

Week 1 Day 4

Accessing and manipulating spatial data

Learning objectives

- know how to access free spatial data online
- able to manipulate spatial data within QGIS
 - Subset point data either using variable names or by location
 - Create buffers around points or lines of interest
 - Extract raster data for a subset of points
 - Summarise raster data

Length : 1 day

Updated : September 2018

[Creative Commons Attribution license](#)

4.1. Free spatial data resources

There is a vast amount of freely available spatially-referenced data online. We are not able to list all these resources here, but will introduce you to the resources that we are using in these practical sessions. These resources are not for a specific country, but provide us with spatial data that can be useful in many countries to help visualise data, and produce maps for the purpose of guiding control programmes. Following this course, you may wish to contact organisations within your country to find out if they have any spatial data resources that you can access e.g. government departments, research organisations, the environmental sector, the health sector etc.

Here we give you a brief overview of the data sources for administrative boundaries (polygons), human population distribution (points, rasters), climate variables (rasters) and infrastructure (points, lines, polygons).

Administrative boundaries

Administrative boundaries are the areas in which countries are subdivided for the purposes of administration. Conventionally, we refer to the administrative boundary levels as follows:

- Level 0 is always the country border/boundary
- Level 0 is subdivided into Level 1 e.g. States, regions
 - Level 1 is subdivided into Level 2 e.g. districts, municipalities
 - Level 2 is subdivided into Level 3 etc

For example the map below (Figure 4.1) denotes the level 0, 1 and 2 administrative boundaries of Mexico



Figure 4.1: A map of Mexico's administrative boundaries. Level 1 (country) = black thick black outlines, level 2 (states) = white outline, level 3 (municipalities) = thin grey outline.

These administrative boundaries can often change over time, so you should always try to access the most recent administrative boundary shapefiles directly from your own government. Often, there are freely available global administrative boundary resources such as Global Administrative Areas (GADM, www.gadm.org) and Natural Earth Data (<http://www.naturalearthdata.com/>), but whilst these resources try to be contemporary, it is not guaranteed. The GADM website allows you to download shapefile for individual countries, or you may also download a shapefile containing all of the countries in the world. Shapefiles are generally available for administrative levels 0, 1 and 2, and may often include smaller divisions. For each country you are able to download a single zipped folder, and this folder contains all of the levels available for the selected country. The Natural Earth Data website allows you to download boundary data at three different scales (1:10m, 1:50m, 1:110m), see Figure 4.2. This ratio represents the distance that each unit on the map represents. In this instance, 1:10m means that 1 unit on the map represents 10 million units in the real world e.g. 1cm represents 10 million cm = 100km, which is the finest scale available. The coarsest scale is 1:110m, meaning that 1cm represents 1,110km.

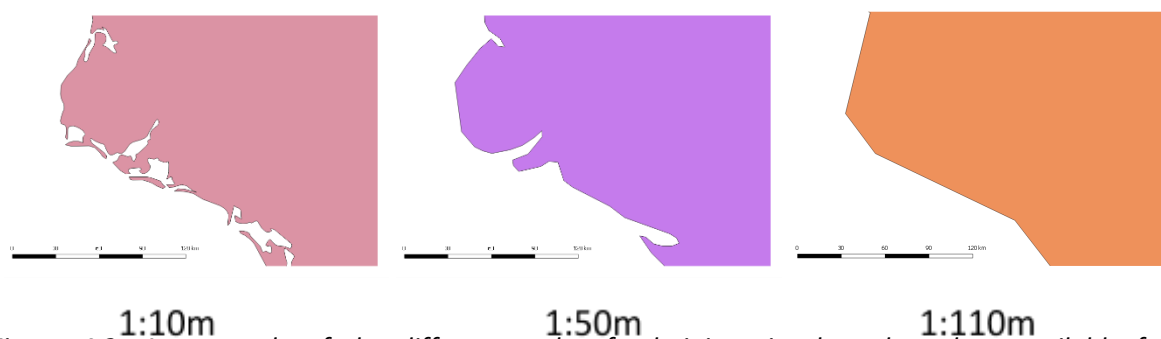


Figure 4.2: An example of the different scale of administrative boundary data available from www.naturalearthdata.com

These files can be found under the 'Cultural' files category, and you can either select to download a shapefile containing administrative level 0 or level 1 for all of the countries of the world e.g. the 1:10m scale files can be found here:

<http://www.naturalearthdata.com/downloads/10m-cultural-vectors/>

Population distribution

Knowledge of the human population is important when considering a public health intervention. Most countries undertake a national population enumeration census approximately every 10 years, but often this information is not released to the public at the geographical level of interest and alternative population data is required. In this course we look at two global population data resources, but again we encourage you to contact your own government to try to obtain the most relevant population distribution data for your country. The global population resources we are using are the locations of the most populated places, again obtained from Natural Earth Data (<http://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-populated-places/>) plus population distribution rasters from the WorldPop project (www.worldpop.org.uk). The populated places data consists of 7,322 points which represent administrative area 0 and 1 capitals, major cities and towns, plus some smaller towns. The shapefile also contains some estimates of population size, although these may not be contemporary. Whilst it is useful to know where the major urban areas are located, these data do not tell you anything about the size of the geographic area covered by these towns and cities, and further the data do not provide any information on the distribution of the population within rural areas.

An alternative to using point-level population data is to use resources that try to describe the continuous distribution of the population using raster data such as the WorldPop project. WorldPop has combined census data, survey data and information obtained from satellite imagery to model the distribution of the population at 100m spatial resolution. Estimates are available for 2015, plus predictions for 2020 have also been generated.

Climate variables

There are many datasets that represent climate data such as temperature and rainfall, which vary greatly both in their spatial scale and frequency in which they are produced. These data are usually derived from meteorological satellite imagery, sometimes in combination with ground-based meteorological stations. In this course we will be using WorldClim data (<http://www.worldclim.org/>), which is a publically available resource of average monthly climate data produced at a 1km spatial resolution. The current WorldClim data is representative of the period 1960-1990, and contains raster data of mean, minimum and maximum temperature, rainfall, plus other variables combined from these measures. Whilst this dataset is useful for learning more about how climate variables vary geographically, this data cannot be used to determine what the weather was like at a particular timepoint, nor can it be used to explore changes in climate over time. Data is downloaded in the form of tiles rather than for specific countries from here <http://www.worldclim.org/tiles.php>. You may need to download multiple tiles to cover your entire area of interest. Note that there are many sources of climate data available. We are using WorldClim as an example dataset, but as it is based on historical data it may not be the best resource for your own work.

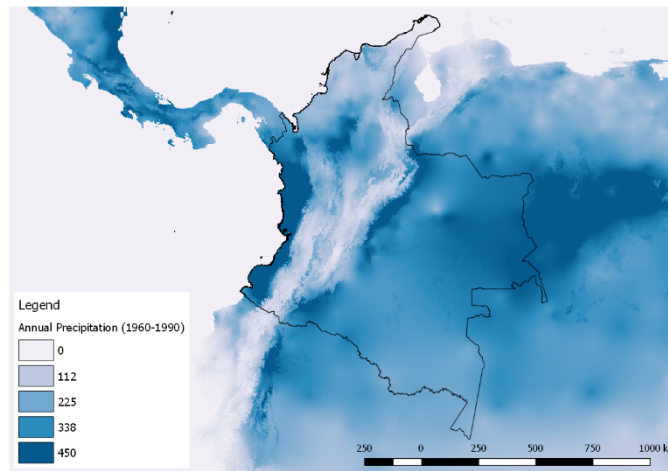


Figure 4:3: An example of precipitation data available from www.worldclim.org

Infrastructure

We have already introduced OpenStreetMaps as a resource for giving geographical context for your data by adding it as a basemap in QGIS. It is also possible to download the data that is displayed as shapefiles, which you can then use to improve your map. For example, the road data that we used in Day 1 was downloaded from OpenStreetMaps. In addition to roads, you can use OpenStreetMaps to obtain other point, line and polygon data relating to the surrounding infrastructure such as the location of local landmarks and amenities, rivers, lakes, parks etc. There are several ways in which you can import OSM data into QGIS, such as using the Export feature on www.openstreetmap.org, and some of the inbuilt Vector tools within QGIS (Figure 4.4). We will give a brief demonstration during the practical, and you can access guides on how this can be done online e.g. <http://wiki.openstreetmap.org/wiki/QGIS>.

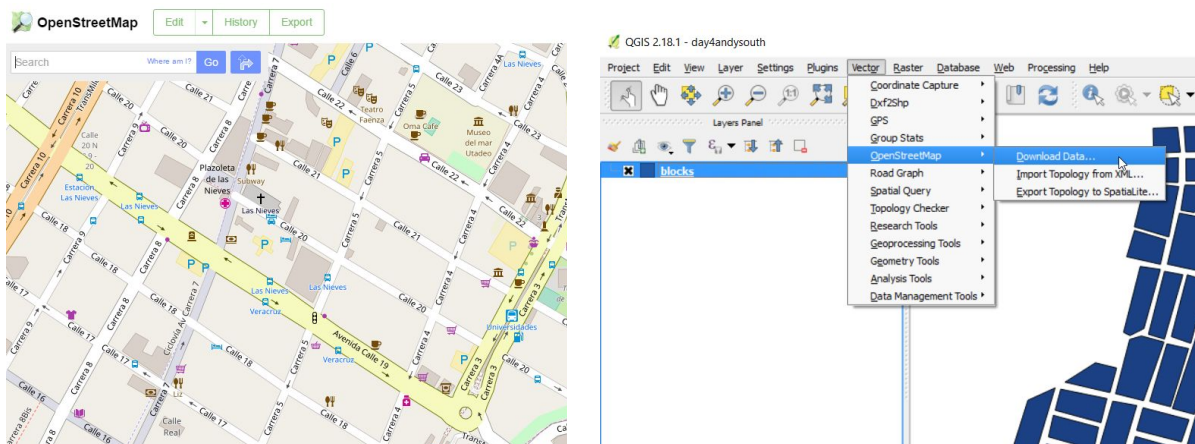


Figure 4:4: Openstreetmaps is a crowdsourced resource for geographical information, and QGIS has inbuilt tools to allow you to download OSM shapefiles


Aedes aegypti and *Ae. albopictus* data

As well as these general resources, you may also be able to find data more relevant to your programme in the published literature, as many academic journals are now encouraging researchers to make their data publicly available. For example, we will be using data published in the paper

The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence, was published in 2015 (<http://www.nature.com/articles/sdata201535>) and all of the geo-located occurrence data here can be downloaded as a CSV using this link: <http://datadryad.org/resource/doi:10.5061/dryad.47v3c> . More details of this data are given below.

Data Descriptor | [OPEN](#)

The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence

Moritz U. G. Kraemer , Marianne E. Sinka, Kirsten A. Duda, Adrian Mylne, Freya M. Shearer, Oliver J. Brady, Jane P. Messina, Christopher M. Barker, Chester G. Moore, Roberta G. Carvalho, Giovanini E. Coelho, Wim Van Bortel, Guy Hendrickx, Francis Schaffner, G. R. William Wint, Iqbal R. F. Elyazar, Hwa-Jen Teng & Simon I. Hay 

Scientific Data 2, Article number: 150035 (2015)
doi:10.1038/sdata.2015.35
Download Citation

Received: 30 March 2015
Accepted: 23 June 2015
Published online: 07 July 2015

Dengue virus Ecological epidemiology


Entomology Risk factors

Data from: The global compendium of *Aedes aegypti* and *Ae. albopictus* occurrence

Kraemer MUG, Sinka ME, Duda KA, Mylne A, Shearer FM, Brady OJ, Messina JP, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Wint GRW, Elyazar IRF, Teng H, Hay SI

Date Published: February 16, 2017
DOI: <http://dx.doi.org/10.5061/dryad.47v3c.2>

Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the [Dryad Terms of Service](#). To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.  [open data](#)

Title	Ae. aegypti and Ae. albopictus occurrences
Downloaded	118 times
Description	This datafile contains a comprehensive list of occurrences of <i>Ae. aegypti</i> and <i>Ae. albopictus</i> from 1960-2014.
Download	aegypti_albopictus.csv (3.406 Mb)
Details	View File Details

Figure 4.5: Screenshots of the *Aedes* occurrence paper (left) plus online data source (right).

4.2. Manipulating spatial data in QGIS

In this section we will be demonstrating different ways of manipulating spatial data partly using data obtained from the free spatial resources mentioned above. We have included the data being used in this practical session on your pen drive. You may want to experiment with downloading similar data from the internet yourself after the session.

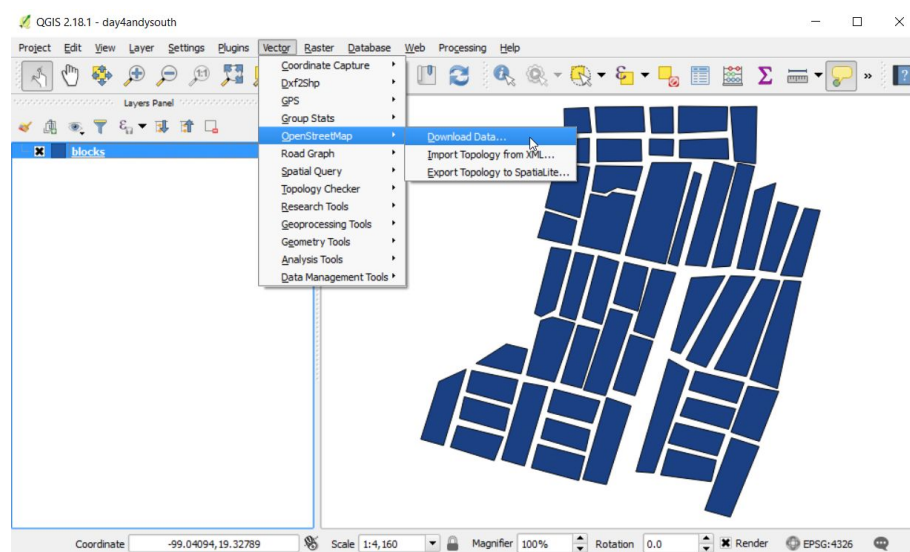
In our first example we will look at the blocks and houses data from Day 1, see how we can add named streets from OpenStreetMaps. We shall then look at selecting houses with disease cases and identifying areas within 50m of these to help with prioritising interventions.

- open QGIS and click 'Project, New' (as we did yesterday).
- Save this blank project as Day4yourname.qgs using 'Project, Save As ..'.
- Add a polygons layer by Layer, Add Layer, Add Vector Layer and browse to find : **QGIS_training\data\spatial\polygons\blocks.shp**

4.2.1. Downloading OpenStreetMap data

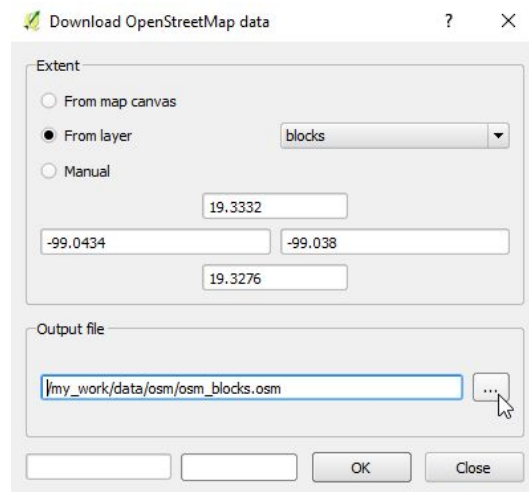
Previously, we displayed OpenStreetMap data as a basemap i.e. a background image. Whilst basemaps are very useful, you do not have the flexibility to manipulate the data being displayed (change its appearance, subset the data etc.). We therefore will now show you how to download the OSM data and save the resulting points, lines and polygons as shapefiles.

- Select Vector, OpenStreetMap, Download data.



This should bring up a window in which you can choose the area to download and whether to save the data as a file. You can choose whether to define the area to download from the map canvas (the area currently visible in QGIS), the area of one of the loaded layers, or by manually entering coordinates.

- set the extent from the blocks layer.
- Under 'Output file', click the ... button to browse and identify the file to save as: **/my_work/data/osm/osm_blocks.osm**. Create this folder if it does not exist already (e.g. within the data folder : right click, New, Folder).
- Press OK.



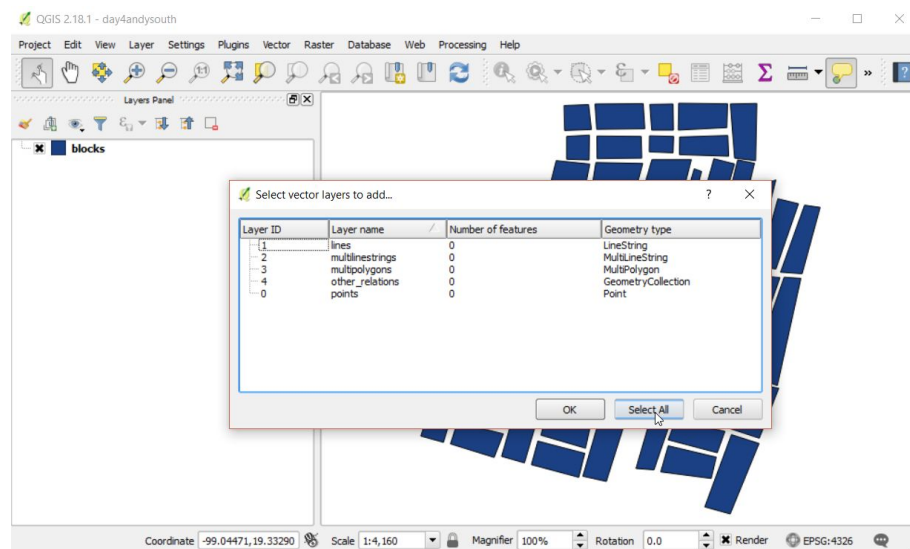
The downloaded data is saved as an OSM format file. This is different from the files we have seen before.

Having downloaded the OSM data we still need to add it to our map in QGIS.

→ select Layer, add layer and navigate to the folder to find the OSM file.

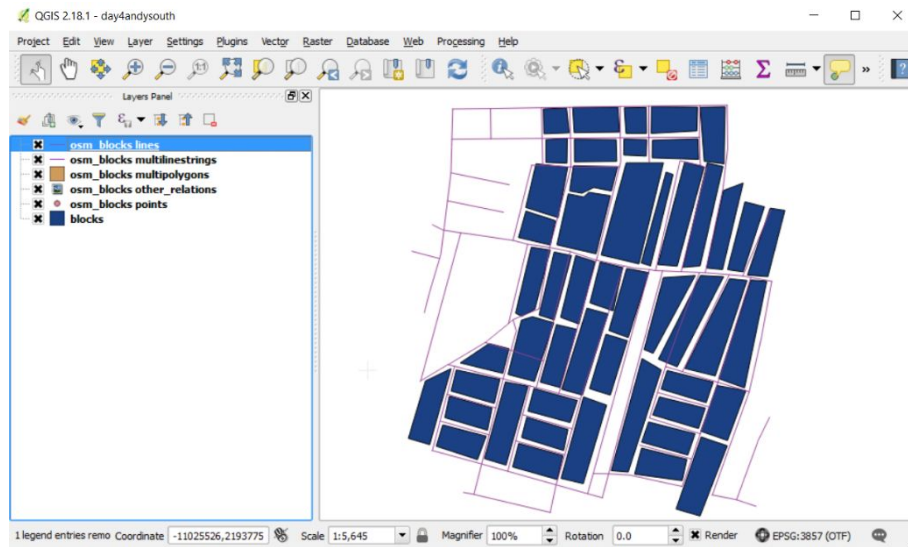
A window will ask you to select vector layers to add. The OpenStreetMap data includes points, lines and polygons.

→ Because we want to look at all of them Click 'select all', then OK.



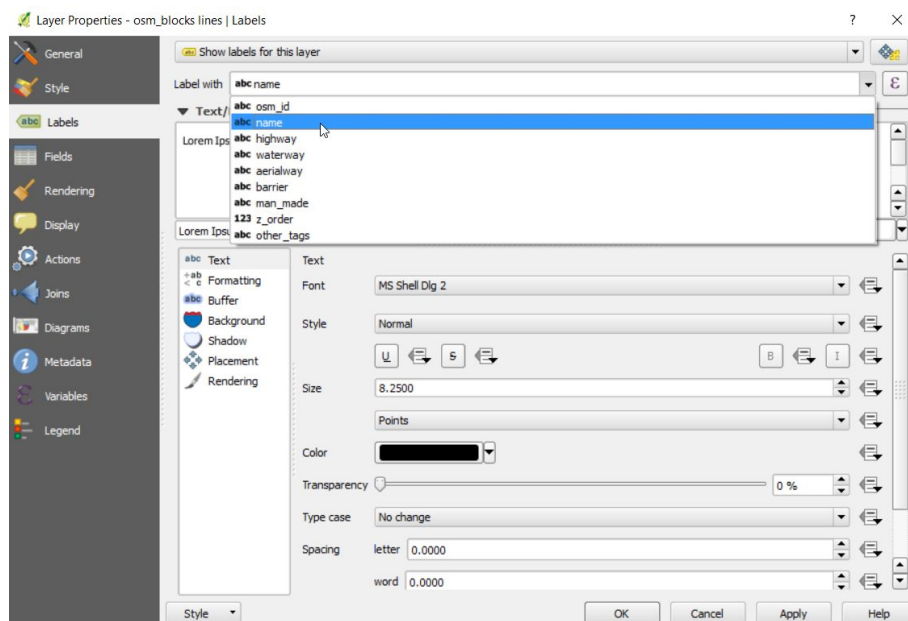
This brings up streets for our area in a lines layer. If you experiment with turning on and off layers you'll see that the other points and polygons layers don't contain anything, so you can remove them (Right click, remove). The lack of other data here was because of the small area we chose, in other areas you can get polygons e.g. indicating buildings or parks and/or points e.g. indicating health centres, shops and post boxes.

→ Right click on empty layers in the Layers Panel and select remove

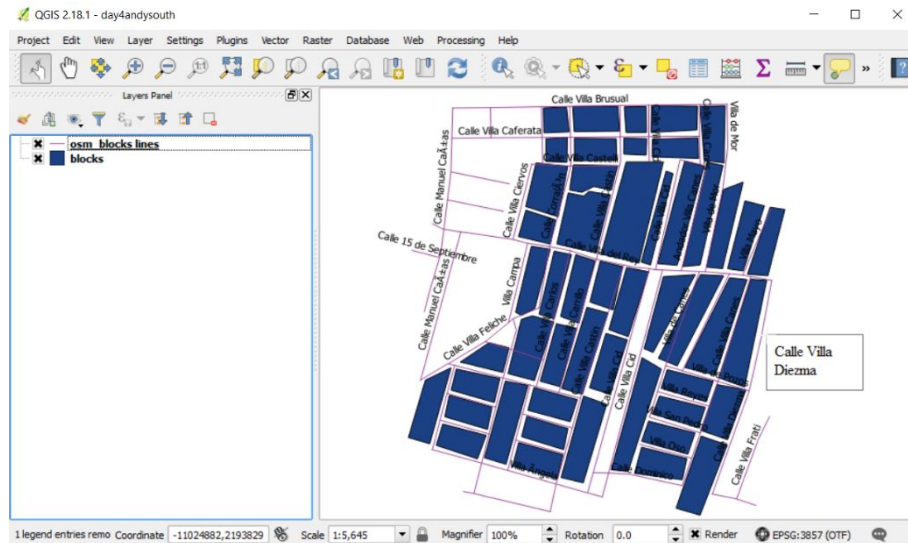


If you open the attribute table for the lines layer you will see that the name column contains street names.

- Turn on labels for the streets by Right clicking on osm_blocks_lines, Properties, labels
- choose 'Show labels for this layer'
- set 'Label with' to the column 'name'



This should display the street names alongside each street :



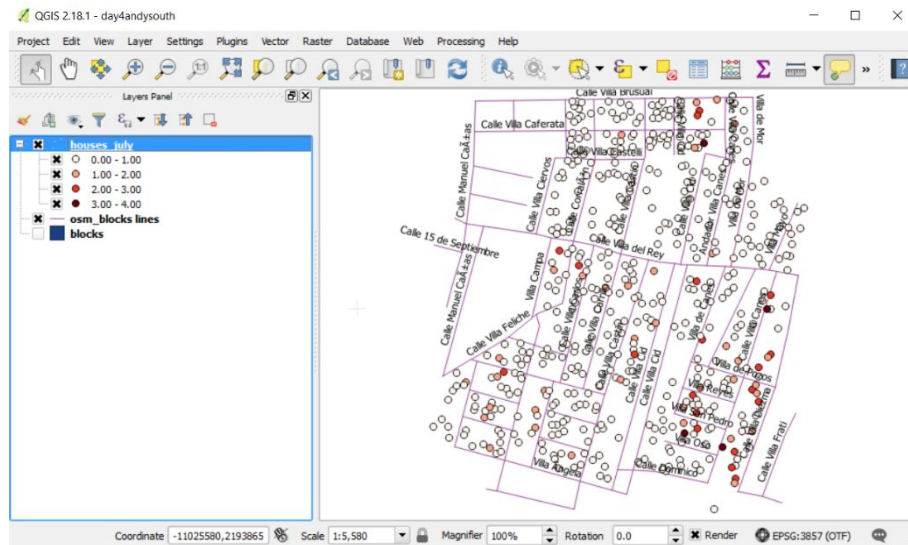
You could also save the streets as a lines shapefile by right clicking and save-as. Both this and using the osm file mean that you can use the OpenStreetMap data later even if you don't have an internet connection. Using OpenStreetMap as a background map as we did before requires an internet connection.

We will now consider an example where Zika cases have been reported and we wish to use the GIS to find where to target interventions nearby.

- Add the example houses layer to the existing project from :
QGIS_training\data\spatial\points\houses_july.shp (Hint : Layer, Add layer).

If you open the attribute table (after right-clicking on the layer name) you will see that there is a column named 'cases' containing the number of Zika cases per house.

- Use properties, style to set the colouring of the points to see which houses have cases and which don't. Try using the reds colour scheme and 4 classes. The points with no colour are those with no cases.



4.2.2. Selection tools in QGIS

You can select features within layers using a number of tools including clicking on features, drawing a polygon around features or selecting according to attributes.

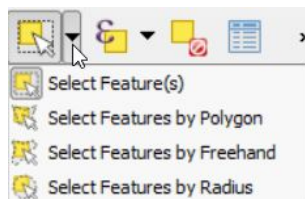
The selection buttons are the yellow ones towards the right of the toolbar :



This can help you remember that when you select features they become highlighted in yellow on the map. Selection can also be accessed from the toolbar by View, Select.

Selection by cursor

The series of commands below allow you to select features by directly clicking on them (Select Features) by drawing a polygon (click to draw each side of a polygon and double-click to finish) or by radius (click and drag to create a circle). Importantly for these cursor selection to work you must have the layer selected by clicking on it's name in the Layer Panel on the left.



You can remove any selections you have made using the button below, or View, select, deselect ... or Ctrl+Shift+A.



Note that you can also use the cursor to select features by first opening the attribute table and then clicking on the row number on the left hand side. The row will become highlighted in blue and again the feature will be highlighted in yellow in the map view. You can use Ctrl and Shift to select multiple rows as you would in a spreadsheet.

houses_july : Features total: 500, filtered: 500, selected: 2

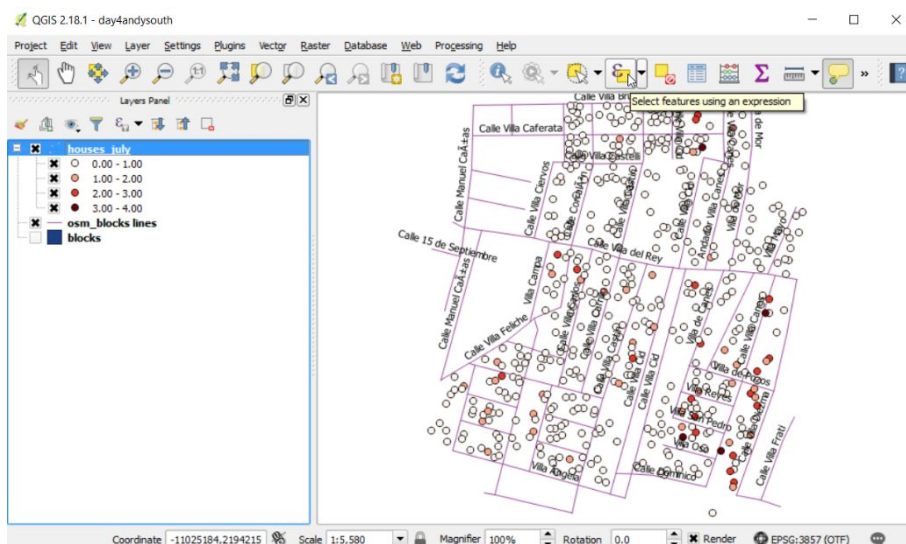
	UniqueID	Longitude	Latitude	BlockID	HouseID	Interventi	Num_people	Cases	Date
1	B1H1	-99.0414143200...	19.33294233999...	1	1	1	5	1	01/07/2016
2	B1H2	-99.0411139200...	19.33295863000...	1	2		5	0	01/07/2016
3	B1H3	-99.0412862499...	19.33302373000...	1	3		6	0	01/07/2016
4	B1H4	-99.0412417099...	19.33292614000...	1	4		7	0	01/07/2016
5	B1H5	-99.0413536599...	19.33305786000...	1	5		3	0	01/07/2016
6	B1H6	-99.0412937400...	19.33288608999...	1	6		1	0	01/07/2016
7	B1H7	-99.0413082400...	19.33312703000...	1	7		1	0	01/07/2016
8	B1H8	-99.0412495199...	19.33311232999...	1	8		3	0	01/07/2016
9	B1H9	-99.0413753200...	19.33312347000...	1	9		8	0	01/07/2016
10	B1H10	-99.0413528600...	19.33285516000...	1	10		9	0	01/07/2016
11	B2H1	-99.0404130600...	19.33291431000...	2	1		11	0	01/07/2016
12	B2H2	-99.0406528299...	19.33292819000...	2	2		2	0	01/07/2016
13	B2H3	-99.0405739699...	19.33307780000...	2	3		3	0	01/07/2016
14	B2H4	-99.0408030499...	19.33297617999...	2	4		5	0	01/07/2016

Show All Features

Selection by attribute

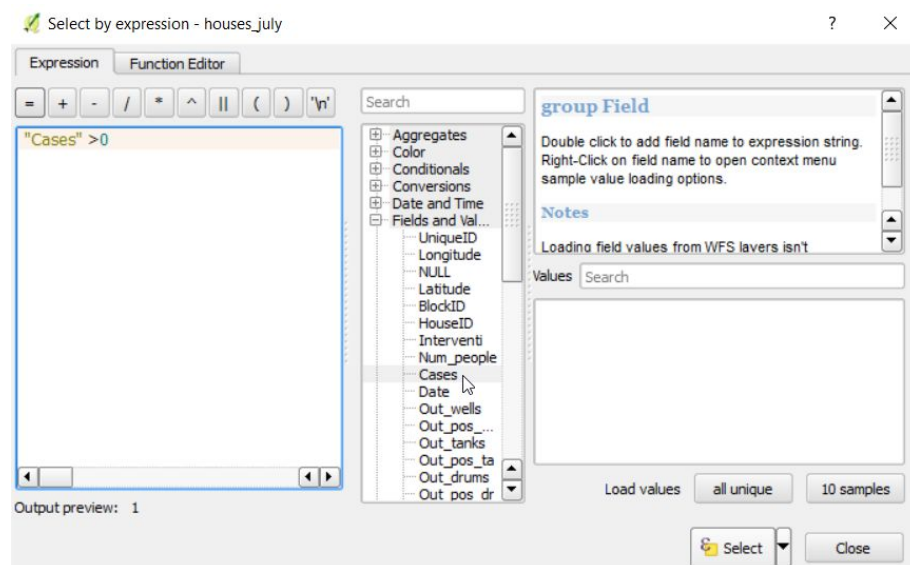
Probably more useful is selecting data according to their attributes i.e. the data stored in the Attribute table. To do this :

- click 'View, Select, Select features by expression' or click the button shown below on the toolbar, or in the tools displayed at the top of the attribute table.



This tool allows you to subset your data according to the values stored within the attribute table. We do this by writing an *expression*, which is simply a formula that uses mathematical symbol similar to that which we use in Excel. Here we will demonstrate how to select only the points that have one or more cases.

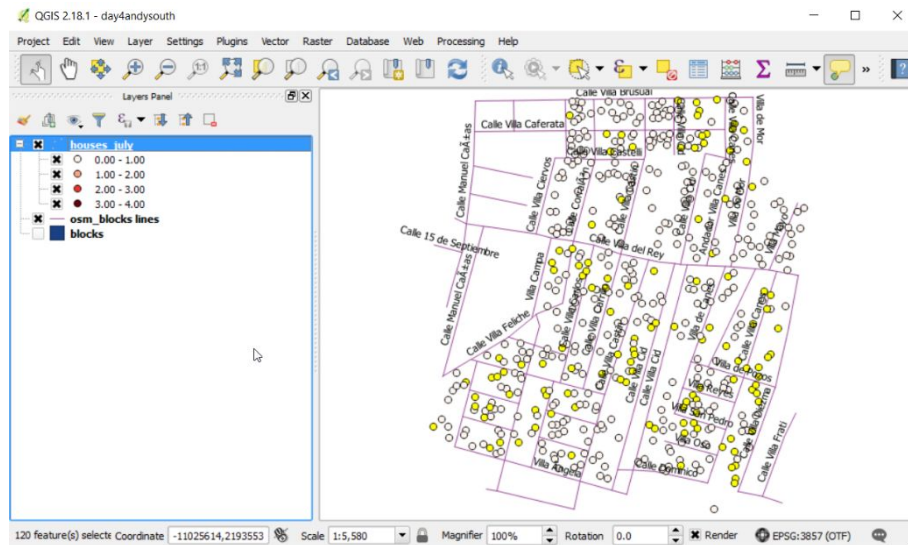
- Within the central box, locate 'Fields and Values' and click on the + next to it to expand the menu.
- Double click on Cases and it gets added to the Expression box.
- To inform QGIS that you only want the points that have at least one case, type > 0. So that the expression is "Cases" > 0.
- Click Select.



You can check whether this expression is valid by checking the information next to the text 'Output preview' at the bottom left of the window. If there is a problem it will say 'Expression is invalid' in red.

- Once you have checked this, click the Close button.

On your map you should see that selected points are bright yellow, hiding all the dark red points which shows that our selection has worked.



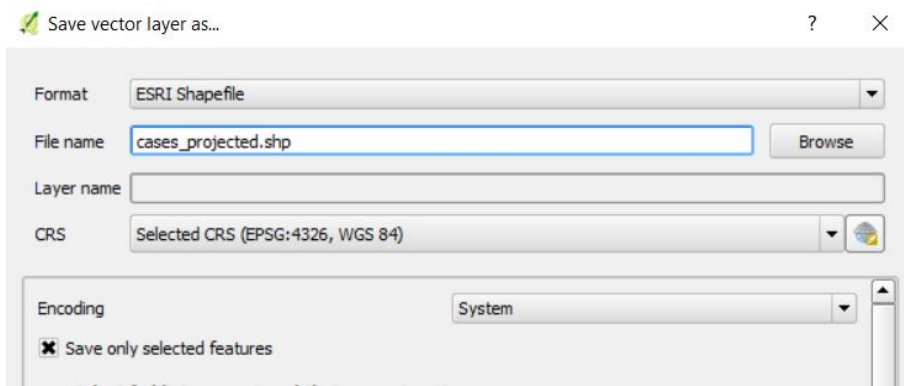
Open the attribute table and you can see the rows selected in blue are indeed the ones with cases.

houses_july : Features total: 500, filtered: 500, selected: 120

UniqueID	Longitude	Latitude	BlockID	HouseID	Intervent	Nun_people	Cases	Date	Out_wells	Out_pos_w	Out_tanks	Out_pos_ta	Out_drums
8345	-99.0407733299	19.3329503900	3	5		7	0	01/07/2016	1	0	1	1	1
8346	-99.0399533399	19.3330156000	3	6		6	0	01/07/2016	0	0	1	1	1
8347	-99.0422228000	19.3329571000	3	7		8	0	01/07/2016	1	0	1	1	1
8348	-99.0401384000	19.33289912000	3	8		4	0	01/07/2016	0	0	1	1	1
8349	-99.0400122100	19.33293237999	3	9		3	0	01/07/2016	0	0	1	1	0
83410	-99.0400891000	19.33313254999	3	10		2	0	01/07/2016	1	0	0	0	0
8441	-99.0393843199	19.33284667999	4	1		3	0	01/07/2016	0	0	1	1	0
8442	-99.0393251099	19.33311360000	4	2		4	2	01/07/2016	0	0	0	0	0
8443	-99.0398029499	19.33287740999	4	3		5	0	01/07/2016	1	0	1	1	0
8444	-99.0398028700	19.33291197000	4	4		4	2	01/07/2016	0	1	1	1	0
8445	-99.0398269399	19.33306911000	4	5		3	0	01/07/2016	0	0	1	1	1
8446	-99.0398615000	19.33301530000	4	6		1	0	01/07/2016	0	0	1	1	1
8447	-99.0398675000	19.33311367000	4	7		5	1	01/07/2016	0	0	1	1	1
8448	-99.0398536099	19.33289309999	4	8		4	0	01/07/2016	0	0	1	1	1
8449	-99.0398665700	19.33280338999	4	9		6	2	01/07/2016	0	0	0	0	0
84410	-99.0391612200	19.33293926000	4	10		7	1	01/07/2016	1	0	1	1	1
84411	-99.0390530799	19.33264280000	5	1		3	0	01/07/2016	0	0	1	1	1
84412	-99.0390582999	19.33282183000	5	2		4	2	01/07/2016	0	0	1	1	0

It may be useful to know that you can save your selected features as a separate shapefile. In this example, this will create a separate layer of case locations (with their associated data).

- ➔ To do this, right-click on your houses_july layer and select Save As.
- ➔ Name your file cases_july.
- ➔ Browse to find the folder \QGIS_training\my_work\data\points
- ➔ Click the '**Save only selected features**' box so that only houses with at least one case will be saved (following your selection in the earlier step).



→ Click OK.

You should now have a new shapefile of points, added automatically as a layer to your project.

→ Click the Save button, just to ensure the edits to your project are captured at this point.

4.2.3. Buffer creation

Now we want to find the areas that are close to houses with cases and we can do this by creating a 'buffer'. A buffer is a zone of specified width. For example, a buffer of 5km around a single point is simply a circle of radius 5km, centred on the single point. You can see more examples of buffers in Figure 4.6.

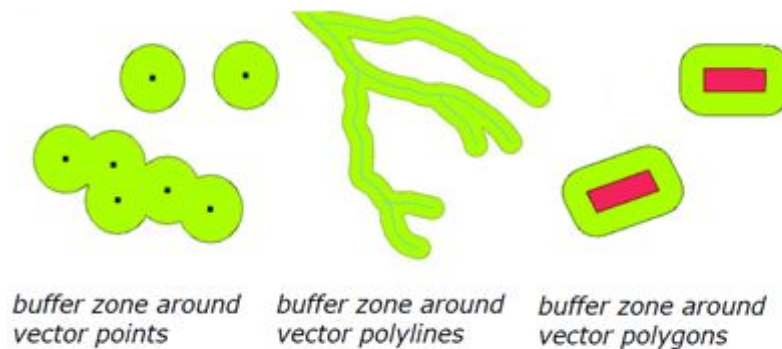
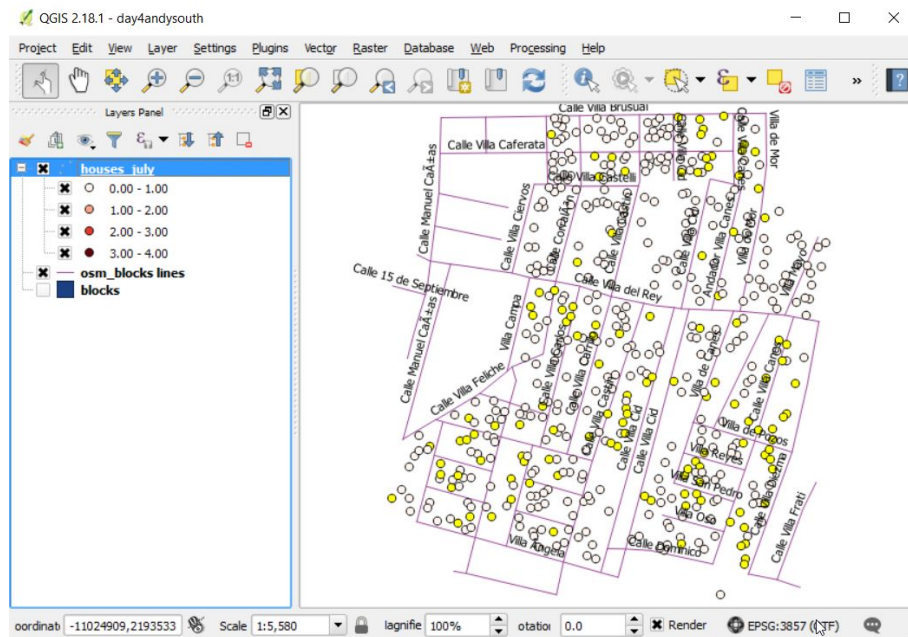


Figure 4.6: Examples of buffers

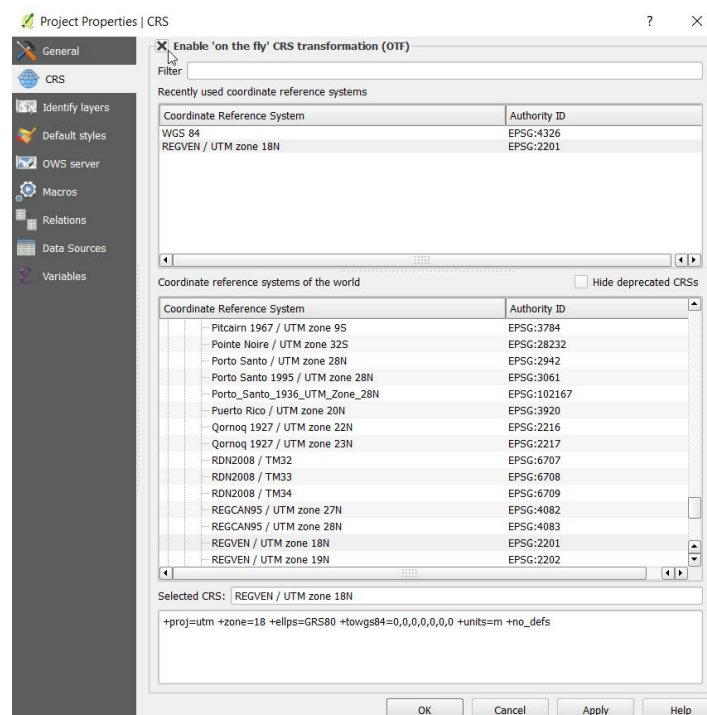
In order to create buffers using distance measures such as metres or kilometres, the shapefile which you are creating a buffer around needs to be in a projected coordinate system. If you try to create a buffer in a geographic coordinate system, such as longitude and latitude, you will need to specify the buffer distance in decimal degrees, which may be difficult to interpret. Therefore, before you create your buffer you need to convert your layer to a projected coordinate system such as UTM (Universal Transverse Mercator) where one unit represents one metre. To do this, you need to complete the following steps below.

Firstly, we need to check the settings within QGIS ensure that it can cope with displaying layers with more than one coordinate reference system (CRS).

→ Click the 2nd icon in from the bottom right and it will bring up a 'Project Properties CRS' window.



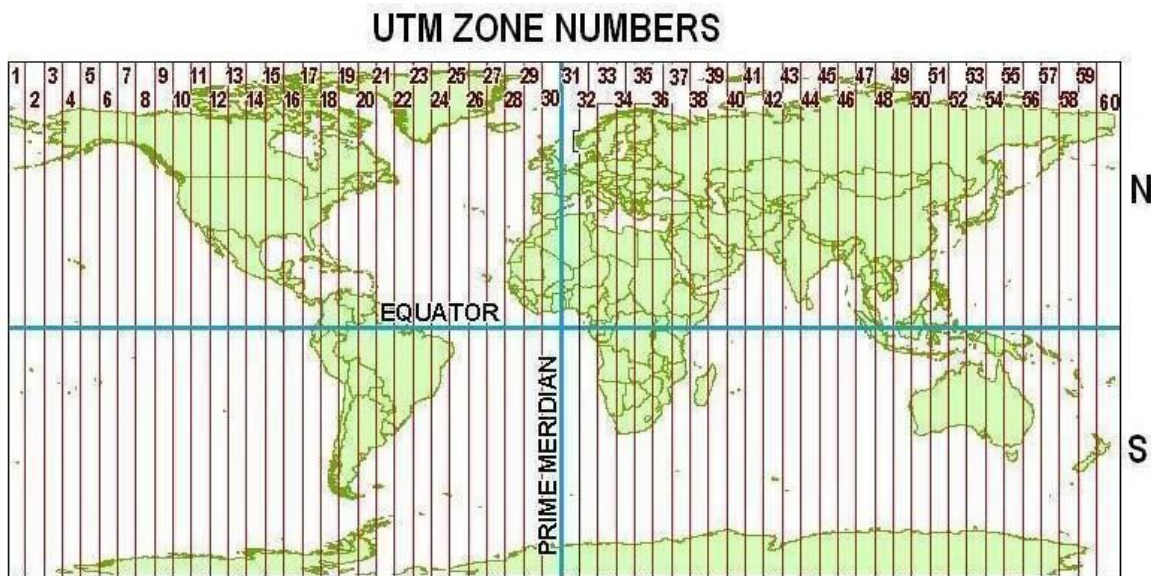
→ Select checkbox at the top to 'Enable 'on the fly' CRS transformation (OTF)'



→ Click OK.

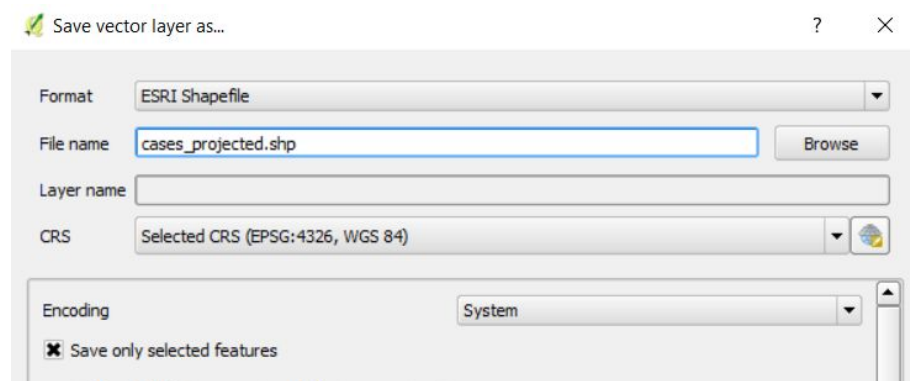
Now we will save the file cases_july in the new projection system. We will be using the projected coordinate system Universal Traverse Mercator (UTM), which splits the globe into 60 zones, and each zone is then split into a further two zone (N or S) depending on whether the zone is north or south of the equator. Here is a UTM zone map from:

http://earth-info.nga.mil/GandG/coordsys/grids/universal_grid_system.html



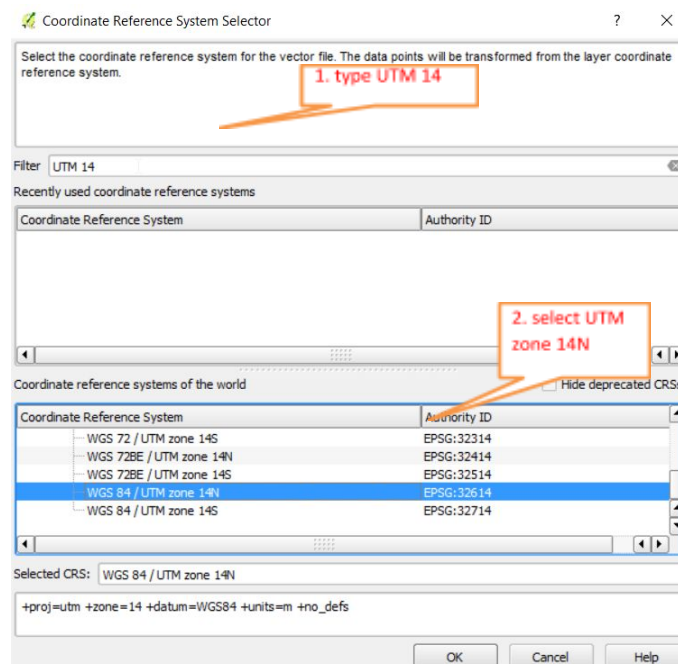
UTM zone 14N covers Mexico city.

- ➔ Right click on cases_july as we have done before and select Save As...
- ➔ Name your file cases_projected.shp (the projected indicates that we have saved in a different Coordinate Reference System)
- ➔ Browse to find the folder \QGIS_training\my_work\data\points
- ➔ Click the globe symbol to the right of the CRS box to change the coordinate system that the new file will be saved in.

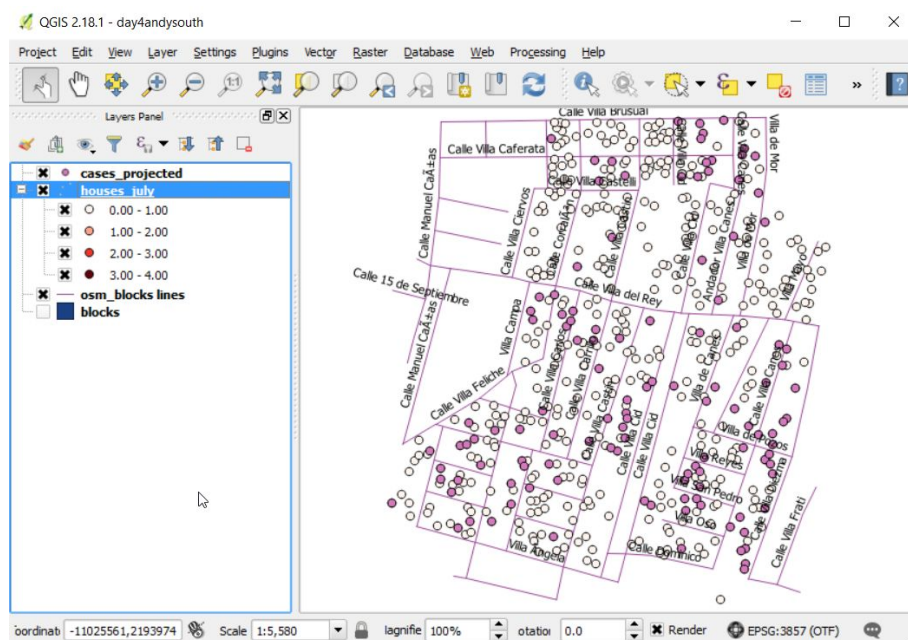


This will bring up a CRS selector window.

- Type 'utm' into the Filter box
- select 'WGS84/UTM zone 14N' from below 'Coordinate reference systems of the world'
- Click OK, then click OK again to save the file and display it on your map.



cases_projected should be added to the map like this :



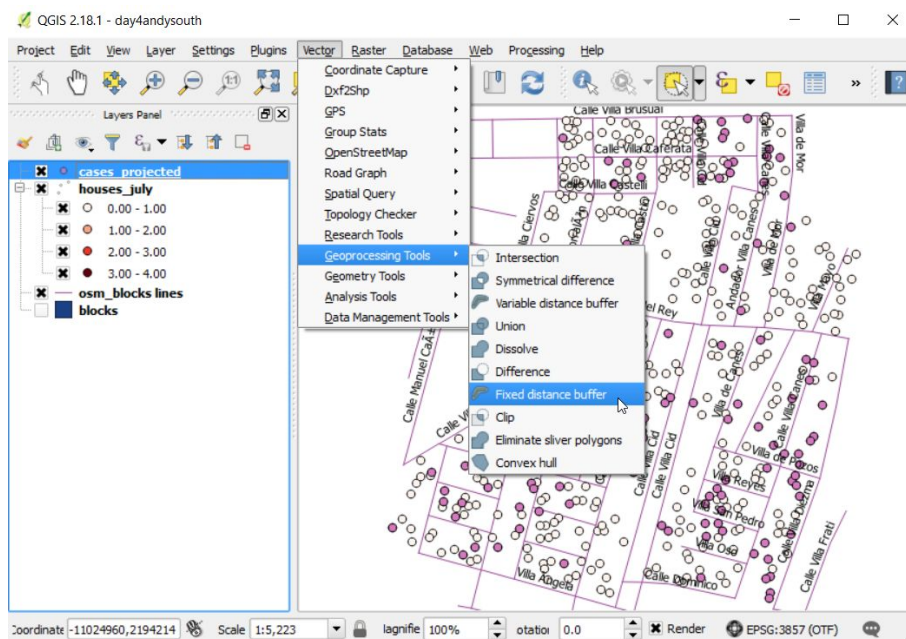
Note that you could combine steps in the future if desired – e.g. Saving only the selected features and saving them in UTM projection both within the Save As options.

- Note what happens if you turn On The Fly on and off. Use Zoom to full extent to see position of points when OTF is off.

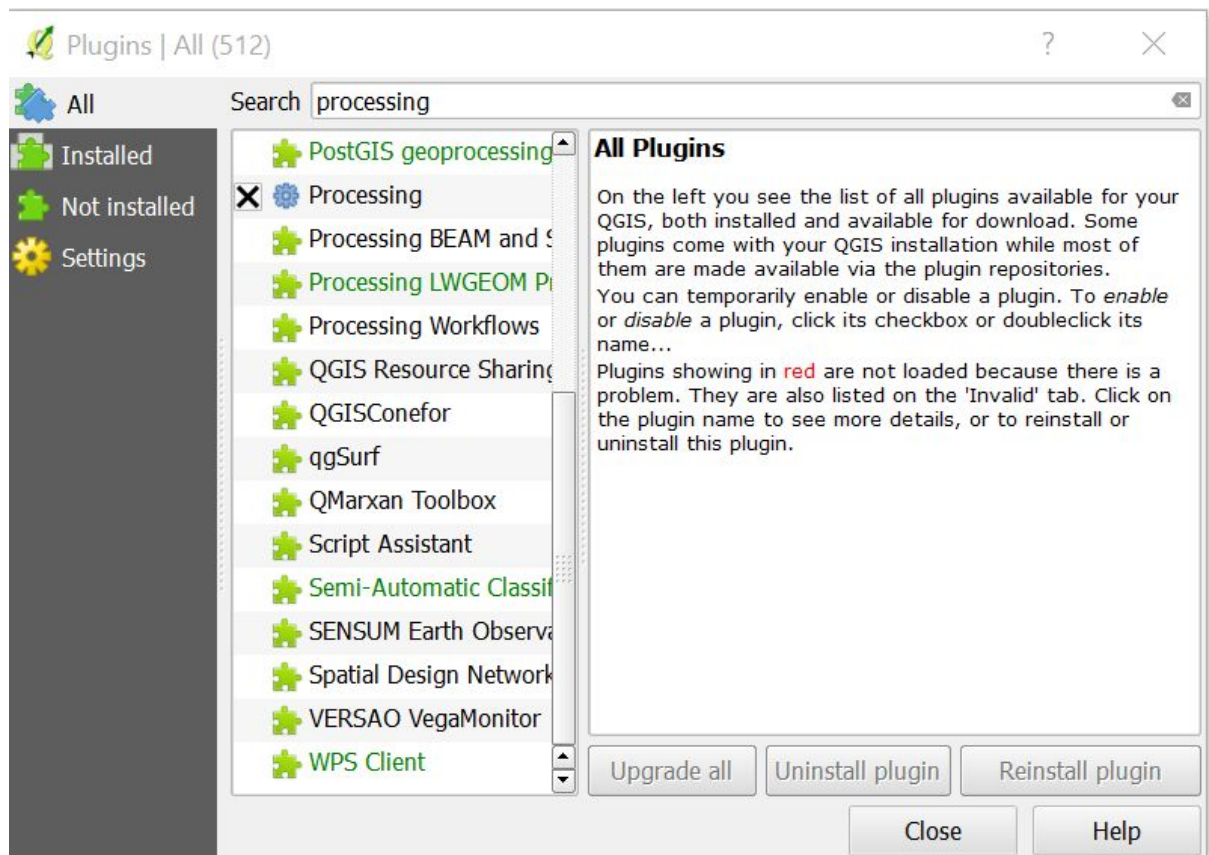
We will demonstrate creating a buffer, by adding a buffer of 50m around the Zika cases. Such a buffer could be used to direct vector control interventions. You could easily repeat this for 150m which is more conventional for vector control interventions, or any other distance. We chose 50m for this example simply because our example area is very small with the sampled houses close together and when we tried a buffer of 150m it covered the whole area.

To create buffers, we are going to use Geoprocessing Tools. These are located under Vector in your menu.

→ Click 'Vector, Geoprocessing Tools, Fixed distance buffer'.



If for some reason they do not appear in your Vector menu, go to Plugins, Manage and Install Plugins and search for Processing by typing in the search box. Make sure that this is installed and selected.

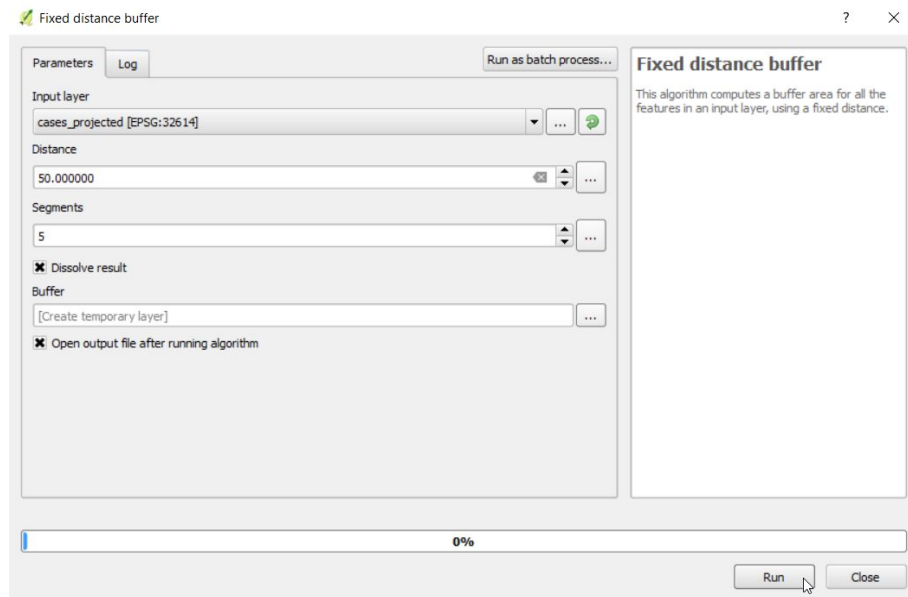


Once the option is there,

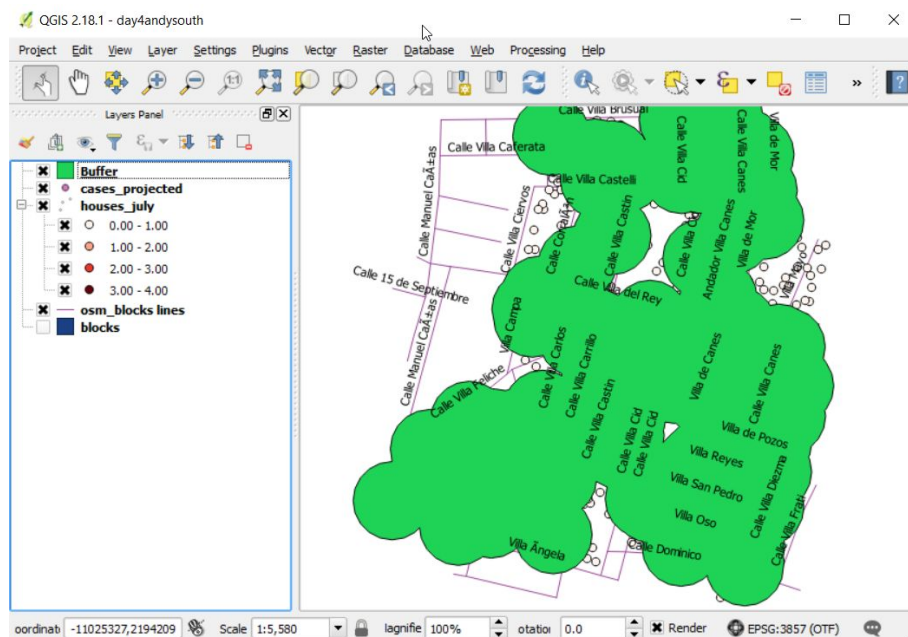
- click 'Vector, Geoprocessing Tools, Fixed distance buffer'.

In the window :

- set Input layer to cases_projected
- set distance to 50
- click the 'Dissolve result' box. This dissolves any boundaries between the buffers if they overlap.
- click Run



The resulting Buffer should look like this. You see that because we selected 'dissolve boundaries' the circles have been merged into one shape. This shape represents all the area that is within 50m of a house with a recorded case.



Hint: If your buffer does not look like this but covers the whole window, you may not have changed your projection to UTM and have created a buffer of 50 decimal degrees instead of metres. If you have many separate circles, you did not select the dissolve boundaries option.

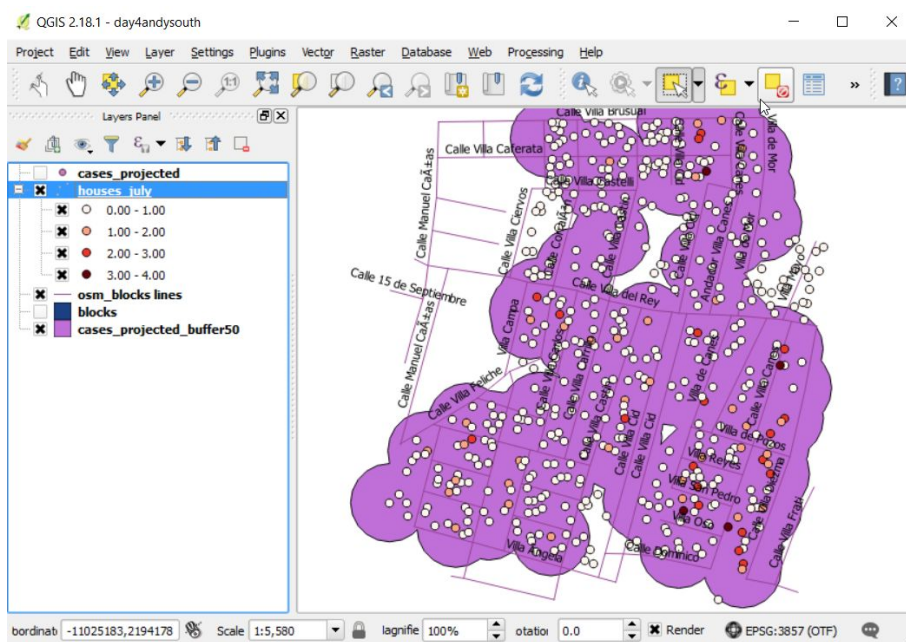
Once you have a buffer that looks like the above image,

- right click on the buffer layer and save it as a shapefile.
- Call it cases_projected_buffer50
- save in the folder my_work/data/polygons.
- remove the temporary buffer layer.

Remember you can change the order of your layers to see things more clearly.

- Drag cases_projected_buffer50 to the end of the list so that the houses can be seen on top.

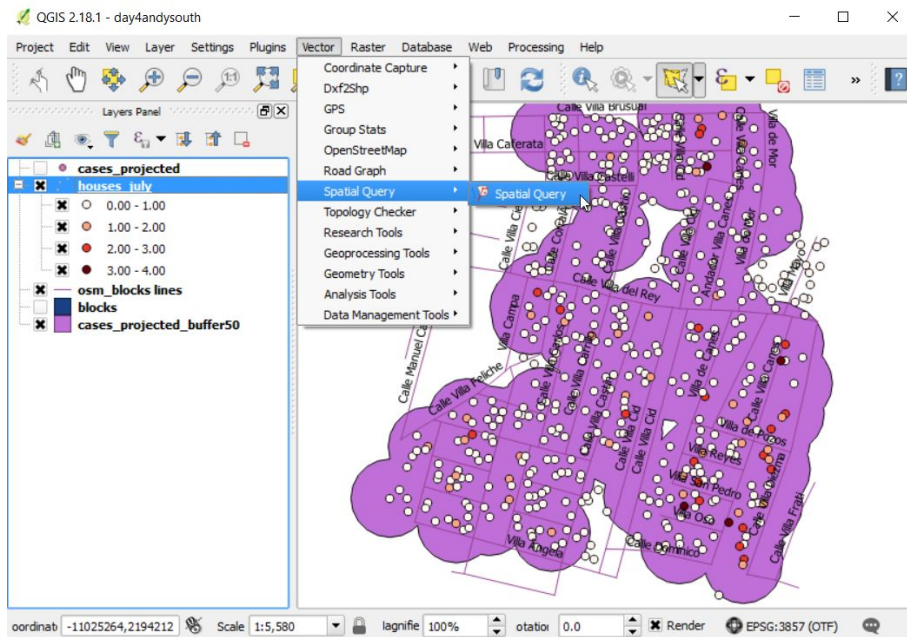
- Use the deselect button to remove any selections that are active from before.



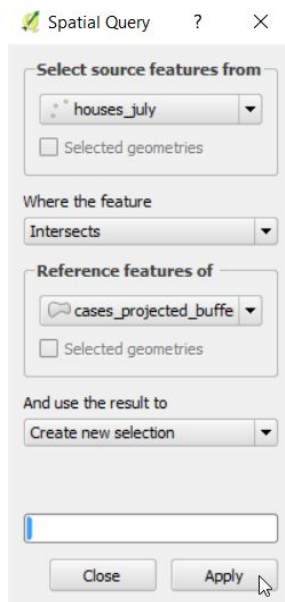
4.2.4. Spatial query

We can see which houses are within the buffer but what if we wanted to calculate which houses are within the buffer ?

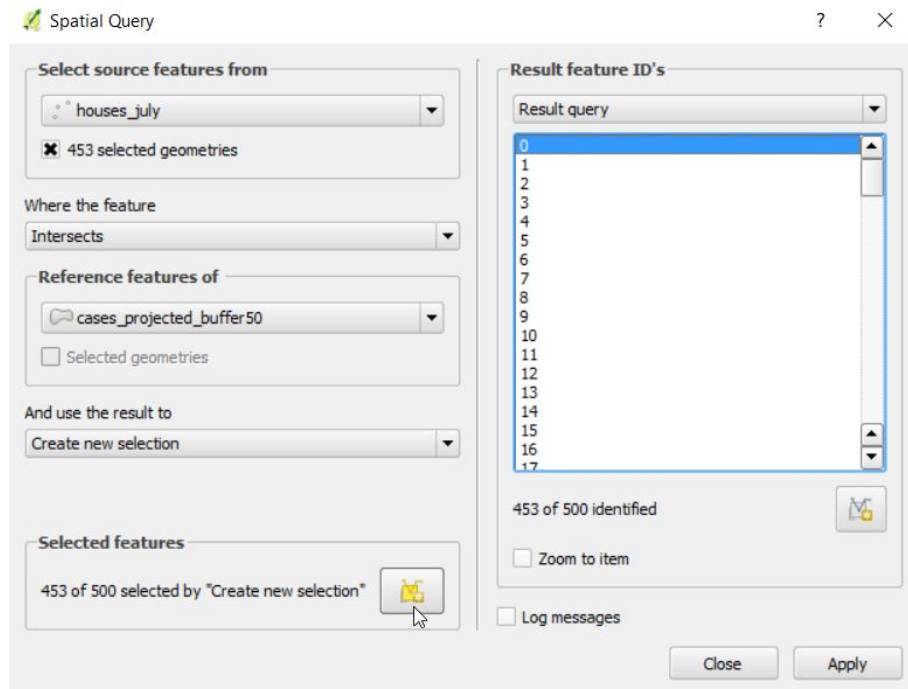
- choose Spatial query from the Vector menu :



- Select source features from : houses_july
- Where the feature : Intersects
- Reference features of : cases_projected_buffer50
- Click Apply



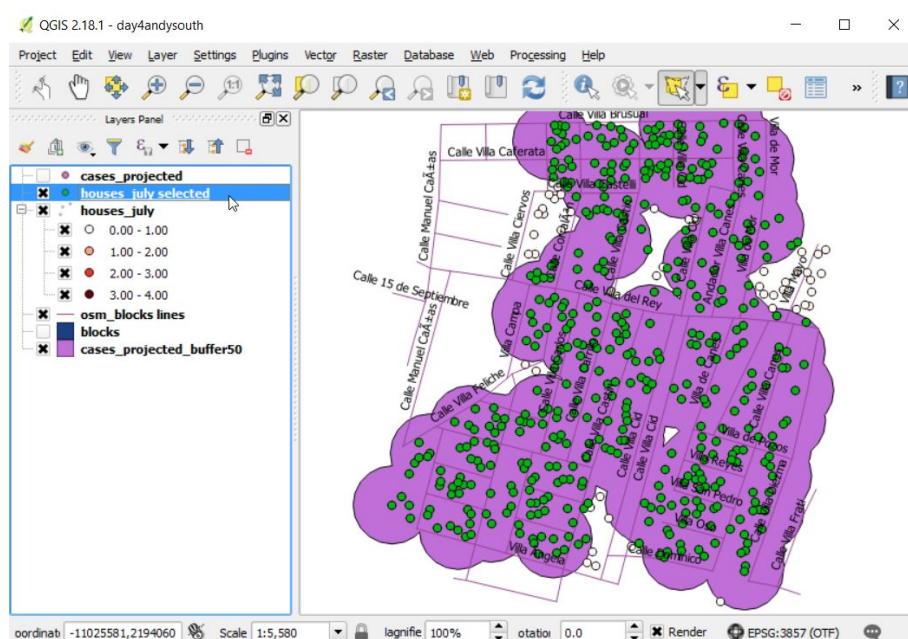
The window should show you that 453 of the 500 houses are selected as being within the buffer :



Hint: If your number of houses is different, you may be working on a selection of houses. Are there yellow points shown? Go back and unselect everything then repeat the above steps.

➔ click the yellow button by 'Create new selection' at the lower left

It will create a new layer and add to the map. You can then close this window. You should see houses_july_selected has been added as a layer in your project :



In a vector control operation a map like this with a 150m buffer could be used to target control interventions.

→ To finish, save the project so that you can return to it later.

This was a brief example of some of the vector analyses that can be done. There are many more. Whilst these analyses can be hugely useful, it is also true that QGIS can be hugely useful for displaying maps even if you don't get to use the analysis tools.

4.3. Raster tools

As explained in section 1.2.2, raster data in GIS represents features above, on, or below the earth's surface, for example to map data from satellites. Raster data takes the form of a grid of cells also referred to as pixels. The cells/pixels have coordinates associated with them which position them within the map canvas. The cells/pixels also have data associated with them, consisting of one value per cell/pixel with different values being represented by different colours. For example in Figure 4.7 below we see elevation represented as a raster, where each cell/pixel contains a value representing the elevation at that location, plus a land usage raster where each cell contains a value associated with a land usage category. In this course we will use elevation and climate data as examples to explore some of the tools available for manipulating raster data.

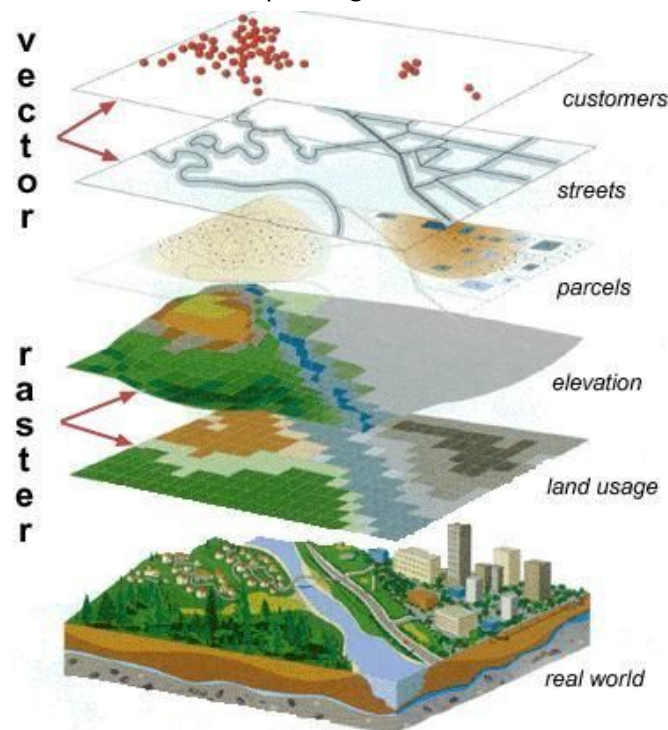


Figure 4.7: Examples of different vector and raster layers

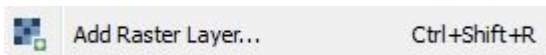
In this section we will be demonstrating raster tools using climate data obtained from WorldClim (1km resolution data). Whilst you may want to download these off the internet yourself, we have also included the data being used in this practical session on your pen drive. The files on your pen drive have been edited since being downloaded to cover Colombia only. We will be showing you

how to extract raster data for specific coordinates. We will do this using the *Aedes* occurrence data we introduced in Section 4.1.

→ open day4raster.qgs (from your materials/day4 folder on your USB)

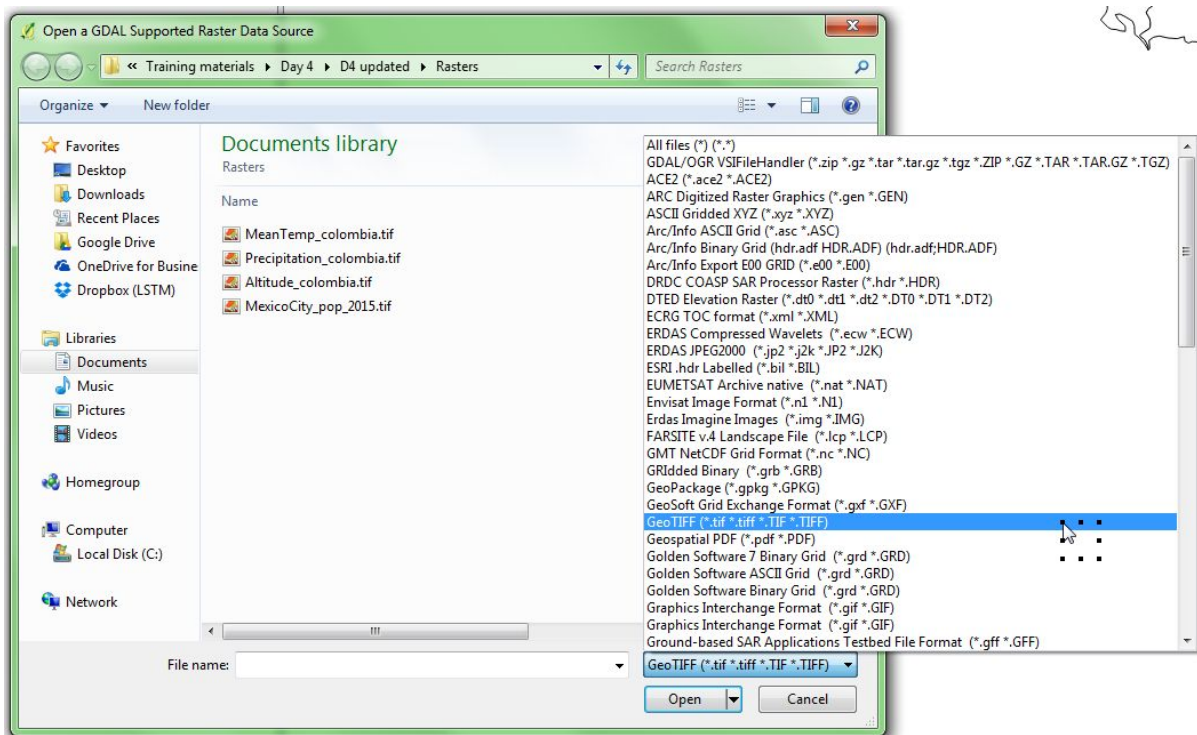
This is a project ready for us to add raster data. It already contains some point records for *Aedes* mosquitoes in Colombia (aedes_colombia), extracted from the database described in Section 4.1 and world country boundaries downloaded from NaturalEarth (ne_50m_admin_0_countries).

4.3.1. Loading raster data

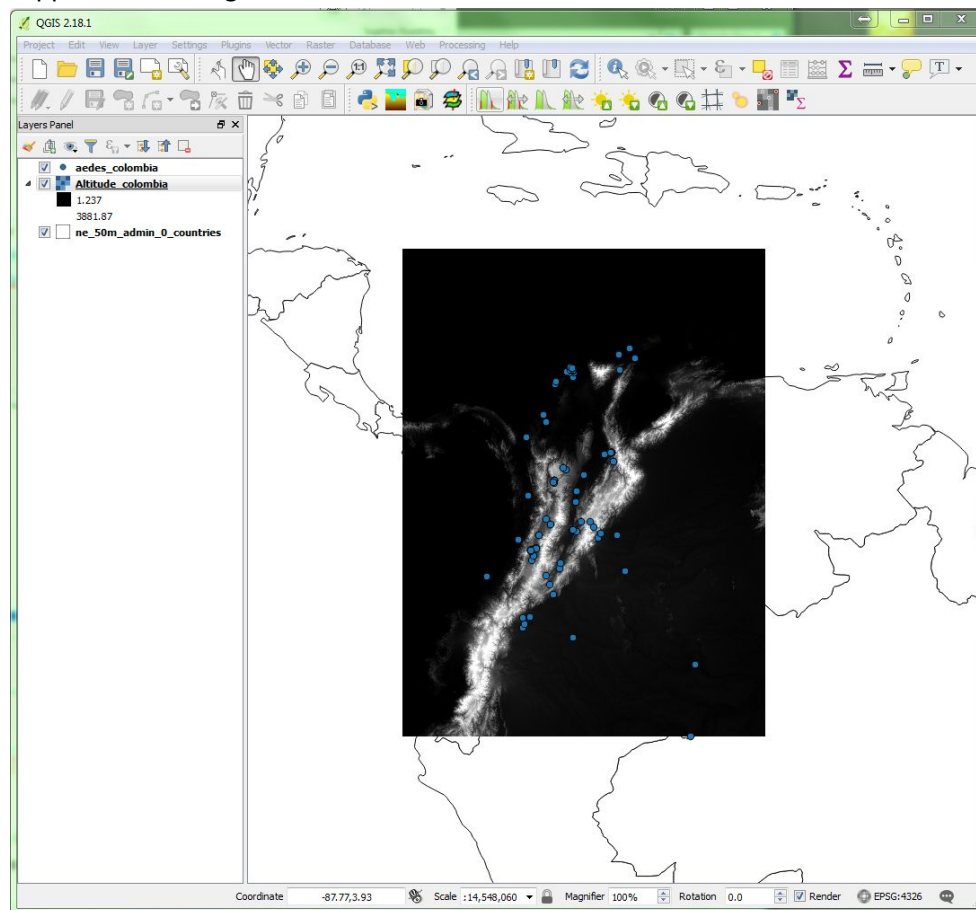


→ To add raster data to your map, simply use the “Add Raster Layer” option under “Layer” in the menu bar. By default this should take you to a window “Open a GDAL Supported Raster Data Source”, where you can navigate to your folder containing the raster data (data/spatial/raster).

Raster data comes in many different formats. QGIS uses something called GDAL to read and write raster data. GDAL stands for Geospatial Data Abstraction Library. It supports more than 100 formats of raster data. In this course we have stored the raster data as GeoTIFF files. The files that you therefore need to add into QGIS have the extension .tif. You can filter the file names by using the dropdown menu above the Open button at the bottom of the window, change it from [GDAL] All files to GeoTIFF (*.tif, *.tiff, *.TIF, *.TIFF) to narrow down your selection. Select “Altitude_colombia.tif” and click “Open” to add this to your map.

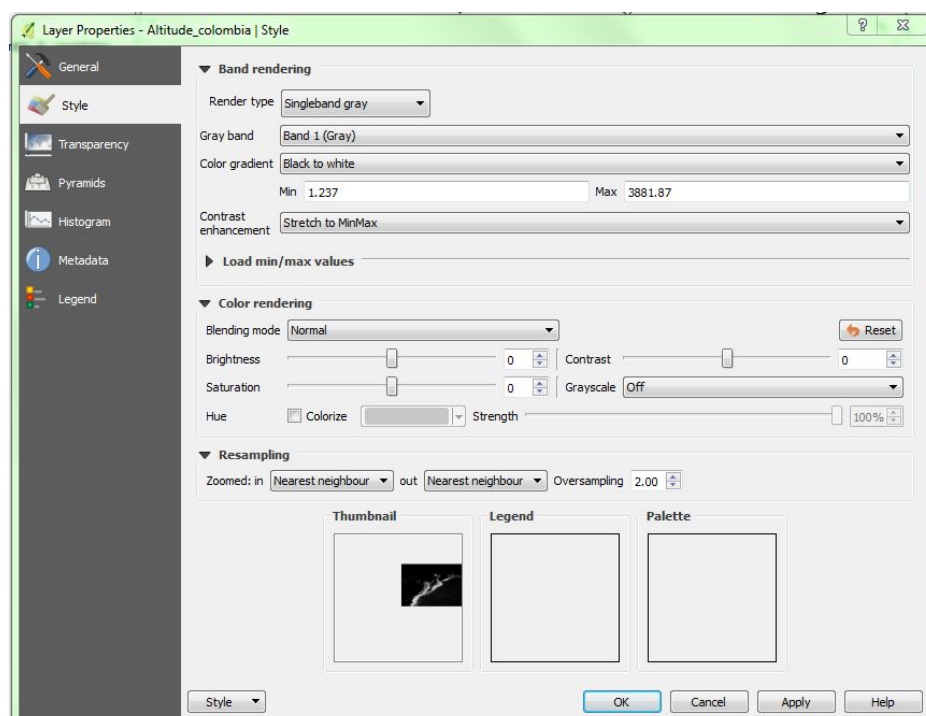


It will first appear something like this:



4.3.2. Raster layer properties

Similar to editing vector layers, you can alter the properties of the raster file via the “Properties” option. (Menu bar: “Layer”, “Properties”, or in the Layers panel: right click on your layer and then select properties, or double-click [left] on your layer). This will open a window with various options, we will just look at band rendering for now.



Band rendering

Band rendering is an option under “Style”. What does it mean? The word ‘band’ in this context was originally used in reference to bandwidths of colour, and ‘rendering’ their interpretation or representation. A band can now be thought of more generally as a characteristic.

We have seen in our example that our raster layer has appeared as a black and white image. We can see in the layer properties window that this corresponds to a ‘singleband gray’ render type, with a colour gradient from black to white. Single band refers to the data associated with the raster cells representing one single characteristic. This is the case with our raster layer, which represents altitude only. If you zoom in and look at this layer, you can get an impression of altitude, but this is not how we are used to seeing elevation represented. It is not immediately intuitive. A more common use of this style is perhaps black and white photographs.

- change the direction of the colour gradient, to go from being black to white, to white to black.
- click ‘Apply’ to action any changes made to the style properties.
- go back and adjust the properties of the other elements in your map if necessary (for example if a black country outline does not show up well on a black raster layer you could change it to a brighter colour).
- if you cannot see everything you expect, to check the order of your layers in the layers panel.

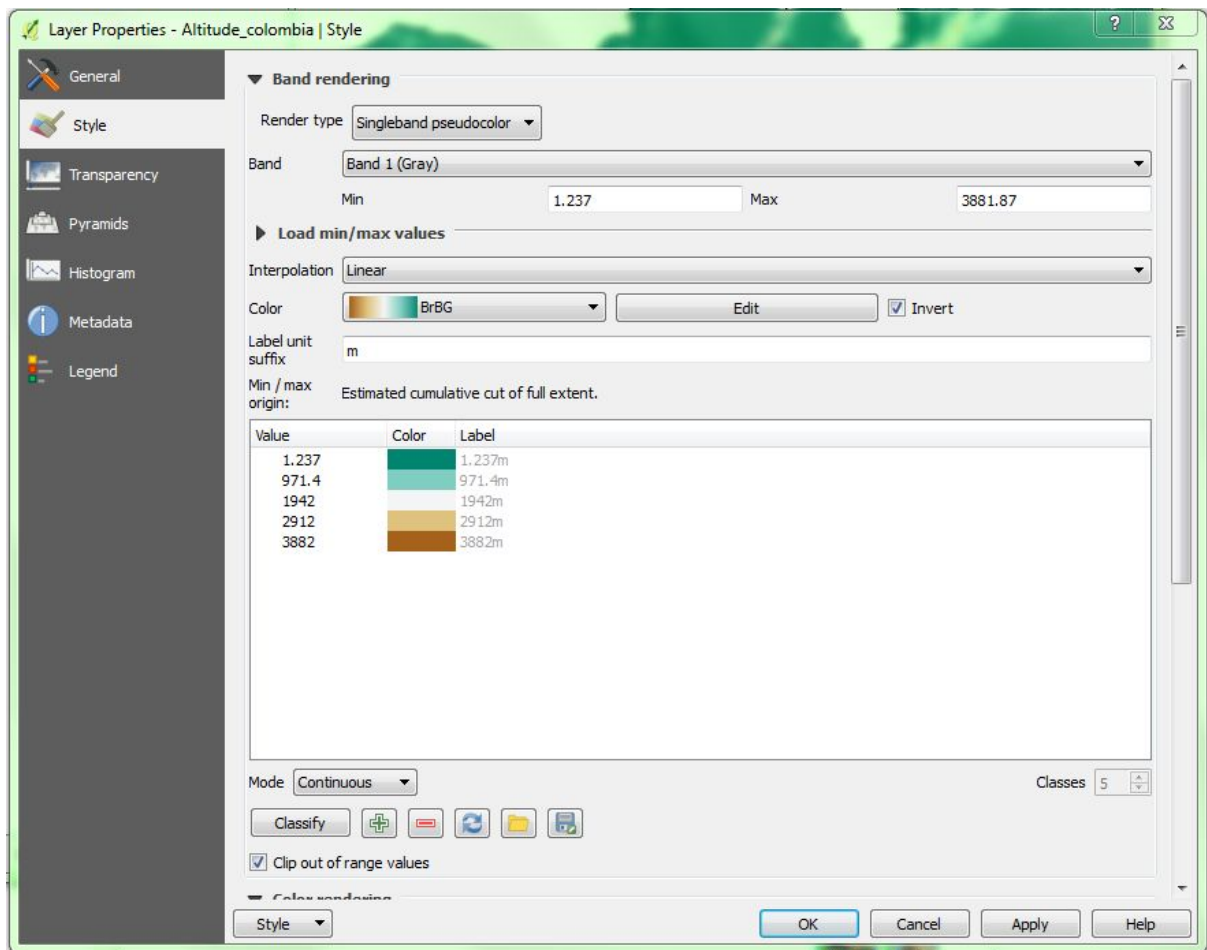
Another single band option is “singleband pseudocolor”. This will give you the option of various colour ramps to apply.

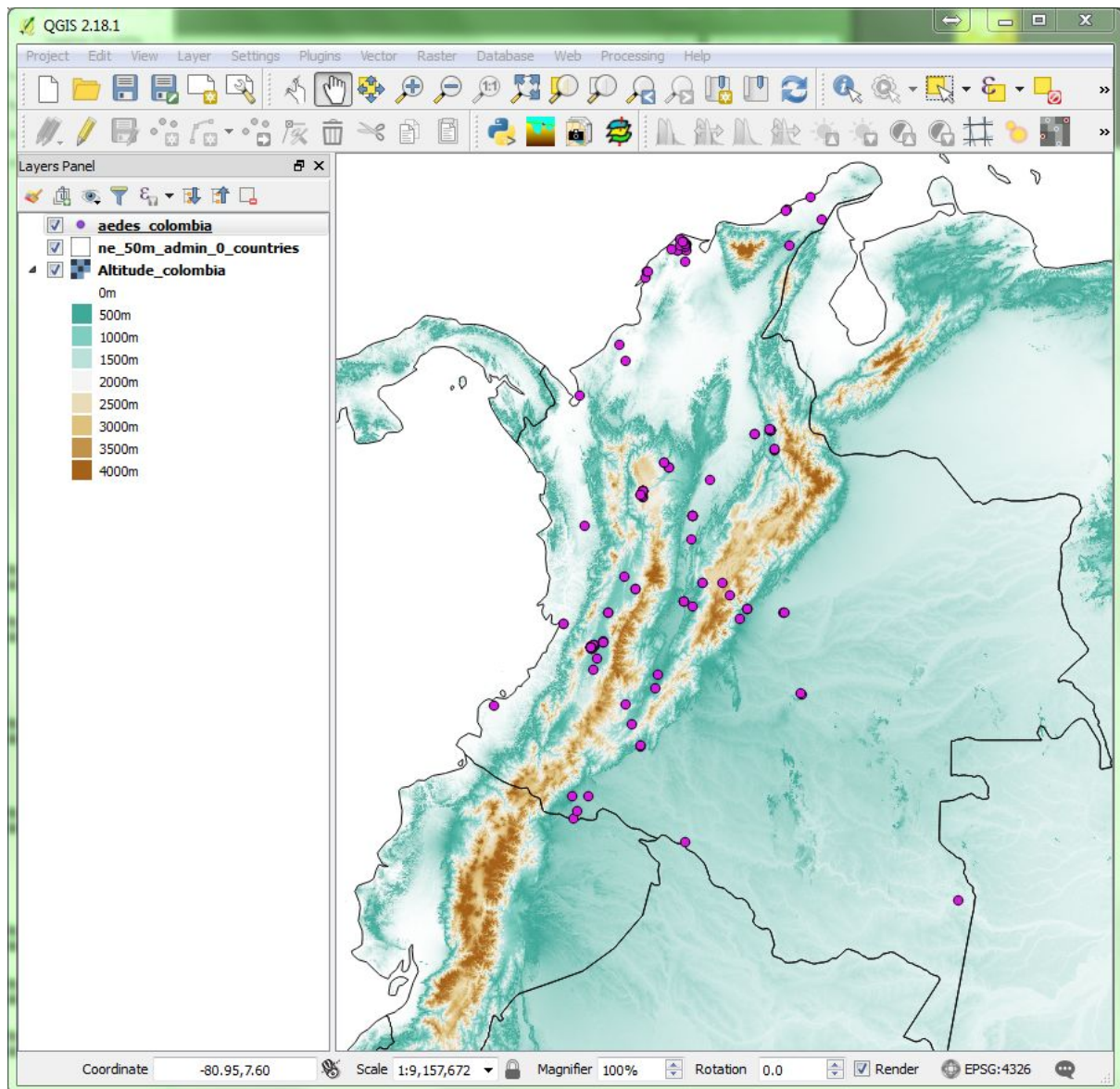
- try BrBG (brown to green), invert it green to brown.
- there is a check box option to ‘clip out of range values’.

The minimum and maximum values remain 1.237 and 3881.87 respectively, as the values associated with the cells of your raster layer. Think about what these values mean, in this example they are altitude figures.

- there is an option to label the units, for example ‘m’ for metres.

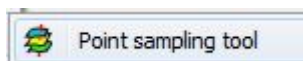
Note that the divisions are classified according to the mode and number of classes selected at the bottom of the panel. These can be changed manually if this is more meaningful, just click and type. The colours can also be adjusted manually by double clicking on them.





4.3.3. Extraction of raster data (point sampling tool)

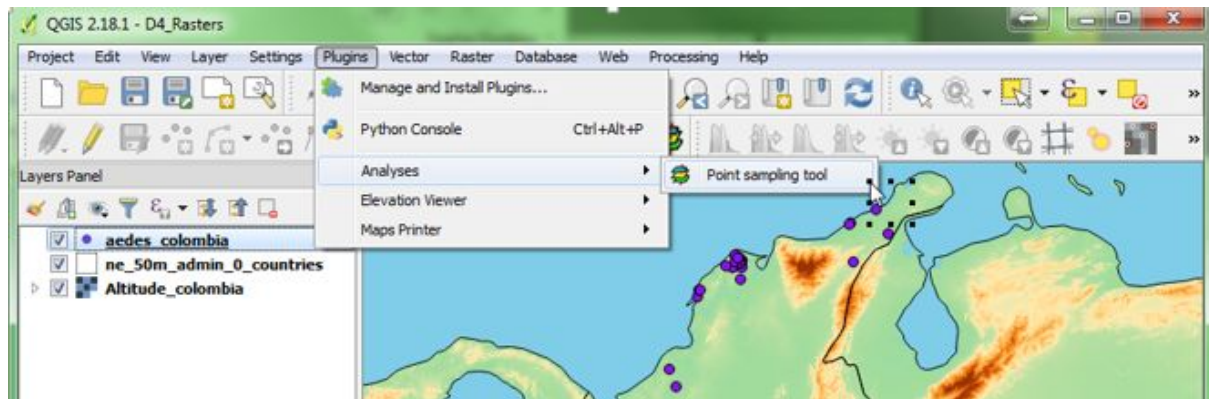
In addition to enhancing the look of your map, it is also possible to access the data associated with the cells/pixels to perform analyses. Continuing on with our example of altitude, it could be interesting to know what elevations the mosquitoes were found at. There is an altitude value associated with each cell in the grid of raster cells. What we want to do in this case is to extract the values which coincide with location points of the entomology data. To do this we will use the “Point sampling tool”.



The tool comes in the form of a plugin, as introduced earlier. If the point sampling tool is not already installed in your QGIS, you will need to install it.

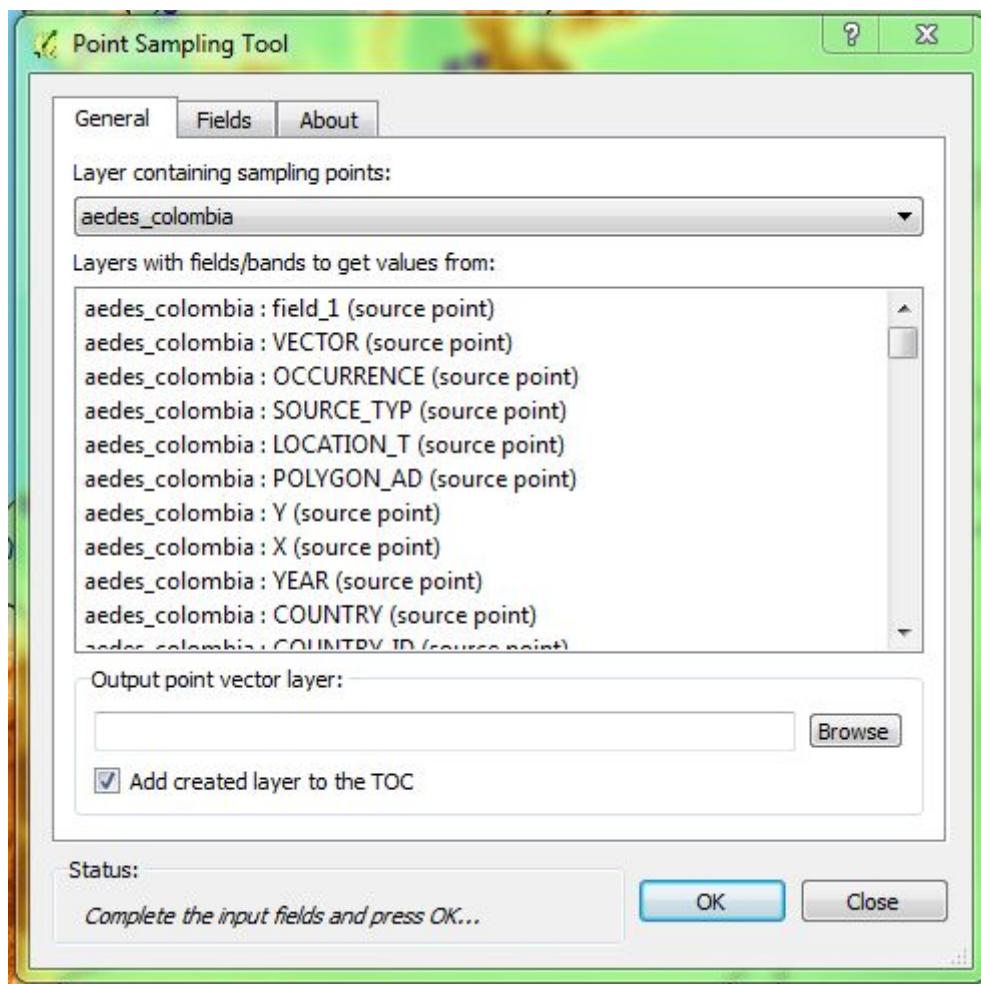
- To install a plugin go to: Plugins, Manage and Install Plugins, type to search for 'point sampling tool' and select the plugin that is returned from the list
- click on Install Plugin.

Once it is installed, you will find it under Plugins in the menu bar, under Analyses, or as a shortcut if you have that toolbar displayed.

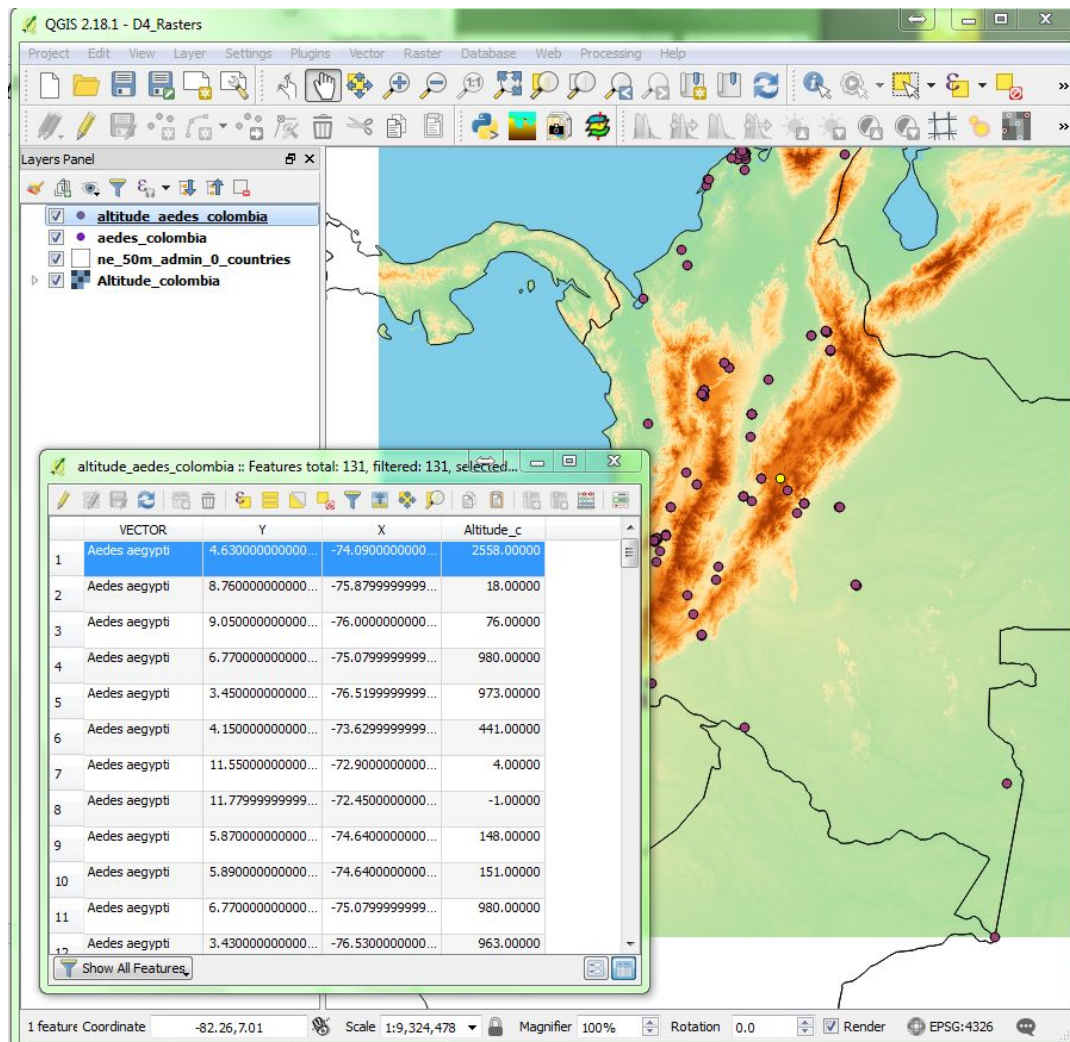


In the Point Sampling Tool window that appears:

- Our first choice will be to specify which layer the sampling points are to come from. In this example our points are the `aedes_colombia` layer, the vector layer which has the locations of the mosquitoes of interest.
- During this step we will be creating a new point layer, so within the Point Sampling Tool we also need to select which fields we wish to include in this new layer, in addition to the extracted altitude. Select the X and Y coordinates and the vector name field from the vector layer, plus the altitude raster layer. Note that you can use Ctrl+click or Shift+click to select multiple fields.
- Lastly, browse to where you would like to save the new shapefile that will be created and give it an appropriate name (for example `Altitude_Aedes_Colombia`).
- Note there is a tick box 'add created later to the TOC' (Table of Contents), which is ticked by default. Click OK.
- The status of processing is displayed at the bottom of the tool window. When it confirms 'OK. The new layer has been added to the TOC, you can then close the tool window.

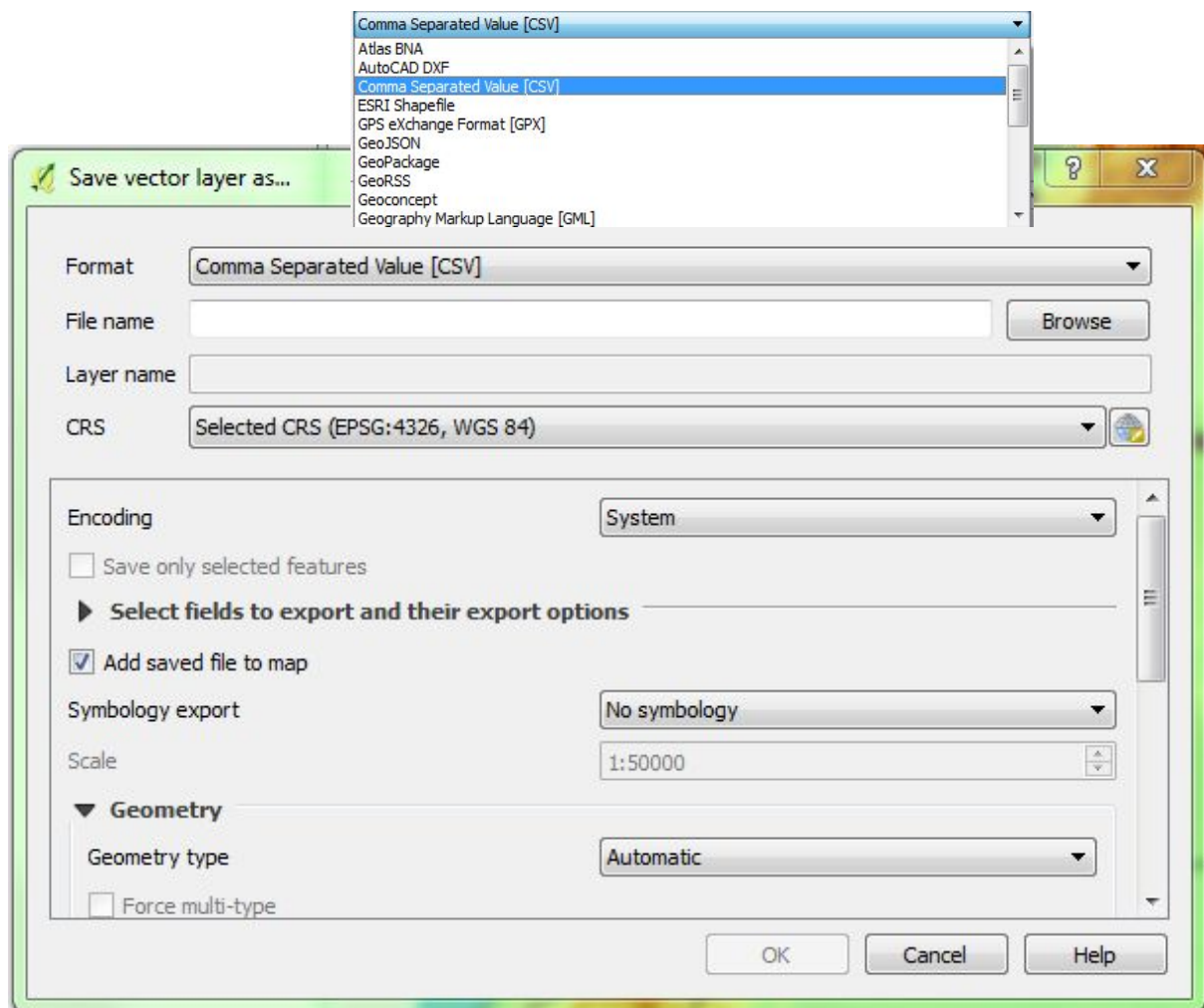


Look at the attribute table of your new layer to see the results of your extraction. Try selecting a line in the table and see if the highlighted point is in an area you would expect according to the background colour of your band rendering.



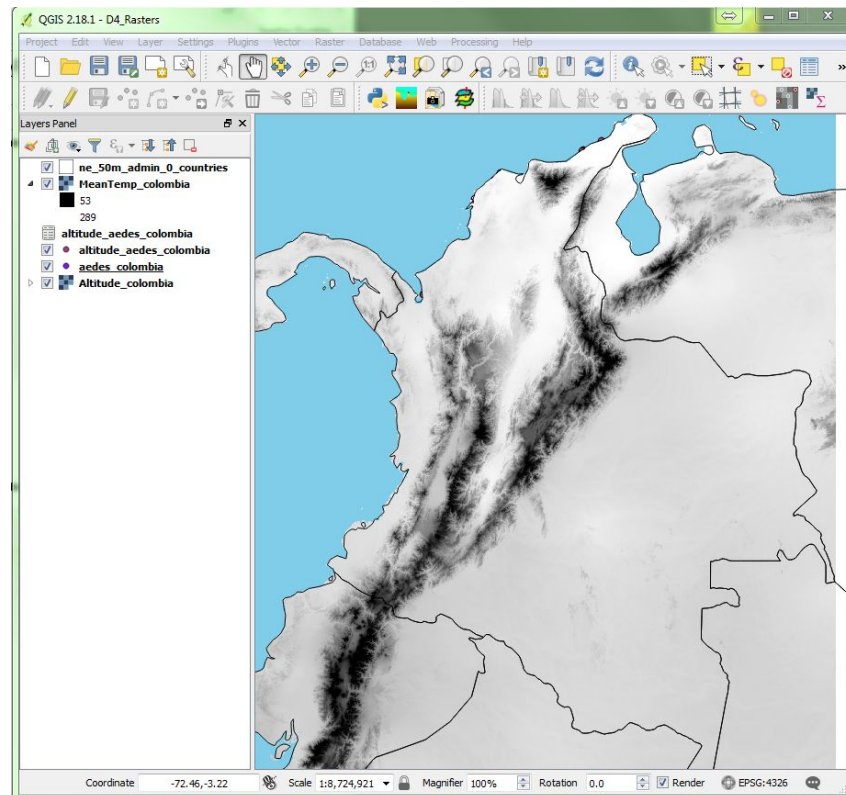
This tool can be used successively to extract information from different raster files and compile it in one shapefile. For example, once you have the altitude at your points of mosquito presence, you could then extract data from other raster layers on temperature or rainfall in the same way and add these values to the same shapefile.

Should you then wish to have this data in Excel, it is possible to save the attribute table as a CSV file. To do this, simply right-click on the layer in question, choose “Save As” and name the file using the “Save vector layer as” window pictured here, browsing to the desired file location and selecting Comma Separated Value [CSV] as your desired format from the drop down menu at the top. Once saved, check in the documents folder where you saved it and you should see that an Excel CSV file has been created. Note that if the “Add saved file to map” box was checked, you will also see the new csv file as a layer in the layers panel.



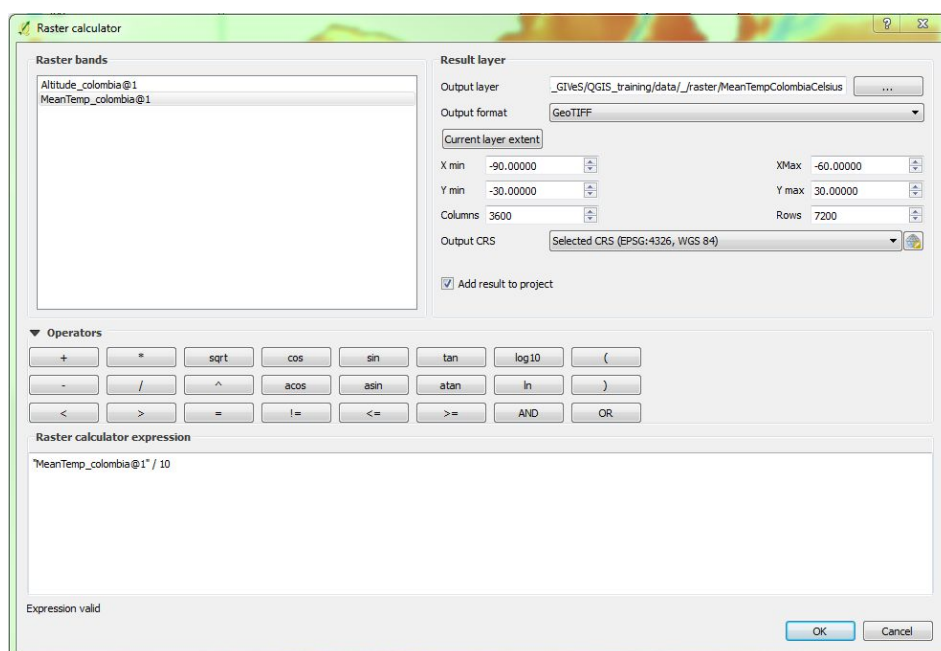
The same procedures can be repeated to display and extract information from other raster layers. For example, we have the file “data/spatial/raster/MeanTemp_colombia.tif”, which contains the mean temperatures for Colombia in June.

Added to the map, it will first look something like the below, a black and white image as before. By changing the colour ramp in the band rendering, the warmer and colder regions can quickly be appreciated.

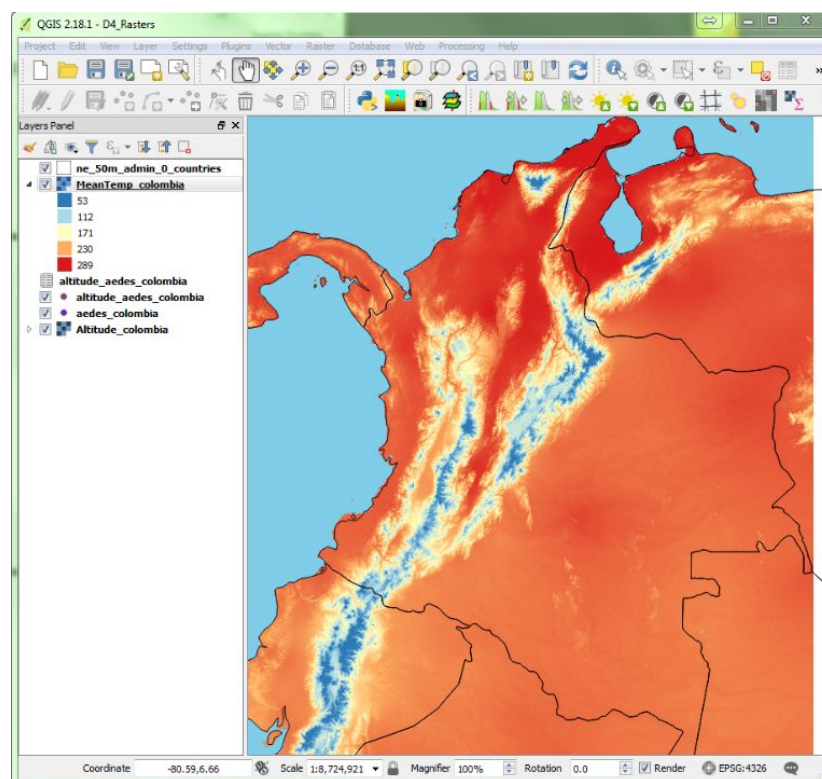


Optional: Using the raster calculator

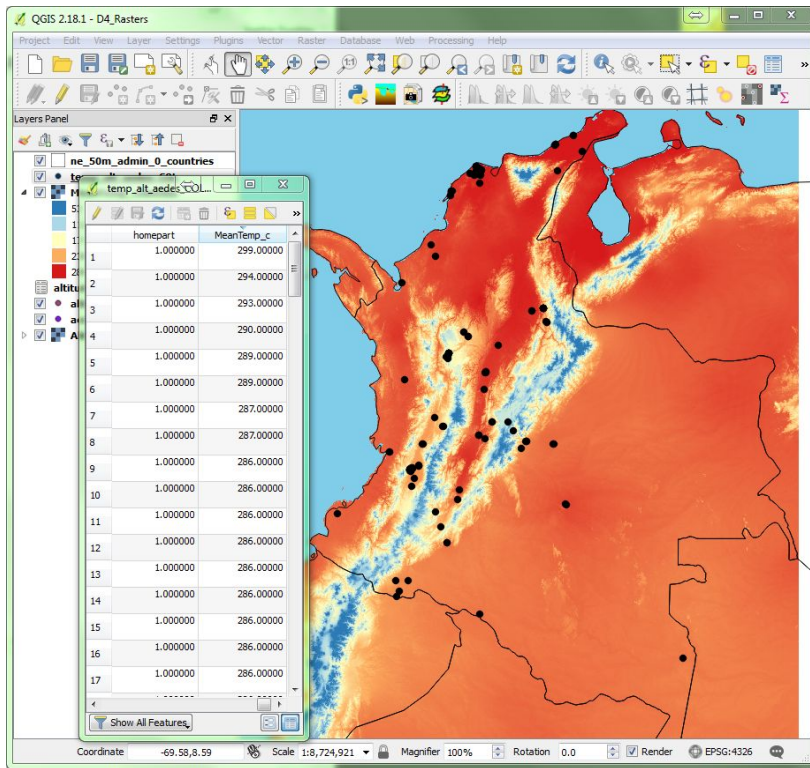
The temperature data downloaded from WorldClim are in Celsius (°C) multiplied by ten (to allow for a reduced file size). To avoid confusion when interpreting this data later on, it is possible to convert the temperature figures to Celsius. This can be done for the entire raster file using the “Raster Calculator” (Menu bar: Raster, Raster Calculator).



- When you open up the Raster calculator window, the available raster bands will be displayed in the raster bands section on the left. We have two raster layers open currently, altitude and temperature.
- Double-click on the band you want to manipulate, and it will be added to the calculator expression window below.
- You can then use the operator buttons available, or type in your desired calculation in this raster calculator expression window. In our case we want to simply divide by ten (/10) . The formula will read “MeanTemp_colombia@1”/10.
- Now specify your output layer, browse to your raster data file (QGIS_training, data, spatial, raster), and give your new file the name “MeanTemp_colombia_celcius”.
- Hit Save and Ok. A black and white image will be added to your project.
- The properties of this can be changed by right clicking on the layer and entering the Style section as before.



By using the point sampling tool, the temperatures for our entomological locations can also be extracted. Note that you will now have the option to add them to your file containing the altitude data. Once successfully extracted, you will be able to visualise them in the attributes table of your new shapefile.



This data can also be saved as a CSV file as before, to be viewed in Excel.

temp_alt_aedes_COL.csv [Read-Only] - Excel

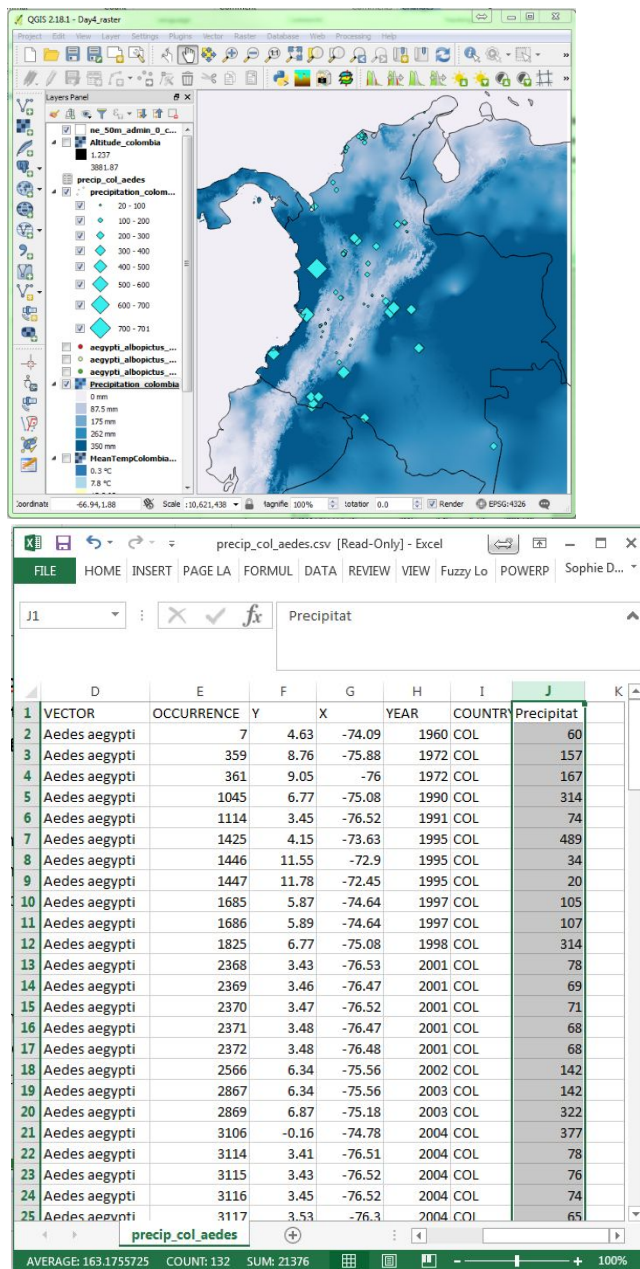
	A	B	C	D	E	F	BR	BS
1	X	Y	VECTOR	Y	X	Altitude_c	MeanTemp_c	
2	-74.09	4.63	Aedes aegypti	4.63	-74.09	2558	135	
3	-75.88	8.76	Aedes aegypti	8.76	-75.88	18	274	
4	-76	9.05	Aedes aegypti	9.05	-76	76	270	
5	-75.08	6.77	Aedes aegypti	6.77	-75.08	980	232	
6	-76.52	3.45	Aedes aegypti	3.45	-76.52	973	238	
7	-73.63	4.15	Aedes aegypti	4.15	-73.63	441	246	
8	-72.9	11.55	Aedes aegypti	11.55	-72.9	4	294	
9	-72.45	11.78	Aedes aegypti	11.78	-72.45	-1	299	
10	-74.64	5.87	Aedes aegypti	5.87	-74.64	148	278	
11	-74.64	5.89	Aedes aegypti	5.89	-74.64	151	277	
12	-75.08	6.77	Aedes aegypti	6.77	-75.08	980	232	
13	-76.53	3.43	Aedes aegypti	3.43	-76.53	963	239	
14	-76.47	3.46	Aedes aegypti	3.46	-76.47	947	239	
15	-76.52	3.47	Aedes aegypti	3.47	-76.52	972	238	
16	-76.47	3.48	Aedes aegypti	3.48	-76.47	947	240	
17	-76.48	3.48	Aedes aegypti	3.48	-76.48	952	240	
18	-75.56	6.34	Aedes aegypti	6.34	-75.56	1475	216	
19	-75.56	6.34	Aedes aegypti	6.34	-75.56	1475	216	
20	-75.18	6.87	Aedes aegypti	6.87	-75.18	709	244	
21	-74.78	-0.16	Aedes aegypti	-0.16	-74.78	191	253	
22	-76.51	3.41	Aedes aegypti	3.41	-76.51	953	240	

temp_alt_aedes_COL

Exercise 4.3

To practice what you have learned in this section:

- Add the provided raster file of Colombia's rainfall (Precipitation_colombia.tif) to the map.
- Change the style in keeping with this data to visualise the information it contains.
- Extract the rainfall figures for the *Aedes* occurrence points in Colombia.
- Save the resulting attribute table as a CSV file.



4.3.4. Summarise raster data (zonal statistics)

We have seen that we can extract raster data from a given point. Now we will look at an example of what can be done for data associated with polygons.

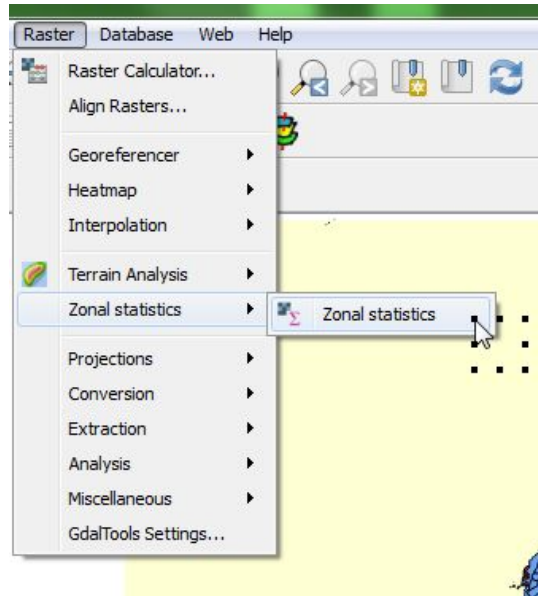
For example, imagine that instead of knowing the altitude or rainfall for a given mosquito collection (point), we want to know the conditions for that whole region or district (a polygon). The area of a polygon will contain more than one pixel, and therefore have more than one number associated with it. To return a value/s for a whole district for example, therefore requires the data to be summarised in some way.

There is another plugin called **“Zonal statistics”**. This is another raster tool, which as its name suggests, can perform calculations for a given area (e.g. a polygon).

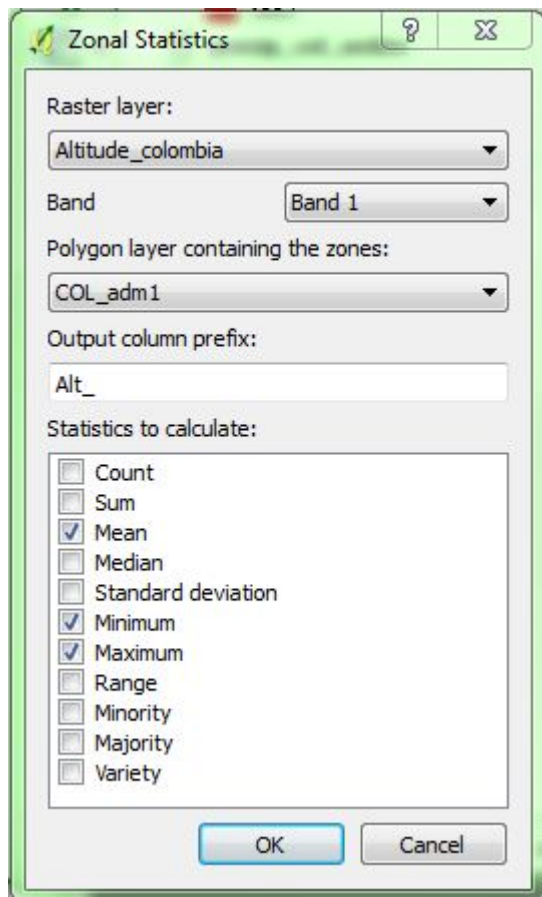
For our Colombia example, add the vector layer COL_adm1 (from your USB training_QGIS/data/spatial/polygons). These will be the administrative polygons we use to summarise the raster data by, the 32 Departments of Colombia.



Zonal statistics should already be installed as a ‘core plugin’ of your system. Find it via the menu under Raster, Zonal Statistics, Zonal statistics, or look for the icon in your raster toolbar. If for some reason it is not installed, use the Manage and Install plugins option in the plugins menu to search for and install the tool.



If you open the tool, you will be taken to the Zonal Statistics window (see below). This shows you the list of statistics available to calculate in a series of check boxes. You will also see the other options you will need to specify.



The tool in itself is quite simple to use, but you will need to think about what calculations will actually be of added value to you. There are some statistics selected by default for example, but it may not make sense to apply these to your raster data. We need to decide what it is that we want to know for each of the Departments.

Let's return to our altitude data as an example. This file has altitude values for each cell/pixel of the raster grid. Does it make sense to count these? Not really. Sum them? Also no. Knowing the minimum and maximum altitude for each department could be informative. You could also add the mean.

Once you know what we want to know, you can go ahead and do it.

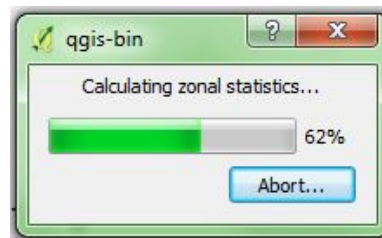
- select your raster layer from the drop-down options available.
- The next option is band selection. Our raster layer altitude has one single band (characteristic), altitude, therefore there is only one option here, Band 1.
- select the polygon layer, COL_adm1.

There is then space to type an "Output column prefix". The statistics generated by the tool will be added to the attribute table of the vector layer, with the statistic or an abbreviation of the statistic as a column header. Use the prefix option to make the headings more specific and more meaningful. Imagine you may want to come back to this information as a variable in the future and know what it

represents, or that you plan to calculate multiple statistics for the same admin layer – how will you know which is which?

- As we are looking at Altitude, type “Alt_” as the output column prefix.
- Next tick and untick the boxes to select which calculations to perform.
- Click OK.

QGIS may show the progress of the calculations with a small message box. This will disappear when they are complete.



- To see the results of the calculation, open the attribute table of your polygon layer, COL_adm1. The columns will have been added to the right-hand side of the table.

	TYPE_1	ENGTYPE_1	_NAME_	_RNAME_	Alt_mean	Alt_min	Alt_max
1	Comisar�a	Commissiary			145.2386744899...	48.0000000000...	342.0000000000...
2	Departamento	Department			990.2173924727...	-1.0000000000...	3824.0000000000...
3	Intendencia	Intendancy			373.9143135019...	74.0000000000...	5083.0000000000...
4	Departamento	Department			48.66818700114...	-1.0000000000...	431.0000000000...
5	Departamento	Department			272.4508180755...	-3.0000000000...	2186.0000000000...
6	Departamento	Department			2266.368654706...	119.0000000000...	5147.0000000000...
7	Departamento	Department			166.9436959104...	-1.0000000000...	2053.0000000000...
8	Departamento	Department			1586.644946357...	151.0000000000...	4471.0000000000...
9	Intendencia	Intendancy			354.3196935795...	88.0000000000...	3482.0000000000...
10	Intendencia	Intendancy			259.5121876788...	84.0000000000...	3822.0000000000...
11	Departamento	Department			1500.003719431...	-6.0000000000...	5185.0000000000...
12	Departamento	Department		El Cesar	470.8770142005...	-5.0000000000...	4995.0000000000...
13	Departamento	Department			219.7355777364...	-7.0000000000...	3248.0000000000...
14	Departamento	Department			1892.242233212...	149.0000000000...	4081.0000000000...
15	Comisar�a	Commissiary		Guania	151.2527847378...	76.0000000000...	751.0000000000...
16	Comisar�a	Commissiary			230.1053133117...	140.0000000000...	796.0000000000...

Remember that should you want to look at this data in Excel, you can use the “save vector layer as” option to generate a CSV file as before.

	H	I	J	K	L	M	N	O	P
1	CCA_1	TYPE_1	ENGTYPE	NL_NAME	VARNAME	Alt_mean	Alt_min	Alt_max	
2		ComisarA	Commissiary			145.2387	48	342	
3		Departam	Department			990.2174	-1	3824	
4		Intendenc	Intendancy			373.9143	74	5083	
5		Departam	Department			48.66819	-1	431	
6		Departam	Department			272.4508	-3	2186	
7		Departam	Department			2266.369	119	5147	
8		Departam	Department			166.9437	-1	2053	
9		Departam	Department			1586.645	151	4471	
10		Intendenc	Intendancy			354.3197	88	3482	
11		Intendenc	Intendancy			259.5122	84	3822	
12		Departam	Department			1500.004	-6	5185	
13		Departam	Department		El Cesar	470.877	-5	4995	
14		Departam	Department			219.7356	-7	3248	
15		Departam	Department			1892.242	149	4081	
16		ComisarA	Commissiary		Guania	151.2528	76	751	

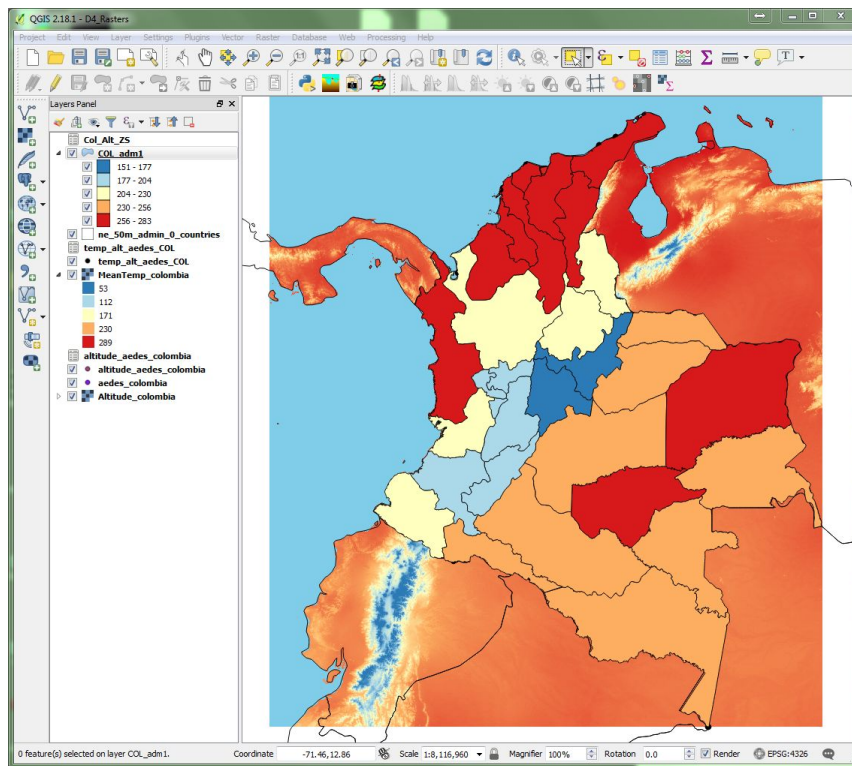
We can use the temperature file as another example. Note that the temperature data we have for Colombia is already in the form of a statistic, it is the mean temperature per grid cell for June. Bear this in mind when planning and interpreting your data. For example, should you select the minimum and maximum statistics as an option, know that you will not get the actual minimum and maximum temperatures for the given area, but the lowest and highest mean temperatures. The tool can only work with the data it has available. It is therefore important to know your dataset and understand what calculation you are asking the tool to perform.

To summarise the raster temperature data, simply follow the same steps as before for altitude. Note the importance of the output prefix when performing statistics for more than one dataset for the same area.

Remember that once you have this data in the attribute table of your chosen vector layer, you can also display these summary statistics by polygon by entering into the style section of layer properties.

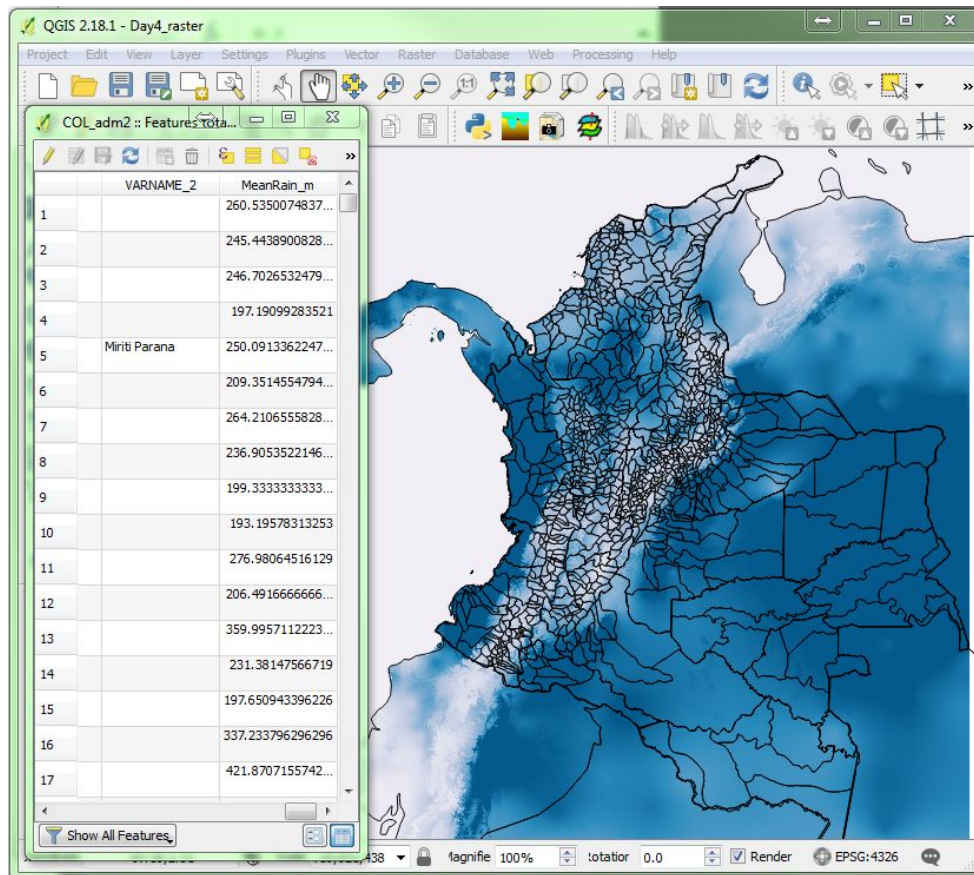
For example, below is a map of mean temperature per department, displayed with a red-yellow-blue colour ramp. The divisions used for the temperature categories were arbitrary. Your map will only be as meaningful as you make it with your choices. Think about what might be useful operationally.

Think about what you are trying to show in the end. For example, what relationship does your climate data have to your entomological data?



Exercise 4.4

Use the zonal statistics tool to calculate the mean June rainfall for the Admin Level 2 for Colombia. The COL_adm2 shapefile is available in your USB (training_QGIS/data/spatial/polygons).



This is the end of what we wanted to cover today, well done for getting this far. In case there was extra time available we have added an extra supplementary section on viewing international data.

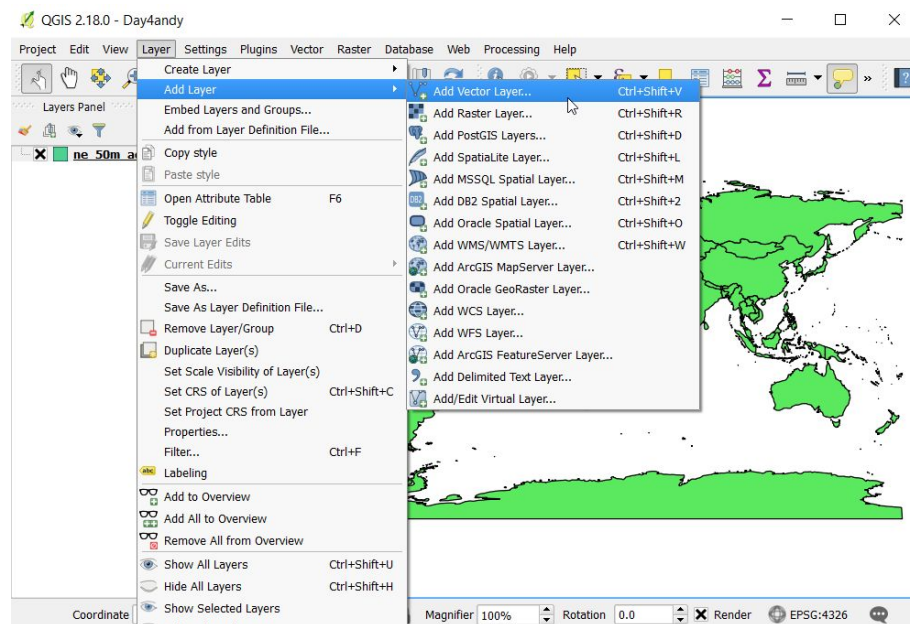
Optional Section

In this optional section we will look at manipulating data over a large geographical scale. We will be using the same or similar techniques that we introduced in previous sections (e.g. selection tools), but will be demonstrating them on a different dataset. Start by adding country boundaries. These can be downloaded for free from the Natural Earth Data (<http://www.naturalearthdata.com/>) website or the Global Administrative Areas (GADM) website (<http://gadm.org/>). In both of these resources, administrative level 0 is an international code for country level boundaries, which are subdivided into level 1 (states, regions etc) and further subdivided into level 2 etc. For example, in Colombia, administrative level 1 boundaries are 'departamentos' and administrative level 2 boundaries are 'municipios'.

We have downloaded the shapefiles for the country boundaries of the world from Natural Earth Data (www.naturalearthdata.com/downloads/50m-cultural-vectors/), and have included them in the teaching data folder. This file is in the WGS84 geographic coordinate system.

\Data\Spatial data\Polygons\ne_50m_admin_0_countries.shp

We shall now add this layer into QGIS. As in previous days, from the top menu bar you should select 'Layer, Add Layer, Add Vector Layer' and browse to find the file. This should bring up a map of the world :



Right click on the layer name within the Layer Panel and select 'Open Attribute file'. This brings up the attribute data for the map. Scroll down to the bottom and you will see that there are 241 countries with 1 row for each country, look at the columns to see the values that are stored for each country.

ne_50m_admin_0_countries :: Features total: 241, filtered: 241, selected: 0

	scalerank	featurecla	labelrank	sovereignty	sov_a3	adm0_dif	level	type	admin	adm0_a3
224	1	Admin-0 coun...	3.000000	Ukraine	UKR	0.000000	2.000000	Sovereign cou...	Ukraine	UKR
225	1	Admin-0 coun...	3.000000	Ukraine	UKR	0.000000	2.000000	Sovereign cou...	Ukraine	UKR
226	1	Admin-0 coun...	4.000000	Uruguay	URY	0.000000	2.000000	Sovereign cou...	Uruguay	URY
227	1	Admin-0 coun...	2.000000	United States ...	US1	1.000000	2.000000	Country	United States ...	USA
228	1	Admin-0 coun...	3.000000	Uzbekistan	UZB	0.000000	2.000000	Sovereign cou...	Uzbekistan	UZB
229	6	Admin-0 coun...	6.000000	Vatican	VAT	0.000000	2.000000	Sovereign cou...	Vatican	VAT
230	1	Admin-0 coun...	6.000000	Saint Vincent ...	VCT	0.000000	2.000000	Sovereign cou...	Saint Vincent ...	VCT
231	1	Admin-0 coun...	3.000000	Venezuela	VEN	0.000000	2.000000	Sovereign cou...	Venezuela	VEN
232	3	Admin-0 coun...	6.000000	United Kingdom	GB1	1.000000	2.000000	Dependency	British Virgin I...	VGB
233	3	Admin-0 coun...	6.000000	United States ...	US1	1.000000	2.000000	Dependency	United States ...	VIR
234	1	Admin-0 coun...	2.000000	Vietnam	VNM	0.000000	2.000000	Sovereign cou...	Vietnam	VNM
235	1	Admin-0 coun...	4.000000	Vanuatu	VUT	0.000000	2.000000	Sovereign cou...	Vanuatu	VUT
236	3	Admin-0 coun...	4.000000	France	FR1	1.000000	2.000000	Dependency	Wallis and Fut...	WLF
237	1	Admin-0 coun...	4.000000	Samoa	WSM	0.000000	2.000000	Sovereign cou...	Samoa	WSM
238	1	Admin-0 coun...	3.000000	Yemen	YEM	0.000000	2.000000	Sovereign cou...	Yemen	YEM
239	1	Admin-0 coun...	2.000000	South Africa	ZAF	0.000000	2.000000	Sovereign cou...	South Africa	ZAF
240	1	Admin-0 coun...	3.000000	Zambia	ZMB	0.000000	2.000000	Sovereign cou...	Zambia	ZMB
241	1	Admin-0 coun...	3.000000	Zimbabwe	ZWE	0.000000	2.000000	Sovereign cou...	Zimbabwe	ZWE

Show All Features

Global *Aedes aegypti* and *Aedes albopictus* occurrence data have been obtained from a review published in 2015 by Kraemer et al. (<http://www.nature.com/articles/sdata201535>). These data

consist of all known occurrences between 1960-2014, with coordinates measured in longitude and latitude, and the data can be freely downloaded here <http://datadryad.org/resource/doi:10.5061/dryad.47v3c>. We will now add this data to our map. Please keep in mind that the data represent all known occurrences. The dataset does not contain any information on locations where surveys were conducted and no *Aedes* were found, therefore we cannot be certain that the absence of data in the dataset represents absence of *Aedes*. We therefore need to take care when interpreting these data, and the resulting maps.

The data are downloaded as a .csv file which has > 42,000 rows representing sites across the globe. Variables in the csv include the vector species identifies, the source of the data, the year the vector was identified, the country it was identified in and the longitude (X) and the latitude (Y) of the location where it was identified. The complete file is very large (>3MB), and as such may cause an error when you try to add it into QGIS. We therefore added an extra continent column, selected only those countries within the Americas, and saved the resulting file to your data folder as :

\Data\Entomology data\Presence data\aegypti_albopictus_americas.csv

We shall now add this CSV into QGIS.

Within QGIS select 'Layer, Add Layer, Add Delimited Layer'

Browse to

\Data\Entomology data\Presence data\aegypti_albopictus_americas.csv

Note that the X & Y fields under *Geometry definitions* are detected automatically so after checking these default options you can press OK.

Create a Layer from a Delimited Text File

File Name: C:/Dropbox/QGIS Training/Data/Entomology data/Presence data/aegypti_albopictus_americas.csv Browse...

Layer name: aegypti_albopictus_americas Encoding: UTF-8

File format: ☒ CSV (comma separated values) ☐ Custom delimiters ☐ Regular expression delimiter

Record options: Number of header lines to discard: 0 ☒ First record has field names

Field options: ☐ Trim fields ☐ Discard empty fields ☐ Decimal separator is comma

Geometry definition: ☒ Point coordinates ☐ Well known text (WKT) ☐ No geometry (attribute only table)

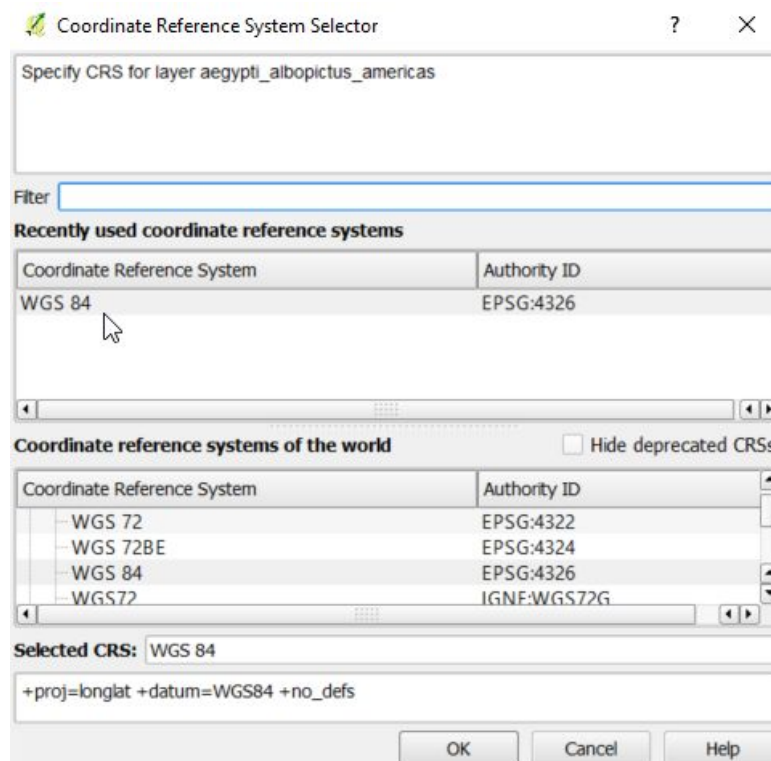
X field: X Y field: Y ☐ DMS coordinates

Layer settings: ☐ Use spatial index ☐ Use subset index ☐ Watch file

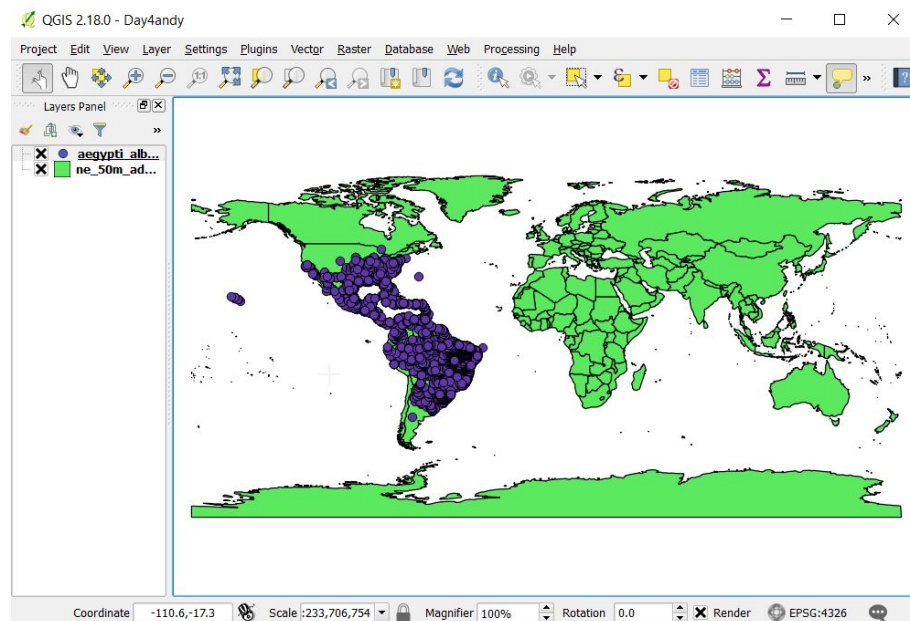
	field_1	VECTOR	OCCURRENCE_ID	SOURCE_TYPE	LOCATION_TYPE	POLYGON_ADMIN	Y	X	YEAR
1	7	Aedes aegypti	7	published	point	-999	4.63	-74.09	1960
2	10	Aedes aegypti	10	published	point	-999	10.34	-67.04	1960
3	11	Aedes aegypti	11	unpublished	polygon	2	25.49	-80.99	1960

OK Cancel Help

The Coordinate Reference System Selector screen will then appear. As the data are in longitude and latitude, make sure that the WGS84 Geographic Coordinate Reference System is selected, then click OK.



The point data should then be displayed on top of the world map:



Select 'View, Zoom to layer' to zoom in on the points.

Make sure you now save this point data as a shapefile. Recall that this is done by right-clicking on the layer name in the Layers Panel, selecting “Save As” and then saving the data as aegypti_albopictus_americas.shp in the appropriate Data folder, making sure the selected format is “ESRI Shapefile”.

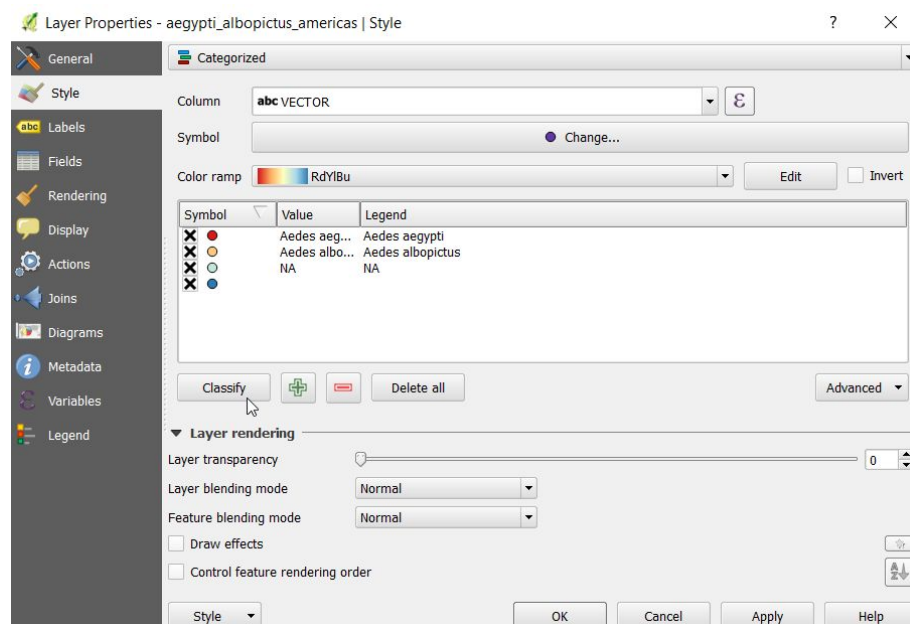
We shall now take a closer look at the data. Right click on the 'aegypti....' layer in the Layers Panel and select 'Open Attribute Table' to look at the attribute data. You should be able to see that there are columns called VECTOR, YEAR and COUNTRY. If you scroll down to record 7261 you should see where the entries in the VECTOR column change from one species to the other.

We shall now change the symbols on the map to identify the two species. Right click on the 'aegypti..' layer in the Layers Panel and select 'Properties'.

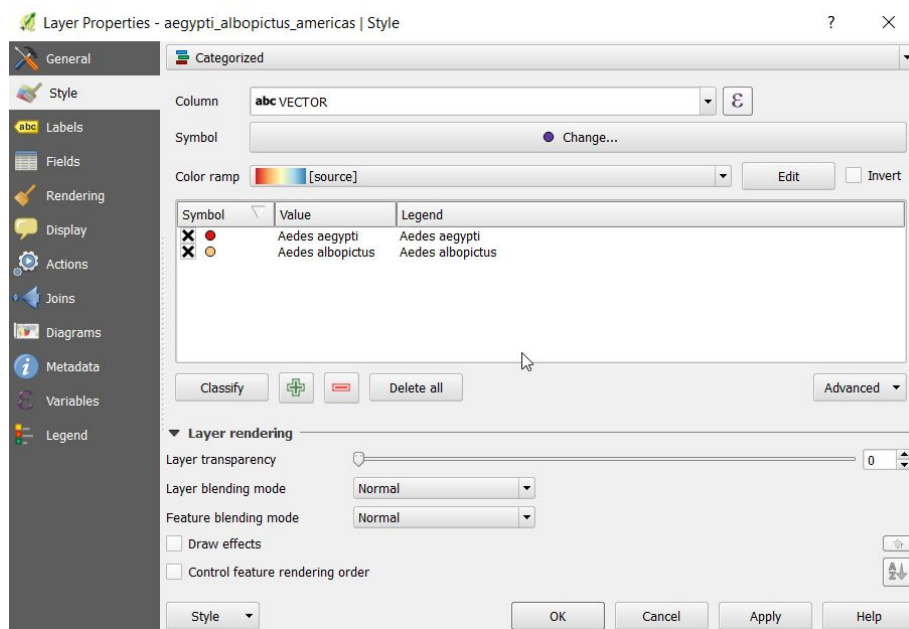
Within the 'Style' tab, change :

- 'Single symbol' to 'Categorized'
- Column to VECTOR
- Color ramp to RdYlBu
- Click 'Classify'

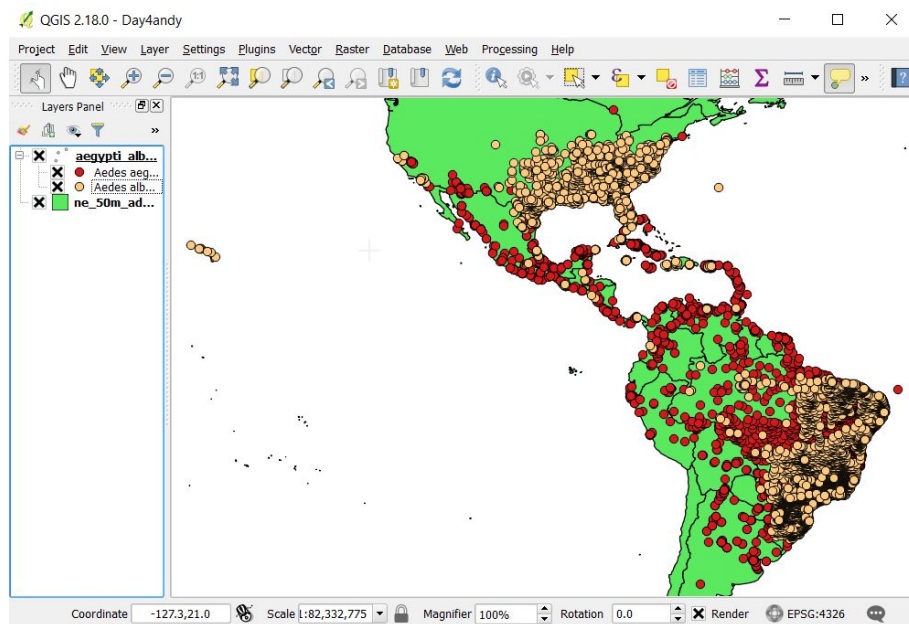
You will see the two species names will be coloured red and yellow, and two other options are green (VECTOR = NA) and blue (VECTOR is missing). We do not want to display points where the vector species is unknown so we will deselect green and blue. You can either uncheck the box next to these two categories, or remove them from the list by clicking on each of them one at a time, then using the red '-' button to delete.



Now you should just have the 2 categories displayed in the category list.

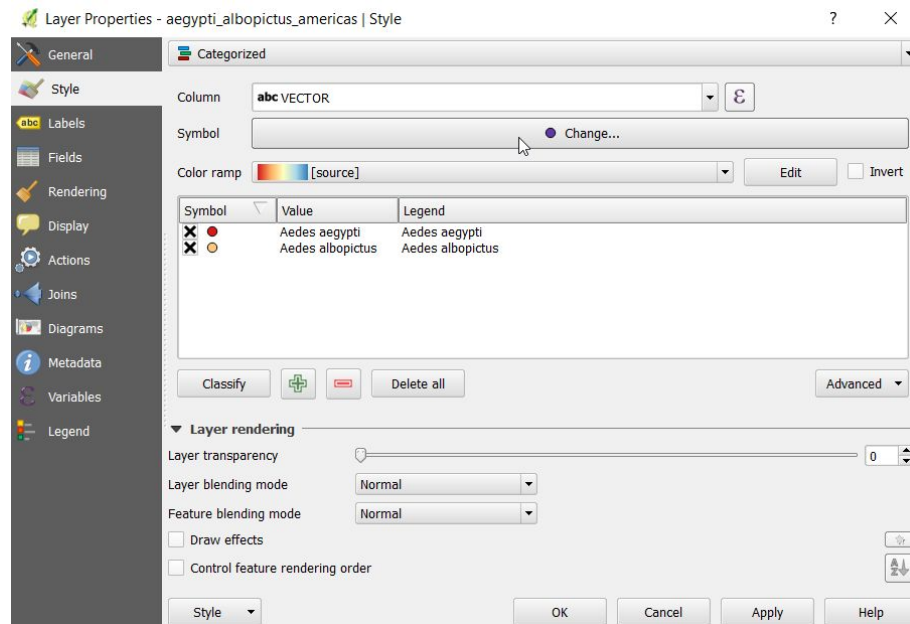


Now look at the map. Initially it looks like *Ae. aegypti* (red) may be more common in the west of the continent, and *Ae. albopictus* (yellow) may be more common in the east. However, there are lots of overlapping points so it is not easy to see exactly what is happening, plus we need to bear in mind that we do not know whether areas with no data have no *Aedes*, or if they have simply not been surveyed.

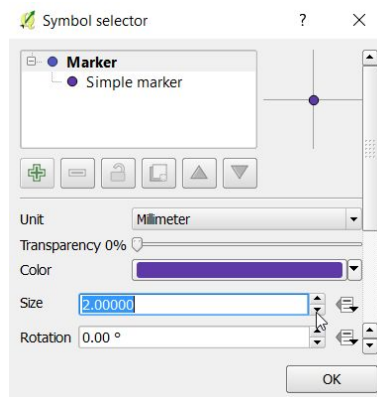


To make things clearer we can make the points smaller and make the background map less distracting.

Select 'Properties' for the point layer again (Right click on 'aegypti...' within the Layers Panel) and click the 'Change..' button.

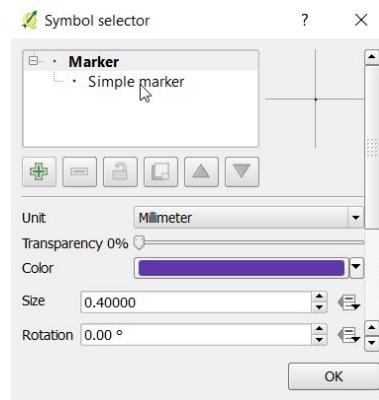


We can decrease the size of the circles from their default of 2 down to 0.4 by clicking the down arrow next to the number representing size, or typing the number directly in to the box. Then click OK.

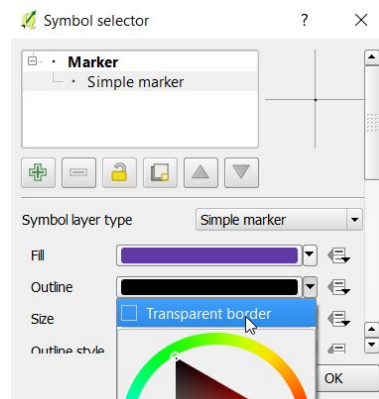


Click 'Apply' from the properties window. This should apply your changes but keep the properties window open in case you want to make any other changes. Now the points are much smaller but they are all black. This is because they have a black outline. We can turn that outline off, also using the Properties window.

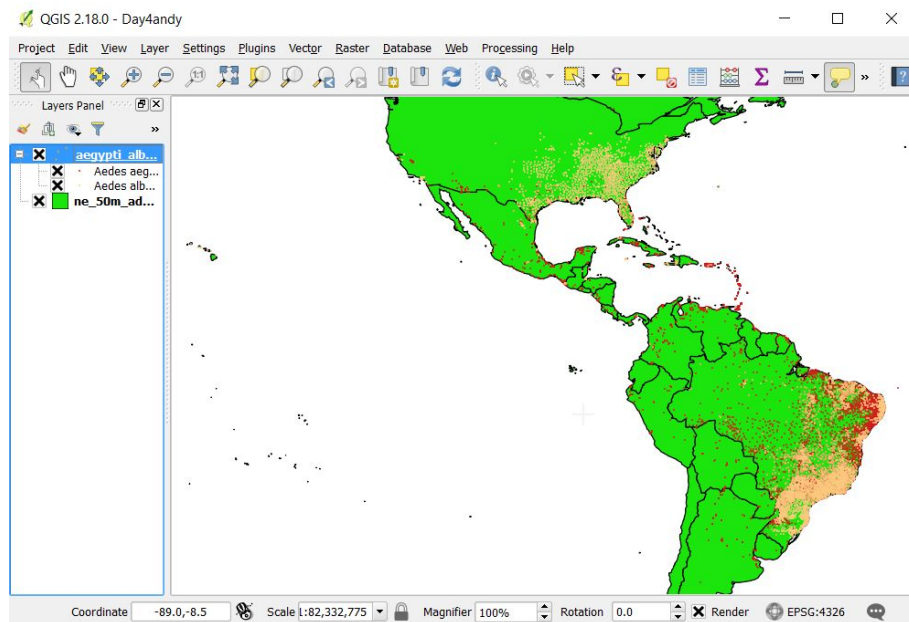
Once again, click the 'Change...' button. Change the properties of the Marker i.e. the circle by clicking on 'Simple marker'.



Click 'Outline' and select 'Transparent border'. Press OK, and in the Properties window click OK to apply the changes and close the window.

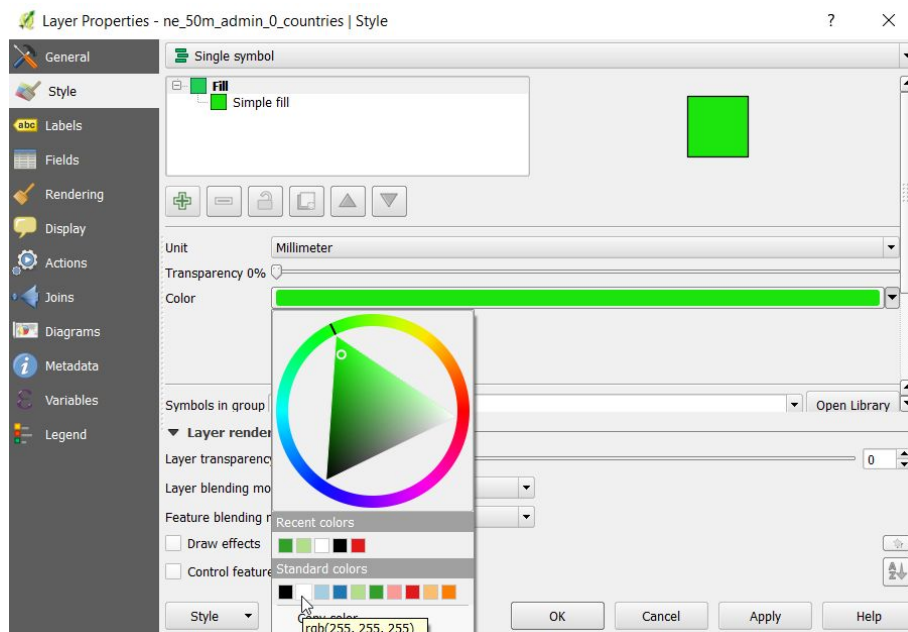


Now we can see that there are some red dots in the east too. But the green background and the thick country border lines are distracting. We want to be able to see the data themselves which are the most important part.

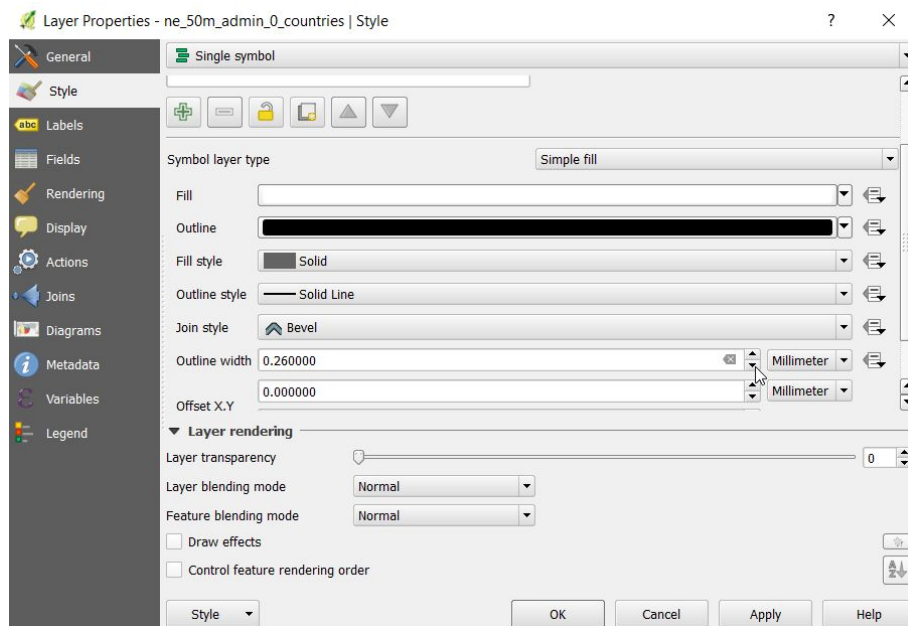


To allow us to see the data more clearly we can turn off the land colour and make the country borders thinner.

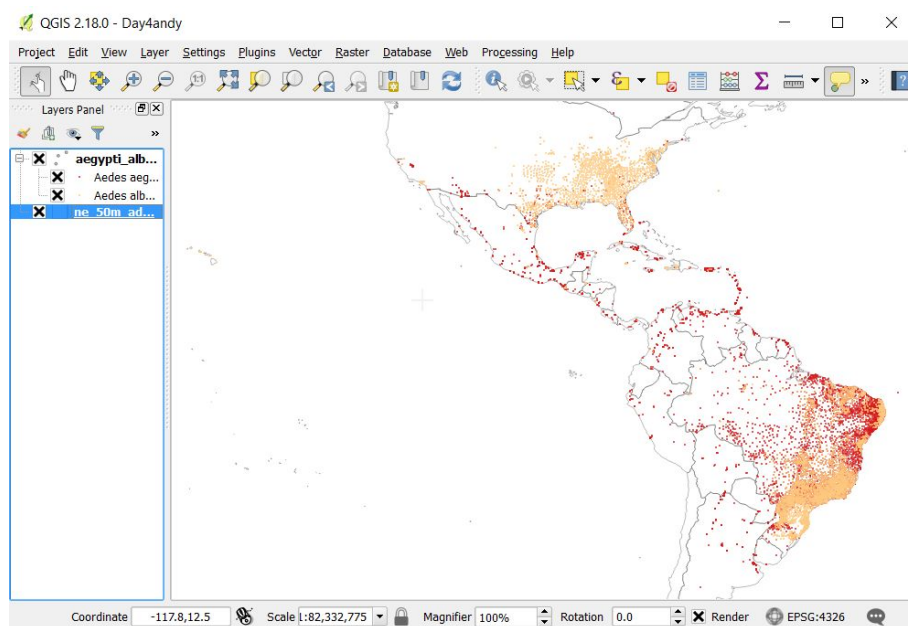
Select 'Properties' for the countries layer (Right click on 'ne_50m...' within the Layers Panel). Change the color to white and click Apply.



To make the country borders thinner, click 'Simple fill' near the top of the window. Use the scroll bar on the right to scroll down until you see 'Outline width'. If you click the down arrow once this will go from 0.26 to 0.06. Click OK.



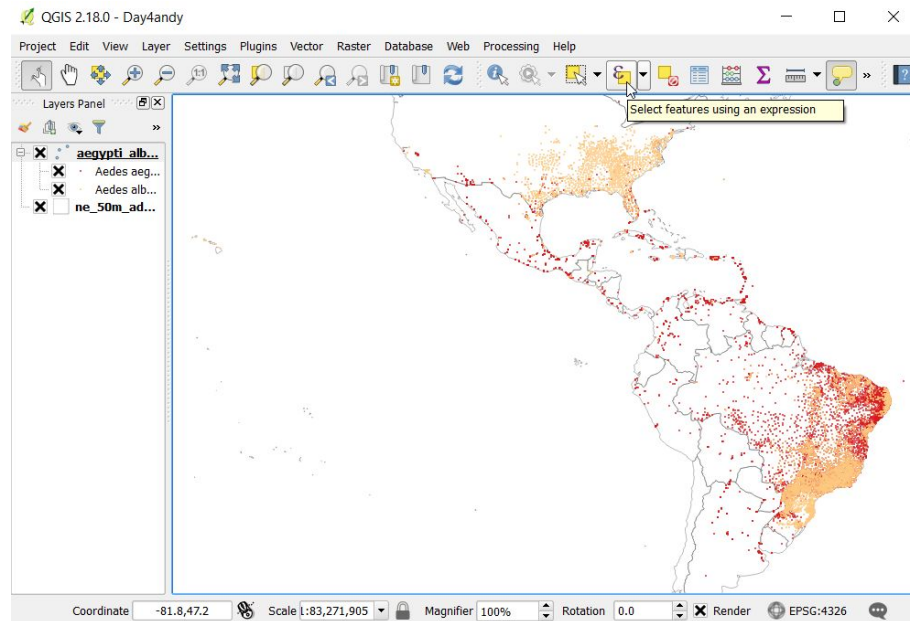
Now we can see more clearly that there are *Ae. aegypti* (red) in the eastern part of South America, but not very many sites in the northern part of South America. This is one example of how changing the symbology of a map can make the data clearer.



Click Project, Save to save the work you've done on the project so far. This would allow you to get back to this point if your computer crashes or you make a mistake later.

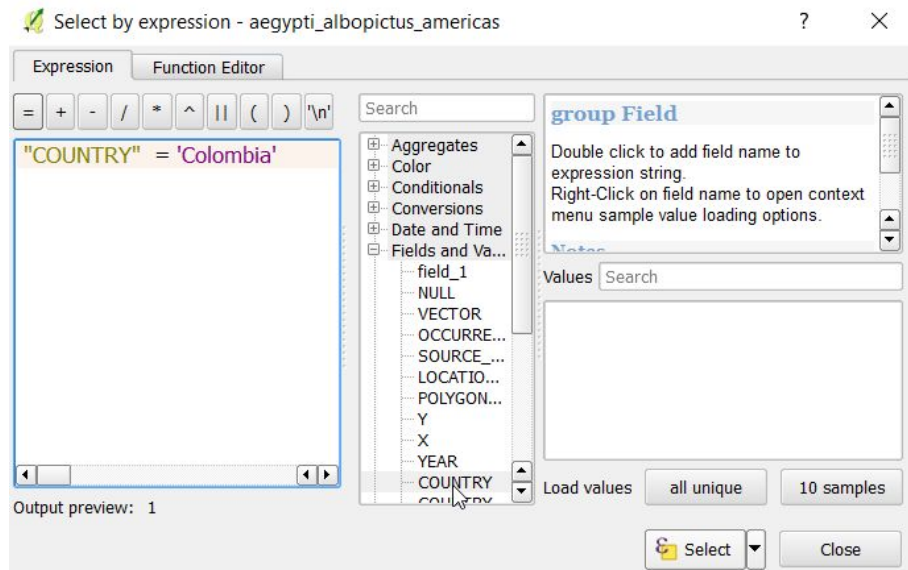
Now we are going to see how we can look at the data more closely by using the different selection tools within QGIS.

Firstly we will look at selecting data according to their attributes i.e. the data stored in the Attribute table. To do this, click the button 'Select features using an expression' on the toolbar, or in the tools displayed at the top of the attribute table.



This tool allows you to subset your data according to the values stored within the attribute table. We do this by writing an *expression*, which is simply a formula that uses mathematical symbol similar to that which we use in Excel. Here we will demonstrate how to select only the points that are within Colombia by using the COUNTRY variable.

Within the central box, locate 'Fields and Values' and click on the + next to it to expand the menu. Double click on COUNTRY and it gets added to the Expression box. To inform QGIS that you only want the points that are labelled as being in Colombia you can either then type = '**Colombia**' making sure that you include the quotation marks, or you can click the = button above the word COUNTRY on the left of the window, then click the 'all unique' button which displays a list of all of the unique countries included in the COUNTRY column of the attribute table, then double click 'Colombia' to add it to the expression. Regardless of which method you use, you should ensure that the expression in the box on the left is
"COUNTRY" = 'Colombia'



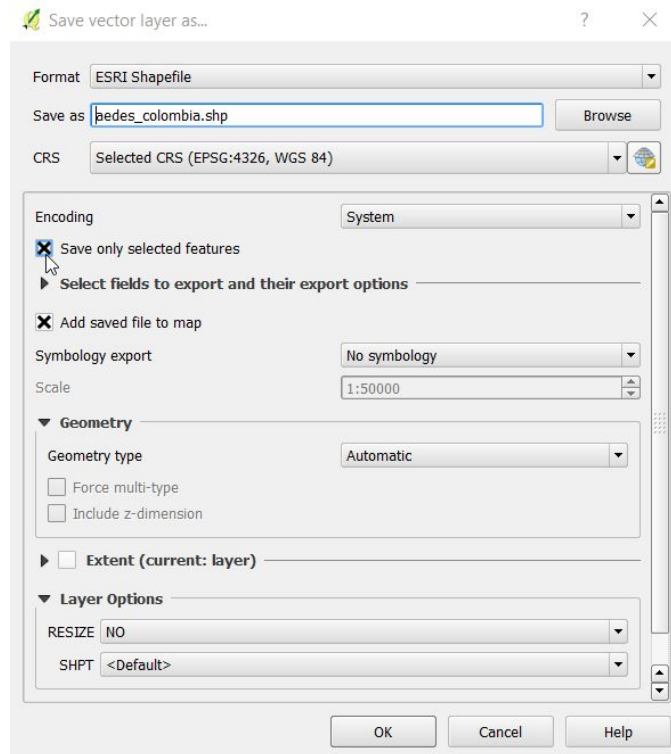
You can check whether this expression is valid by checking the information next to the text 'Output preview' at the bottom left of the window. If there is a problem it will say 'Expression is invalid' in red.

Once you have checked this, click the Close button.

On your map you should see that selected points are bright yellow. To look closer at this, choose View, Zoom to selection from the main menu and you will see the map view zoom in on Colombia. Now the points are very small and difficult to see, we will sort that later.

We will now save just the Colombia occurrence data to a new shapefile.

Right click on aegypti... in the Layers Panel and select 'Save as'. In the 'Save vector layer as' window click the 'Save only selected features' checkbox. Type a name for the file and click OK.



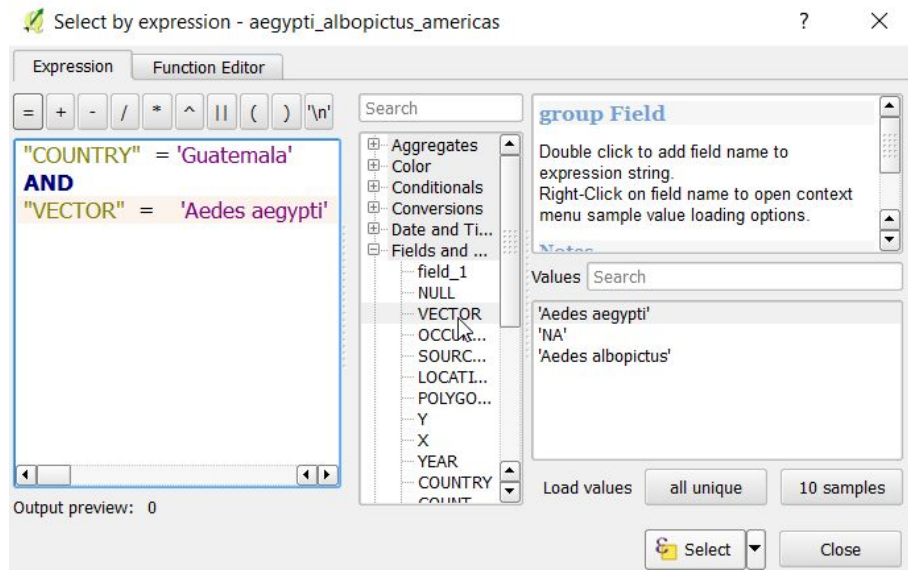
If you open the Attribute table for the resulting Colombia shapefile and scroll down you should find that there are 131 occurrences of which 3 are *Ae. albopictus*.

We can also use the selection tools to subset the data by multiple attributes.

Select the original *aegypti_albopictus_americas* layer and return to the 'Select by' expression window. We shall now add to this expression and select only points that are in Guatemala AND for which the vector species is *Ae. Aegypti*.

Similar to the previous example, add the expression "**COUNTRY**" = '**Guatemala**'

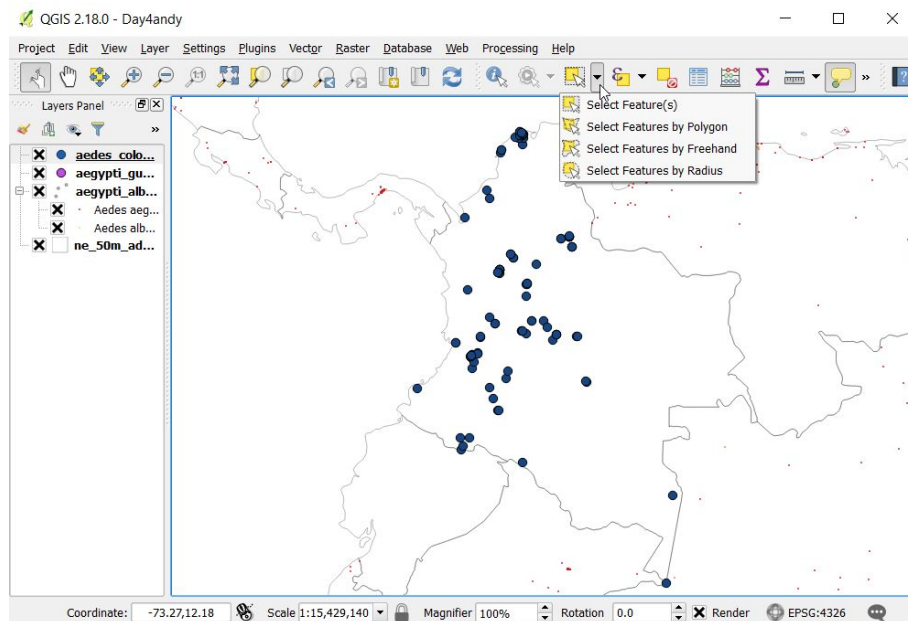
After the initial expression, type AND (all capital letters, no quotation marks). Then, under the Field and Values list, double click on 'VECTOR' from the column names and make sure it is added to the expression. Type = or click the = button, then click 'Load values: all unique'. Double click 'Aedes aegypti' from the list that appears so that it is added to the expression. Check the output preview for errors, then click Close.



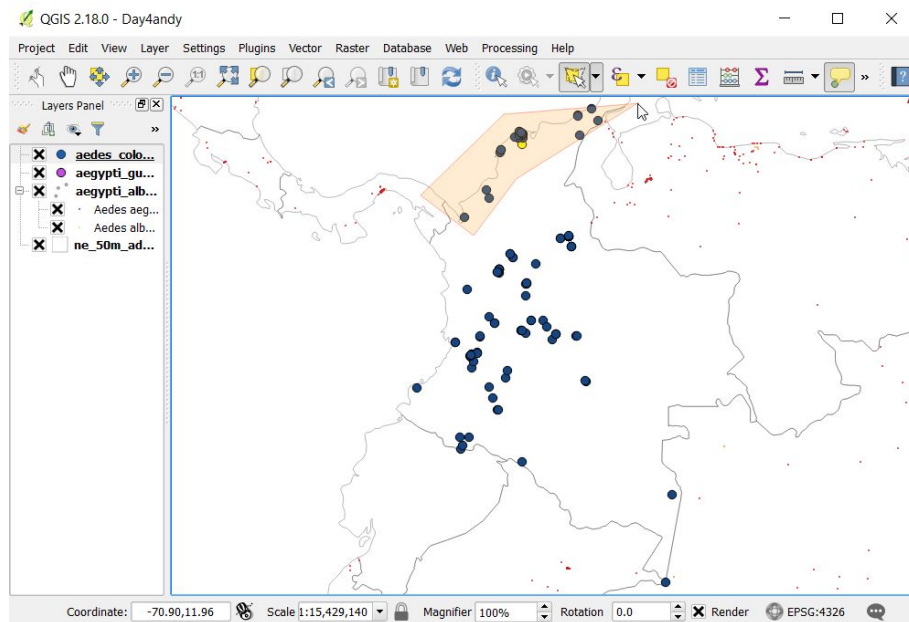
Repeat the 'Save as' routine we did before, making sure the 'Save only selected features' checkbox is checked. Save the file as aegypti_guatemala. If you open the attribute table you should see there are 13 sites for aegypti only.

We can also select points by their location on the map. First select the aedes_colombia layer again and 'View, Zoom to layer'.

Find the Select Feature(s) button on the toolbar and experiment with the different select options. e.g. with the Select Feature(s) option you can use the cursor to select an initial point on the map and then hold down Ctrl to add other points to the selection.



With 'Select Features by Polygon', you can left click to create a polygon surrounding points you wish to select. Right click to finish the polygon and select the points.



Open the attribute table for the `aedes_colombia` layer after you have selected points on the map. You should see that some rows are highlighted in blue. These correspond to the points you have selected on the map. This selection process works in reverse too, you can select rows in the attribute table by left clicking on the first column (containing the row numbers). The points on the map corresponding to these rows will get selected. Click and drag within the first column or click using Ctrl and/or Shift keys to select multiple rows.

aedes_colombia :: Features total: 131, filtered: 131, selected: 41

	field_1	VECTOR	OCCURRENCE	SOURCE_TYP	LOCATION_T	POLYGON_AD	Y	X	YEAR	COUNTRY
1	7	Aedes aegypti	7	published	point	-999	4.6300000000...	-74.09000000...	1960	Colombia
2	359	Aedes aegypti	359	published	point	-999	8.7600000000...	-75.87999999...	1972	Colombia
3	361	Aedes aegypti	361	published	point	-999	9.0500000000...	-76.00000000...	1972	Colombia
4	1045	Aedes aegypti	1045	published	point	-999	6.7700000000...	-75.07999999...	1990	Colombia
5	1114	Aedes aegypti	1114	published	point	-999	3.4500000000...	-76.51999999...	1991	Colombia
6	1425	Aedes aegypti	1425	published	point	-999	4.1500000000...	-73.62999999...	1995	Colombia
7	1446	Aedes aegypti	1446	published	point	-999	11.5500000000...	-72.90000000...	1995	Colombia
8	1447	Aedes aegypti	1447	published	point	-999	11.7799999999...	-72.45000000...	1995	Colombia
9	1685	Aedes aegypti	1685	published	point	-999	5.8700000000...	-74.64000000...	1997	Colombia
10	1686	Aedes aegypti	1686	published	point	-999	5.8900000000...	-74.64000000...	1997	Colombia
11	1825	Aedes aegypti	1825	published	point	-999	6.7700000000...	-75.07999999...	1998	Colombia
12	2368	Aedes aegypti	2368	published	point	-999	3.4300000000...	-76.53000000...	2001	Colombia
13	2369	Aedes aegypti	2369	published	point	-999	3.4600000000...	-76.46999999...	2001	Colombia
14	2370	Aedes aegypti	2370	published	point	-999	3.4700000000...	-76.51999999...	2001	Colombia
15	2371	Aedes aegypti	2371	published	point	-999	3.4800000000...	-76.46999999...	2001	Colombia
16	2372	Aedes aegypti	2372	published	point	-999	3.4800000000...	-76.48000000...	2001	Colombia
17	2566	Aedes aegypti	2566	published	Less than 25km	Less than 25km	6.3400000000...	-75.56000000...	2002	Colombia
18	2867	Aedes aegypti	2867	published	Less than 25km	Less than 25km	6.3400000000...	-75.56000000...	2003	Colombia

Show All Features

Experiment with 'Select Features by Radius'. This is a quick way to select points that are within a circle defined by the cursor.

End of Optional Section