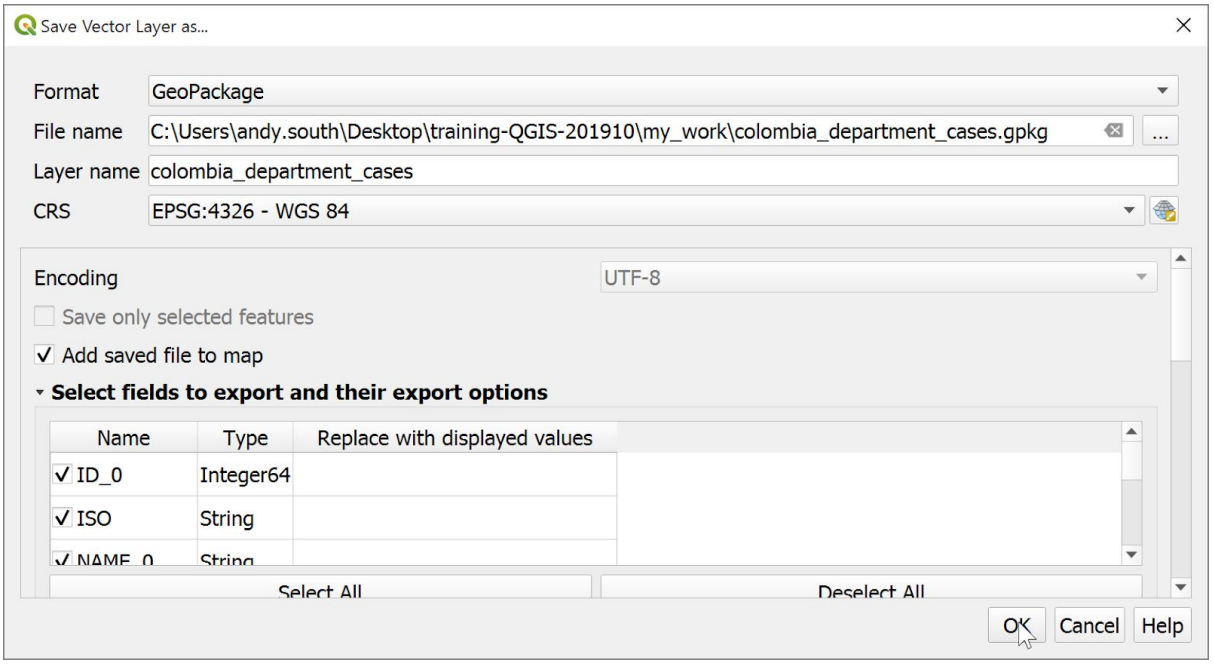
OUR NEW SIMPLE RULE FOR ACCENTS IS : IF THEY AREN’T WORKING CHANGE THE ENCODING TO EITHER UTF-8 or SYSTEM (whichever one it isn’t currently !).

If you find a case when this rule doesn’t work, please let us know !

Now back to the join :

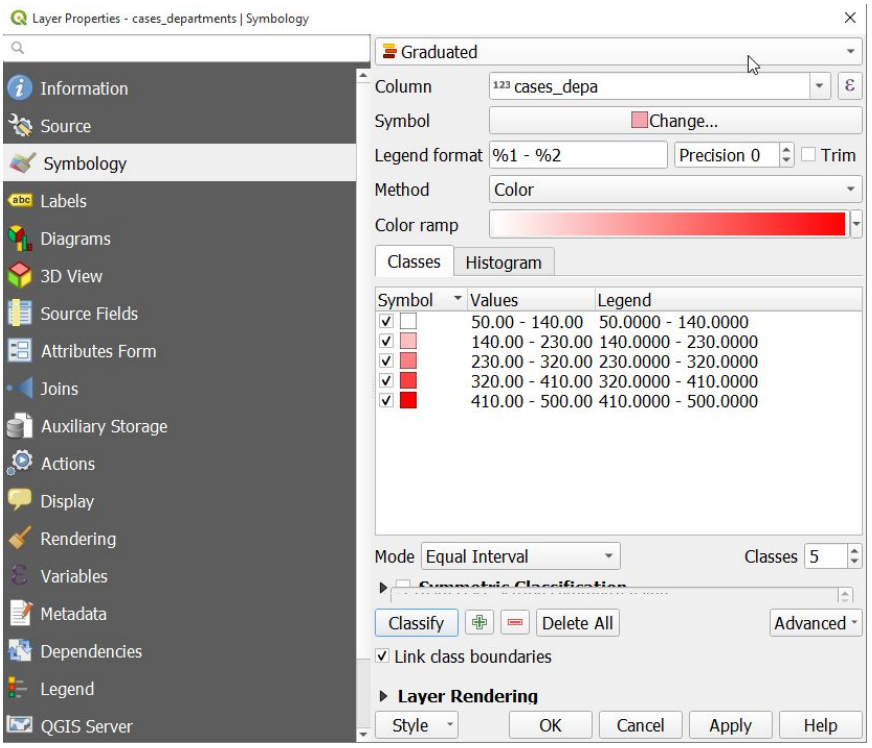
This join is only temporary. We should save as a new file to make the join permanent using Export, Save features as like we did before.

- ➔ Right click on COL_adm1, Export, Save features as
- ➔ in the next window select 'geopackage'.
- ➔ click browse to choose a new filename and location.

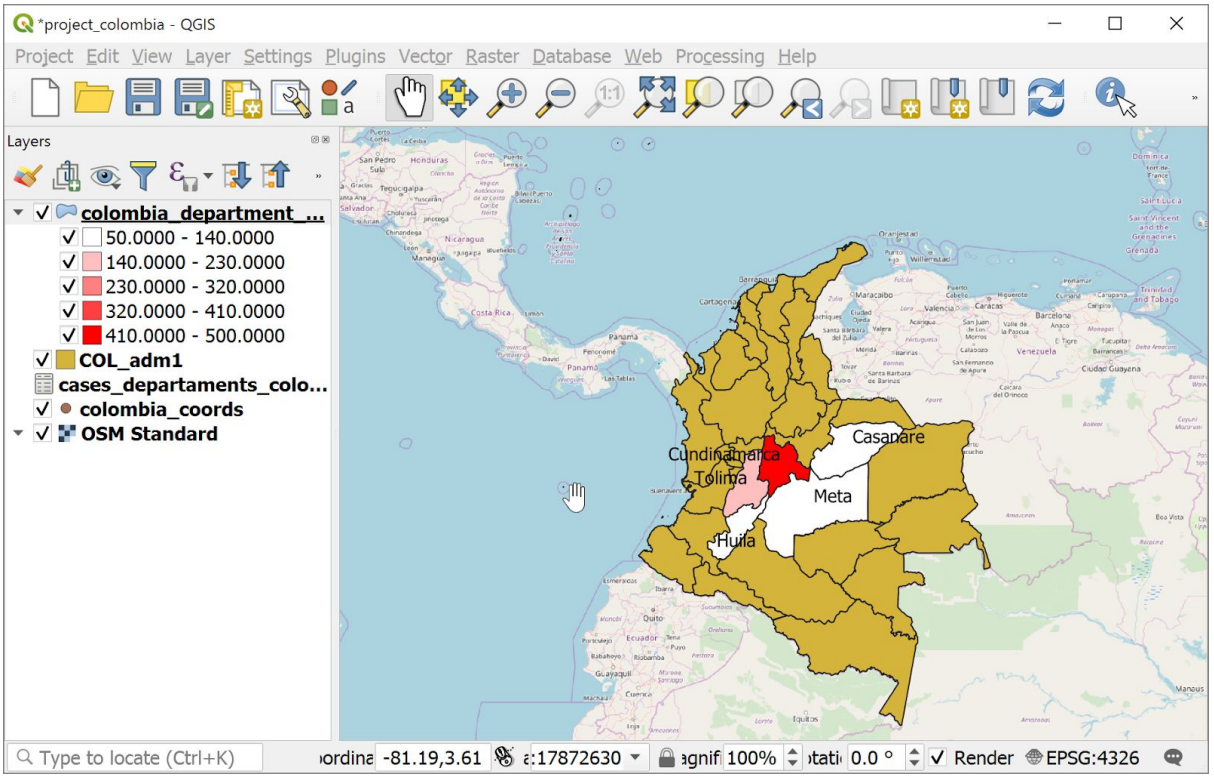


We can check that the join has worked by trying to use the new cases column to set the colour of the departamentos polygons.

- ➔ Right click on your new layer,
- ➔ Select properties, Symbology.
- ➔ Choose 'graduated' and set the column to cases.
- ➔ press Classify



➔ While in properties set the labels to Name_1 to show the names of the departments to get the map below.



Remember that when joining two columns it is very important that the spelling, capital letters and accents of the names are identical. If they are not then the rows will not join.

Appendix 1 : Finding free spatial data online

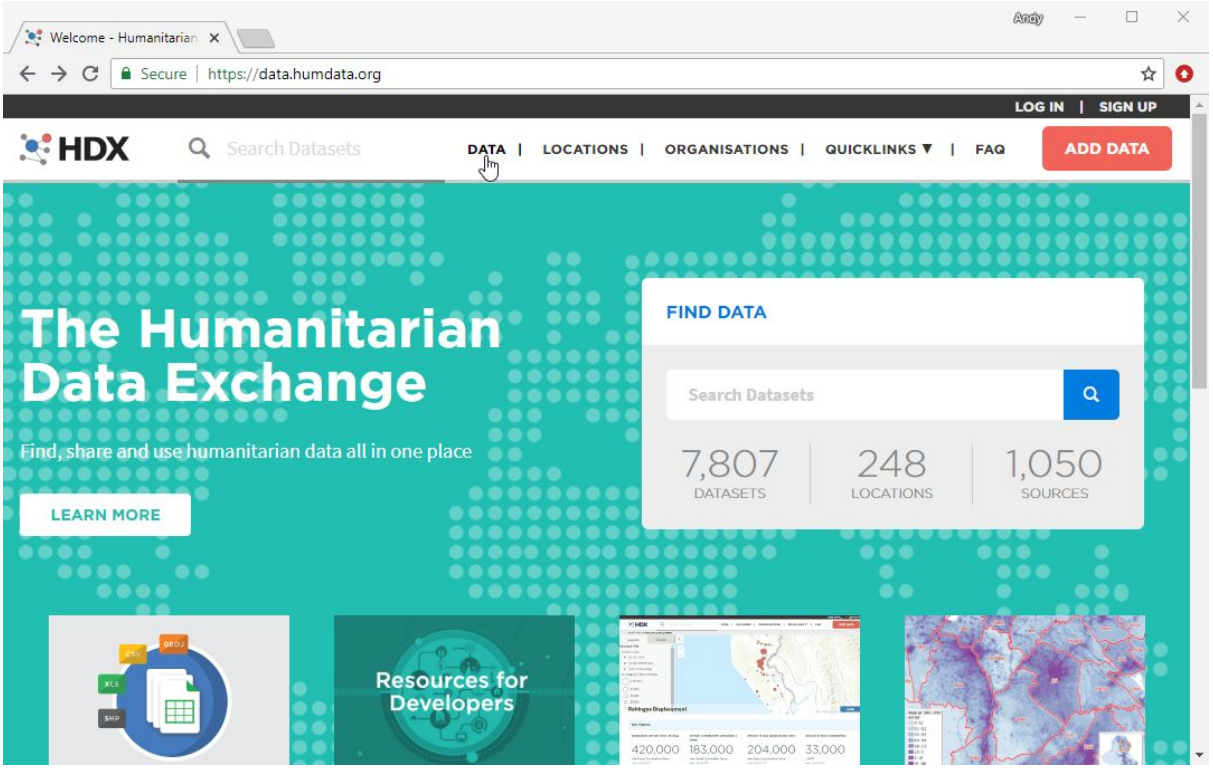
There are lots of freely available spatially-referenced data online. Often the difficulty is in choosing which data are best and of most use to you among a confusing abundance of options that change over time. Data available online are of variable quality and a source that is good one year may become out of date a few years later. Here we will introduce you to some data that may be useful for you as a start. You may be aware of other resources that are better for your purposes. Also we are concentrating on freely available open data, you may have access to other national data sources through your local contacts.

QGIS is able to deal with data in all common spatial data formats, it is unlikely that you will find a spatial data source that QGIS cannot accept.

Because the most difficult thing can be choosing between data sources af variable usefulness we are going to start by pointing you to one website that collates many useful data sources in one place and keeps them updated.

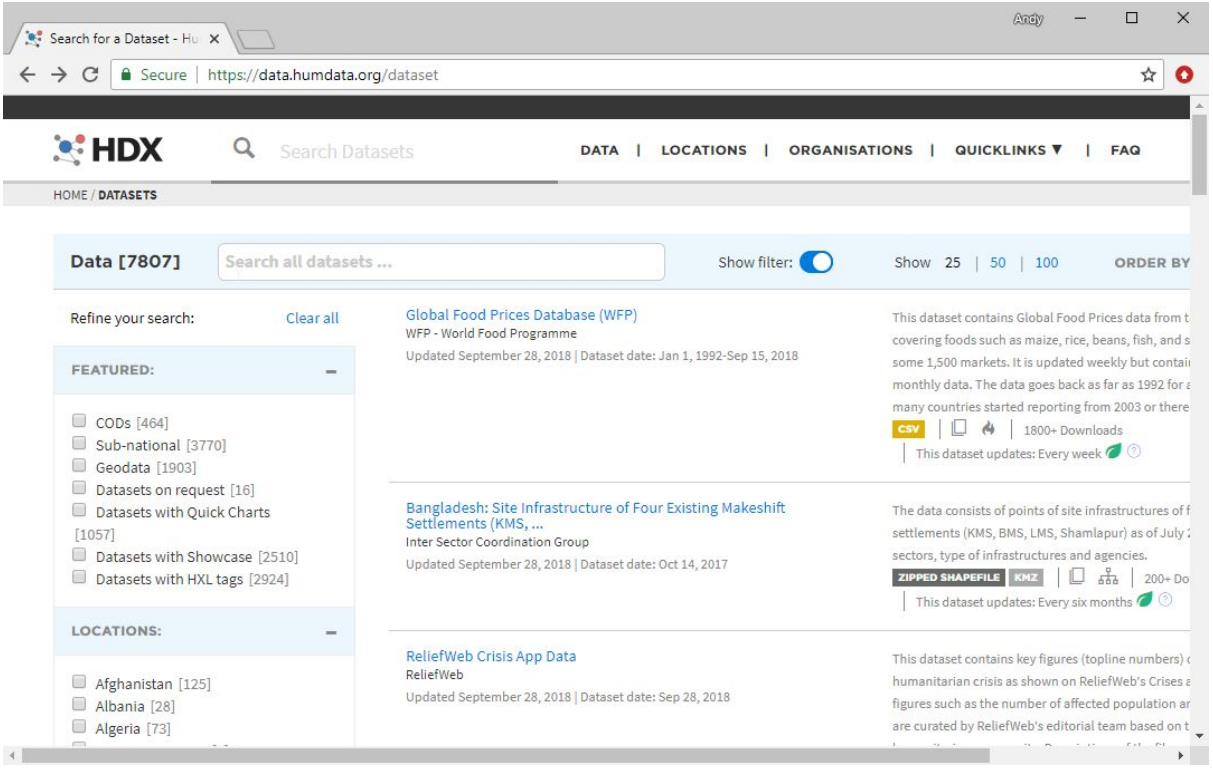
Humanitarian Data Exchange

Humanitarian Data Exchange https://data.humdata.org/ is a UN run website that collates data for humanitarian action all in one place. These data can also be of use to us. Data can be downloaded freely and easily and you don't have to register. (If you do register there are also options for submitting your own data.)



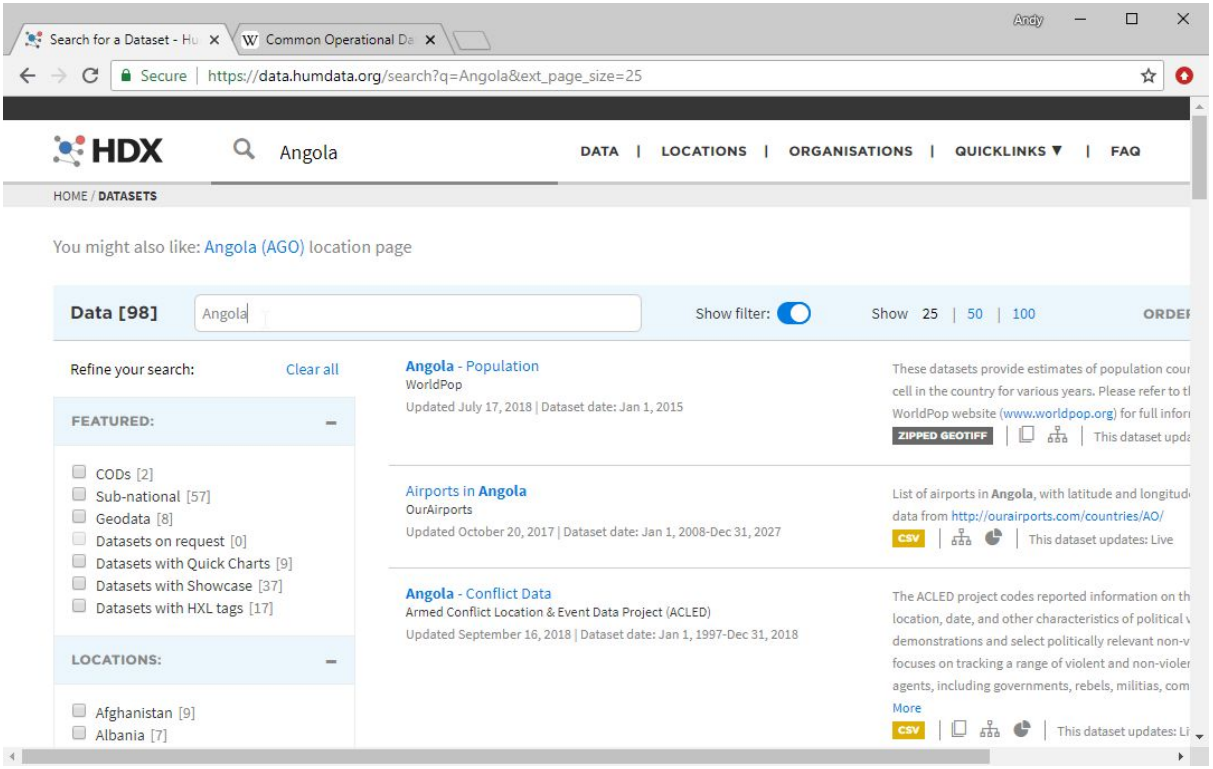
➔ go to https://data.humdata.org/

➔ click on the DATA tab at the top to go to the data explorer



➔ type your country name into the search box and press return

We are using Angola as an example here but we suggest you use your own country.



This will is likely to return a large number of datasets, not all are spatial and not all will be useful to you.

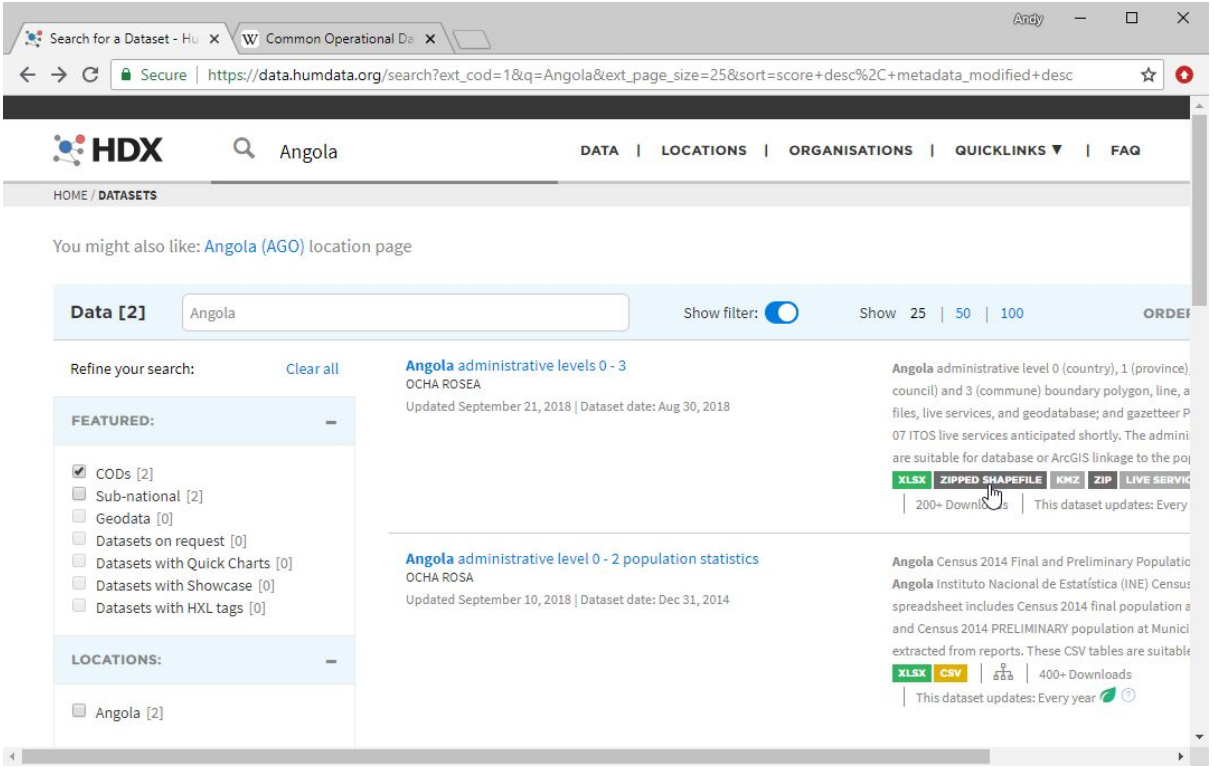
To get the most useful datasets we can look for the ones called **Common Operational Datasets** or **CODs**. These are defined as authoritative reference datasets needed to support operations and decision-making in a humanitarian response. CODs are 'best available' datasets that ensure consistency and simplify the discovery and exchange of key data.

The main CODs are administrative boundaries (COD-AB) and population statistics (COD-PS).

There are a number of advantages of using HDX as a source of administration boundaries over other sources such as GADM or Natural Earth. One is that they are regularly updated and may contain edits from more detailed or more up-to-date national data. The second advantage is that they may have been edited to make sure that they can **join** with population or other datasets. We will cover data joining later.

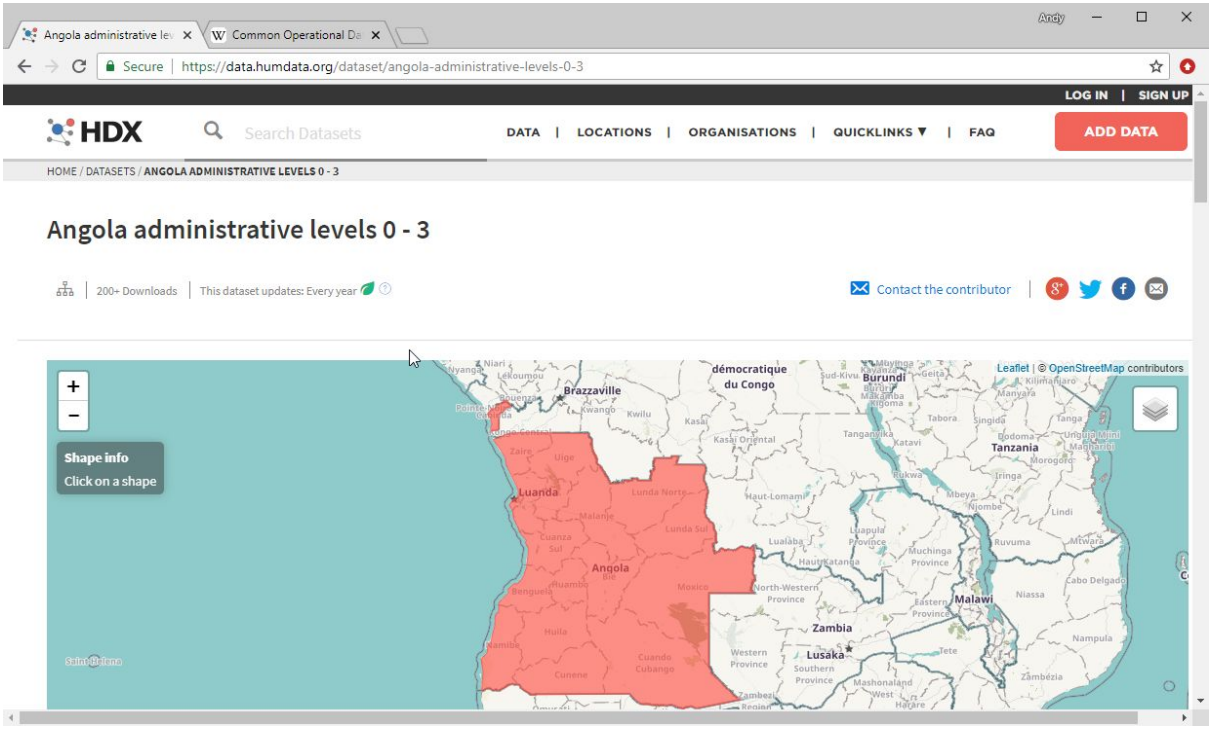
➔ Tick the CODs tickbox on the HDX page to show just the Common Operational Datasets

This should display the administrative boundaries and population datasets.



➔ Select the administrative boundaries polygons (be careful not to select the population statistics which may also refer to administrative levels in the text).

This will take you to a preview map :



You will see that this refers to **administrative levels** 0-3. Administrative (**admin**) levels are an international standard for classifying boundaries that copes with fact that the sub-regions of countries are named differently, e.g. provinces, states, counties, districts, municipalities. Admin levels are hierarchical and get smaller as the admin level number increases.

Admin Level 0 is the country outline
Admin Level 1 is the next level down, e.g. province in Angola
Admin Level 2 e.g. municipality in Angola
Admin Level 3 e.g. commune in Angola

Countries have different numbers of admin levels and not all of the boundaries may be available.

(If admin boundary shapefiles seem not to be available for your country in HDX, another good source to try is the GADM site : **https://gadm.org/download_country_v3.html**).

Humanitarian Data Exchange also has other data layers including processed street and building data from OpenStreetMap, and modelled population rasters from Worldpop.

Appendix 2 : Recap of useful commands and ideas

1. Open QGIS Desktop from Windows start menu or by double clicking on QGIS project file (.qgs or .qgz)

2. QGIS project files (.qgs or .qgz) contain appearance settings and links to data files, they don't contain data - if you send one to a colleague without data it won't work.

3. Data are added to QGIS as Layers, that appear in a Layers Panel on the left.

4. The main types of geographic data are Vector (points, lines and polygons) and Raster (gridded).

5. Vector data formats include Shapefiles (multiple files per layer .shp, .shx, .dbf etc.) and Geopackage (.gpkg).

6. To add vector data to QGIS click **Layer, Add Layer, Add vector layer or drag and drop** a .shp or .geopkg file into the layers panel.

7. The data behind a vector layer are stored in Attribute tables with one row per feature (e.g. point) columns for each attribute.

8. To open an Attribute table, **right click on a layer name in the layers panel, Open Attribute Table**.

9. To change the appearance (Symbology) of a map layer, **Right click on the layer name, Properties, Symbology.**

10. To add extra functionality to QGIS, choose Plugins, Manage and Install Plugins.

11. To add a basemap first install the QuickMapServices plugin then select Web, QuickMapServices, OSM, OSM Standard.

12. To add a text file with coordinates, click **Layer, Add Layer, Add delimited text layer.**Select Point coordinates for the Geometry Definition.

13. The most common Coordinate Reference System CRS (used, e.g. by GPS) is unprojected latitude, longitude using a WGS 84 datum with the code **EPSG 4326**.

14. To export a text file to geographic file. **Right click on a layer name, Export, Save Features As**

15. To add a text file with region names and join to an existing geographic layer. **Layer, Add Layer, Add delimited text layer.**Select No Geometry for the Geometry Definition. **Right click on the layer name of the existing geographic layer, Properties, Joins. click small green +**. (remember that the join is temporary and to export the layer to make it permanent).

16. To arrange a map on a page to show/send to others. **Project, New Print Layout.**

17. **The internet is your friend. Any question you have someone else has probably had it before and blogged about it. Type QGIS [*my problem*] into Google and see what comes back. Also works well to get advice in languages other than English.**

Appendix 3 : Troubleshooting exercise

Even when we follow good practice in data management and have learnt how to use a software package we will inevitably experience times when the software doesn't do what we expect it to and initially we don't know why. Sometimes this will be because we have made a typo, or have chosen the wrong command or because the data that we have got from someone else has errors. The software not doing what we expect can create frustration. However, fear not, this frustration can be a powerful motivator for learning ! Indeed it has been argued that this frustration is a necessary part of the learning process. Developing an ability and confidence to deal with these difficulties will make you better able to use the software to help in your work. The reason why your map won't display may well be because someone else put a comma in the wrong place rather than because of a lack of software knowledge on your part. Developing such troubleshooting ability (e.g. to spot commas in the wrong place) is valuable skill.

In the following section we have developed a series of troubleshooting questions that we would like you to work through one-by-one. These are based upon real issues that learners on our previous courses came to us with.

We will go through the answers to the questions at the end (and the answers are in the manual but we encourage you not to look at them before trying).

The data files for the questions are stored in the folder :
training-QGIS-201910/trouble_shooting_exercises/questions/

The files are named according to which question they relate to.

First, in QGIS, open up a blank project and save as :
training-QGIS-201910/my_work/projects_QGIS/trouble_shooting.qgs.

Questions :

Q1 Add the file q1.csv to QGIS. You are expecting the points to be in Colombia. Do they appear there?

Q2 Add the file q2.csv to QGIS. You are expecting the points to be in Colombia. Why are they not ? How do you fix it ?

Q3 Add the file q3.csv to QGIS. You are expecting the points to be in Colombia. Why are they not ? How do you fix it ?

Q4 Add the file q4.csv to QGIS. (hint your colleague who gave you the file says it has been saved in UTM)

Q5 Add the file q5.csv to QGIS. You are expecting the points to be in Colombia. Why are they not ? How do you fix it ?

Q6 Add the file q6.csv to QGIS. You are expecting the points to be in Colombia. Why are they not ? How do you fix it ?

Q7 The file q7.csv has some regional data without coordinates. Can you join this to the layer COL_adm1 ?

Q8 The file q8.csv contains more data to join to COL_adm1. Why don't fields join correctly ?

Q9 Open the project file q9.qgs, there is a problem with the data layers. What is the problem ? How can you fix it ?

Answers :

Q1 Add the file q1.csv to QGIS.
A1 these points should appear in Colombia as expected.

Q2 Add the file q2.csv to QGIS.
A2 the negative sign was missing from the longitude values, adding the negative sign to all the values in excel or notepad and re-adding the file to QGIS should result in the points being in Colombia

Q3 Add the file q3.csv to QGIS.
A3 The longitude and latitude columns were labelled wrongly in the source file. Changing them around in Excel or notepad will result in the coordinates appearing in Colombia when the new file is added.

Q4 Add the file q4.csv to QGIS. (hint your colleague who gave you the file says it has been saved in UTM)
A4 the coordinates were saved in the SRC 'UTM 18 N' which is the UTM zone for Colombia. You can set the SRC for the layer when you import it or afterwards through Right click, properties, general, SRC.

Q5 Add the file q5.csv to QGIS.
A5 The coordinates are saved as degrees, minutes, seconds, when adding the layer to QGIS, you need to tick the box : 'DMS coordinates'

Q6 Add the file q6.csv to QGIS.
A6: This csv file speaks Spanish ! It has semi-colon as the delimiter between data values and coma as the decimal indicator. Choose File format, Custom delimiters and select 'semicolon', also tick the box : 'Decimal separator is comma' . If you view the text file in notepad or other simple text viewer you can see what the delimiter is.

Q7 The file q7.csv has some regional data without coordinates. Can you join this to the layer COL_adm1 ?
A7 These 3 rows should join OK

Q8 The file q8.csv contains more data to join to COL_adm1. Why don’t fields join correctly ?
A8 The first row is missing a capital letter, the 2nd row is missing an accent and the third row has a spelling mistake, the last letter should be e not a.

Q9 Open the project file q9.qgs, there is a problem with the data layers. What is the problem ? How can you fix it ?
A9 The project file cannot find the data layers because either the project file or the data have been moved. If you click on the message box you should be able to navigate to tell QGIS where the data files are.