GENTLE INTRODUCTION TO GIS IN QGIS SOFTWARE

A step-by-step training module

QGIS

Developed by

Francis Oloo & Godwin Murithi

GEOPSY RESEARCH

September 1, 2020

GEOPSY RESEARCH

P.O. Box 10101-00101 Nairobi-Kenya| www.geopsy-research.org|info@geospy-research.org

Table of Contents

1. Introduction

Geographic Information System (GIS) is a useful tool for preparing, managing, analyzing and presenting spatial data. Furthermore, in this era of open science, ample open access data is available that can easily be retrieved and integrated in open source desktop GIS software, such as Quantum GIS (QGIS).

In this exercise, you will be guided through steps of working with spatial data models including, vector and raster data models. Vector data models represents real world features using points, lines and polygons. Examples of vectors include boreholes and populated places as points, roads and waterways as lines and administrative areas (counties, countries etc.) as polygons.

Raster data models represent real world features in the form of equally sized grid cells arranged in rows and columns with each cell containing an attribute value. They are ideal for continuous data such as elevation and precipitation (Shekhar et al., 1997).

Apart from introducing spatial data models, the module will introduce the learner to basic spatial analysis methods and simple map design in QGIS software. This is meant to be a gently introduction to GIS.

Task Description

The main task in this assignment is to carry out preliminary assessment of access to water (boreholes) in sub-locations in the north of Isiolo County, Kenya. A user of this module should understand that this is only a hypothetical case whose sole purpose is to introduce steps in geospatial data handling, analysis and visualization.

Data

In order to achieve the desired objective which is preliminary assessment of assess to water (boreholes) in Northern parts of Isiolo County in Kenya, we will use different datasets. This includes Vector data files such as Area of Study boundary, Populated places, Boreholes location data, and Sub locations boundaries. In addition, we will use Raster data such as Elevation, Distance to waterways, Population and Land cover. The first step is to download the datasets which can be accessed here (**https://tinyurl.com/yxnmz4fm**). The data is divided into three subfolders which are Raster, Vector and Tabular. Once downloaded, save the data in a Folder easily assessable as required during subsequent operations.

1.1 Learning Objectives

Upon successful completion of this step by step tutorial, the learner should be able to:

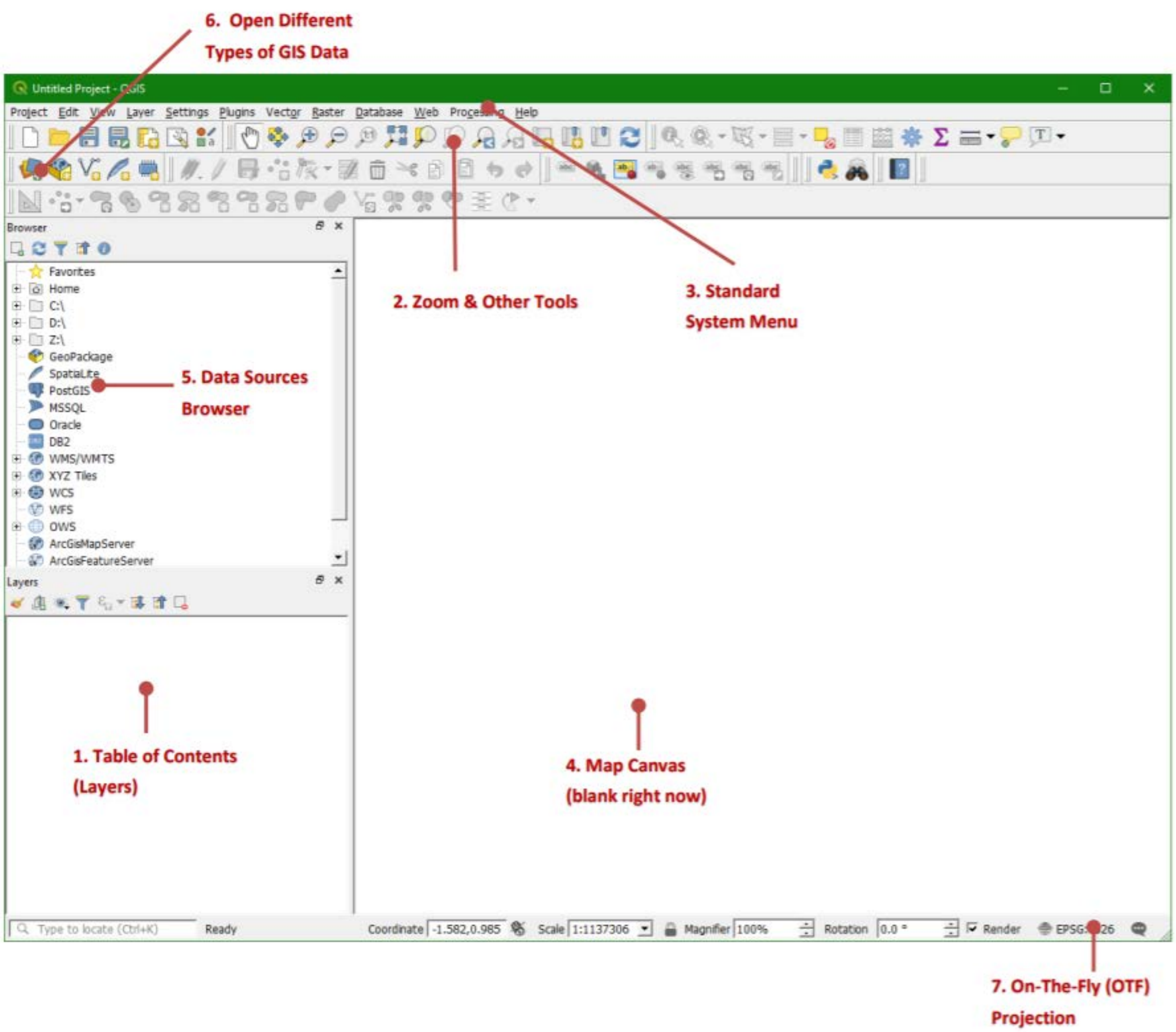
- Add tabular, vector and raster data onto a GIS software and carry out preliminary data preparation.
- Carry out simple spatial data analysis in QGIS using both vector and raster data models
- Create a simple map in QGIS map composer.

2. Getting familiar with QGIS

2.1 Install QGIS

In case you have not installed QGIS software in your computer, an installation file can be downloaded from here https://qgis.org/en/site/forusers/download.html

2.2 Start QGIS



The QGIS main window is made of the Standard System Menu or Main Menu (we will refer to Main Menu in the following pages) which is always on the top of the software's window, and contains standard toolbars like the Zoom, Pan, Select Tools, or others that can be activated clicking in the Main Menu View and then Toolbars.

On the left hand side there is the Table of Content (Layers list) area. This is where the names of the data or layers that are added onto the software will be displayed. To the left of this space, there is a vertical toolbar, the Manage Layers toolbar.

In the center of the window is the Map Canvas, where all maps will be displayed. More information about the QGIS Graphical User Interface can be found here: http://docs.qgis.org/2.18/en/docs/user_manual/introduction/qgis_gui.html

3. Vector data analysis

Commonly, we can originally have geographic data (coordinates and place names) in tabular files like Microsoft Excel or Comma Separated Values (CSV). This part of the tutorial demonstrates the steps of converting tabular data into spatial data models and how to display them in standard GIS software.

Originally, data on places in the study area are in a table with names, geographic coordinates and other attributes. We have to import the table into QGIS and convert the tabular data to a point data model. A common file format for storing vector data models in GIS is called Shapefiles. So in this case the tabular data will be converted to point shapefiles.

It is particularly recommended that your tabular data should either be in CSV format or in a tab delimited text file.

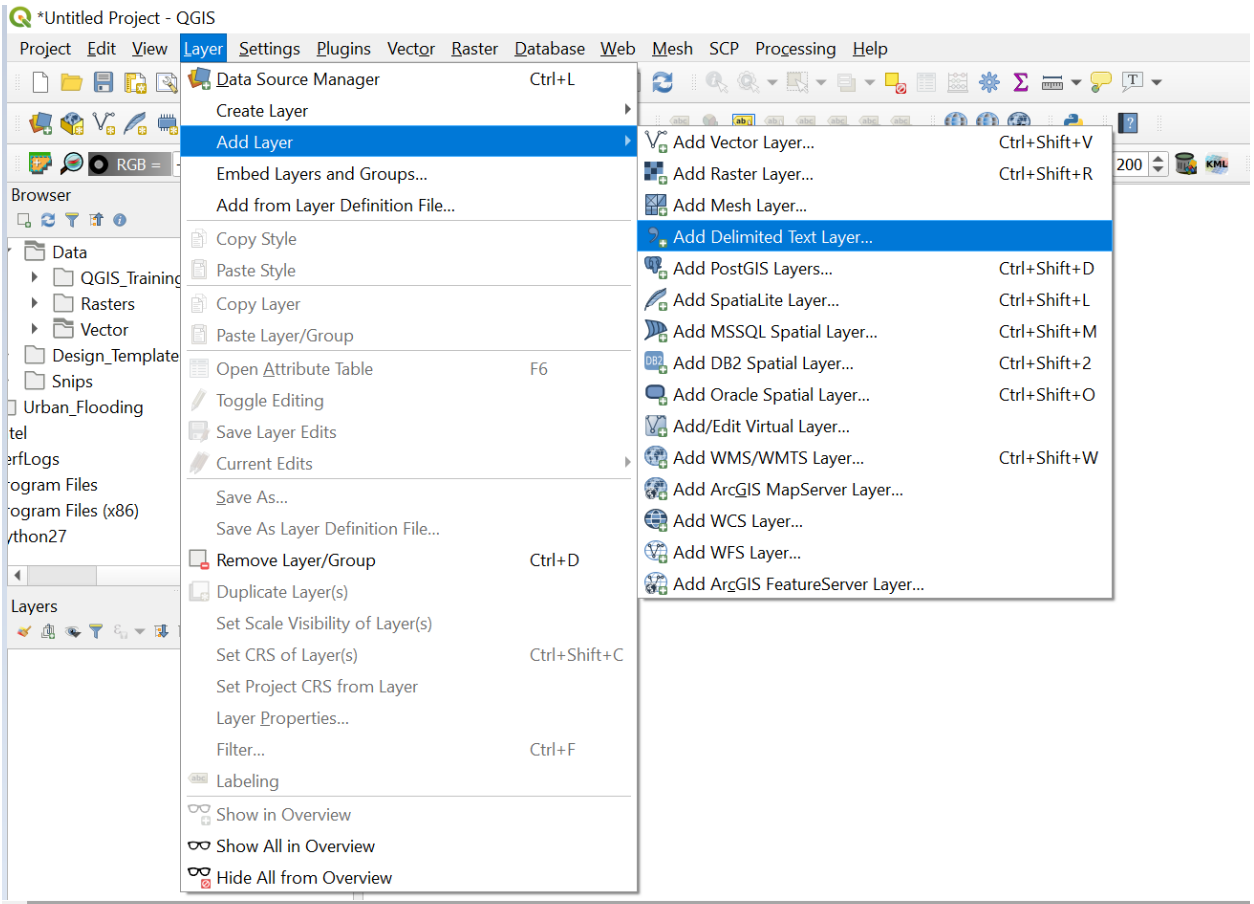
Code	Name	Latitude	Longitude	Class1	Class2	Country
1	Bulesa	1.00273	38.54686	Place	Populated Place	Kenya
2	Gafarsa	0.94054	38.59475	Place	Populated Place	Kenya
3	Malka Galla	1.20097	38.89144	Place	Populated Place	Kenya
4	Merti	1.06725	38.6662	Place	Populated Place	Kenya
5	Sericho	1.14495	39.08681	Place	Populated Place	Kenya

Our tabular data contains 5 features (points). Each point has 7 attributes include code, place name, geographic coordinates (Latitude and Longitude), class (Class1 and Class2) and country. Geographic coordinates are important as they help us to place our points at specific location on earth. Other attributes are also important, as they tell us more (describe) geographic data.

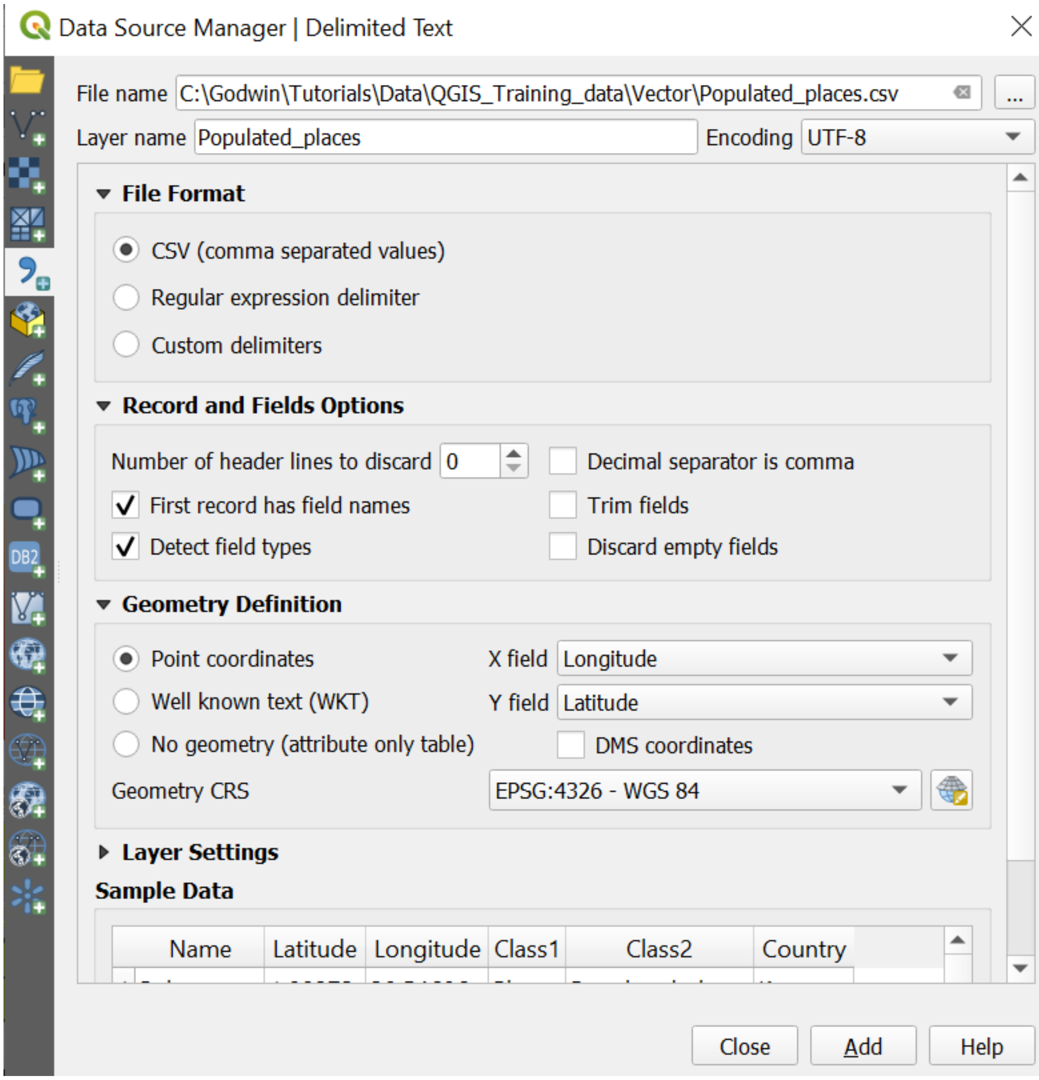
3.1 Import tabular data into the map

In this example, we are going to import a table, containing place names and their corresponding geographic coordinates into QGIS software. The table is named Populated_places.csv. The format is comma separated values (CSV).

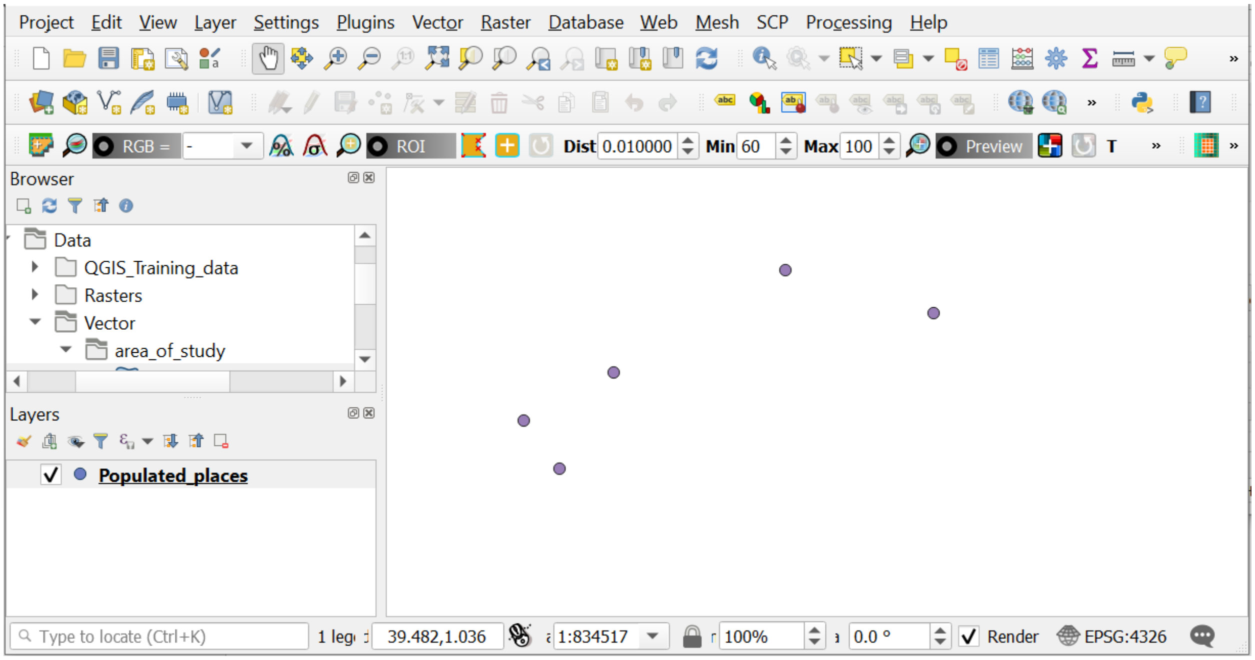
- On the main menu click *Layer* → *Add Layer* → *Add Delimited Text Layer*



- In the new dialogue window use the  button to browse to the folder containing the table we want to add. Check the Geometry definition it should appear similar to the one below. Geometry helps in defining the geographic locations, in this data Longitude is used for the X coordinates and Latitudes are used for Y coordinates.



- Once selected click Add to add the table to the map canvas.
- Your screen should be similar to this

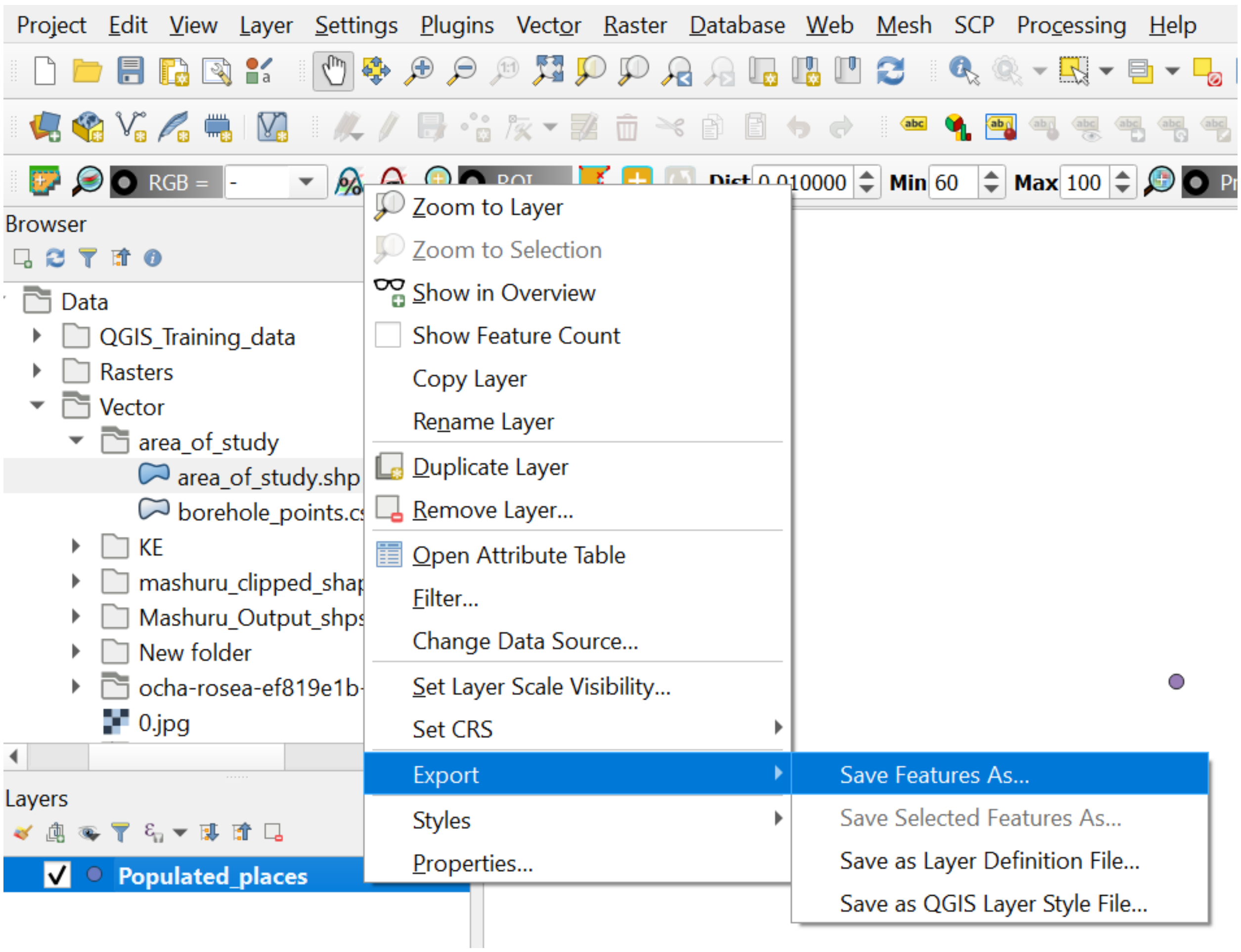


3.2 Export CSV layer as Point shapefile.

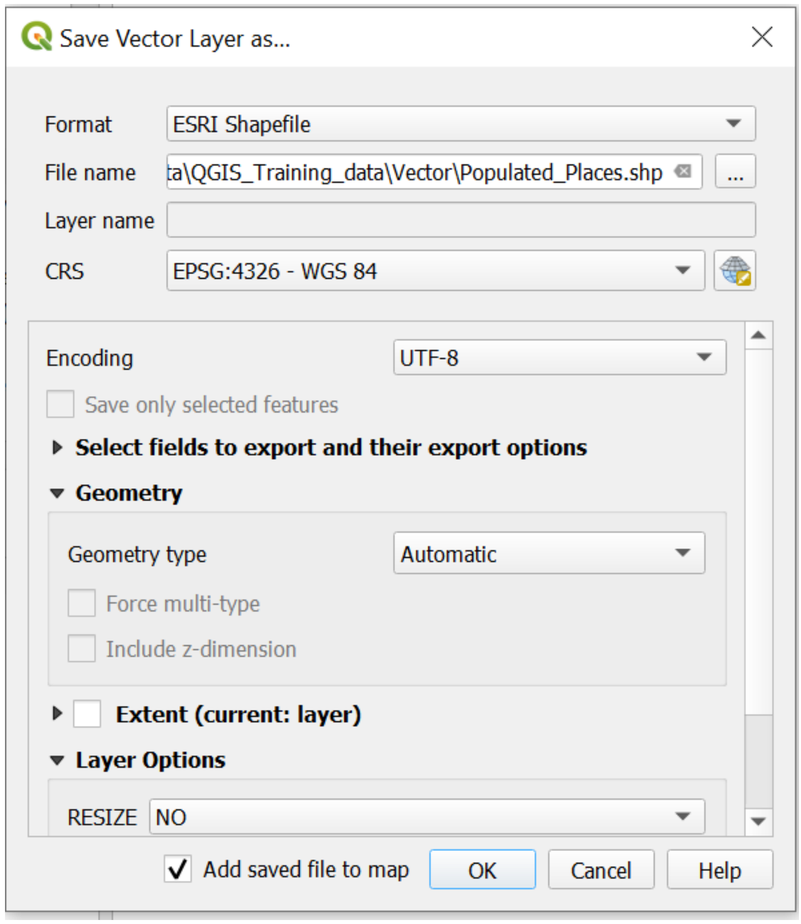
Now we have populated places in QGIS, but the format of the data is comma delimited values (csv) for further analysis we need the data as a point shapefile. Now we are going to save populated places table as a point shapefile.

** Shapefile is a digital vector storage format for storing geometric location and associated attribute information, They can be used in different GIS platforms such as ArcGIS, QGIS and AutoCAD*.*

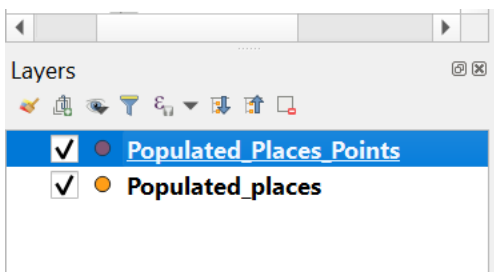
- RIGHT Click on *Populated_places* → *Export* → *Save feature as*



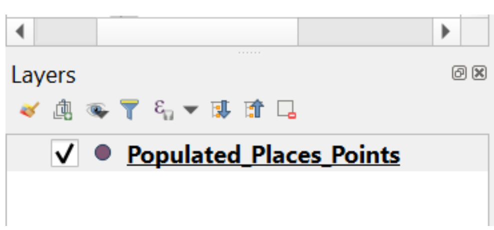
- On the new dialogue window select ESRI Shapefile as the format, use ... button to navigate and select a folder to save the new shapefile. Name the new shapefile *Populated_Places_Points.shp*
- Click Ok.



- The new layer appears on the layers tab as seen below



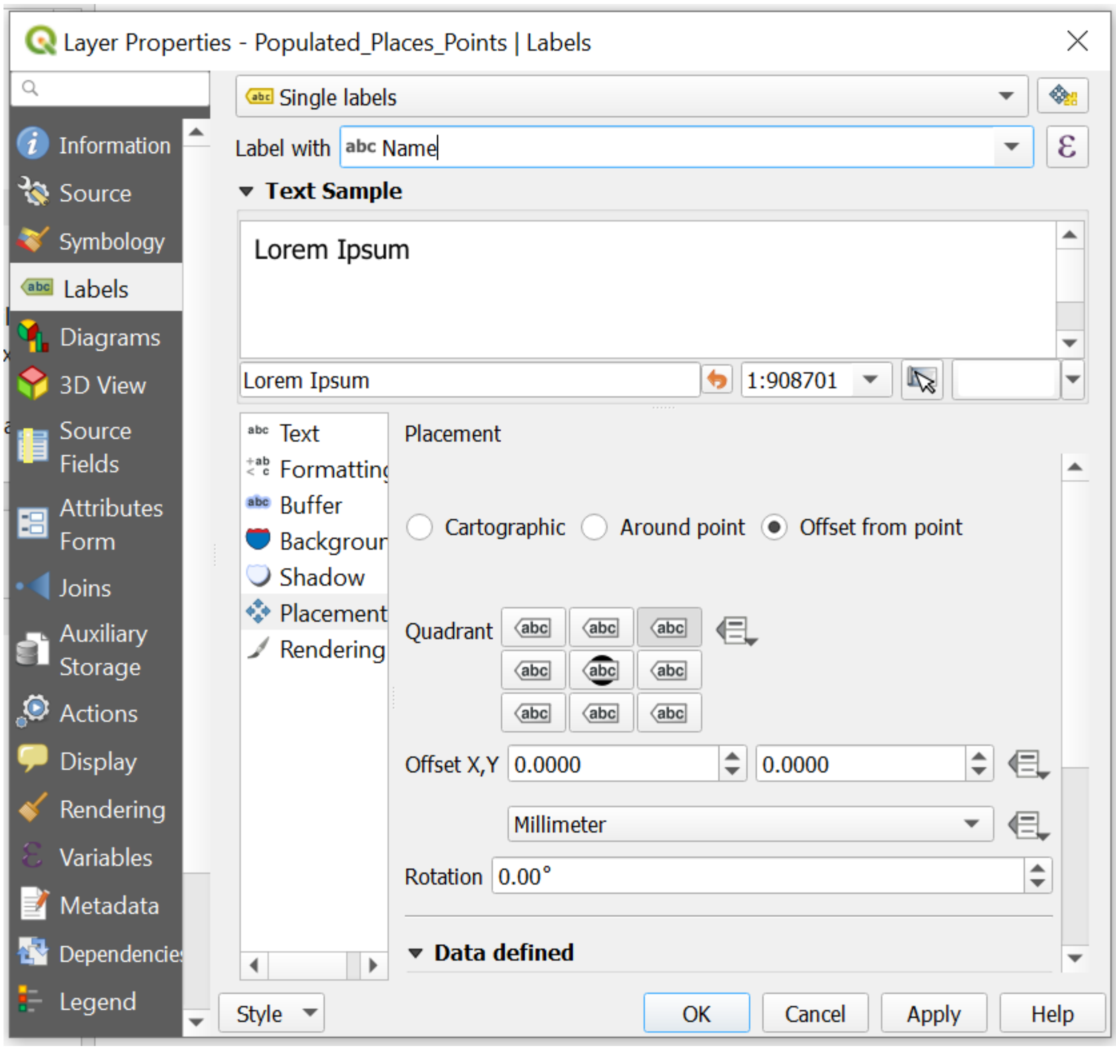
- To remove Populated_places table. Click on *Populated_places* → *Remove layer* → *OK.*
- The layer is removed



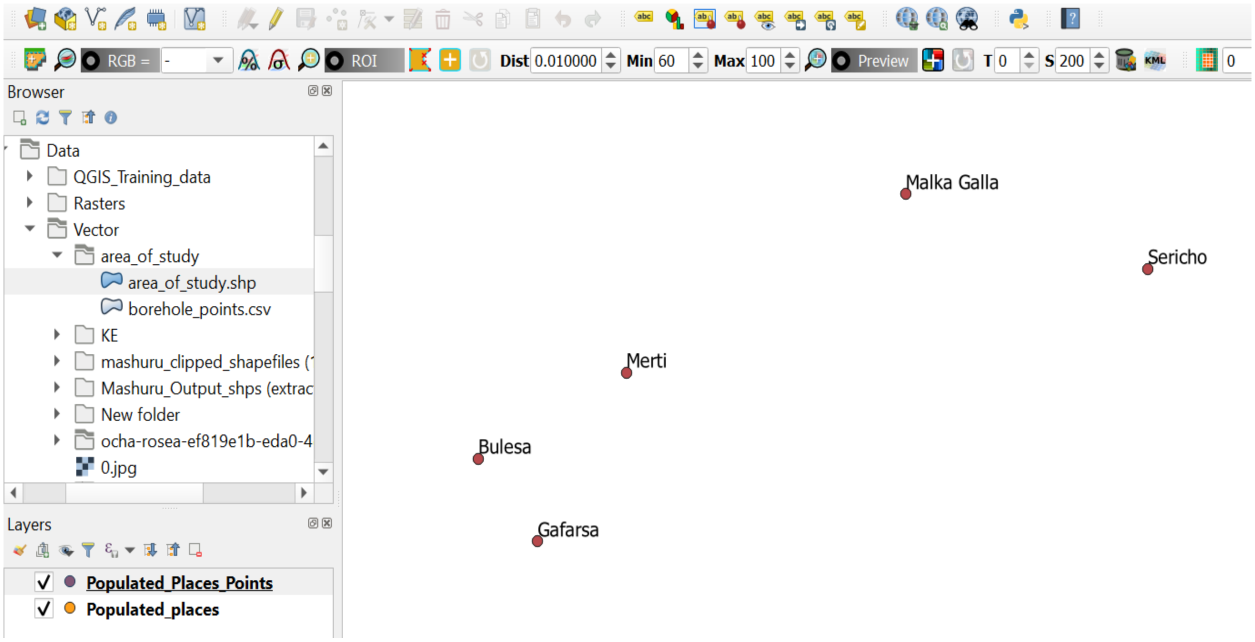
To this point we have added tabular data to QGIS and converted the tabular data to a point shapefile.

Now we need to apply labels to the points on the canvas in order to view the name of the location. To do this we will use the Name column in the attribute table.

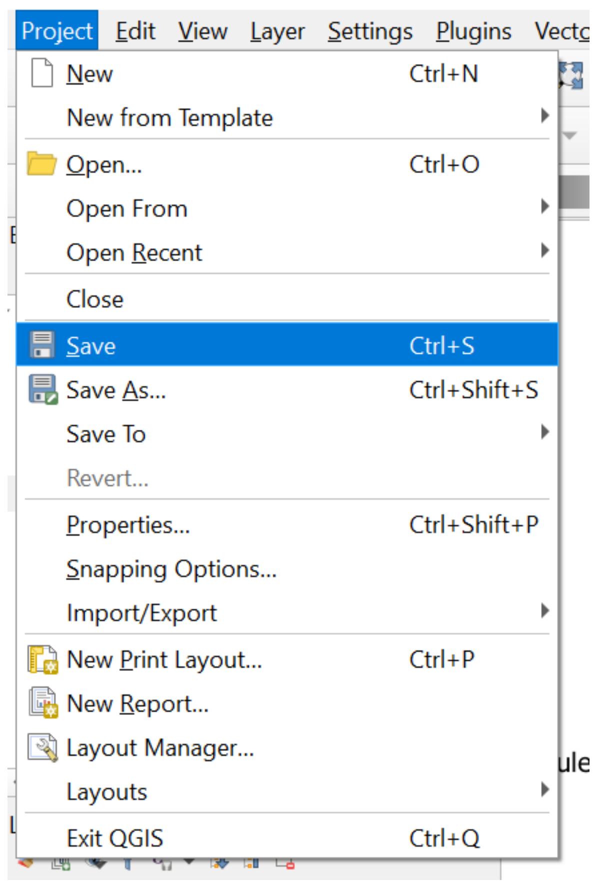
- Click on *Populated_Places_Points → Properties → Labels*
- Select *Single labels* **label with** select ***Name*** click OK



- Your canvas should be similar to the one below



- On far left of the Main Menu click → *Projects* → *Save*



- Save your QGIS work on a working chosen folder.

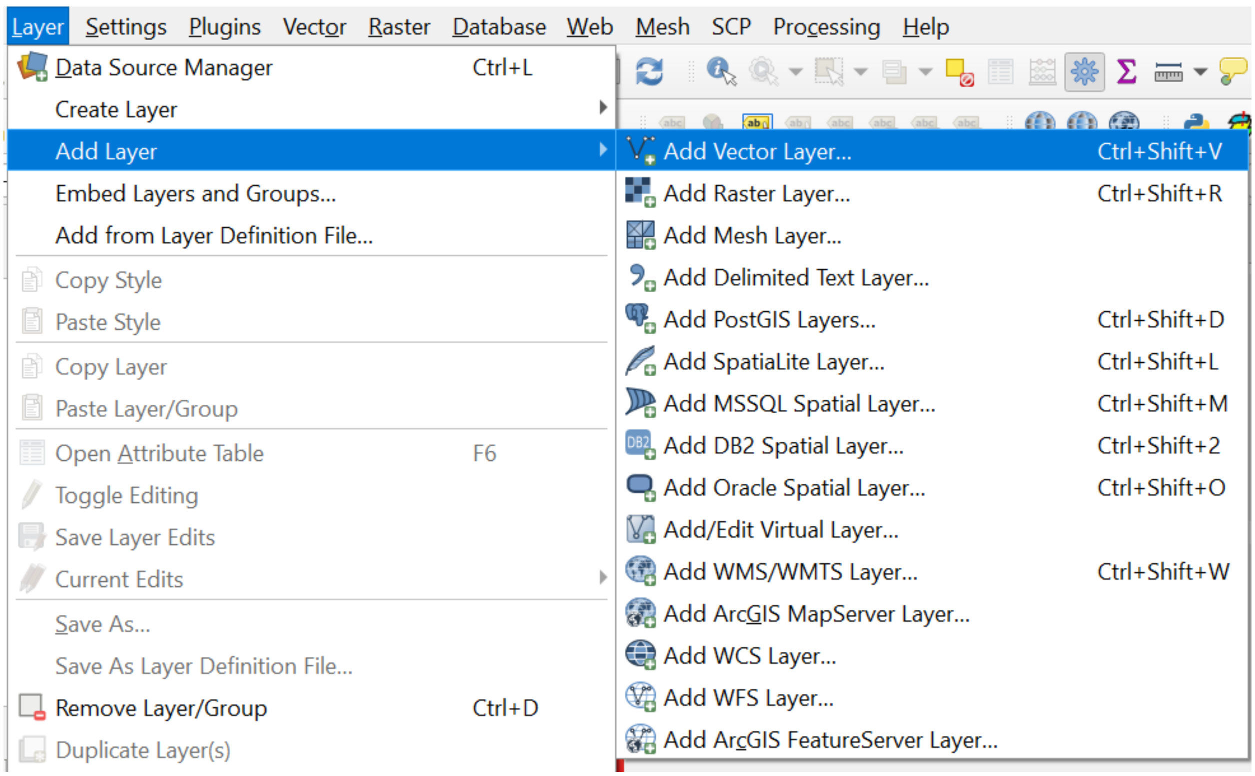
Adding Vector Data

Vector data in GIS include Points, polygons and lines. In this exercise we are going to add points, polygons and lines. **Note that different geographic features are categorized as points, lines or polygon. For instance in our exercise we will add points, which are locations of populated places or boreholes; for polygon, we will add the area of study boundary ; for lines we will add rivers/Stream networks in our study area*.*

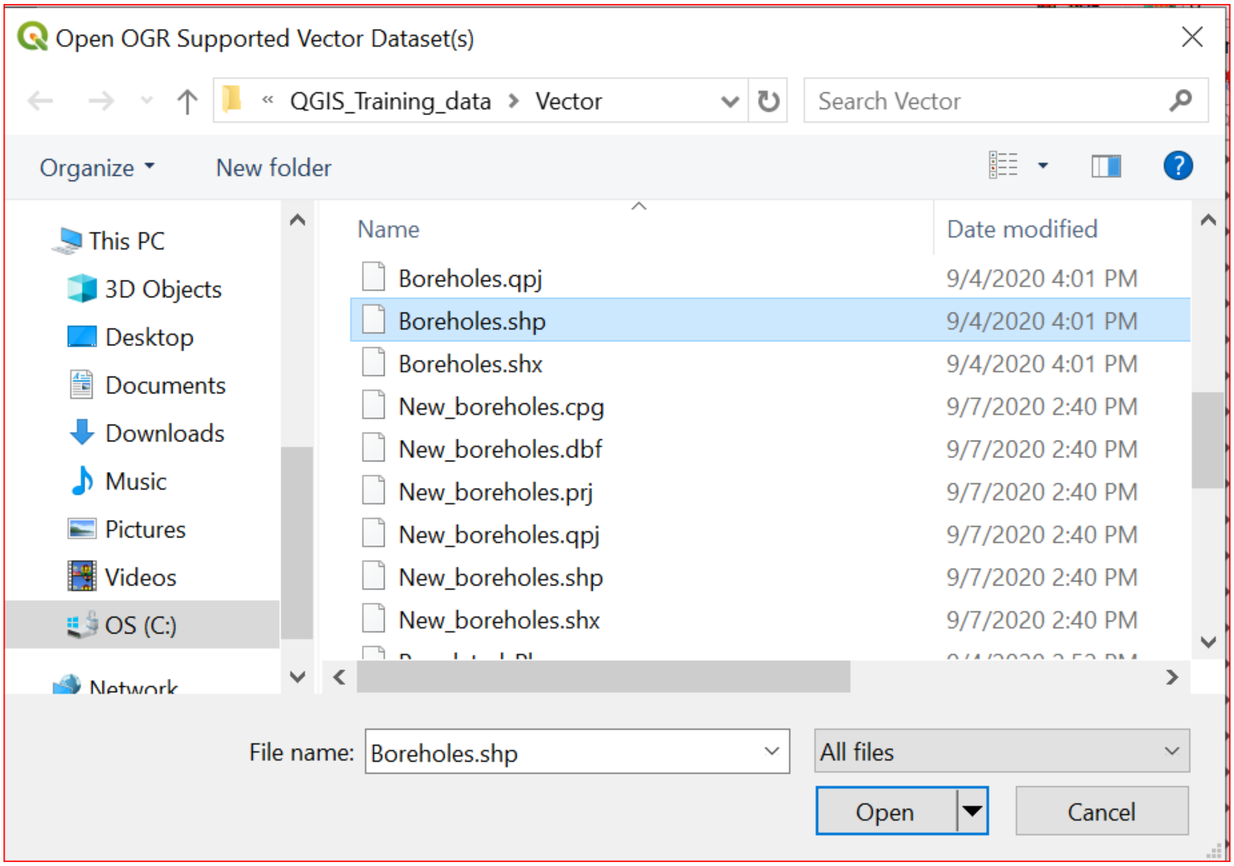
(a) Adding Points

In this exercise we are going to add points into our QGIS. The points are locations of boreholes in the study area. We are going to add using *Add Vector Layer* function in QGIS

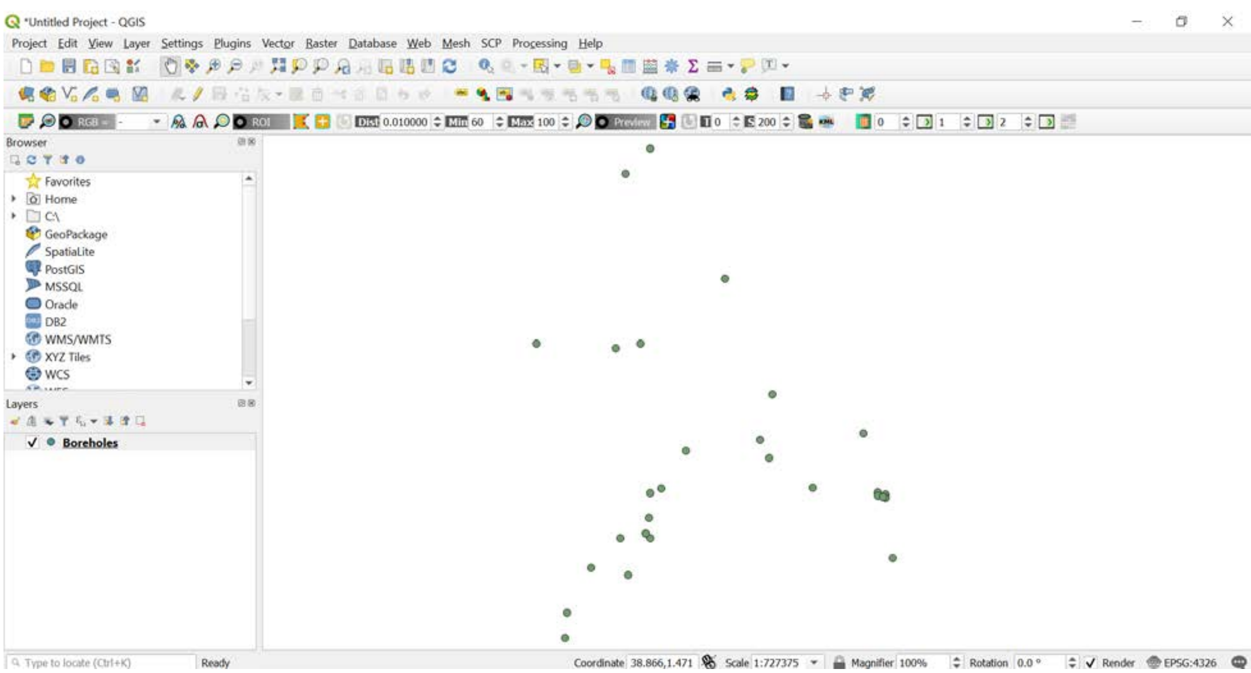
- Start QGIS
- Click *Layer* → *Add* Layer→*Add Vector Layer…*



- In the new dialogue window, use the  button to navigate to the folder with the data
- The point's shapefile is in the folder named *Vector.*
- Navigate to the folder and select *Boreholes.shp* click open.



- In the new dialogue window click add.
- The layer is added to the map canvas, It should appear similar to the one below

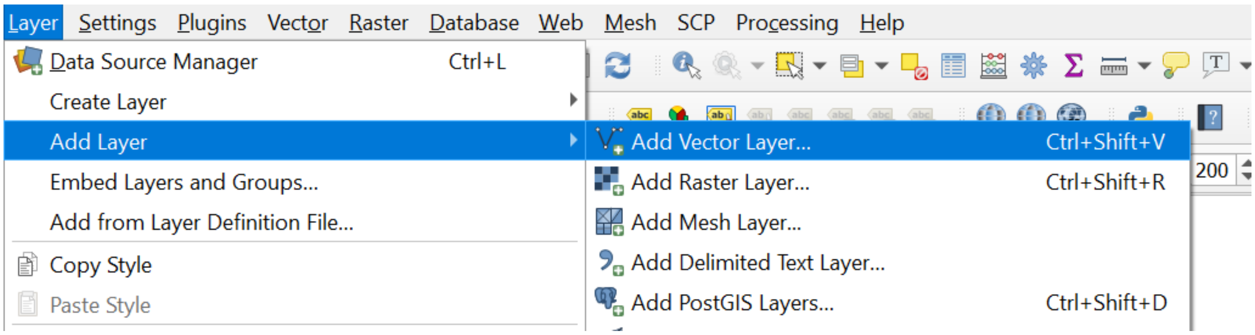


We have successfully added points to our map. In the next section we will add a Polygon.

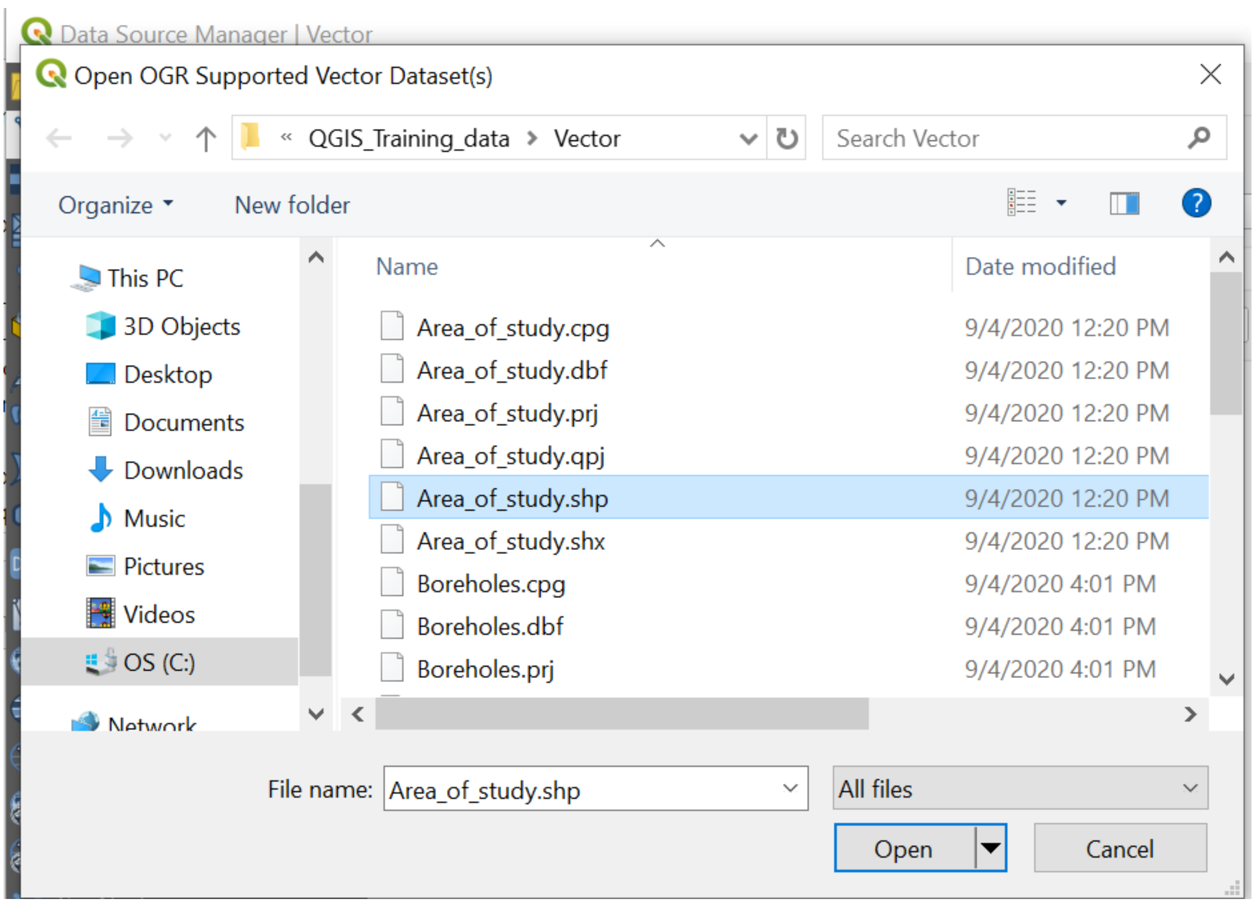
(b) Adding a Polygon

The polygon we intend to add in this section is the Area of study boundary.

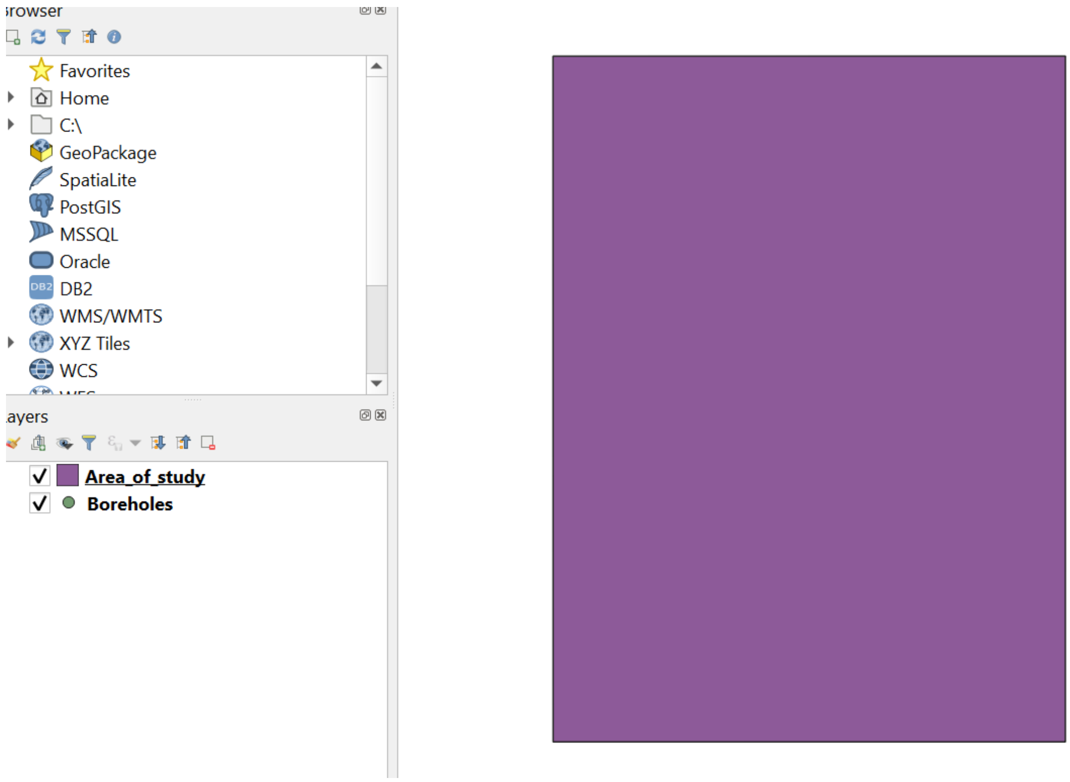
- On the upper tab click *Layers* → *Add Vector Layer…*



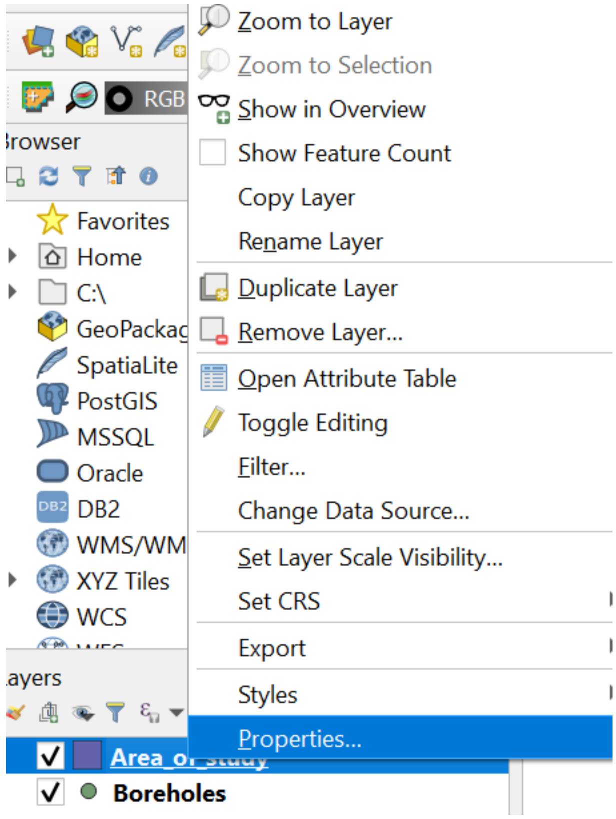
- In the dialogue use the ... button to browse to the folder with the area of study data we want to add. Choose *Area_of_study.shp* since the polygon is a shapefile.



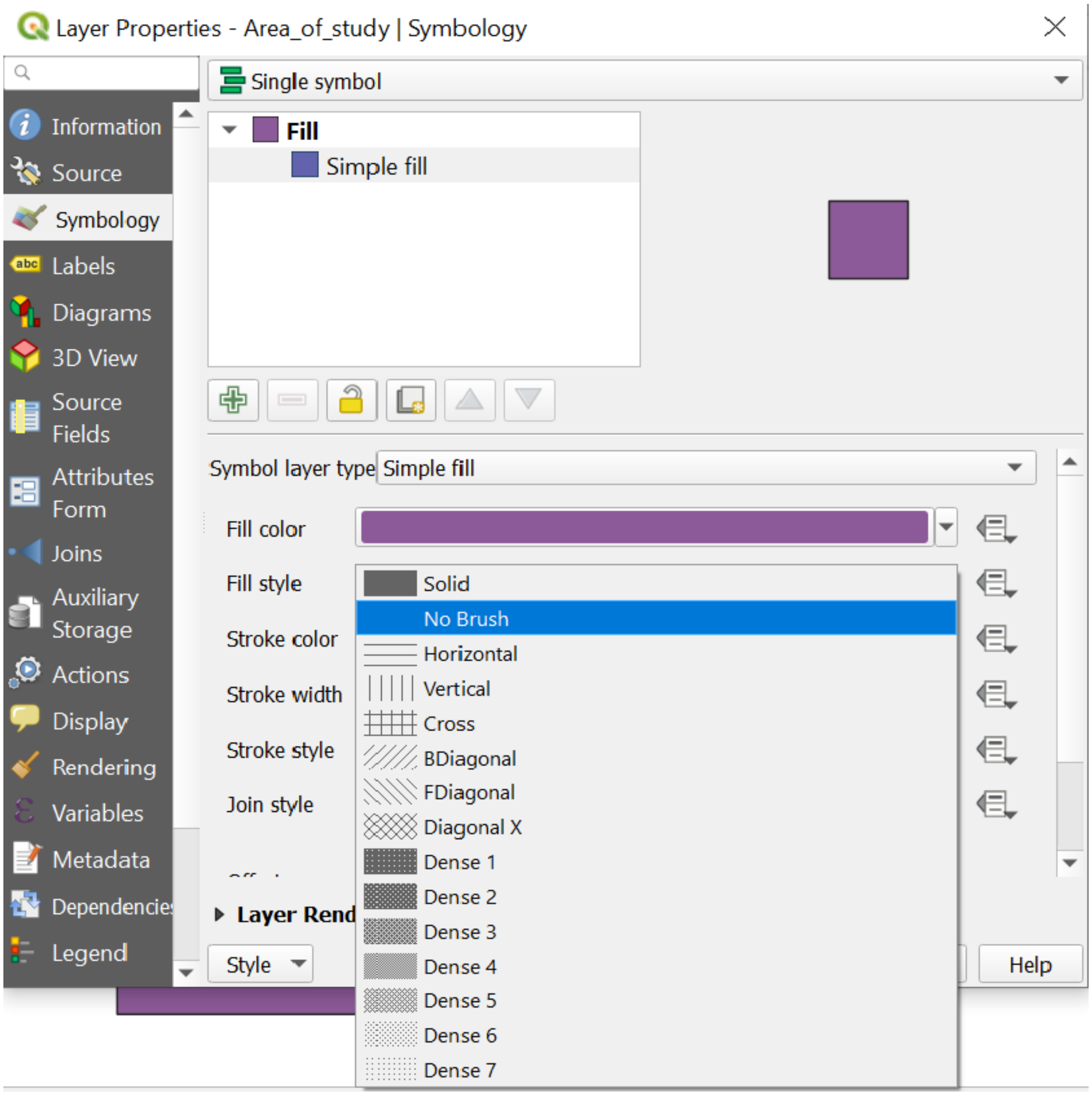
- Once selected click *Open*
- In the next window click *Add.*
- Done that, a rectangle should appear on your canvas that indicates the extent of our study area. **Note the fill color might not be similar to the one appearing below since QGIS randomly assigns a fill color to a new layer but it's okay**.



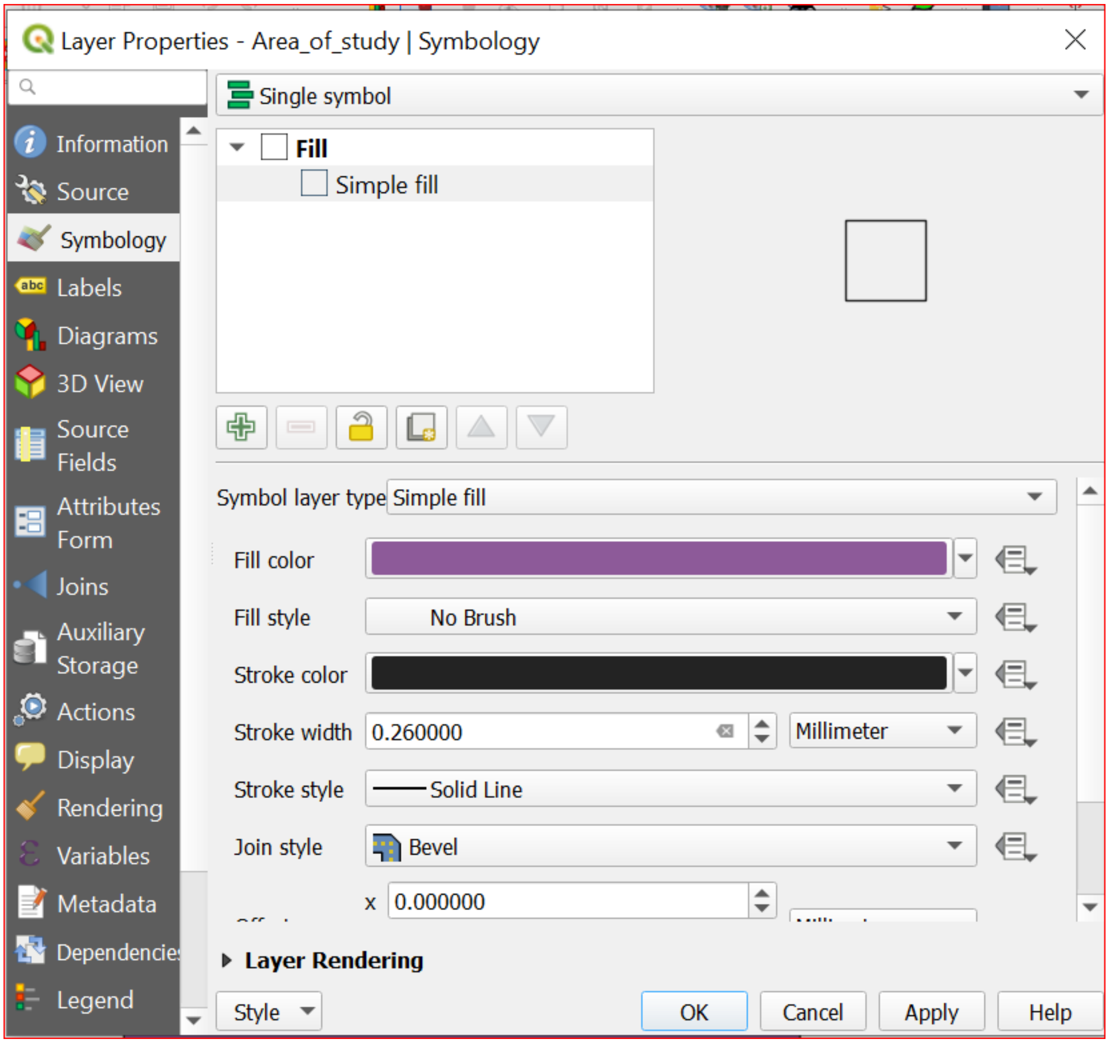
- The new layer overlays the previous one, thus we can’t see the point’s data i.e *Boreholes*. We are going to change the fill color (Symbology) of the Area of study layer in order to be able to view all the layers in the canvas.
- Click on *Area_of_study* → *Properties…*



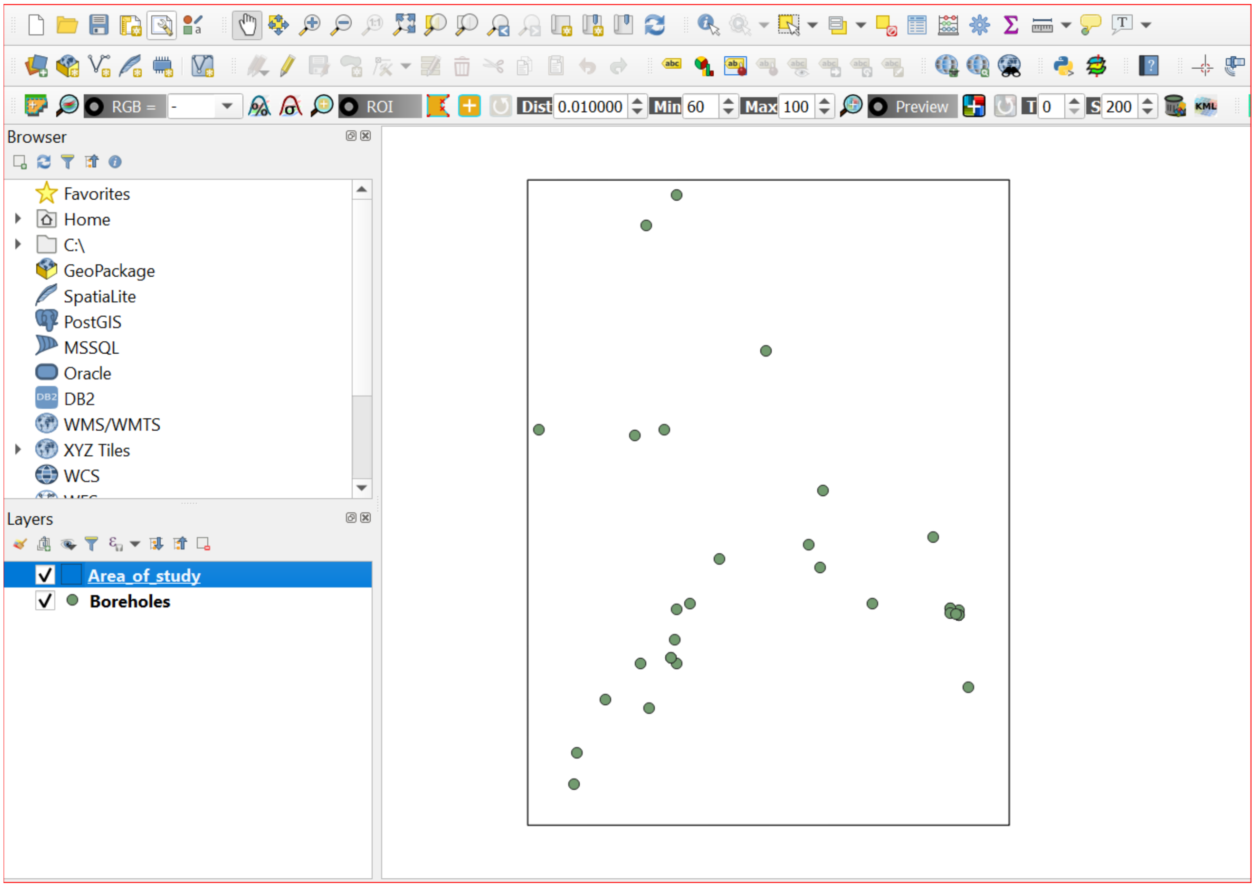
- On the new dialogue symbol Click on *Symbology* → *Simple Fill* → For ***Fill Style*** select ***No Brush***



Once selected, the dialogue window should appear similar to the one below



- On the window click *Apply* then *OK*
- The symbology of the area of study layer changes and now appears as the one below.

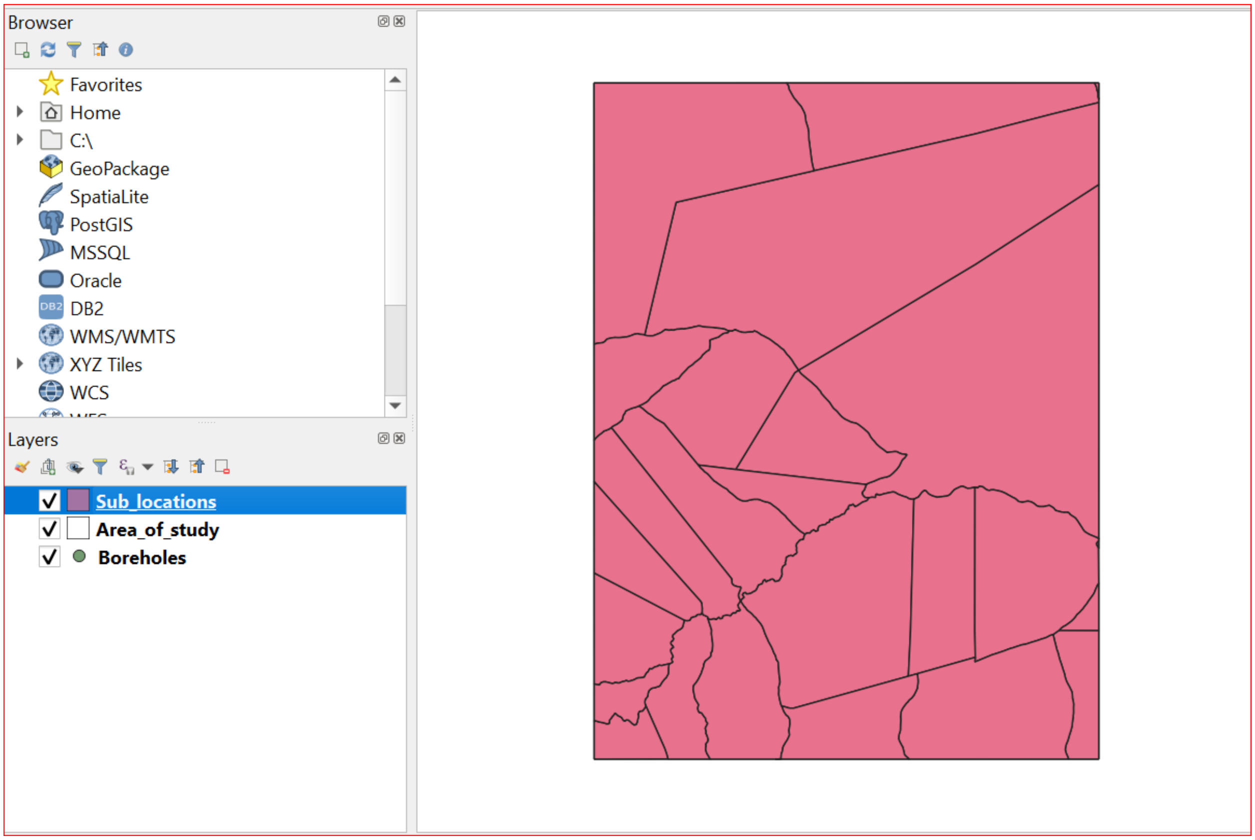


- Now we can see all the layers with the points (boreholes) inside the Polygon (Area of study) boundary.

(c) Adding another polygon data

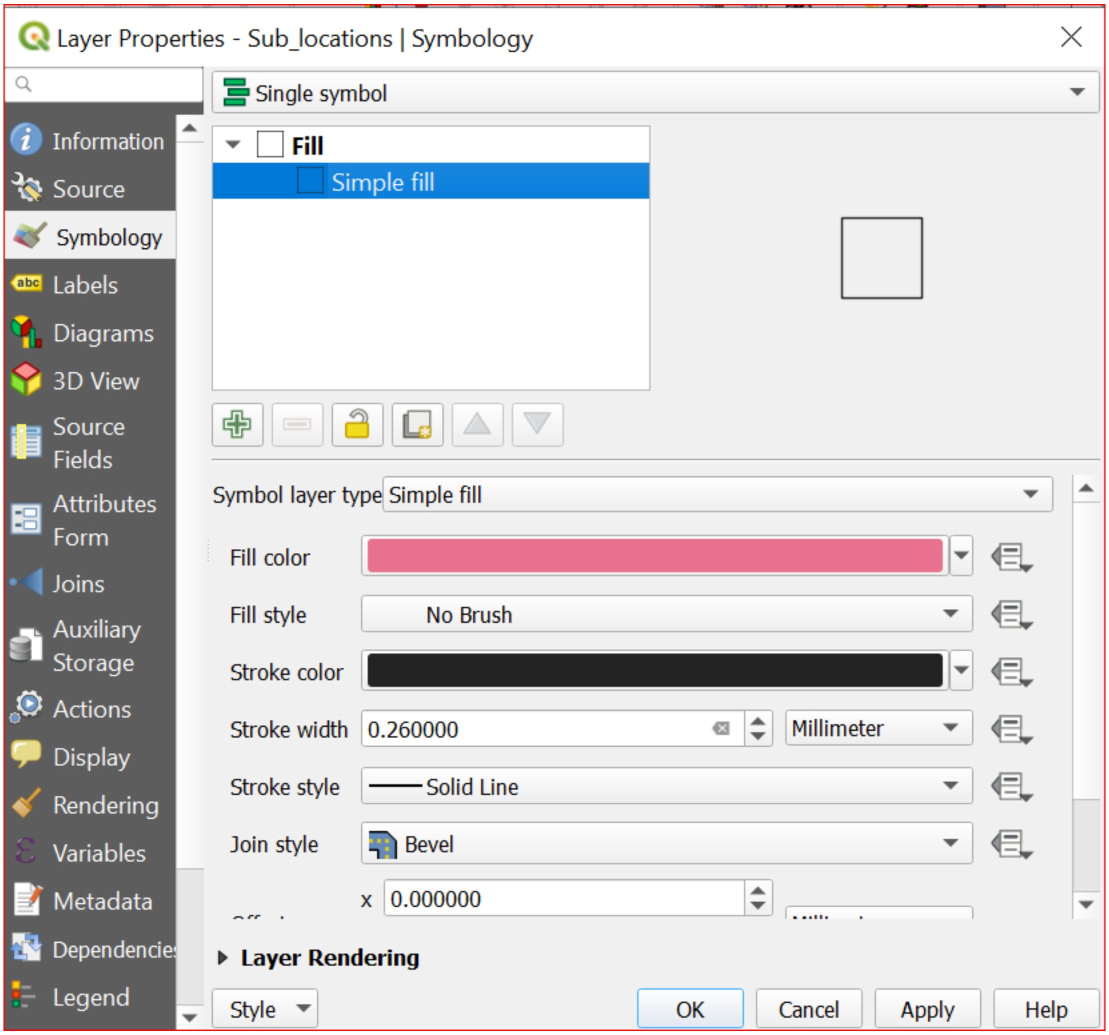
Using the same procedure as we used for the Area of study, we will *add Sub-locations polygon Layer*

- Click Layer → Add layer → Add Vector Layer
- Navigate to the folder with the data (Vector), select *Sub_locations.shp* and Add.
- The Map canvas should appear similar to this one **the color in your canvas may be different since QGIS automatically selects a color palate whenever a layer is added**.



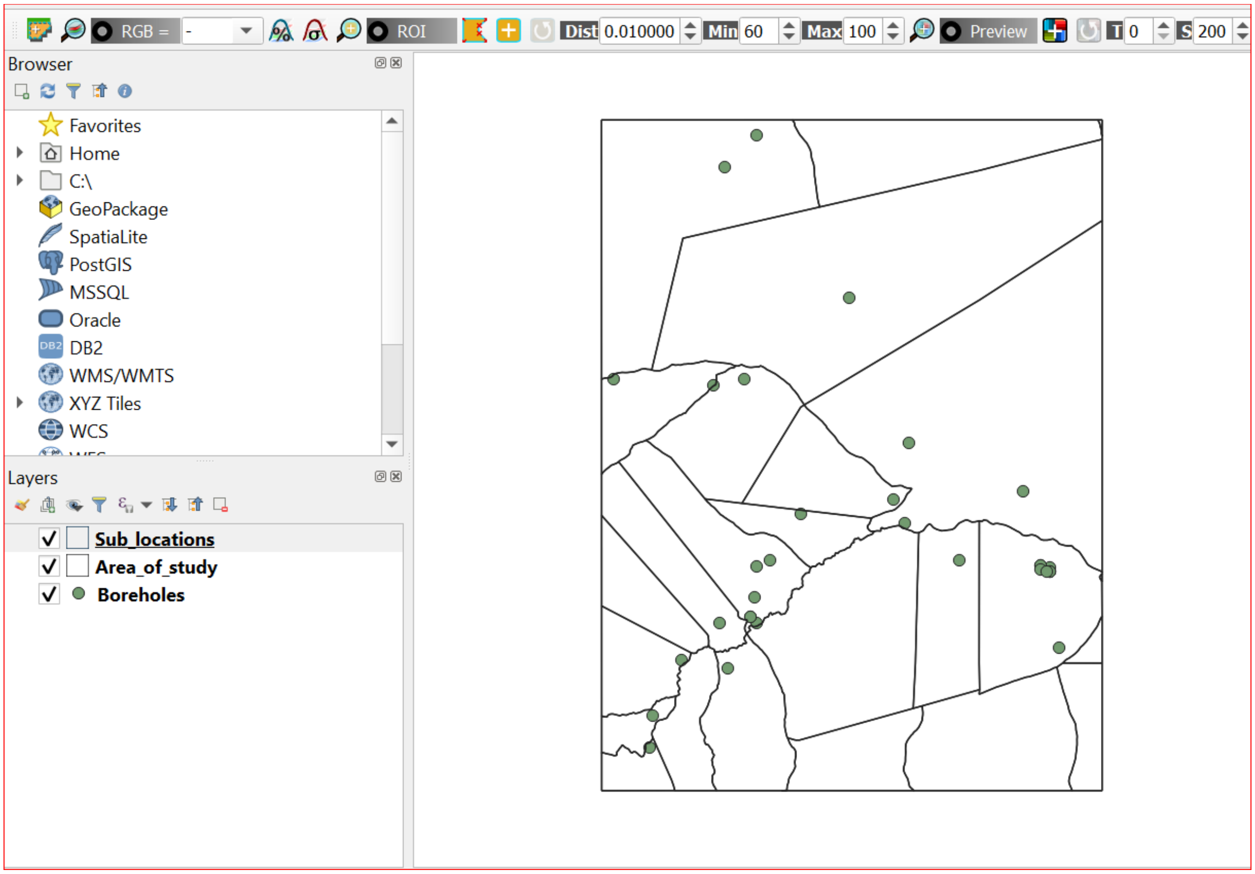
As an exercise using the same procedure we used to change the symbology of Area of study boundary, change the fill color of the Sub locations layer to enable the other feature layers become visible.

- Click Sub_locations →Properties→Symbology→Simple Fill → Fill style then select No Brush



Click Ok and Apply.

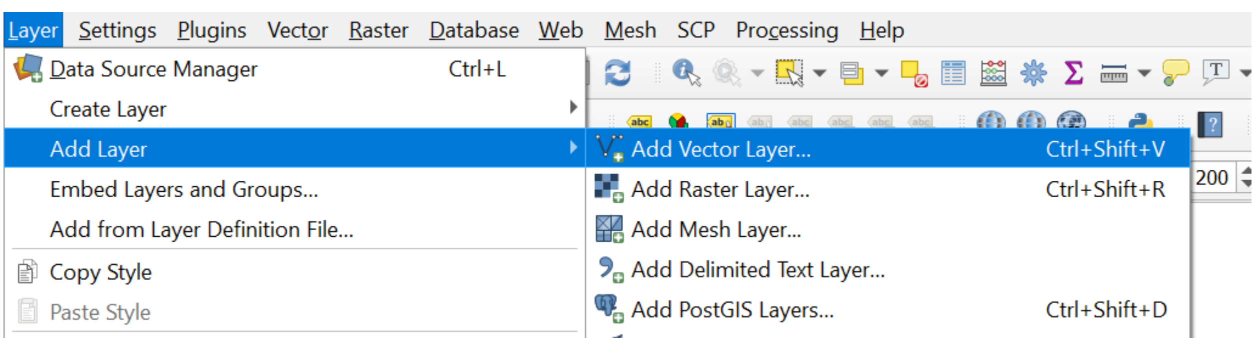
The layer in the canvas should appear similar to the one below



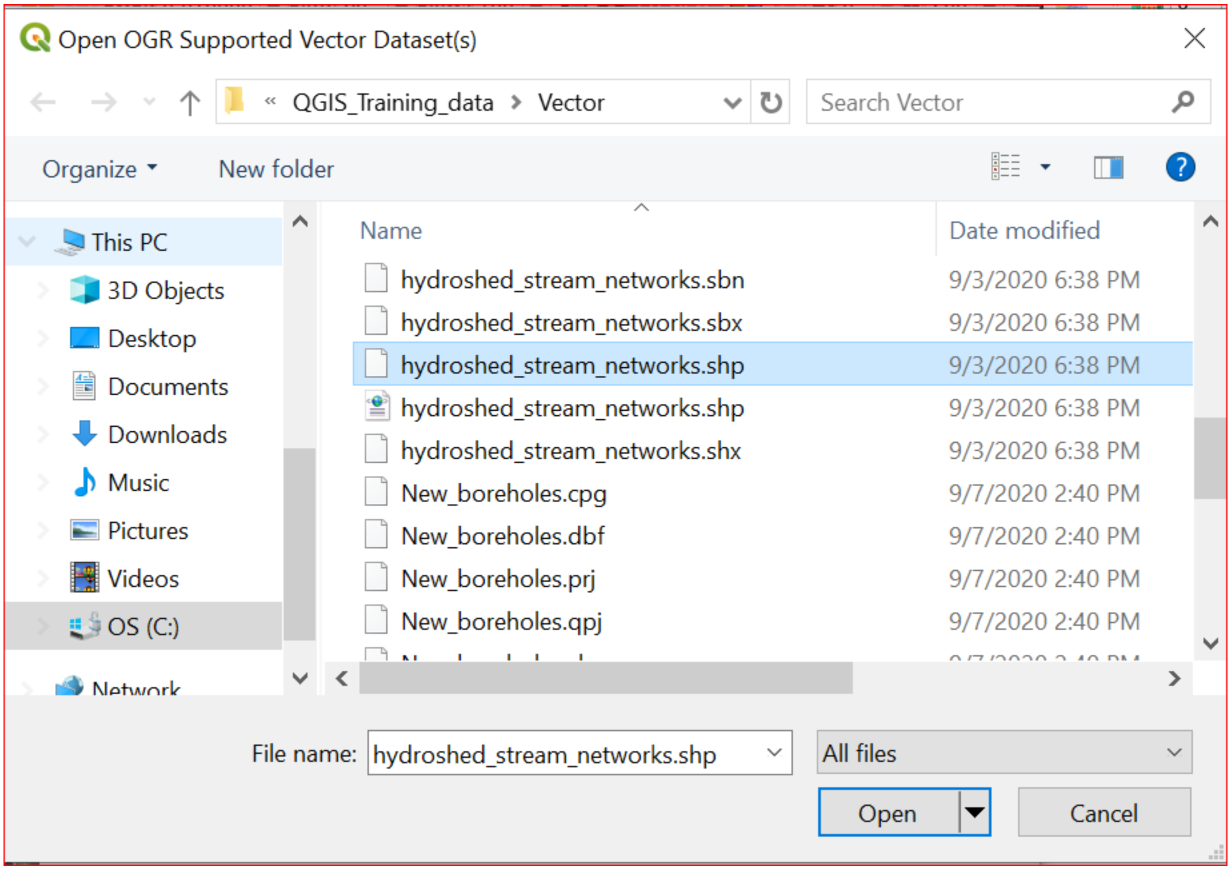
(d) Adding Lines

Here we are going to add lines (rivers shapefile layer) to the map canvas. Rivers in GIS are mostly represented as lines or Polylines in some instances. Since lines fall under vector data format, the procedure for adding them is quite similar with the procedure of adding points and polygons.

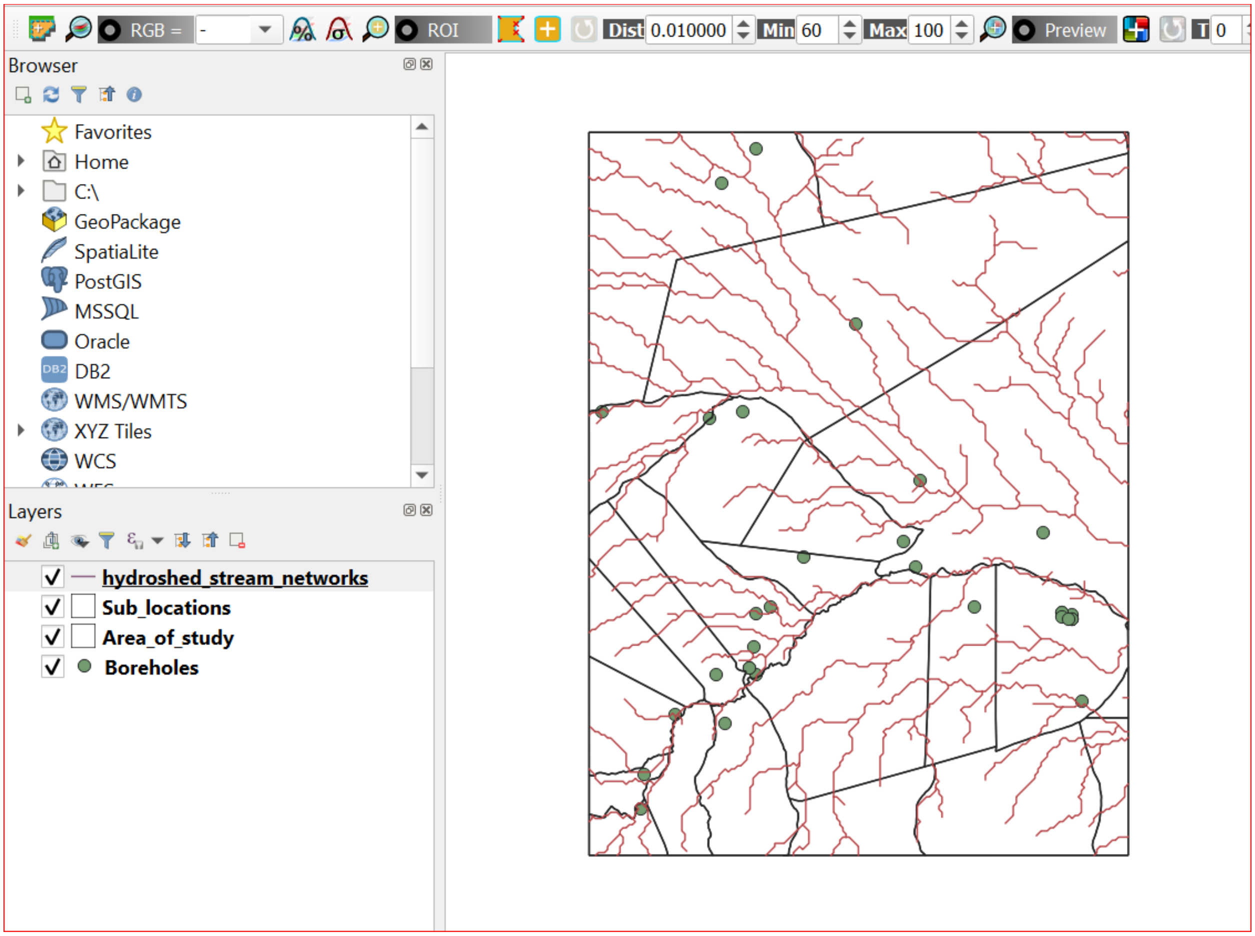
- Click on → *Layer* → *Add Layer* → *Add Vector Layer…*



- On the new dialogue window use the button to navigate to the folder with data. Our data is in *QGIS_TRAINING_DATA/Vector*, select *hydroshed_stream_networks.shp* and click Open



- In the next dialogue window click Add, then Close.
- The new layer is added to the canvas, it appears as shown below **The color of the streams in your canvas may not be identical to the one below but there is no problem as explained earlier**



- We can see the layer is added to the map canvas

Summary

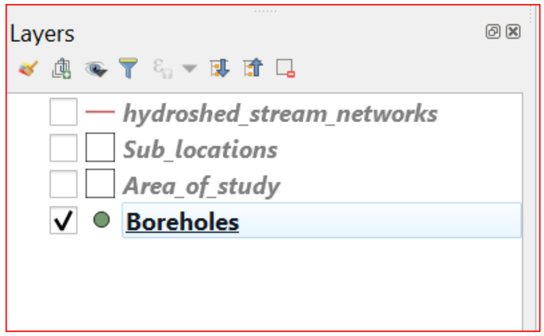
In this section we have worked on adding vector data and displaying it on a map canvas. These includes points, lines and polygons. Furthermore we have learnt how to symbolize or change the Fill Color of polygons.

We now move forward to the next section entailing Basic Vector based Spatial Analysis.

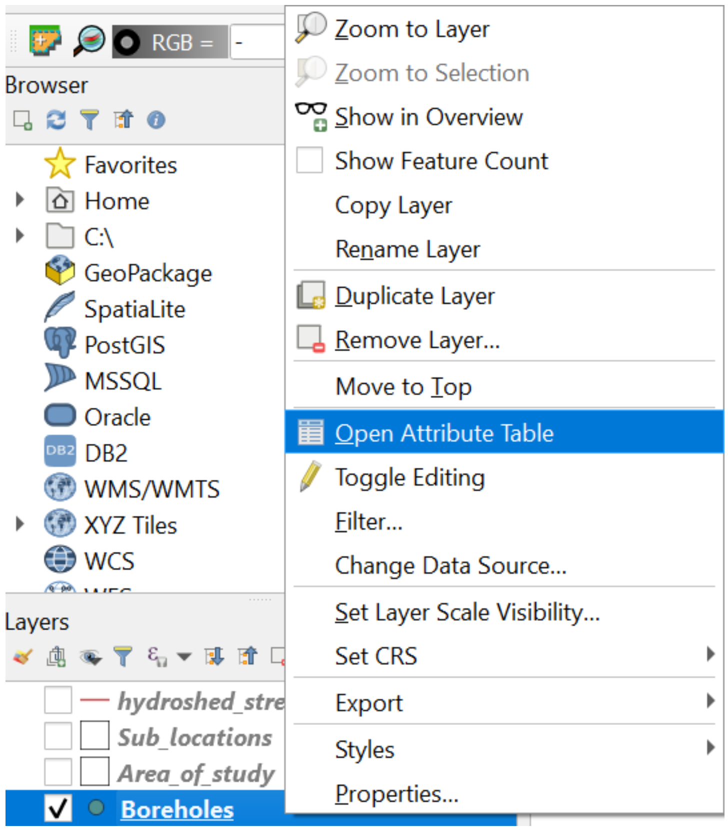
3.3 Basic vector data analysis

In this section we are going to carry out Query analysis on the boreholes, to view and extract operational boreholes from the points we have.

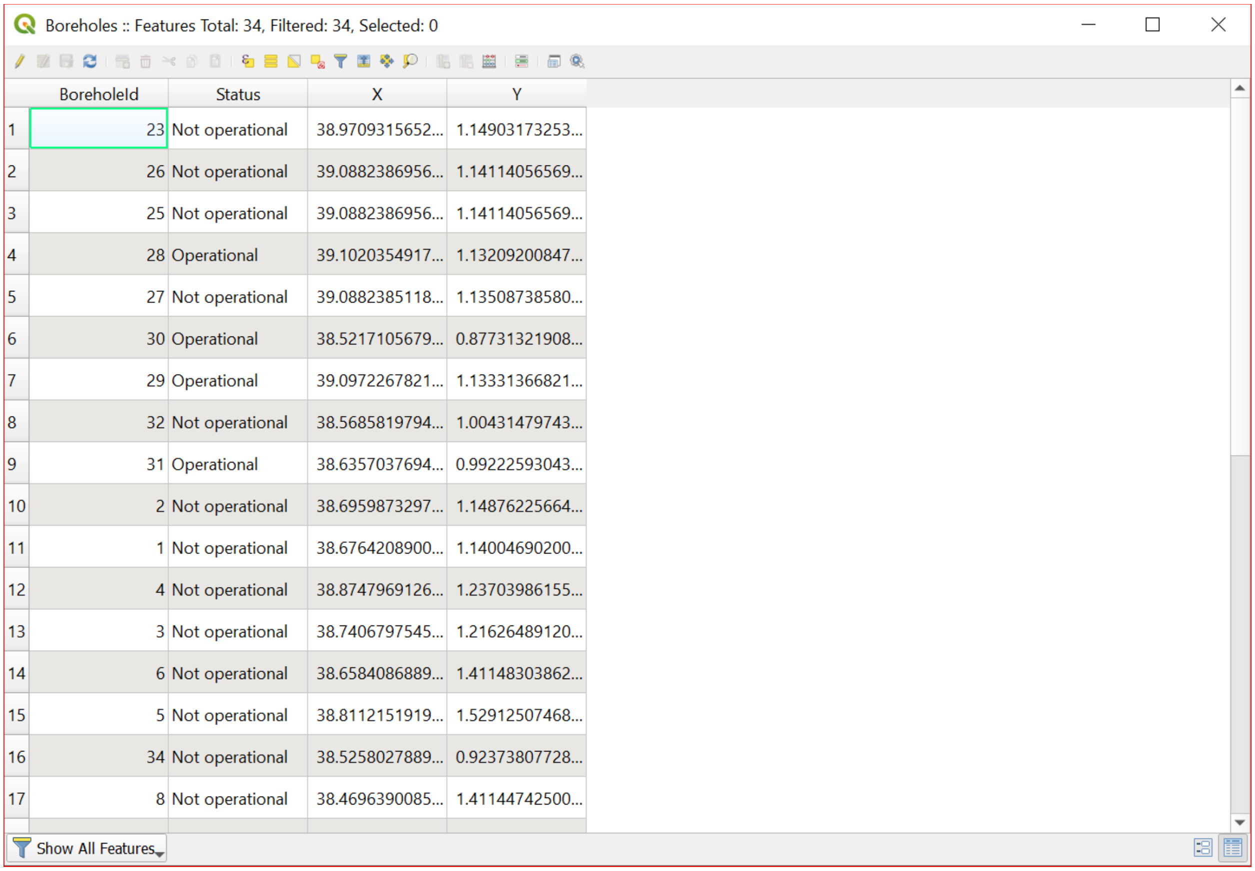
- Next we carry on a Simple query analysis to identify operational and non-operational boreholes.
- Turnoff all the other layers by unchecking the box before layers except Boreholes



- Right Click on *Boreholes* → *Open Attribute Table*



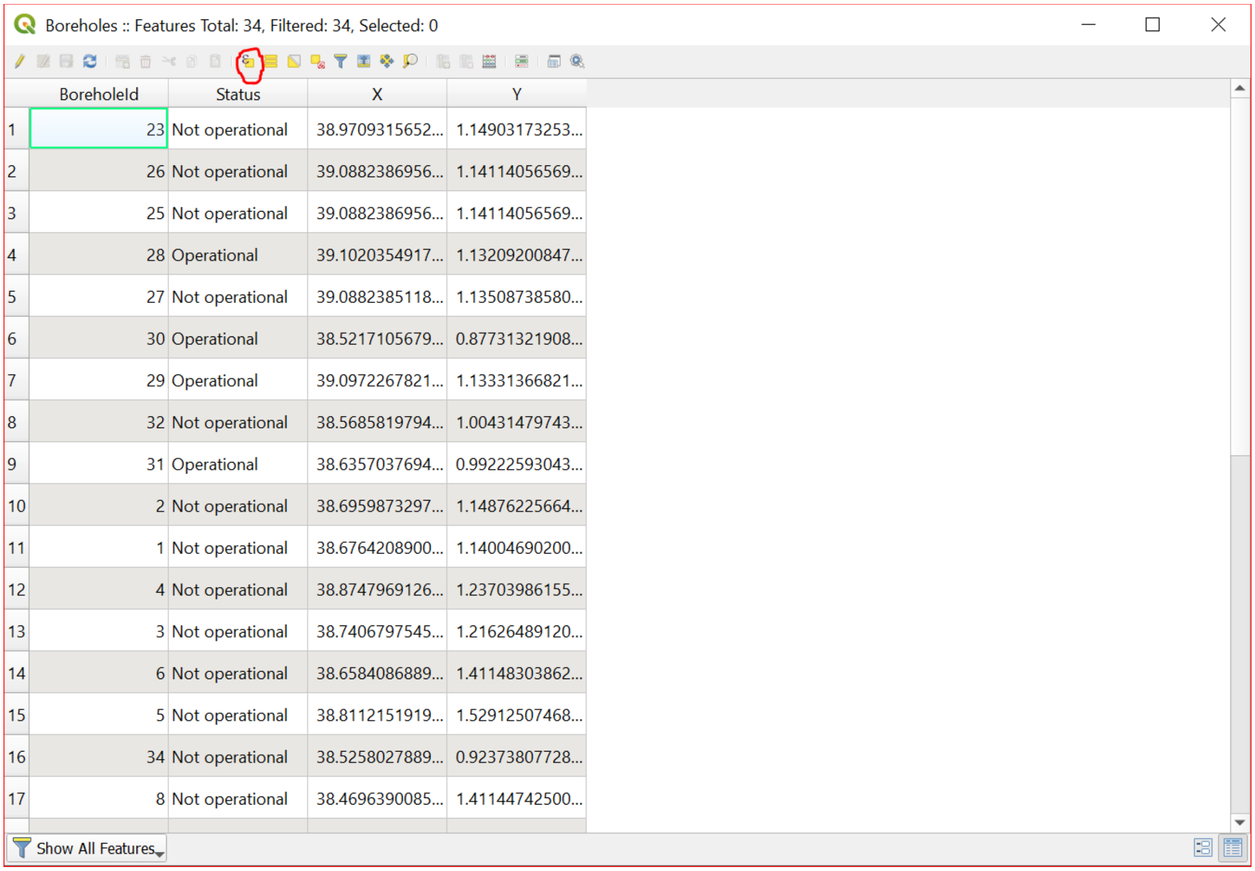
- The attribute table should be similar to the one below.


Boreholes :: Features Total: 34, Filtered: 34, Selected: 0

	BoreholeId	Status	X	Y
1	23	Not operational	38.9709315652...	1.14903173253...
2	26	Not operational	39.0882386956...	1.14114056569...
3	25	Not operational	39.0882386956...	1.14114056569...
4	28	Operational	39.1020354917...	1.13209200847...
5	27	Not operational	39.0882385118...	1.13508738580...
6	30	Operational	38.5217105679...	0.87731321908...
7	29	Operational	39.0972267821...	1.13331366821...
8	32	Not operational	38.5685819794...	1.00431479743...
9	31	Operational	38.6357037694...	0.99222593043...
10	2	Not operational	38.6959873297...	1.14876225664...
11	1	Not operational	38.6764208900...	1.14004690200...
12	4	Not operational	38.8747969126...	1.23703986155...
13	3	Not operational	38.7406797545...	1.21626489120...
14	6	Not operational	38.6584086889...	1.41148303862...
15	5	Not operational	38.8112151919...	1.52912507468...
16	34	Not operational	38.5258027889...	0.92373807728...
17	8	Not operational	38.4696390085...	1.41144742500...

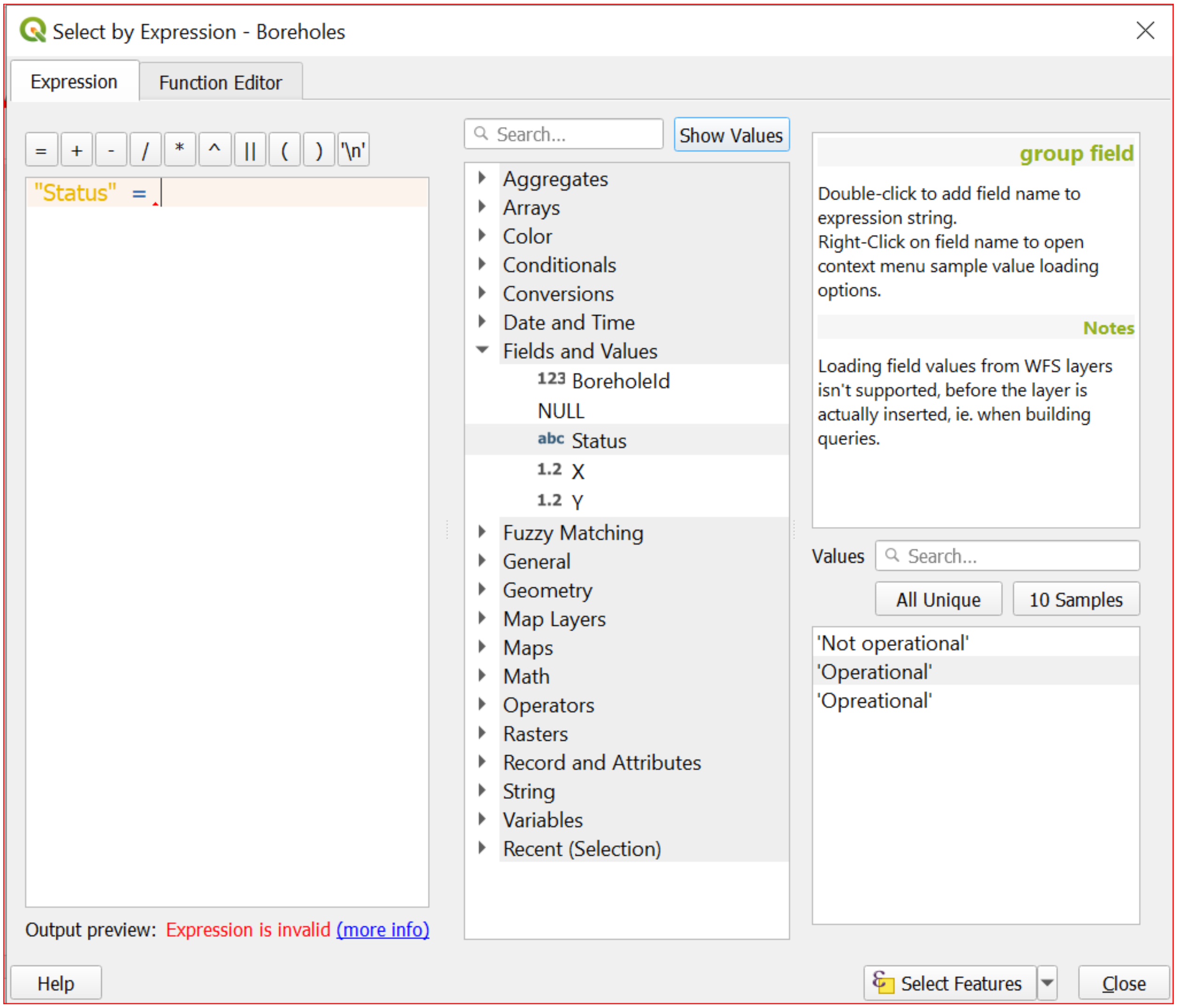
Show All Features

- On the attribute table we can see that there are operational and non-operational boreholes.
- On the Menu bar on the top click on the ***Expression tool.*** (**The tool is circled red in the next screenshot*)*



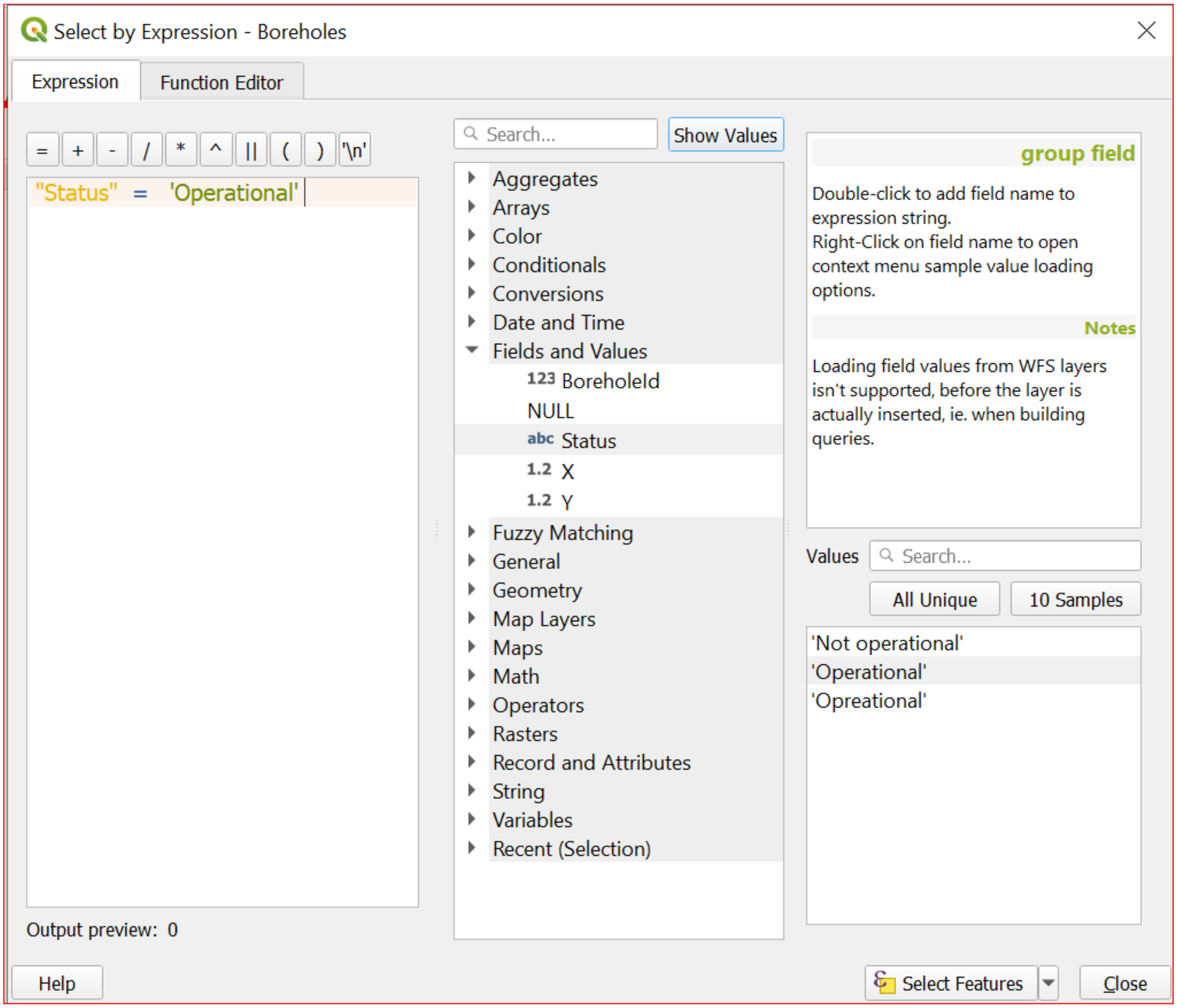
	BoreholeId	Status	X	Y
1	23	Not operational	38.9709315652...	1.14903173253...
2	26	Not operational	39.0882386956...	1.14114056569...
3	25	Not operational	39.0882386956...	1.14114056569...
4	28	Operational	39.1020354917...	1.13209200847...
5	27	Not operational	39.0882385118...	1.13508738580...
6	30	Operational	38.5217105679...	0.87731321908...
7	29	Operational	39.0972267821...	1.13331366821...
8	32	Not operational	38.5685819794...	1.00431479743...
9	31	Operational	38.6357037694...	0.99222593043...
10	2	Not operational	38.6959873297...	1.14876225664...
11	1	Not operational	38.6764208900...	1.14004690200...
12	4	Not operational	38.8747969126...	1.23703986155...
13	3	Not operational	38.7406797545...	1.21626489120...
14	6	Not operational	38.6584086889...	1.41148303862...
15	5	Not operational	38.8112151919...	1.52912507468...
16	34	Not operational	38.5258027889...	0.92373807728...
17	8	Not operational	38.4696390085...	1.41144742500...

- On the new dialogue window expand the Fields and Values section and Select *Status and the equal to (=) operator on the top.*

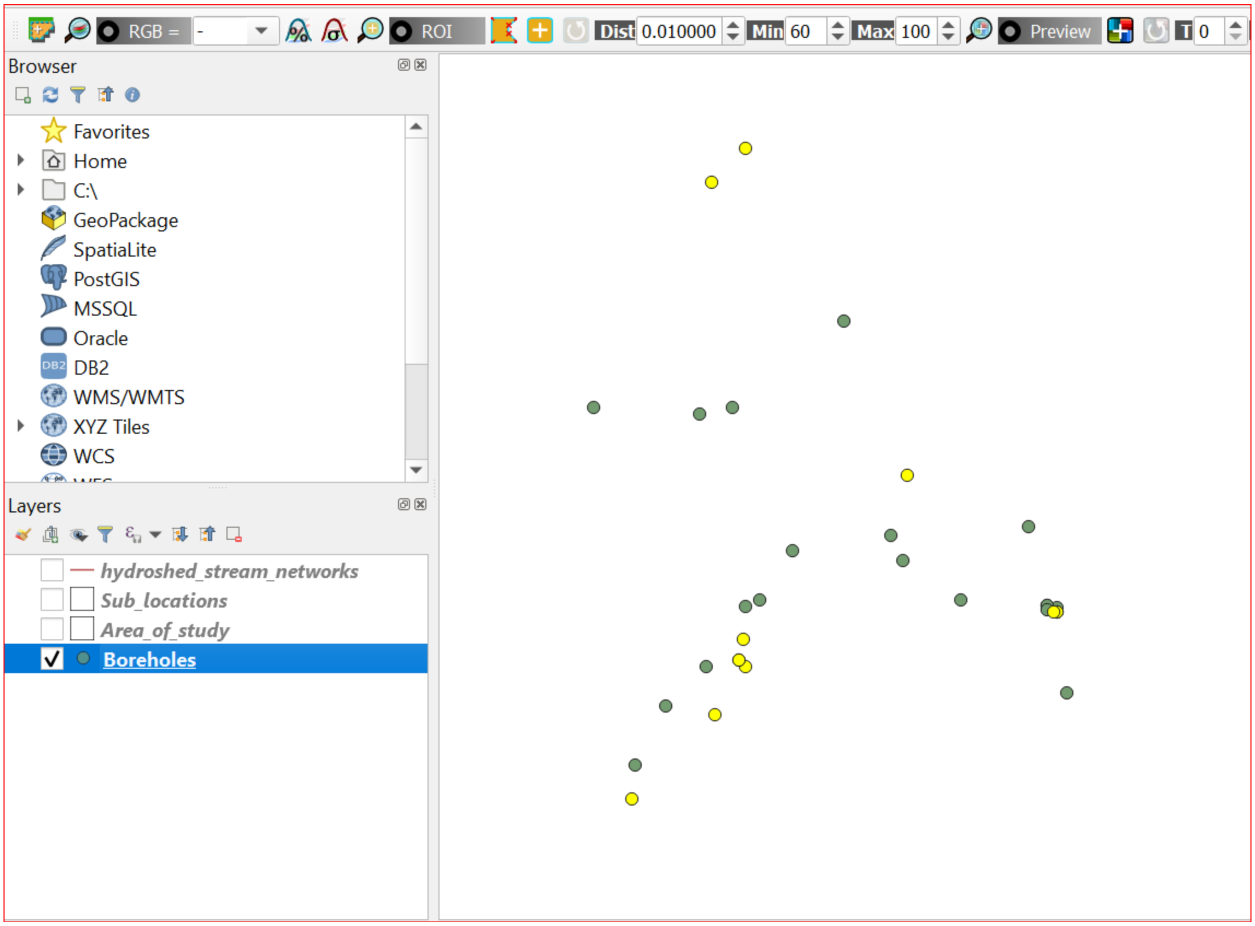


- On the right side of the window, under Values click *All Unique* and double click on *Operational* and *Select Features* then *Close.*

Here we have created an expression that tells QGIS to get us all the boreholes that have indicated Status as Operational

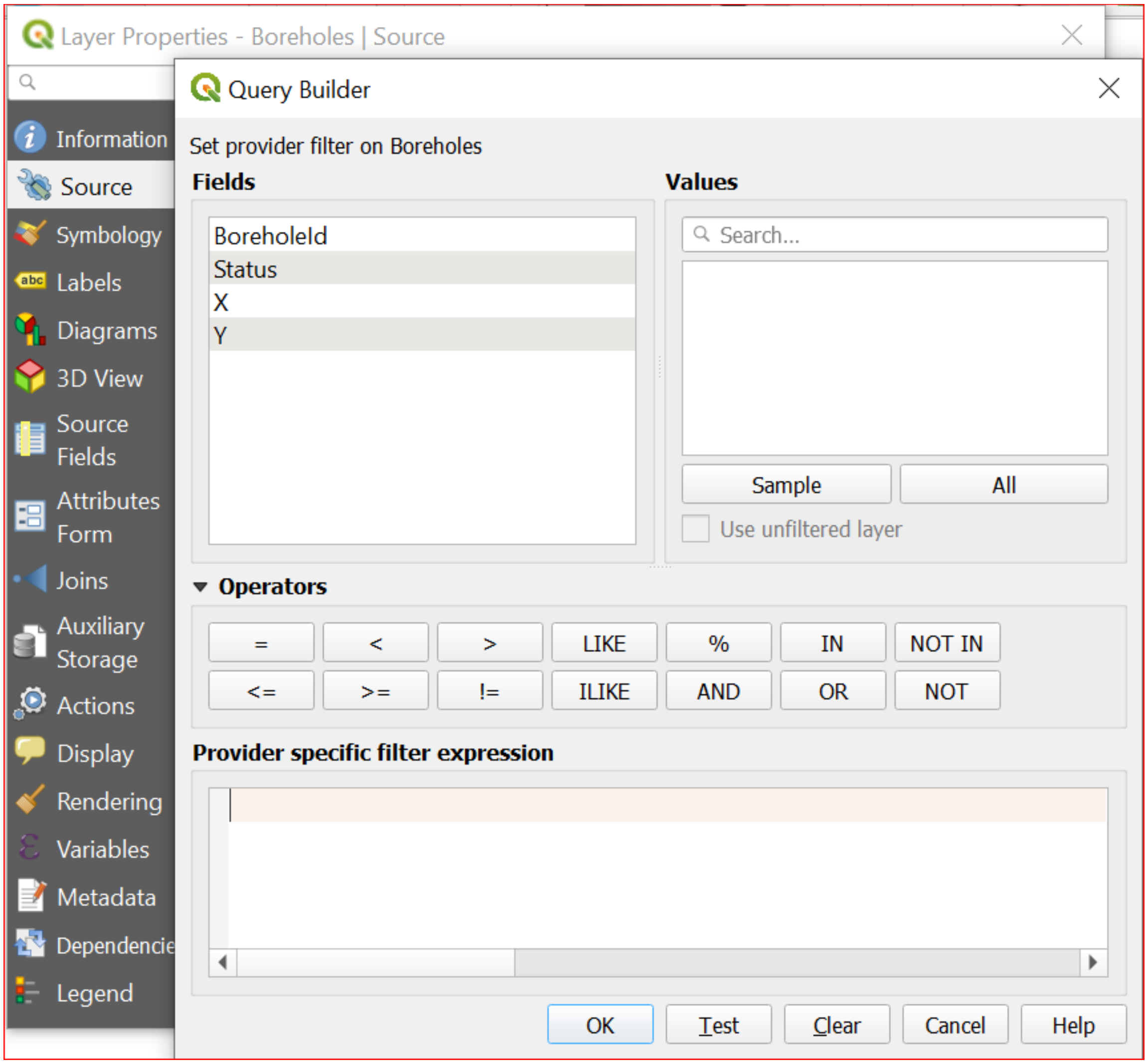


Close the attribute table, on the canvas we can see that some of the boreholes color has changed to yellow. These are the boreholes that are operational. Remember our query was *Status = Operational*; as a result QGIS has selected the operational boreholes and automatically symbolized them with a yellow color.

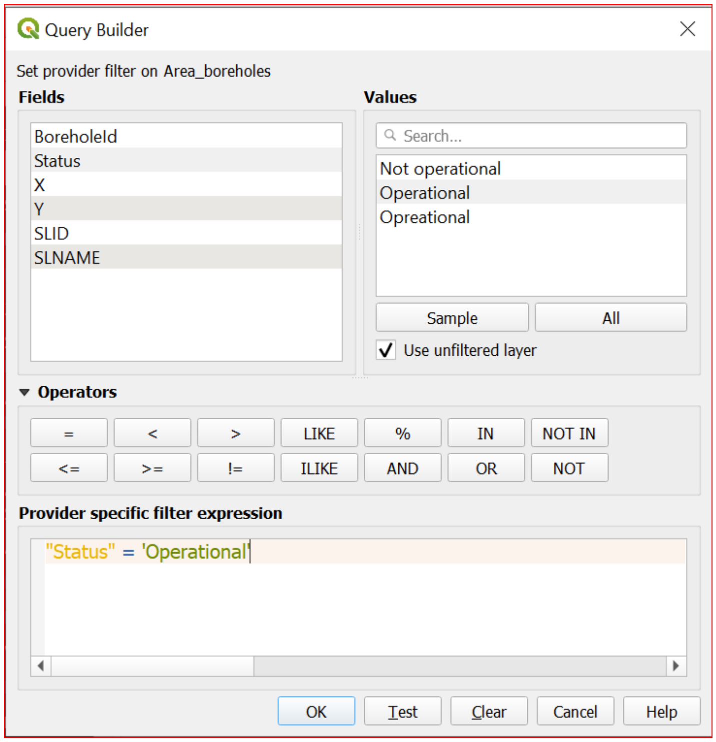


Now we are going to save the operational boreholes to use them for further analysis

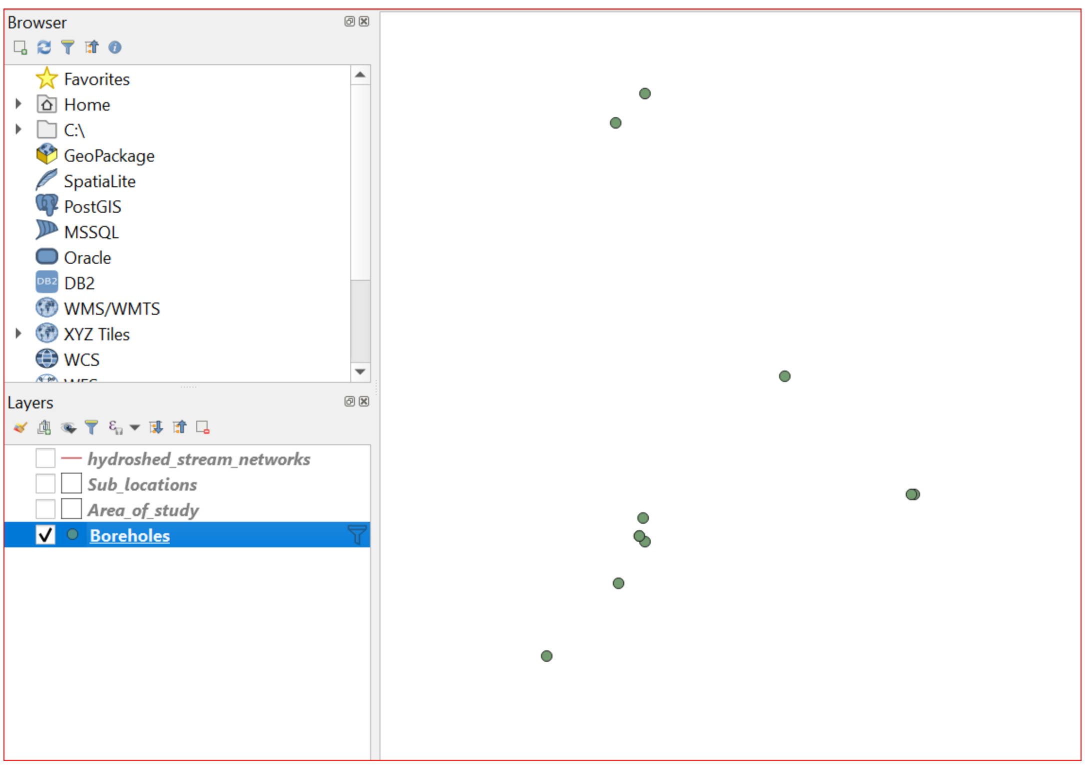
- On the layers tab click on *Boreholes→ Properties →Source→ Query Builder*



- On the Query Builder type the same function we used earlier as show below



- Click Ok on both the current window and previous window to close.
- The other points on the canvas (non-operational boreholes) disappear leaving only the operational boreholes that were extracted from the expression.



Note that on the canvas, we have a layer of operational boreholes only.

Now, export the layer and save as *Operational_boreholes* shapefiles.

NB: Whenever you save files with more than two words in QGIS or ArcGIS, you are required to use an <u>underscore</u> instead of a space while naming the file/shapefile. This is because having a space between words may result to errors during automation processes.

Summary

In this section we have learnt to add vector data to QGIS, change symbology and carry out a simple query on the attributes of features. A query analysis helps us to select an aspect or feature of interest in the attribute table of our feature datasets. For instance we have been able to select operational boreholes from the status attribute of all boreholes in our study area.

Note that in this section, we only queried the attributes of the data. In the next section, we will use geographic aspects of the data to carry out spatial analysis.

3.4 Spatial joins, intersection and distance-based analysis

Spatial analysis is defined as a way of looking at the geographical patterns of data and analyze the relationships between the entities.

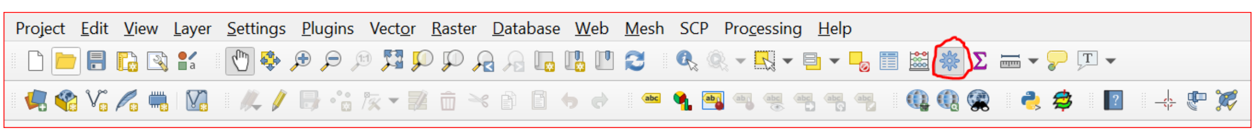
In this exercise, there are 3 main tasks;

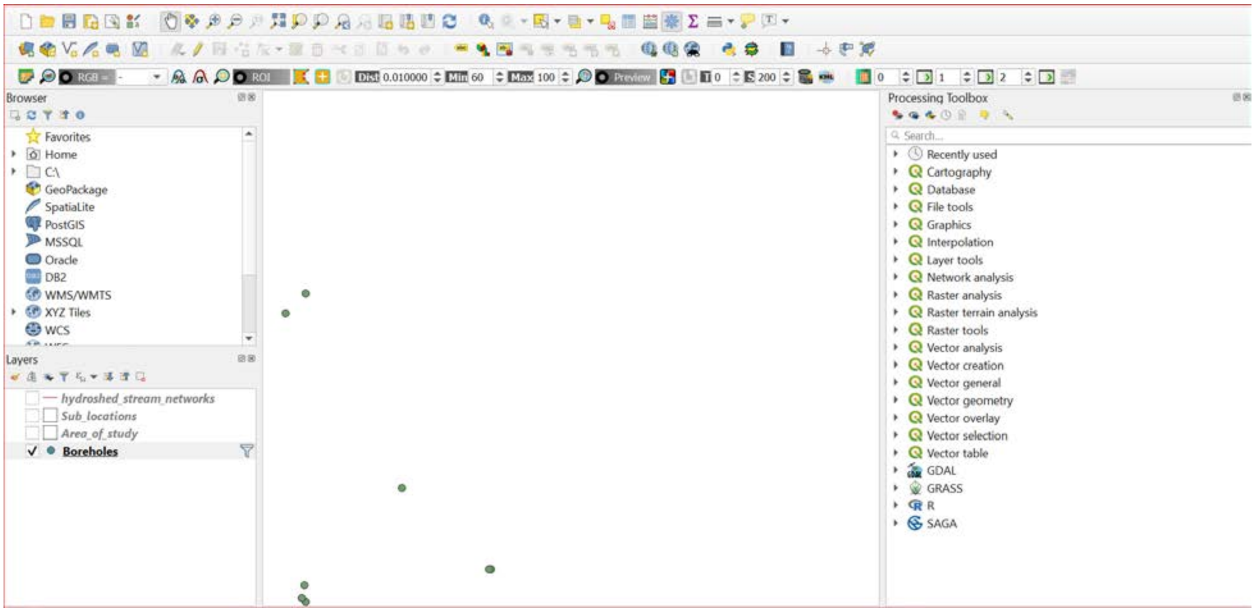
- Adding sub-location name to each borehole,
- Assessing the number of boreholes per each sub-location,
- Assess the distance of boreholes from populated places.

3.4.1. Adding sub-locations names to each boreholes

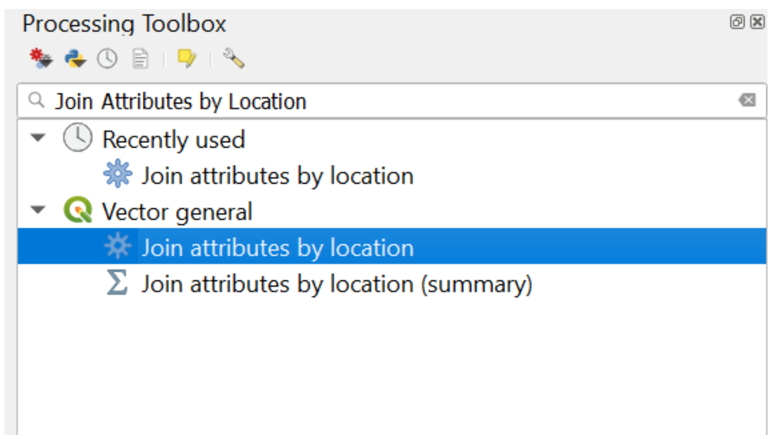
The main goal in this section is to add the sub-location name to each borehole and thus you are required to conduct a spatial join. In QGIS the tool for conducting spatial join is called ***JOIN ATTRIBUTE BY LOCATION.***

- On the top most tab, click on processing tools icon , a new dialogue window appears on the right side of the canvas (The shortcut is Ctrl+Alt+T)

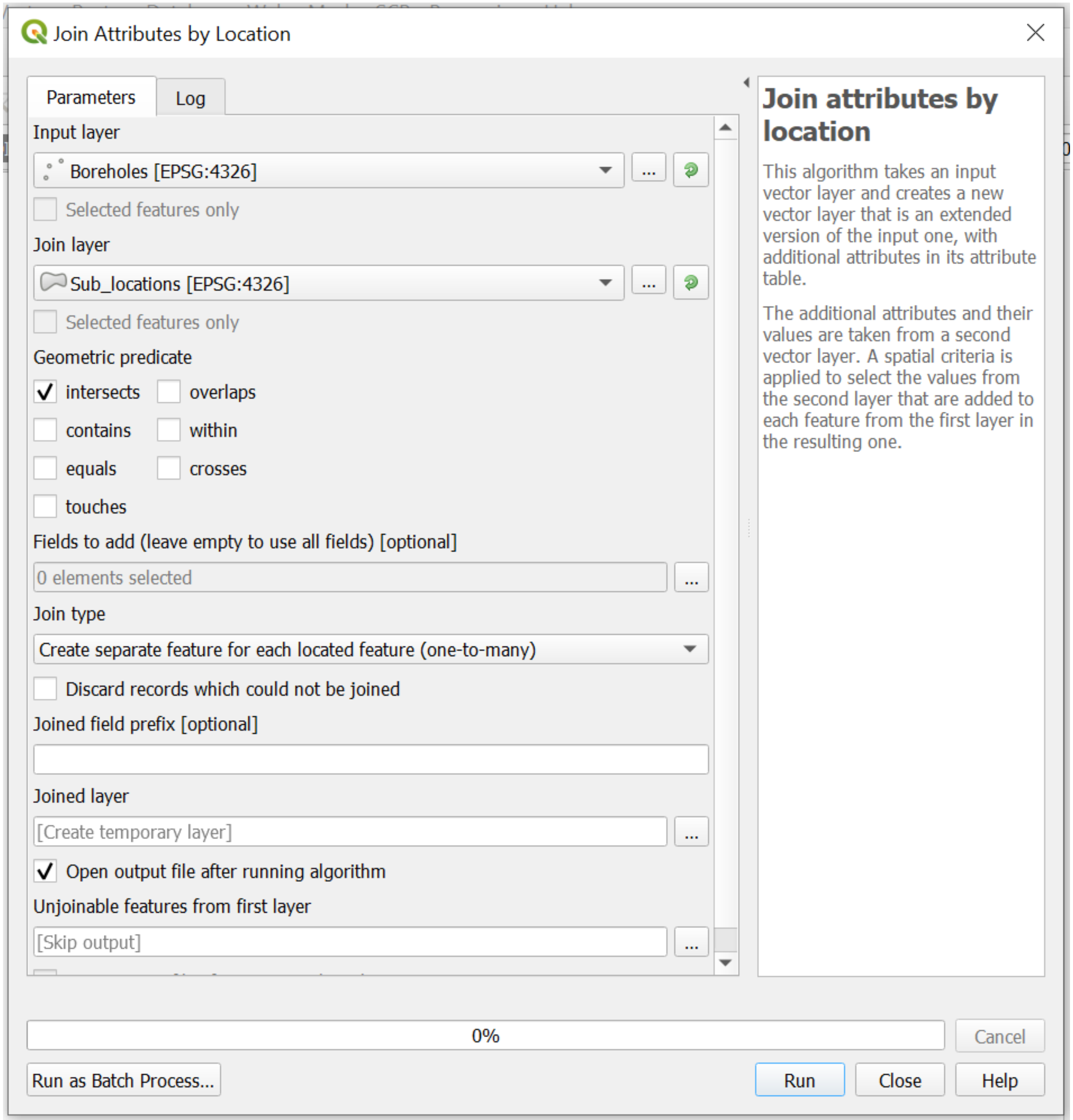




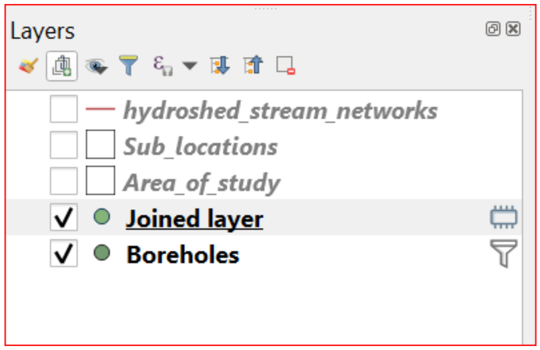
- In the search bar on the processing toolbox, type ***Join Attributes by Location***
- The tool appears as below, double click the tool to open.



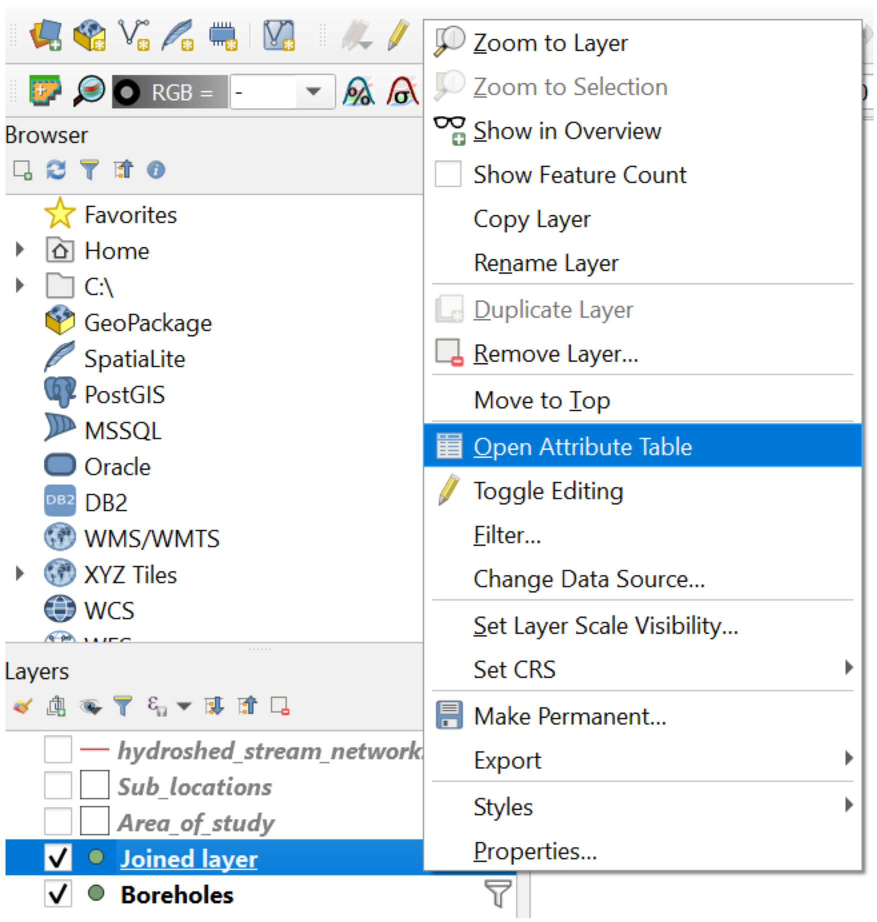
- Input the following Parameters; **Input layer** select ***Boreholes [EPSG: 4326],*** for **Join layer** select ***Sub locations[EPSG:4326]→RUN.***



- After the tool runs successfully click **Close**
- On the layers tab, a new layer named ***Joined layer*** should appear similar to the one below.



- To see if the sub location name is added to the boreholes layer, open and view the attribute table
- Right click on the ***Joined layer*** → ***Open Attribute Table***

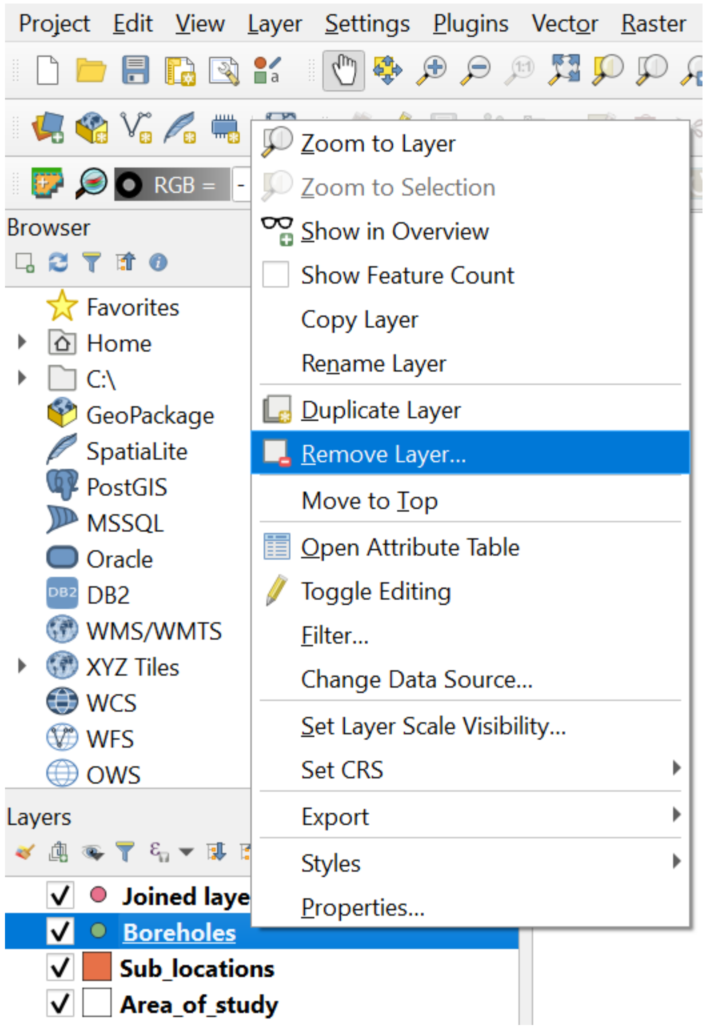


- In the attribute table a new column named **SLNAME** has been added in the far right. **SLNAME** is a short form for Sub-location name.

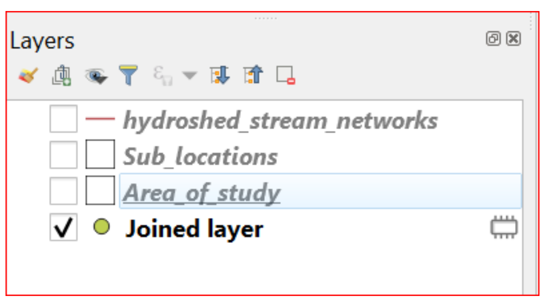
Joined layer :: Features Total: 11, Filtered: 11, Selected: 0

	BoreholeId	Status	X	Y	SLID	SLNAME
1	11	Operational	38.6764296550...	1.05861517711...	402050301.000...	MERTI NORTH
2	10	Operational	38.6737294176...	1.09480676631...	402050301.000...	MERTI NORTH
3	13	Operational	38.6674407221...	1.06766218342...	402050301.000...	MERTI NORTH
4	12	Operational	38.6674407221...	1.06766218342...	402050301.000...	MERTI NORTH
5	29	Operational	39.0972267821...	1.13331366821...	402060201.000...	SERICHO
6	28	Operational	39.1020354917...	1.13209200847...	402060201.000...	SERICHO
7	31	Operational	38.6357037694...	0.99222593043...	402040202.000...	MUCHURO
8	14	Operational	38.8969326961...	1.31900744934...	402050101.000...	MALKAGALLA
9	30	Operational	38.5217105679...	0.87731321908...	402040301.000...	MALKADAKA
10	16	Operational	38.6310100908...	1.71866377271...	402050203.000...	DUMA
11	15	Operational	38.6763321010...	1.76435498136...	402050203.000...	DUMA

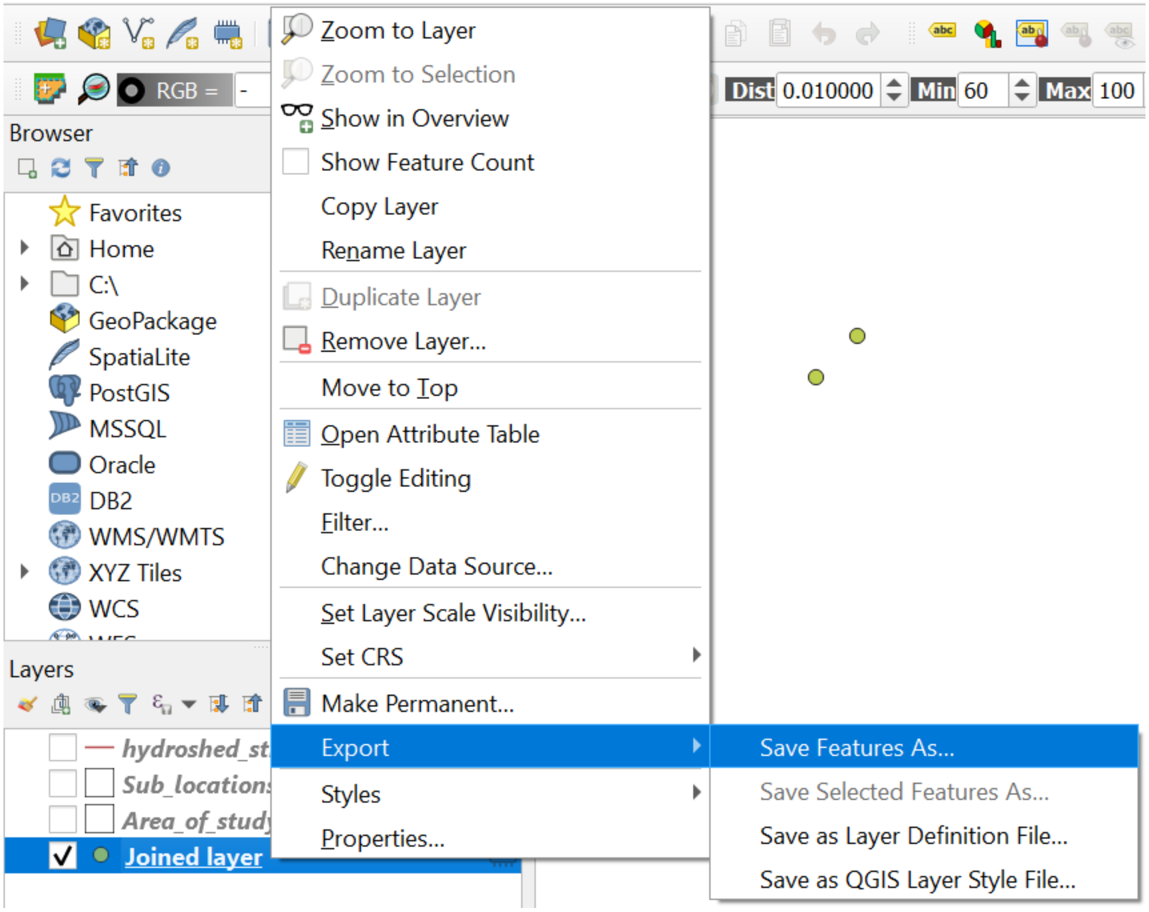
- Since the new boreholes layer has a sub location name for each borehole, save it and remove the previous layer.
- To remove Boreholes layer, right click on ***Boreholes layer →Remove Layer…***



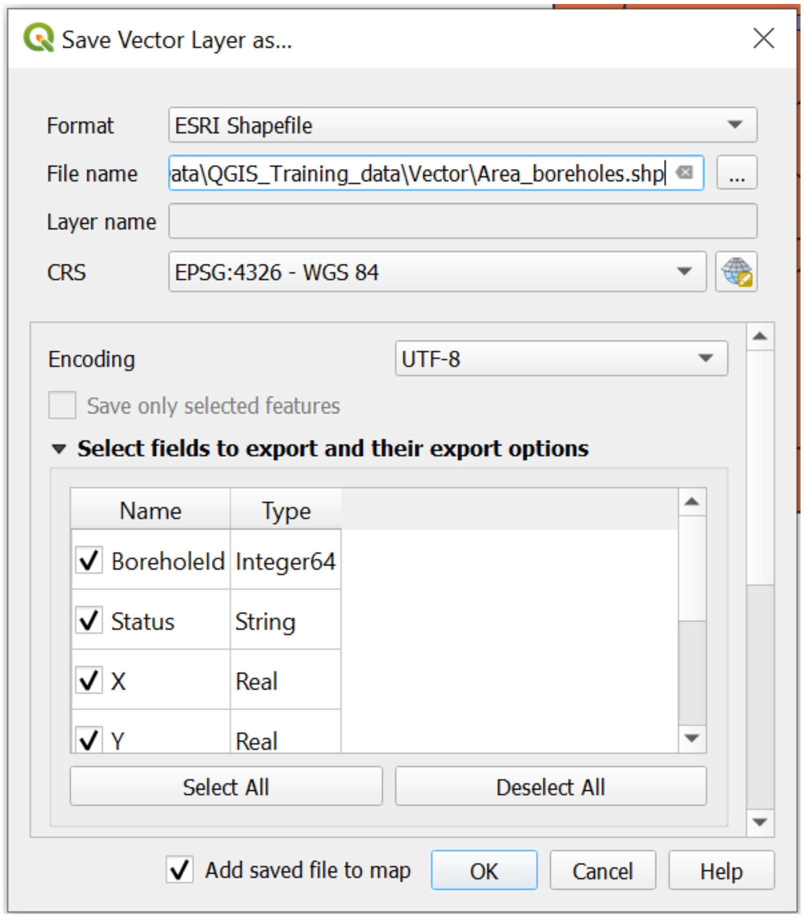
- Click OK in the new dialogue window that appears.
- The layer is removed from the legend.



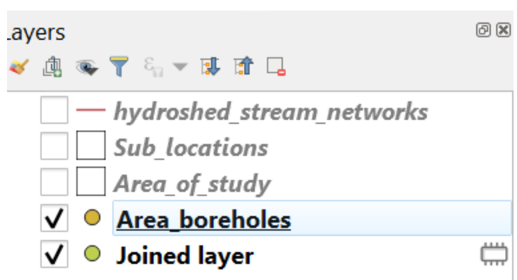
- The new layer ***Joined layer*** is not saved, to continue with the analysis we need to save the layer as a shapefile.
- Right Click on ***Joined layer→Export→ Save Feature As…***



- On the new dialogue window, input the following parameters, for **Format** select ***Esri Shapefile;*** for **file name** type ***Area_boreholes.***



- Click OK, the layer is saved and added to the legend, remove Joined Layer.
- The layers tab should appear similar to this one



Kudos, we have successfully carried out a spatial join, now all the boreholes have the name of the sub location they are located in.

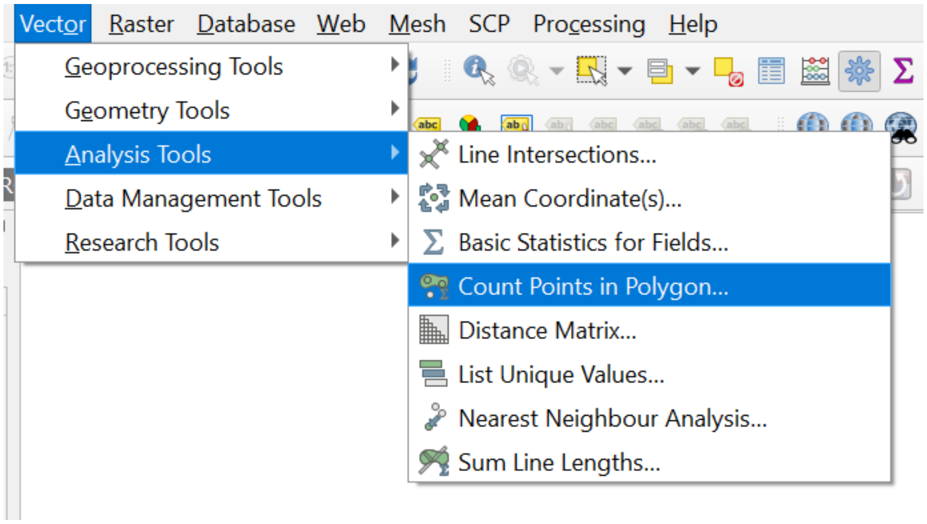
Turn all the layers on at this point.

Now we are going to carry out the next basic spatial analysis, point in polygon analysis. Here we are going to use QGIS to determine the number of boreholes per sub location.

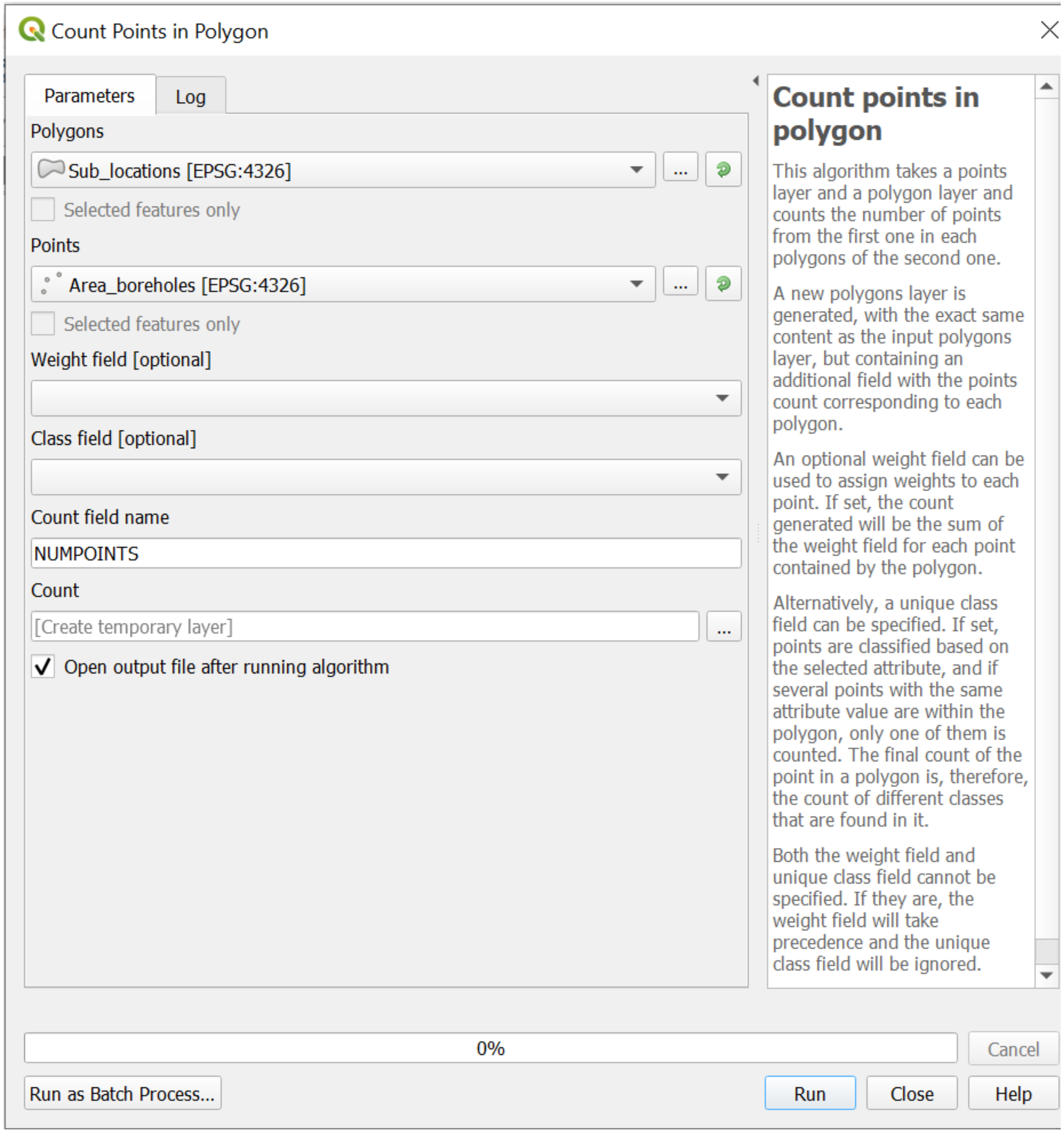
3.4.2 Assess the number of boreholes per Sub location (Point in Polygon Analysis)

In this section we are going to assess the number of boreholes in each sub location and also identify the sub location with the highest and lowest number of boreholes.

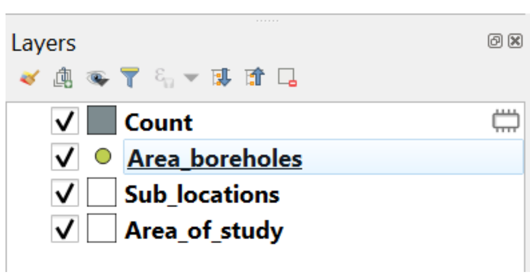
- Click on ***Vector→ Analysis Tools → Count Points in Polygon…***



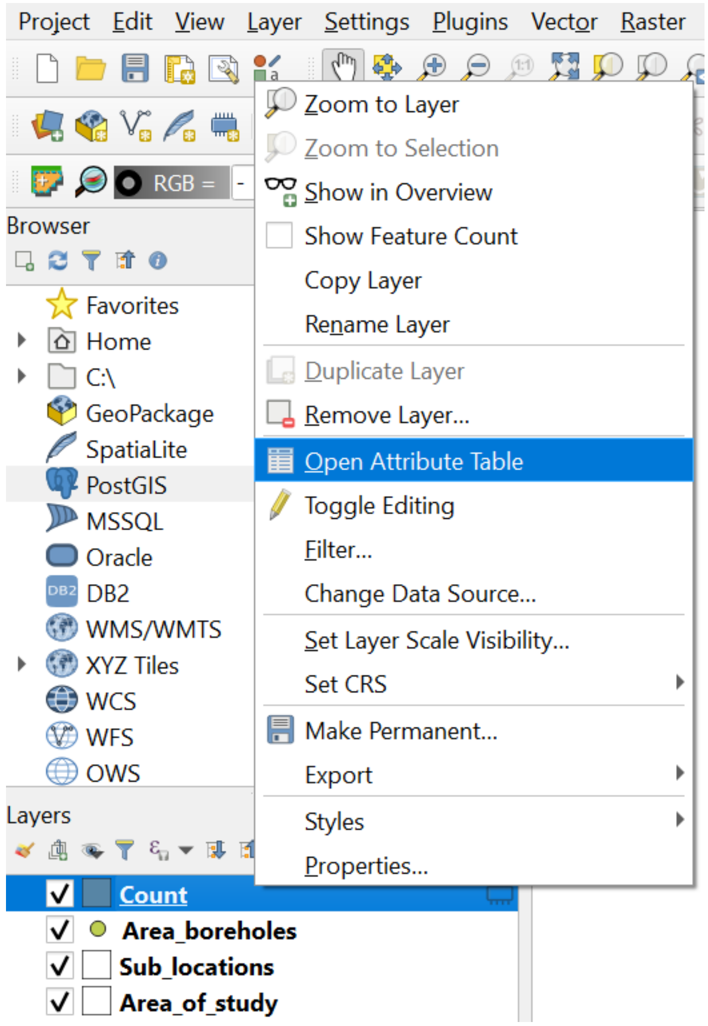
- In the new dialogue window input **Polygon→*Sub_locations[EPSG:4326],*** **Points→*Area_boreholes[EPSG:4326]→Run.***



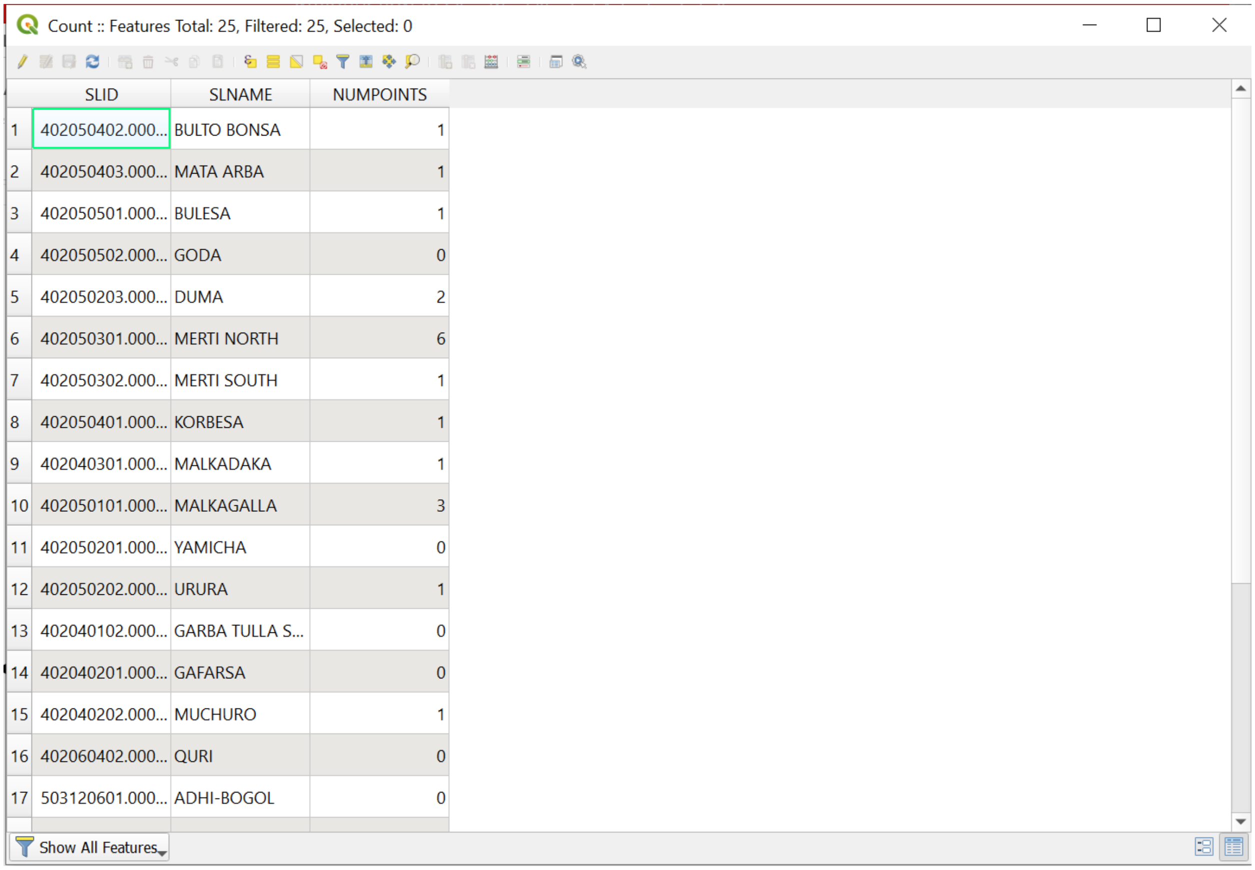
- After running the tool a new layer is created named ***Count.*** It appears on the layers tab as seen below



- To view the number of boreholes per sub-location we now open the attribute table of the new layer (***Count).***
- Right Click on the new layer ***Count→ Open Attribute Table***



- The attribute table has a column for polygon name named SLNAME and a column for points in that polygon named NUMPOINTS. It should appear similar to this one



	SLID	SLNAME	NUMPOINTS
1	402050402.000...	BULTO BONSA	1
2	402050403.000...	MATA ARBA	1
3	402050501.000...	BULESA	1
4	402050502.000...	GODA	0
5	402050203.000...	DUMA	2
6	402050301.000...	MERTI NORTH	6
7	402050302.000...	MERTI SOUTH	1
8	402050401.000...	KORBESA	1
9	402040301.000...	MALKADAKA	1
10	402050101.000...	MALKAGALLA	3
11	402050201.000...	YAMICHA	0
12	402050202.000...	URURA	1
13	402040102.000...	GARBA TULLA S...	0
14	402040201.000...	GAFARSA	0
15	402040202.000...	MUCHURO	1
16	402060402.000...	QURI	0
17	503120601.000...	ADHI-BOGOL	0

- Now we can see the number of boreholes per Sub location. Merti north has the highest number of boreholes with 6 boreholes, while there are several sub-locations with none Close and save the project on a working folder, give it a name of your choice.

Summary

In this section we have carried out basic spatial analysis. Particularly we have conducted a spatial join by adding sub locations names to boreholes such that in the boreholes attribute we can view the sub location name where it is found. We have then carried out a point in polygon analysis, by determining the number of boreholes in each sub location.

4. Basic Raster Analysis and Styling

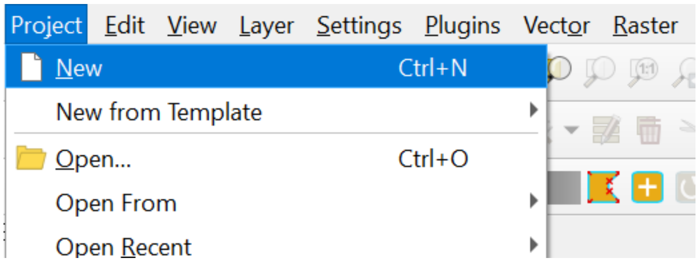
Raster data models are widely used in GIS today. In this section we are going to learn how to add Raster data to a GIS, conduct basic operation and basic spatial analysis. The most widely used Raster format is the GeoTIFF format or TIFF format with an extension (.tif).

4.1 Adding raster data to QGIS.

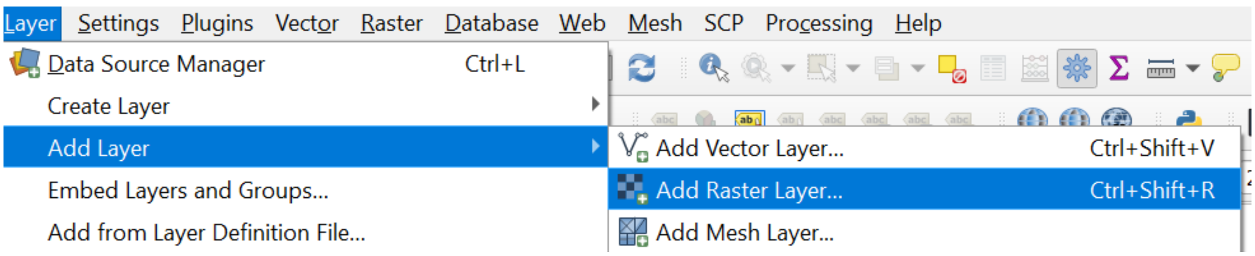
To add raster data to QGIS, use the layer tab on the Main Menu. The process is similar to adding vector data to QGIS, only that here the *Add Raster Layer…* function is used. First we will add *dist_to_water_ways* raster.

- Open QGIS

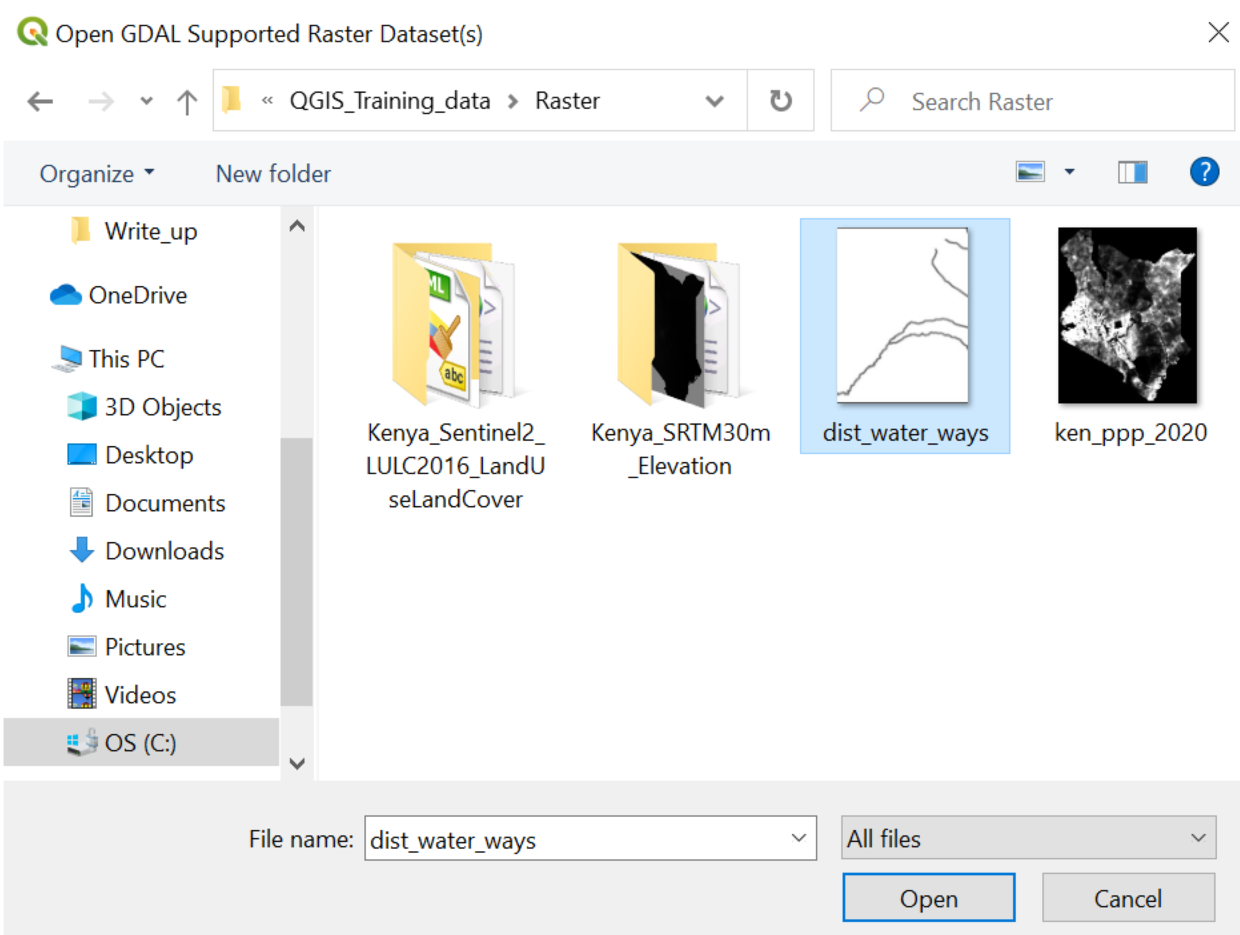
- Start a new project



- Once opened, on the main menu right click on → *Layer* → *Add Layer* → *Add Raster Layer…*

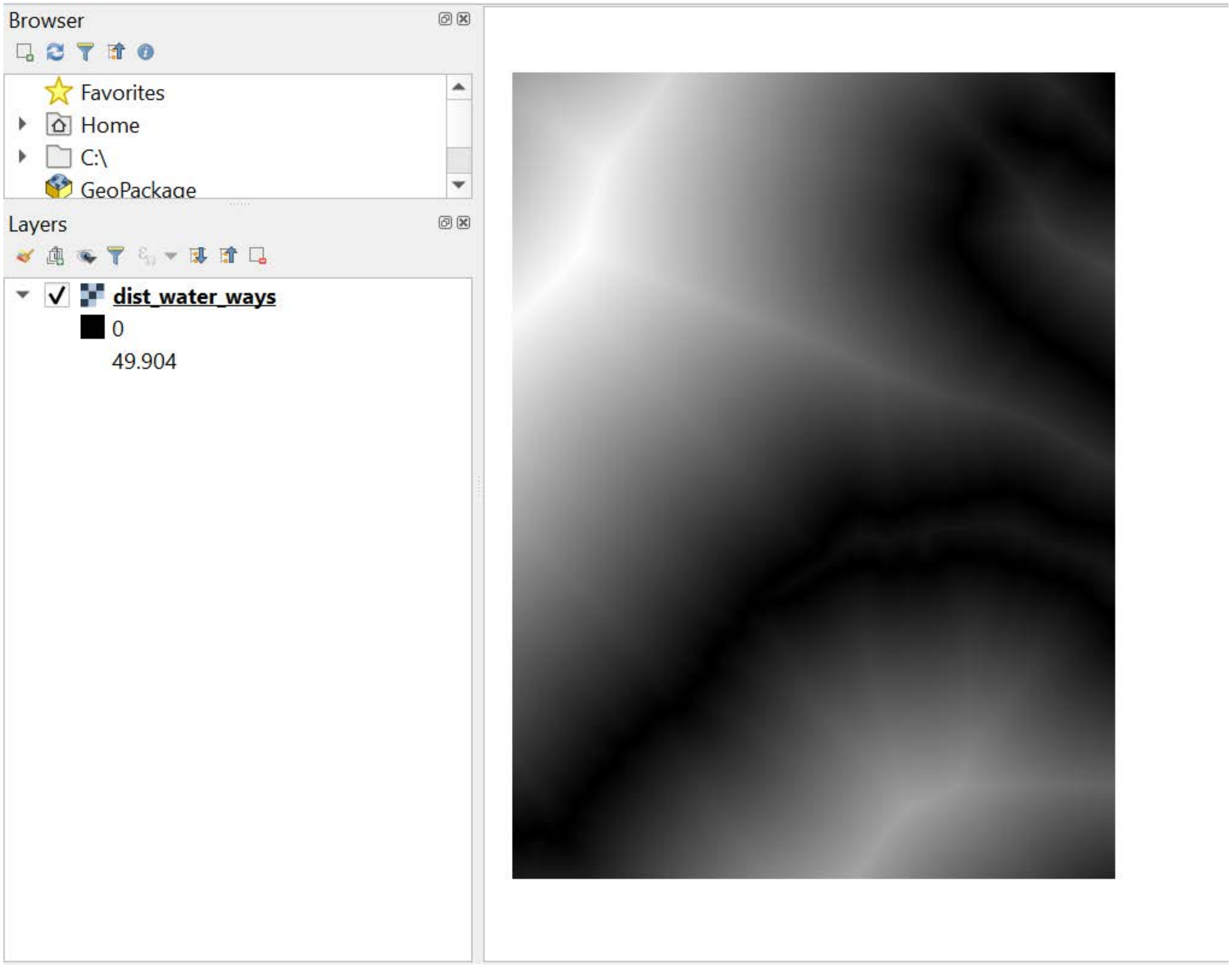


- Navigate to the folder with Distance to water ways using the  button, once selected click add.



- Click Open→ Add, to add the layer into your map

The canvas should appear similar to the one below

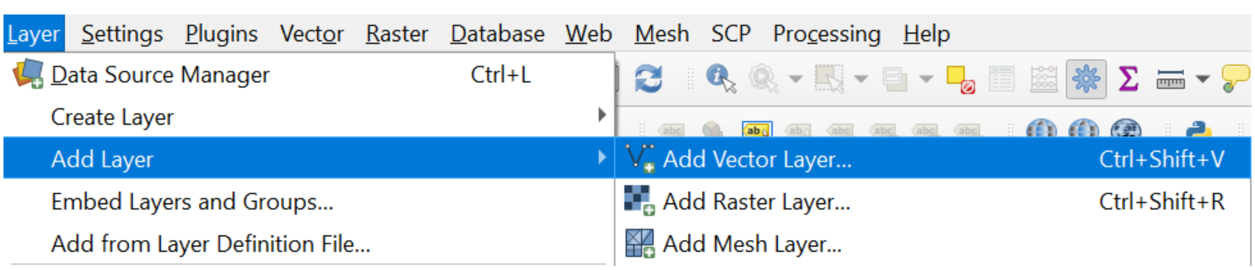


Having added the layer to our map canvas we now move to the next step that is analysis.

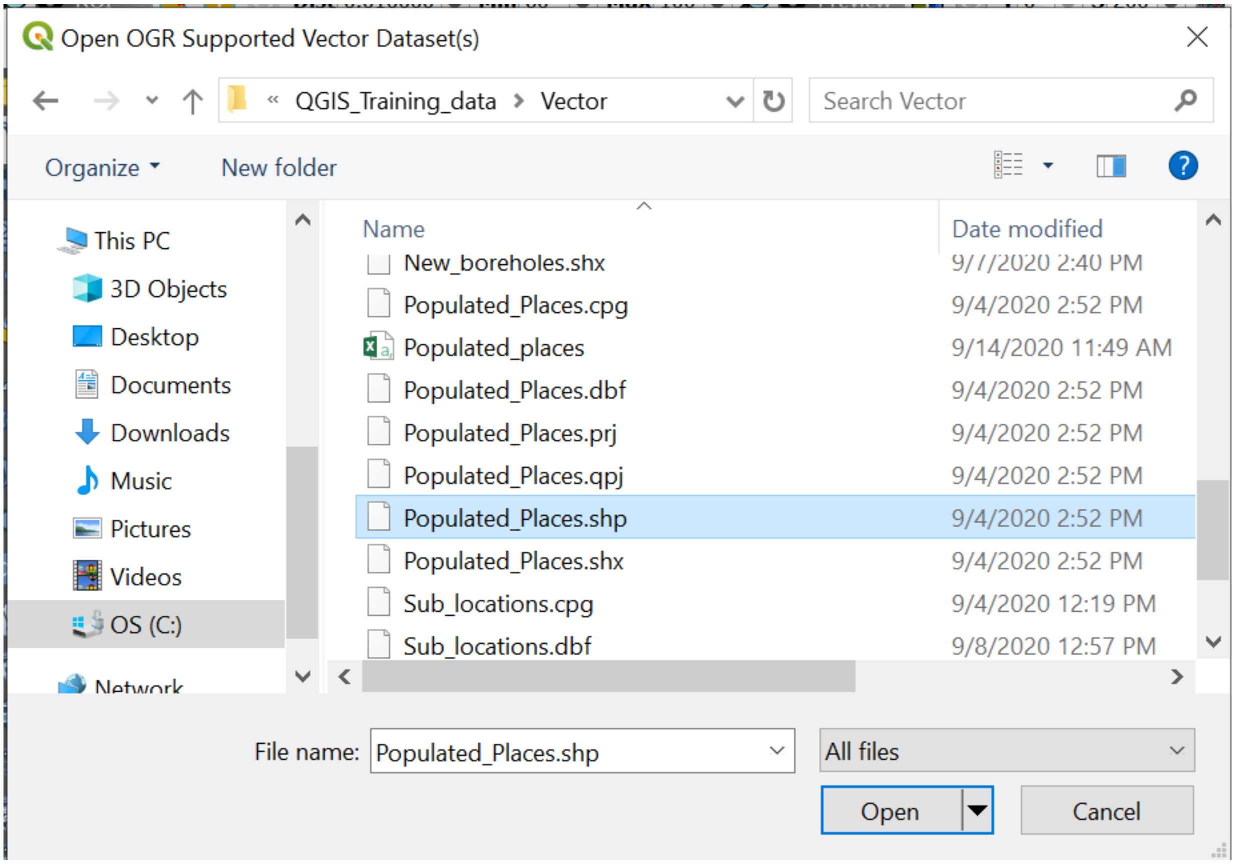
4.2 Distance to waterways

In this section we are going to calculate the distances from populated places to the nearest water ways, this includes rivers, nearest streams and water sources in the area. We will use the layer that we have just added and Populated places data.

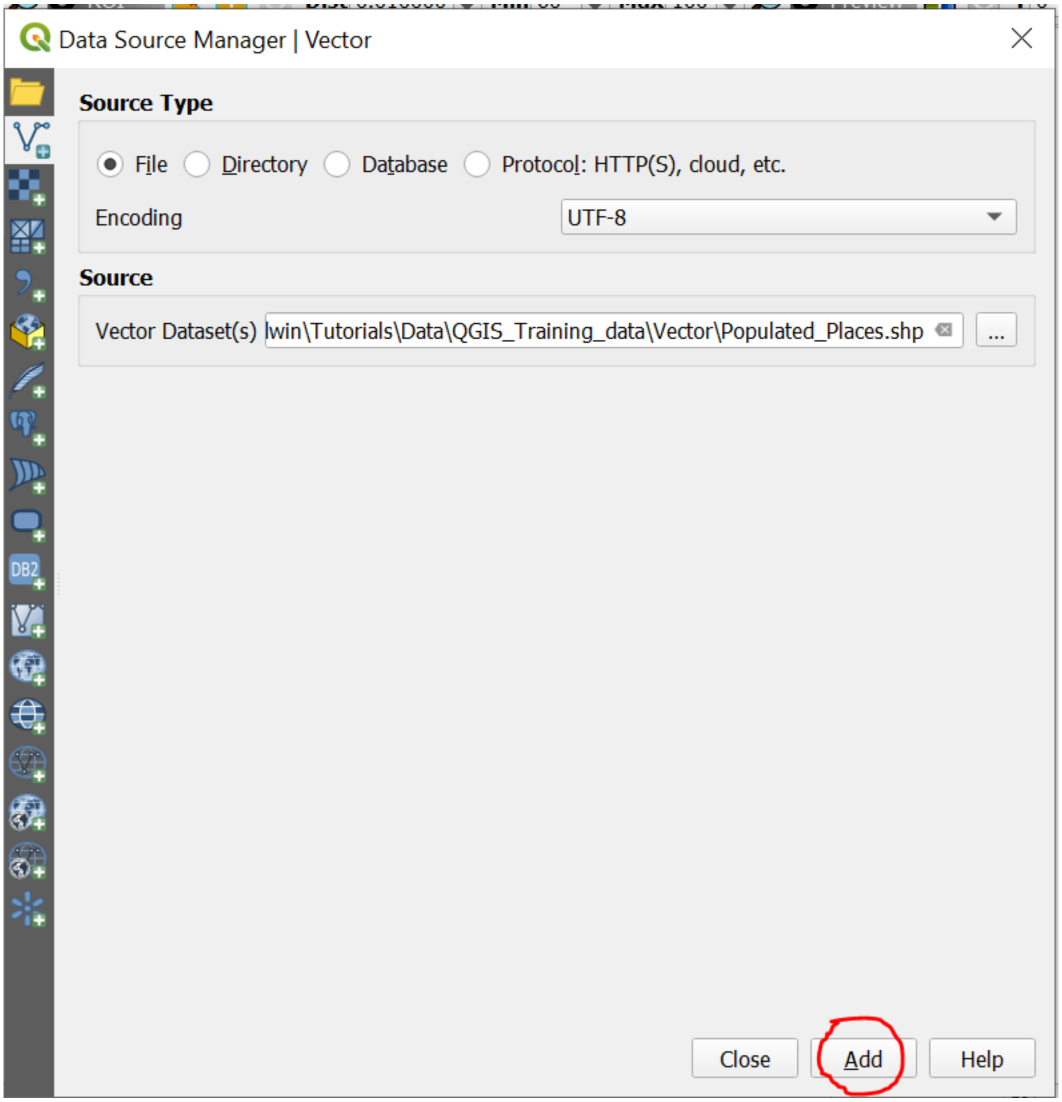
- Next, add the populated places shapefile into our map
- Click *Layers*→ *Add Layer*→ *Add Vector Layer…*



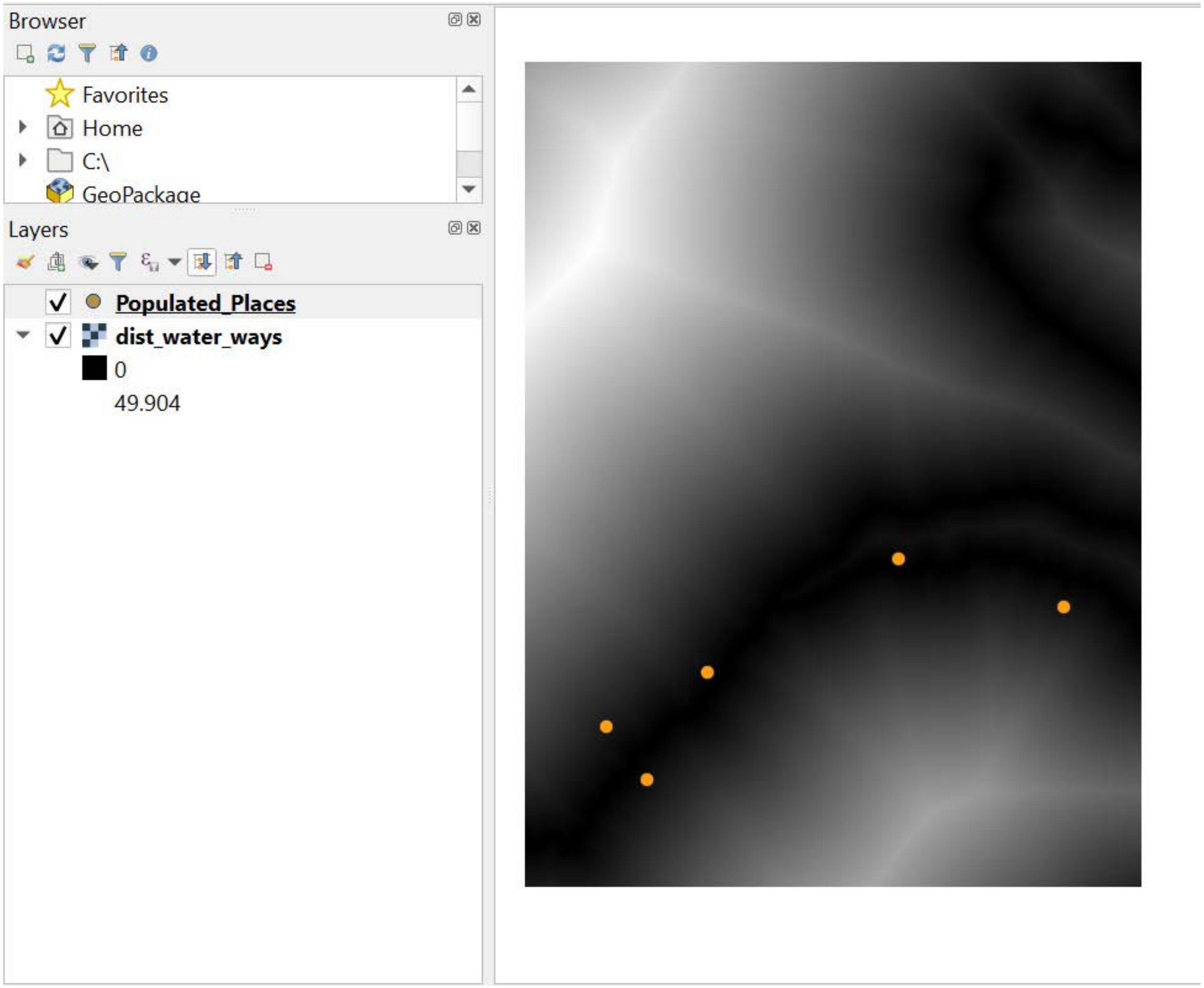
- Navigate to the folder with the populated places shapefiles using the ... button, Once selected the layer click *Open*



- On the new dialogue window click *Add.*

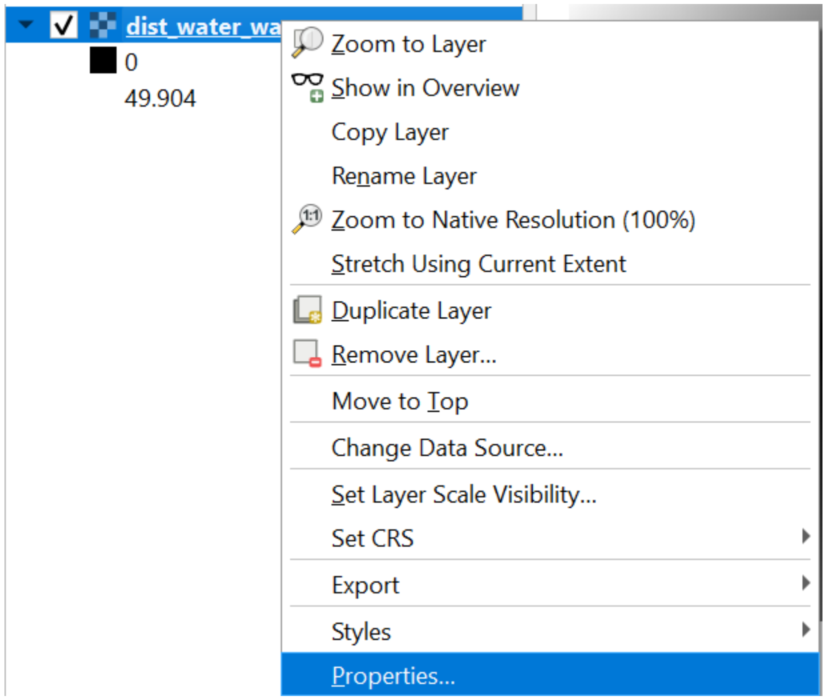


- The layer appears on the canvas

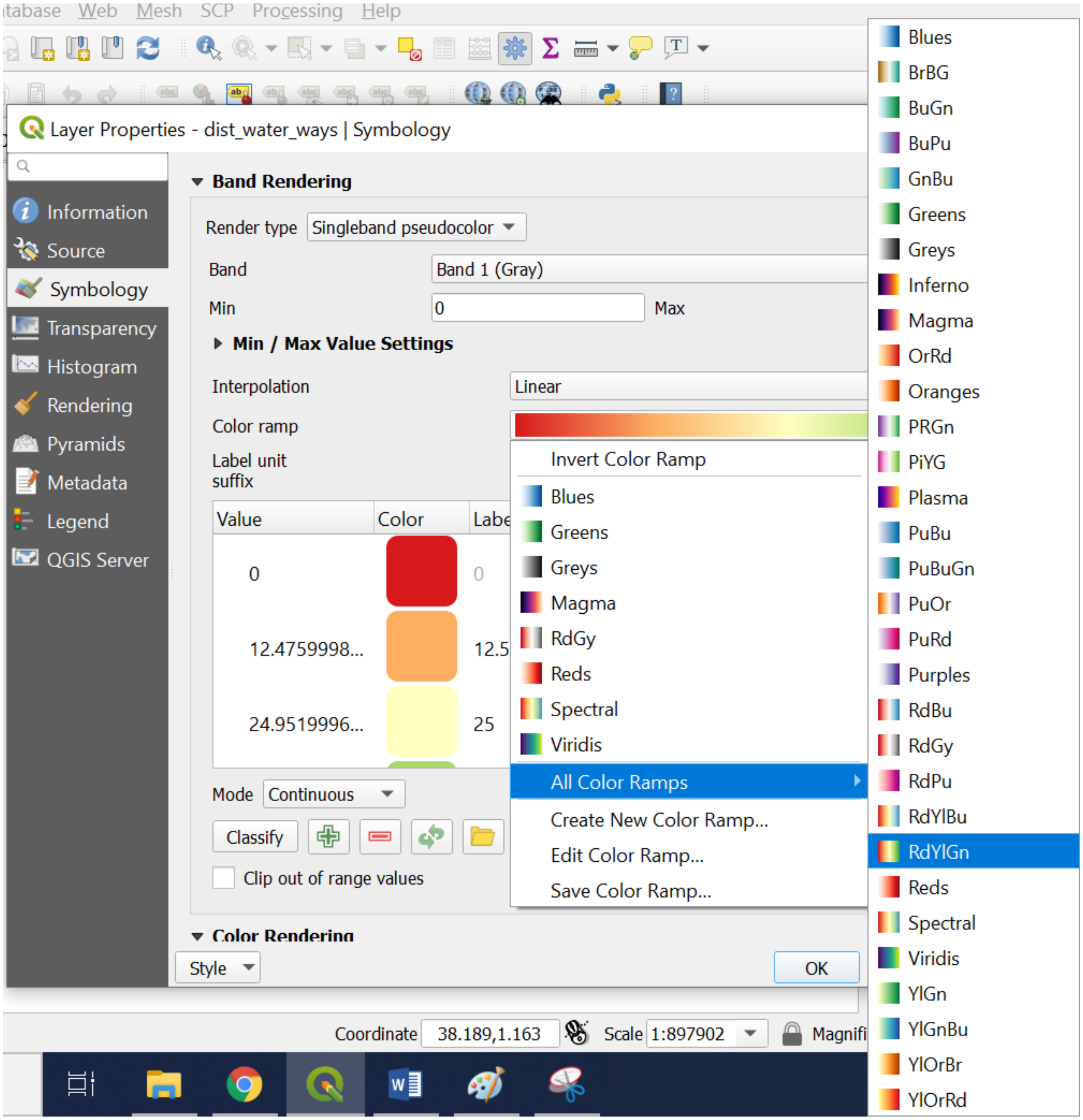


To determine distances to water ways, we will classify distance to waterways layer and apply a suitable symbology in order to identify places closer and further to waterways.

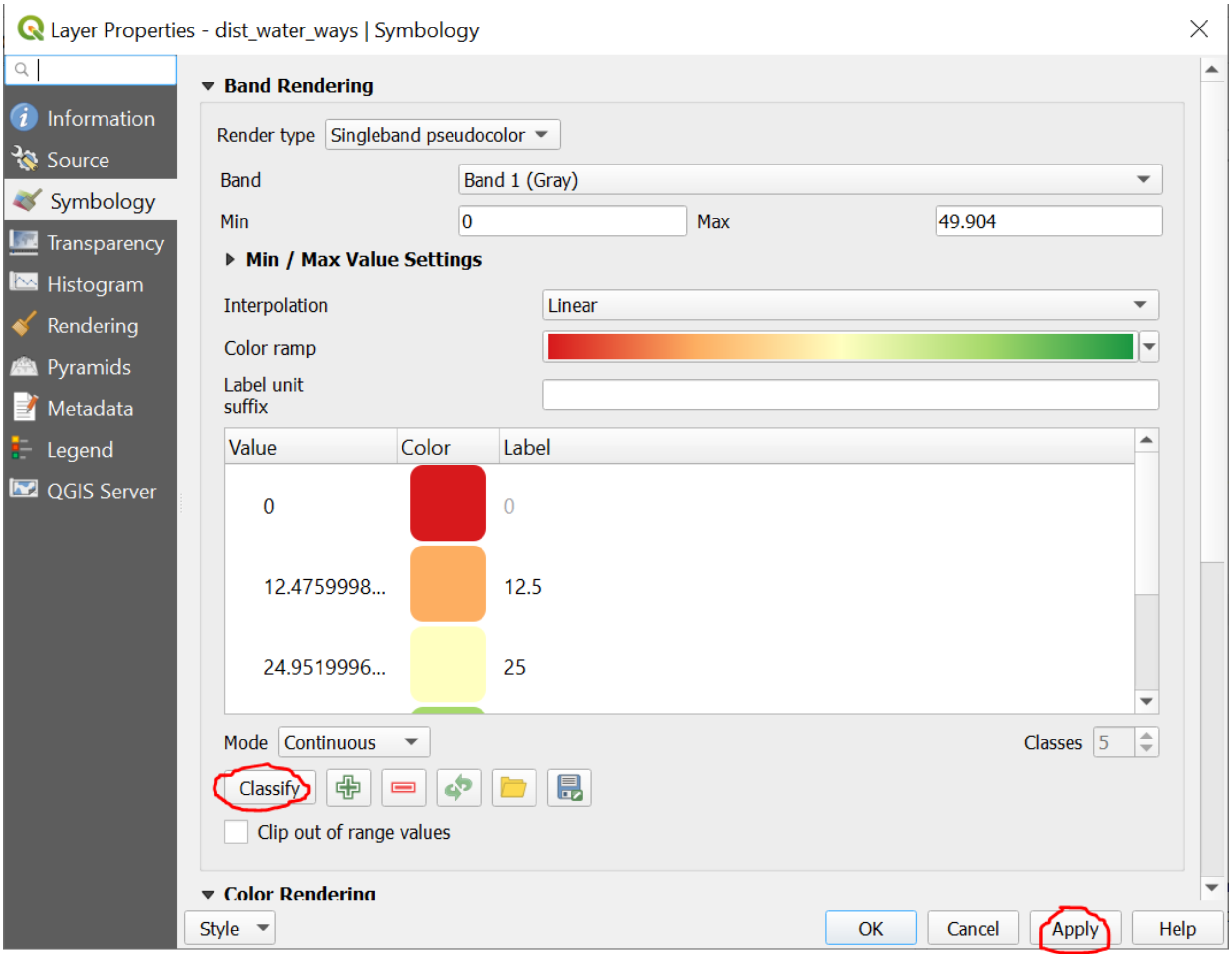
- Right click on *dist_to_water_ways* → *Properties…*



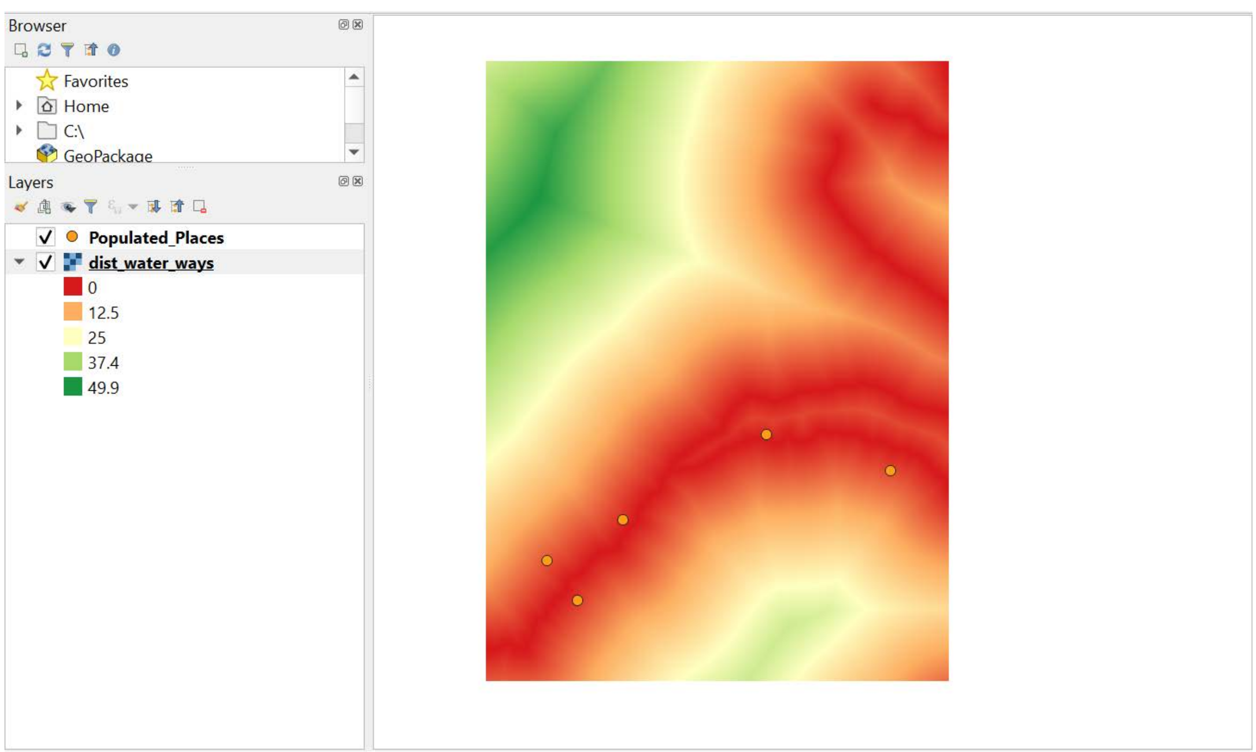
Click on *Symbology* → Render type → *Singleband Pseudocolor* → Color Ramp → select RdYlGn



- Click Classify→ Apply→ *OK*



Note that the symbology and legend of the dist_to_water_ways layer changes.



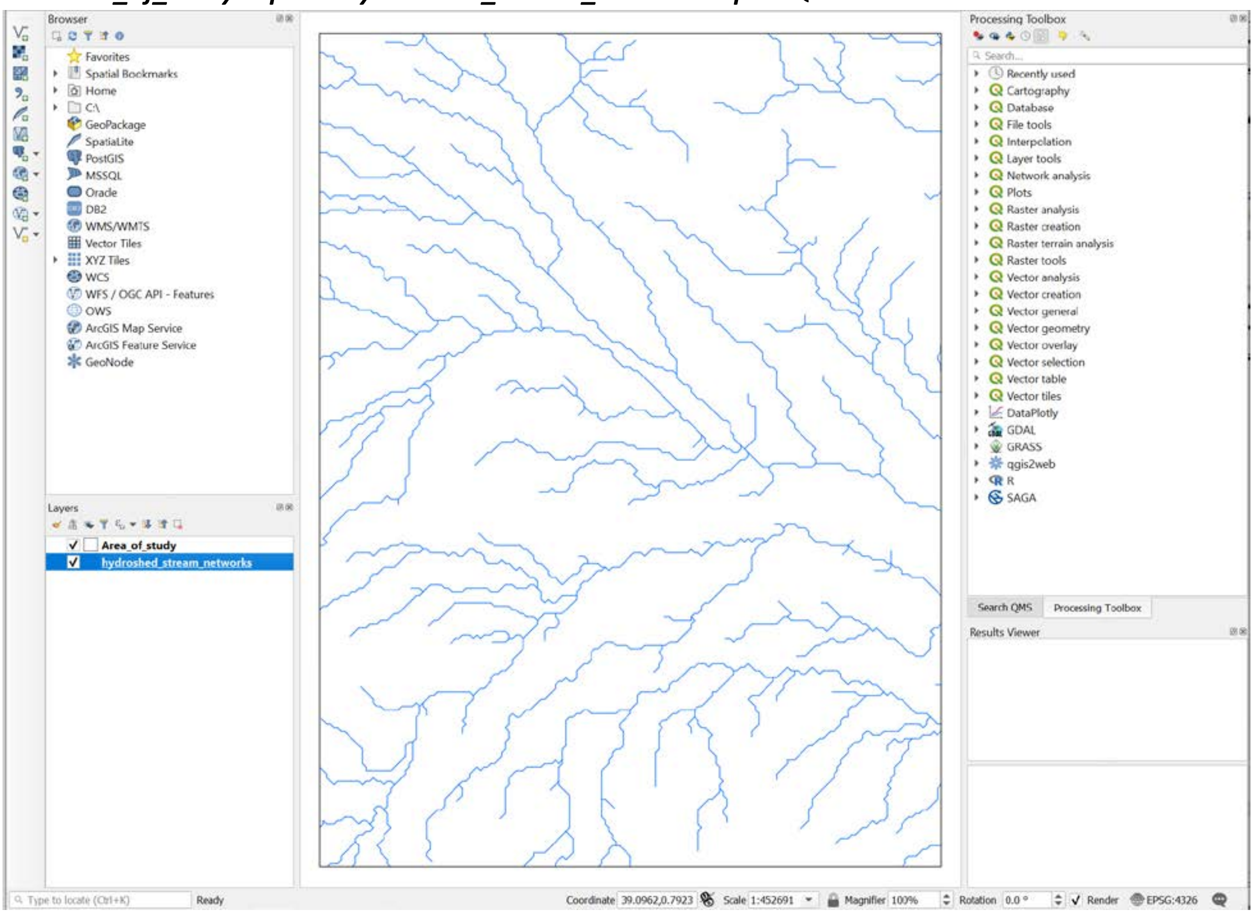
- From the legend, the least value with 0 represents regions closest to waterways while the highest 49.9 which is given a green color represents the furthest areas from the water ways.
- From the canvas we can see that all populated places within our study area are found within the red region, indicating that they are all at a close proximity/nearer to water ways.

4.3 Miscellaneous: Creating a distance raster from vector (lines) data

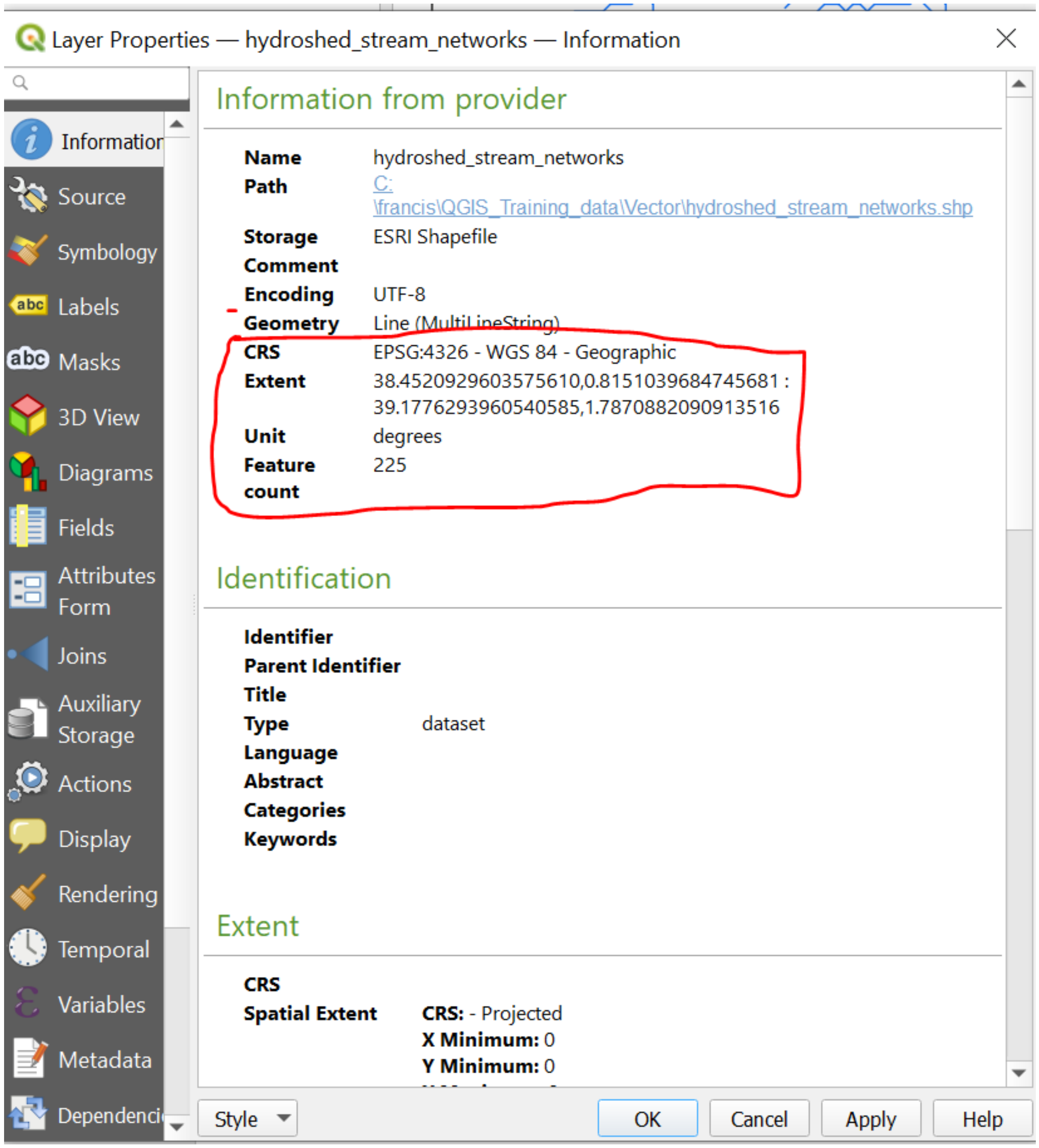
In the previous section, you have used existing distance to water ways raster data. What if you only had the waterways as line (Vector) data, how can you create a raster data, depicting distance to water ways?

To achieve this, we will use ***…/Vectors/hydrosheds_stream_network.shp*** data. This are model derived stream network data.

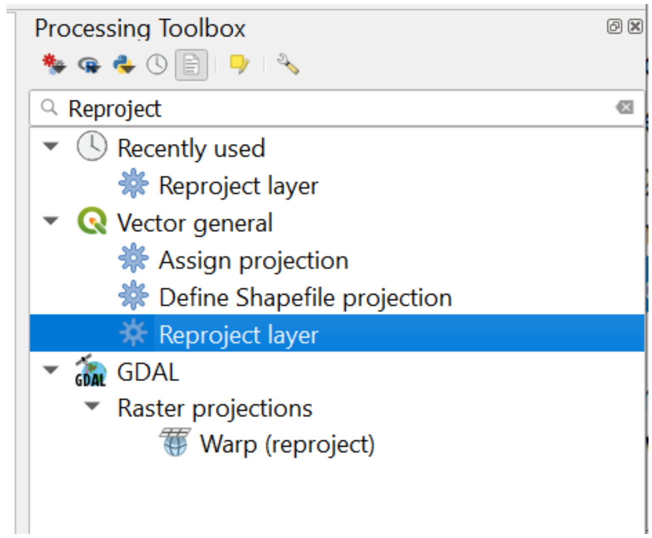
- Add ***Area_of_study.shp*** and ***hydrosheds_stream_network.shp*** to QGIS



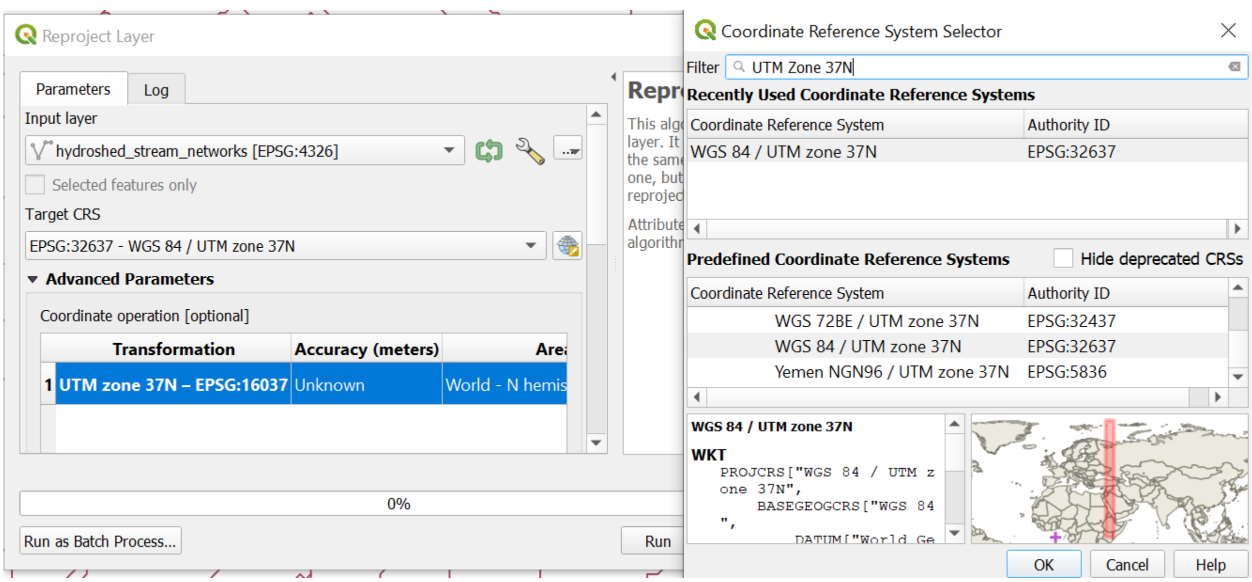
- Before you run any distance-related analysis, you have to confirm that the projection of your data is not in a geographic projection (with distances measured in degrees) but is in an equidistant projection, where distance is measured in metric units e.g. meters, centimeter of kilometer
- Go to properties of the hydroshed_stream_network.shp data and check the projection of the data. ***Right click → Properties → Information***



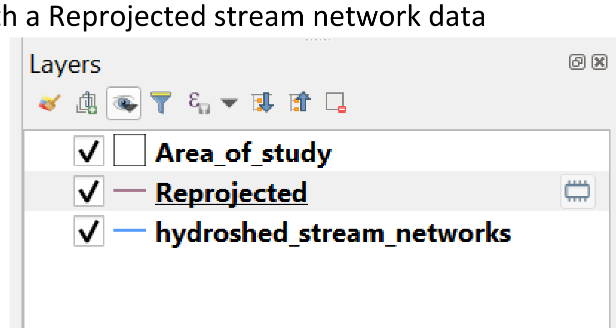
- You can see that the CRS (Coordinate Reference System) is Geographic and the unit of measurements are in degrees. This means that to calculate accurate distance in metric units, we have to Reproject the data to a distance conserving projection like UTM (Universal Transverse Mercator) projection.
- Search for Reproject layer in the processing Toolbox



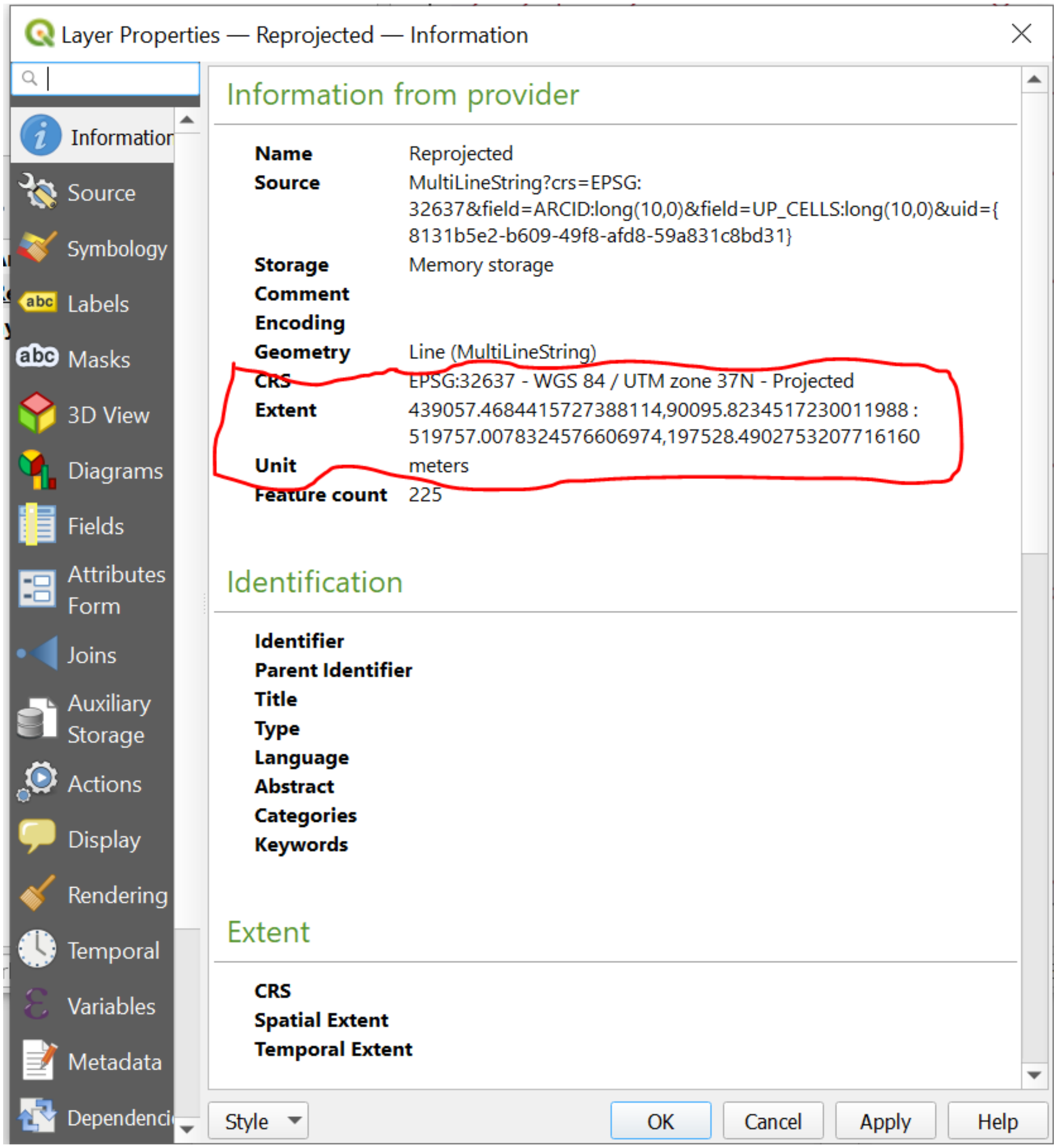
- Click on ***Reproject Layer tool*** → ***Input layer*** (select hydroshed_stream_networks) → ***Target projection*** (select UTM Zone 37N) → Run



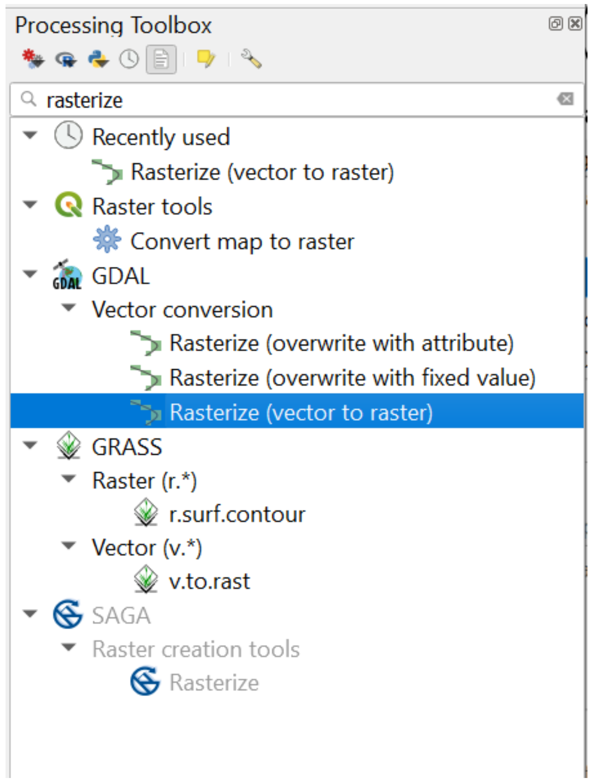
- Note that here, we have saved the reprojected data as a temporary layer (Reprojected), should you want to save the reprojected layer then select save file before you run the Reproject Layer Tool
- You now end up with a Reprojected stream network data



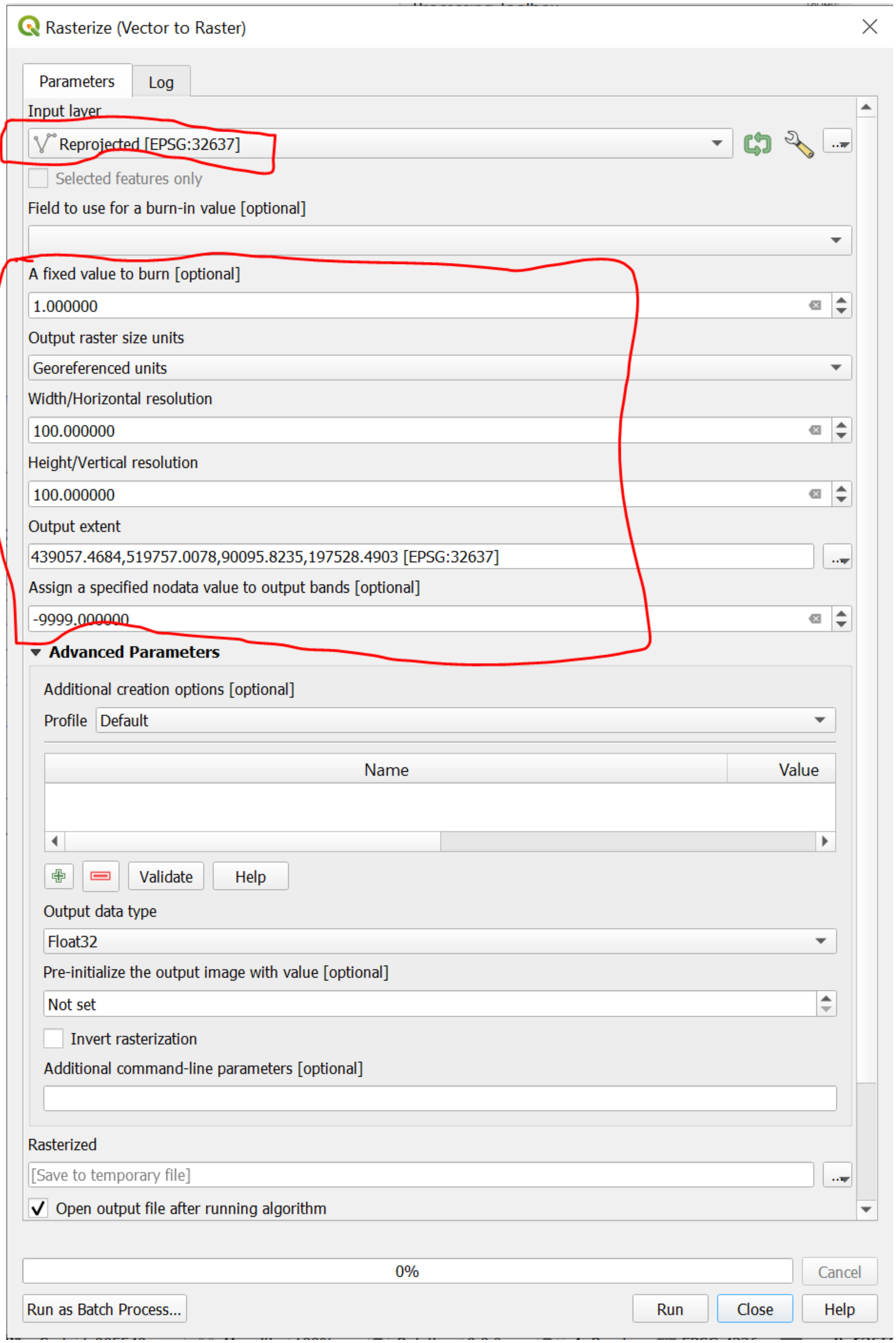
- The projection of the Reprojected layer has now changed to UTM Zone 37N-Projected, and the units are in meters



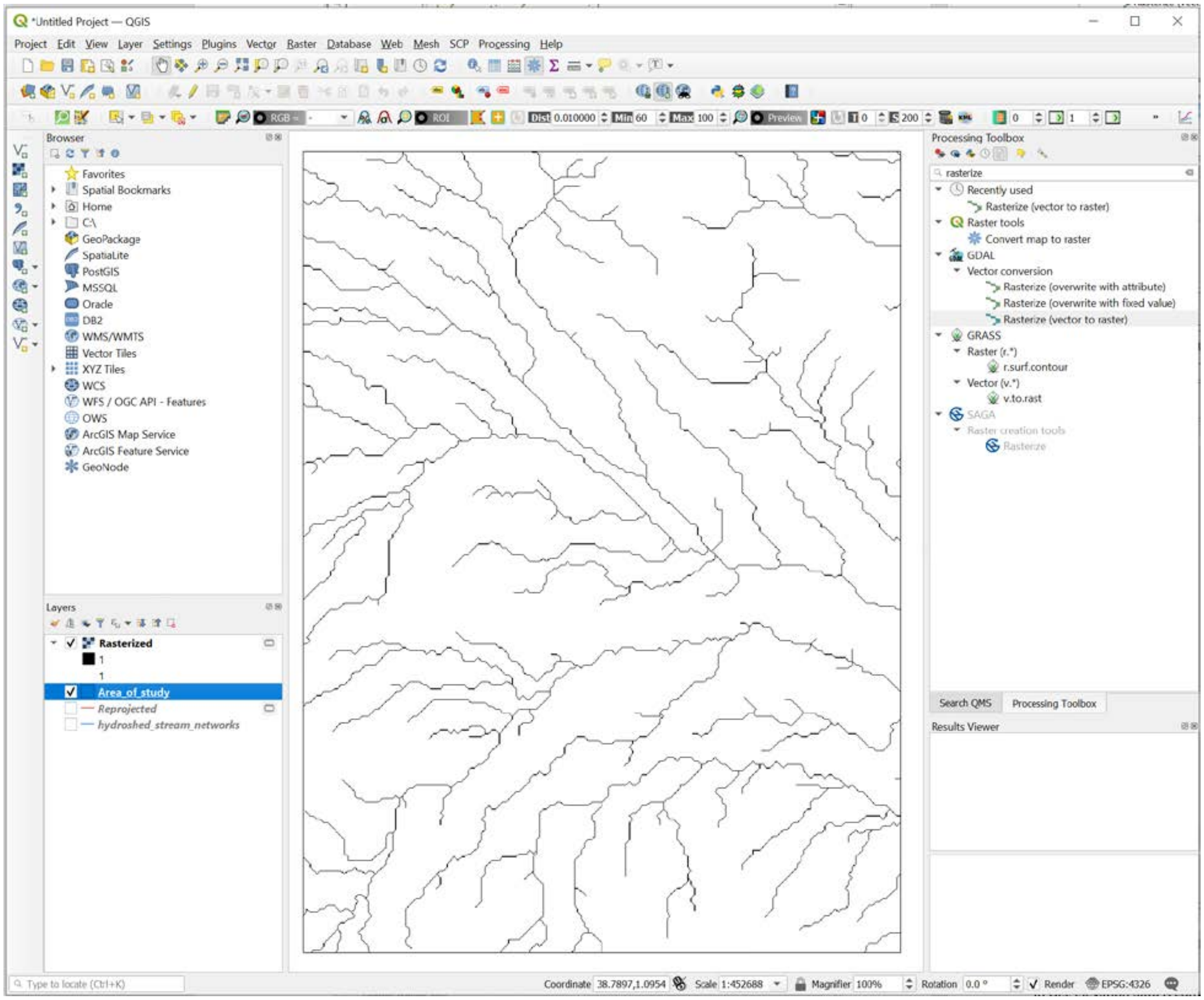
- You can now use the reprojected stream networks to create a distance raster. This will take place in two steps, first we have to create a raster or the stream networks, because they will be used as the target. This process is called rasterizing.
- Secondly you will use the proximity tool to estimate the distances of each pixel from these target cells.
- To rasterize, we use the Rasterize tool that comes under the GDAL Tools, search for Rasterize in the Processing tools



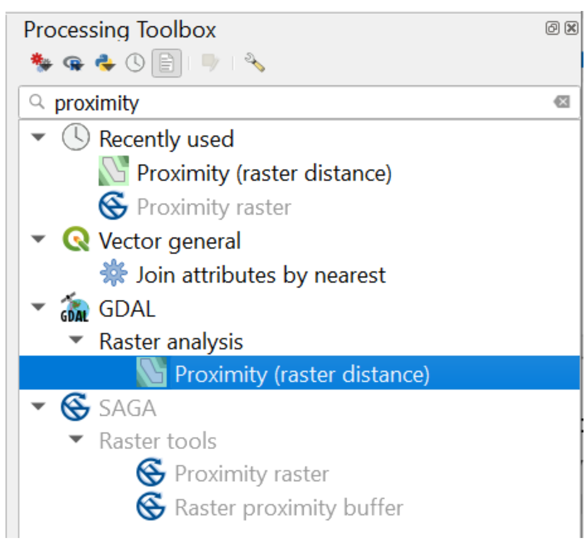
- Click on the tool, and set the parameters accordingly, in particular, the variables that are important/compulsory in this case are
  - Vector layer to rasterize (Reprojected)
  - A fixed value to burn (this is the value that will be assigned to the rasterized stream – Here we have set it to 1)
  - Resolution of the resulting raster, here we have set it as 100m (Georeferenced units)
  - The extent of analysis (Here we selected the extents of Reprojected Layer)
- Once you have set all the necessary parameters, you can run the tool



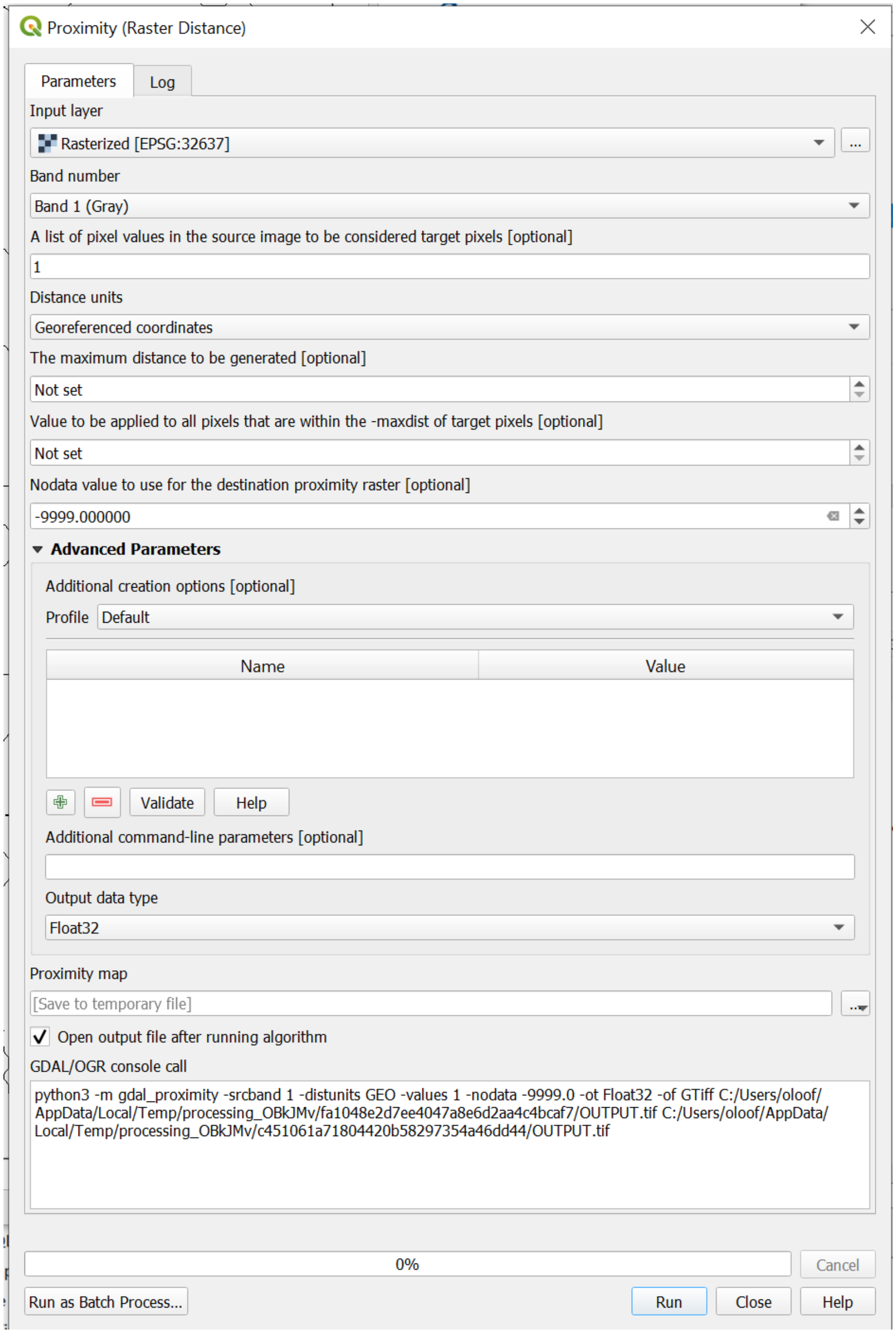
- You then end up with a rasterized stream network



- You are now ready for the final step, which is to create the proximity layer.
- Search for Proximity (raster distance) tool from processing tool bar. This is also a tool within the GDAL tools

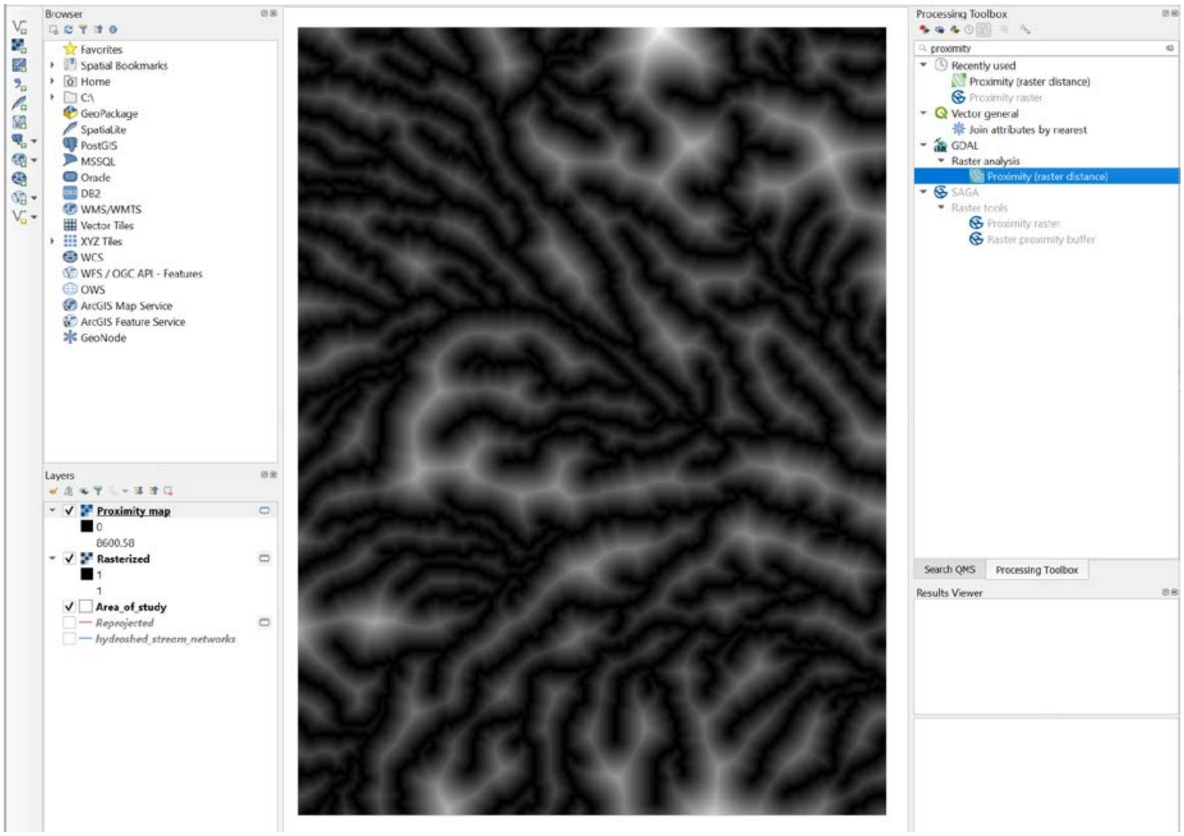


- Set the appropriate parameters, and run the tool. In particular
  - set the correct rasterized input layer,
  - set the pixel value to be used as target (in our case we set it to 1 as this is the value that we had assigned to the target pixels)
  - Set the Distance units to georeferenced coordinates.

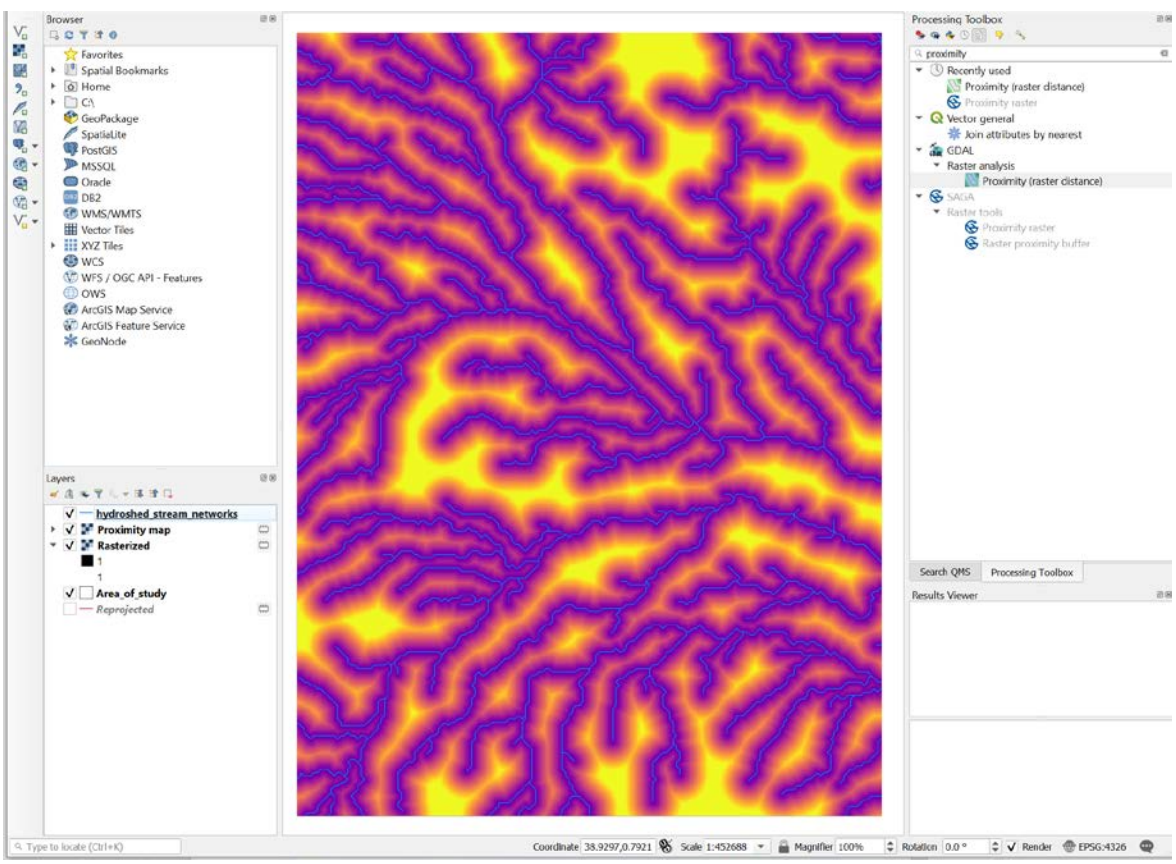


  -

- The result proximity map, showing distances (in meters) from the stream networks



- You can change the symbols of the proximity map accordingly.

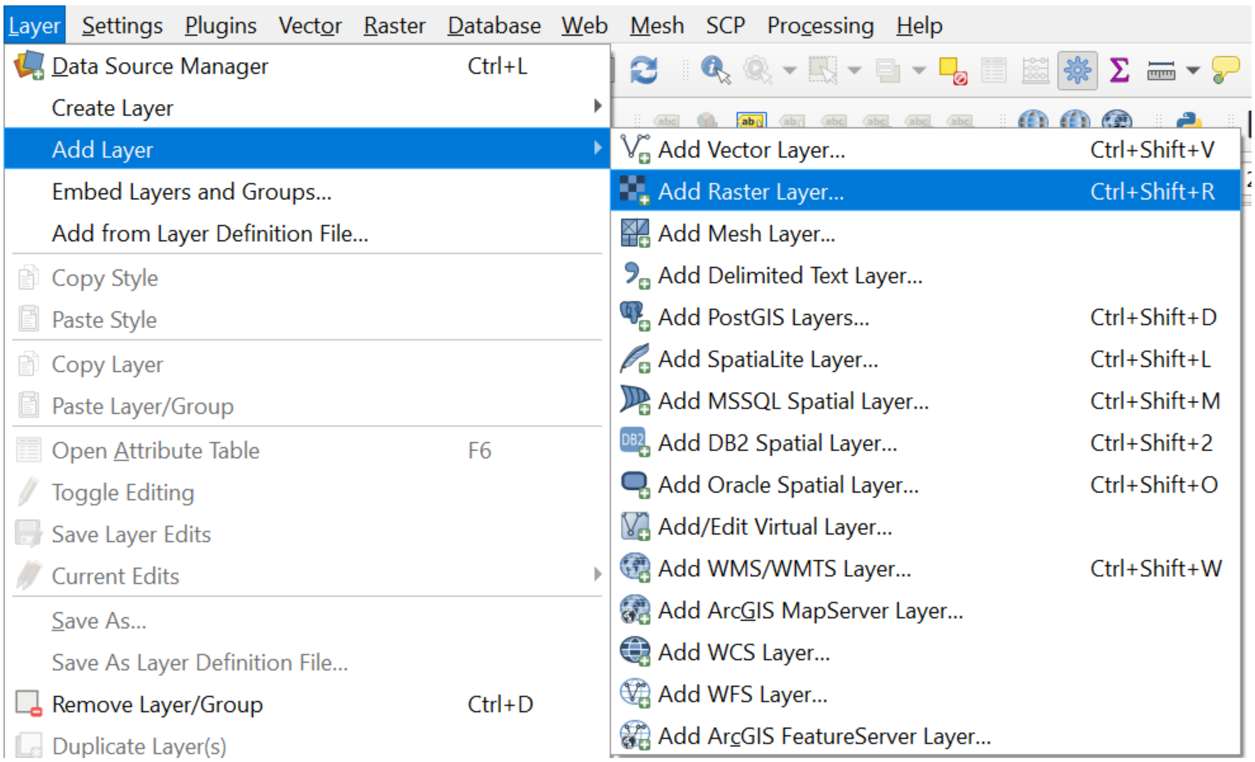


4.4 Working with elevation (raster) data

The elevation of a geographic location is its height above or below a fixed reference point, most commonly a reference geoid, and mathematical model of the Earth's sea level as an equipotential gravitational surface.

In GIS, elevation data is commonly presented in raster format. This is more natural presentation as elevation is continuous. A raster file representing elevation of the ground level known as a Digital Elevation Model (DEM).

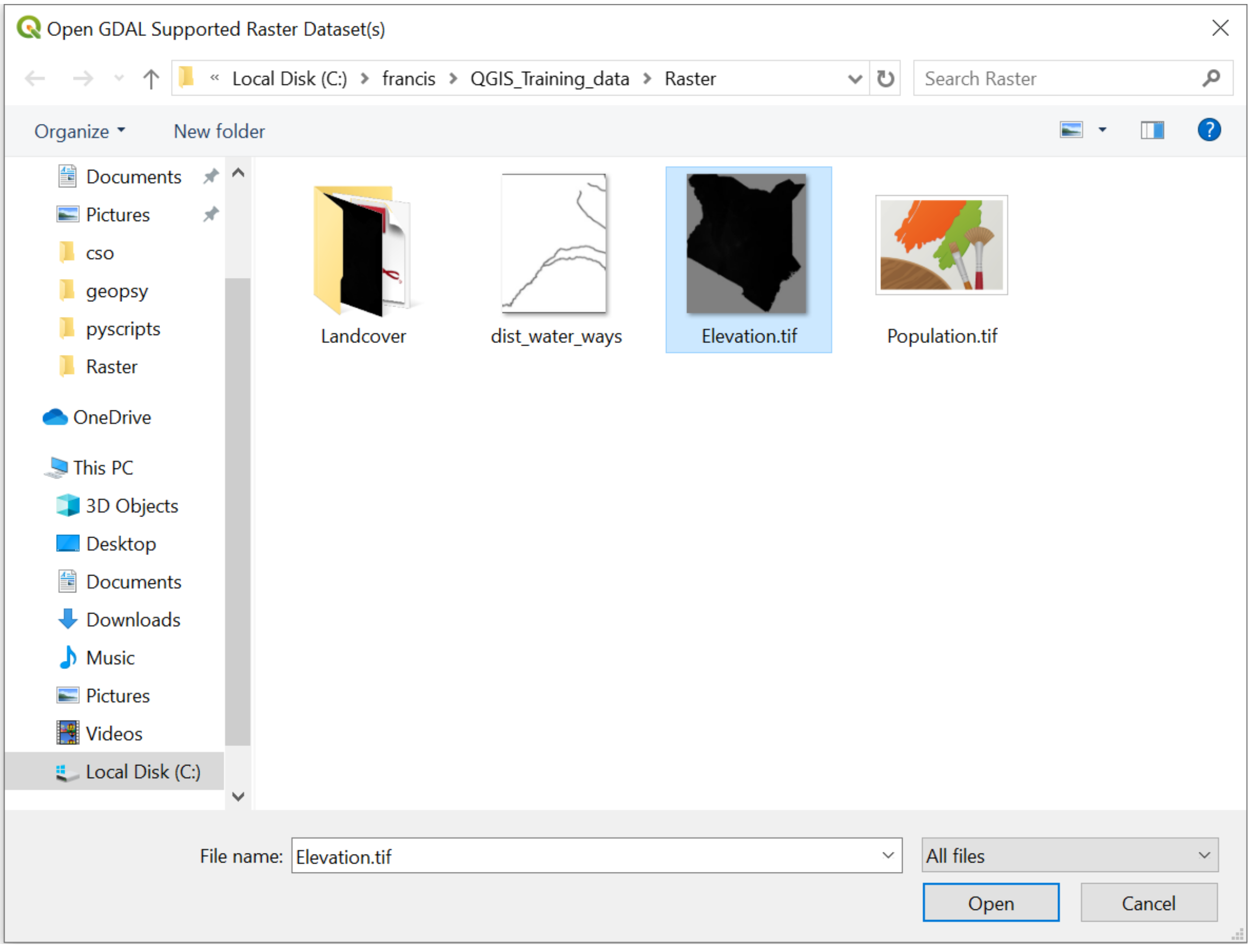
- To add elevation data to QGIS, Click on Layer→*Add Layer*→ *Add Raster Layer*



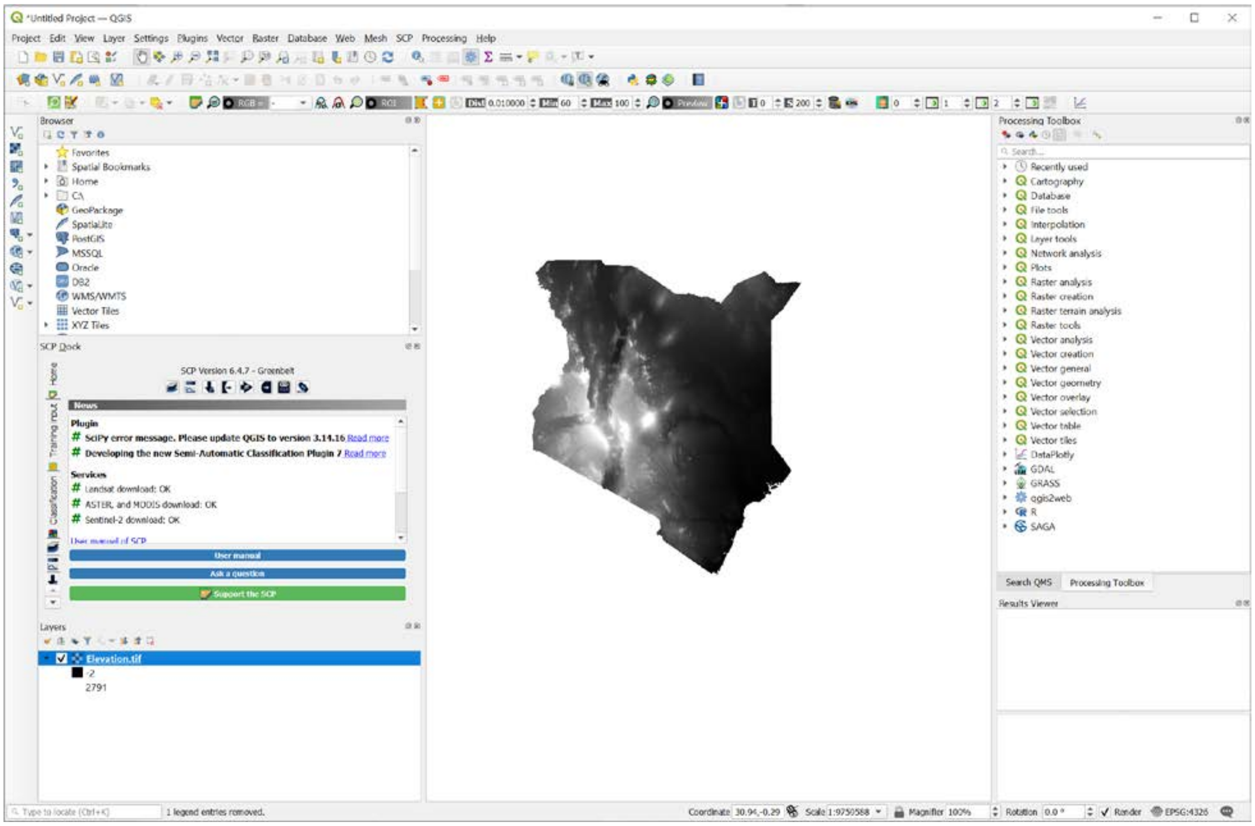
- On the new dialogue click on  button to navigate to folder with raster files

The Elevation data is in ***QGIS_Training_data/Raster*** *folder*

Select ***Elevation.tif,*** *then click on Open → Add*



- Click ***Open→ Add***
- The layer is added to the Canvas. Right click and Zoom to layer as shown below

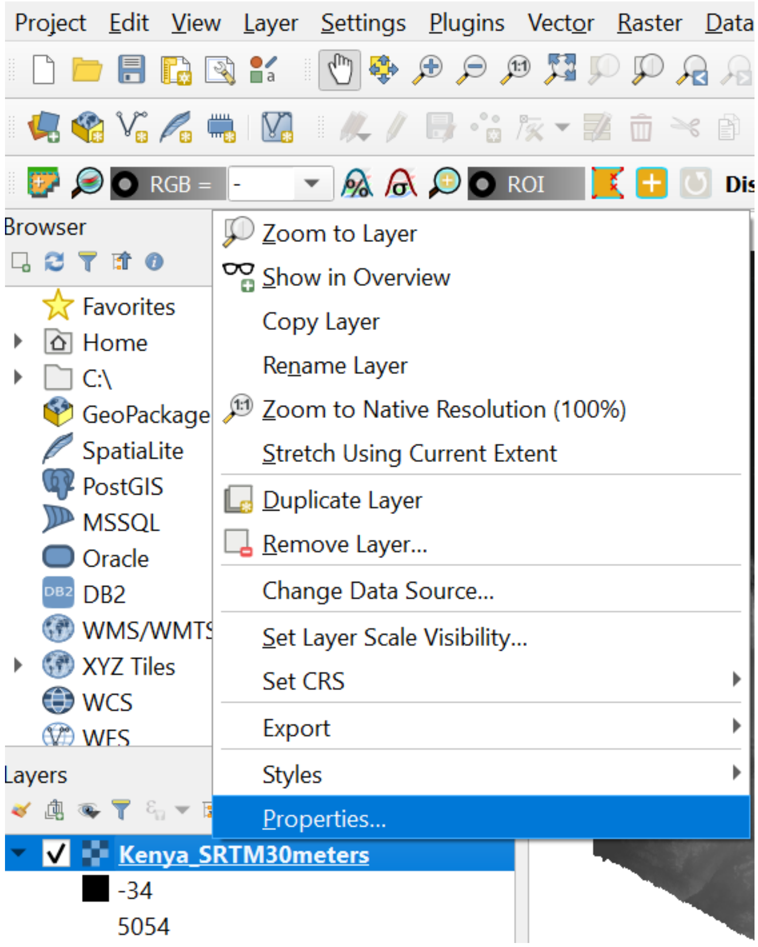


4.5 Symbolizing Raster File

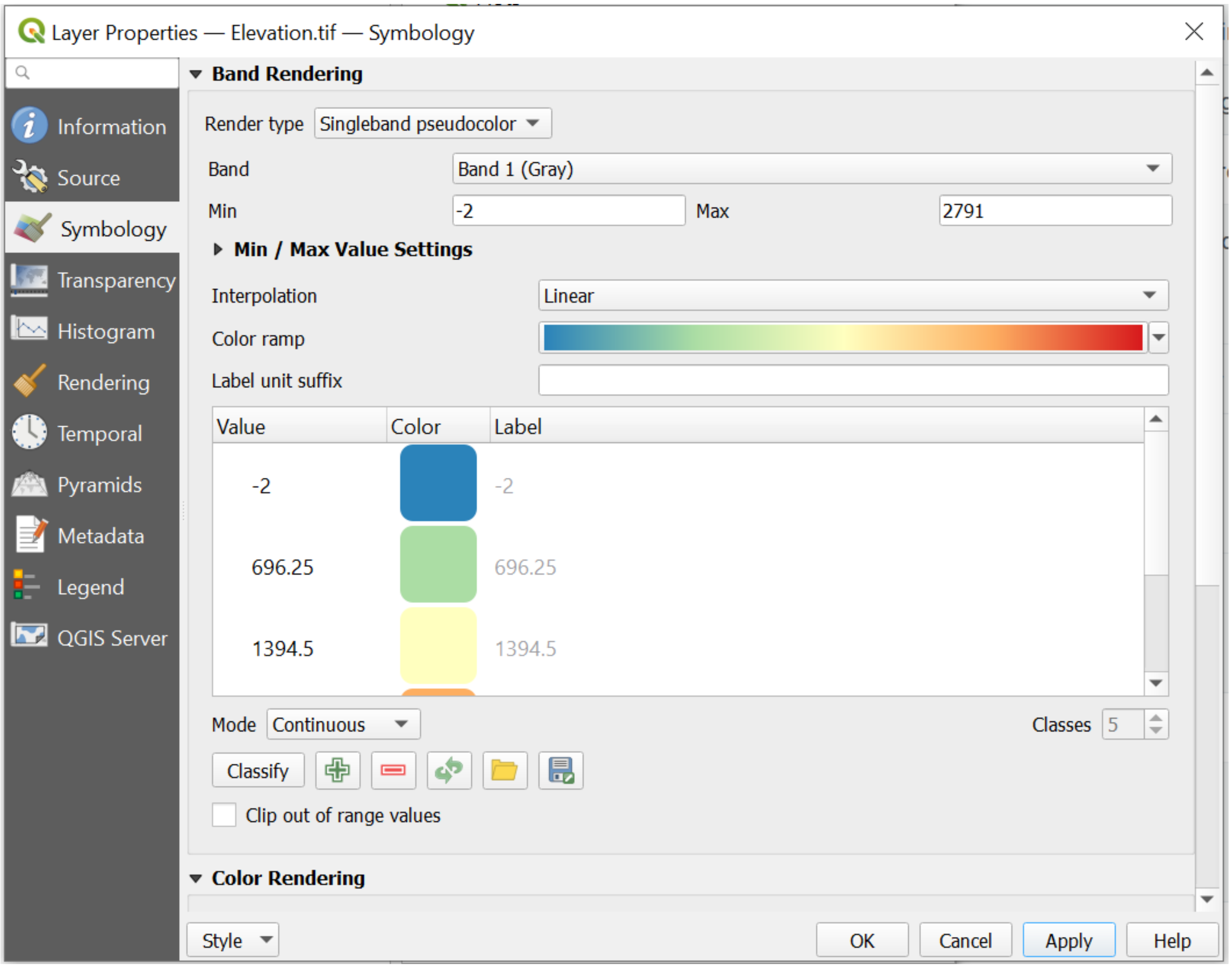
An Elevation Raster simply indicates the height of points on the earth's surface above the mean sea level. In our map we can simply see from the legend that higher (2791m) areas are represented by whiter color while lower areas are represented by a darker color (-2). We are now going to change the symbology (color) in order to apply a different color variation that best suites Elevation representation.

Kindly note that the elevation data in this exercise has been resampled and may not be the accurate representation of the elevation in Kenya

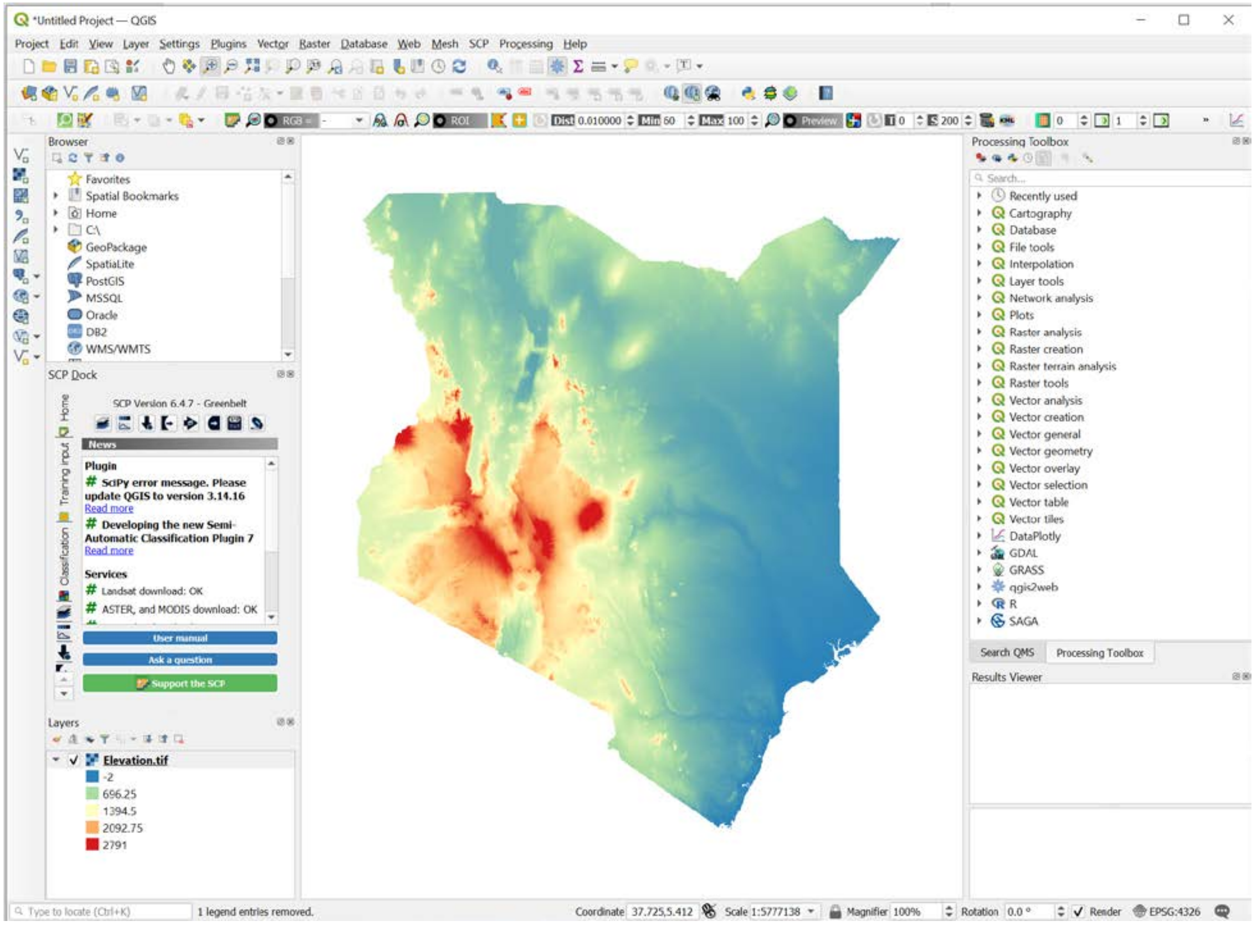
- Right click on the layer ***Elevation.tif* → *Properties…*** *, This will open the dialog box so that you can read and edit some of the properties of the raster data*



- On the new dialogue window that appears click on ***Symbology,*** on Render type select ***Singleband Pseudocolour*** Interpolation ***Linear***. Next to the color ramp click on the view more button  to view more select ***All Color Ramps*** and select one that is suitable.



- Once selected the color ramp, click ***Apply*** and ***Ok***
- The layer on the canvas should appear similar to this one



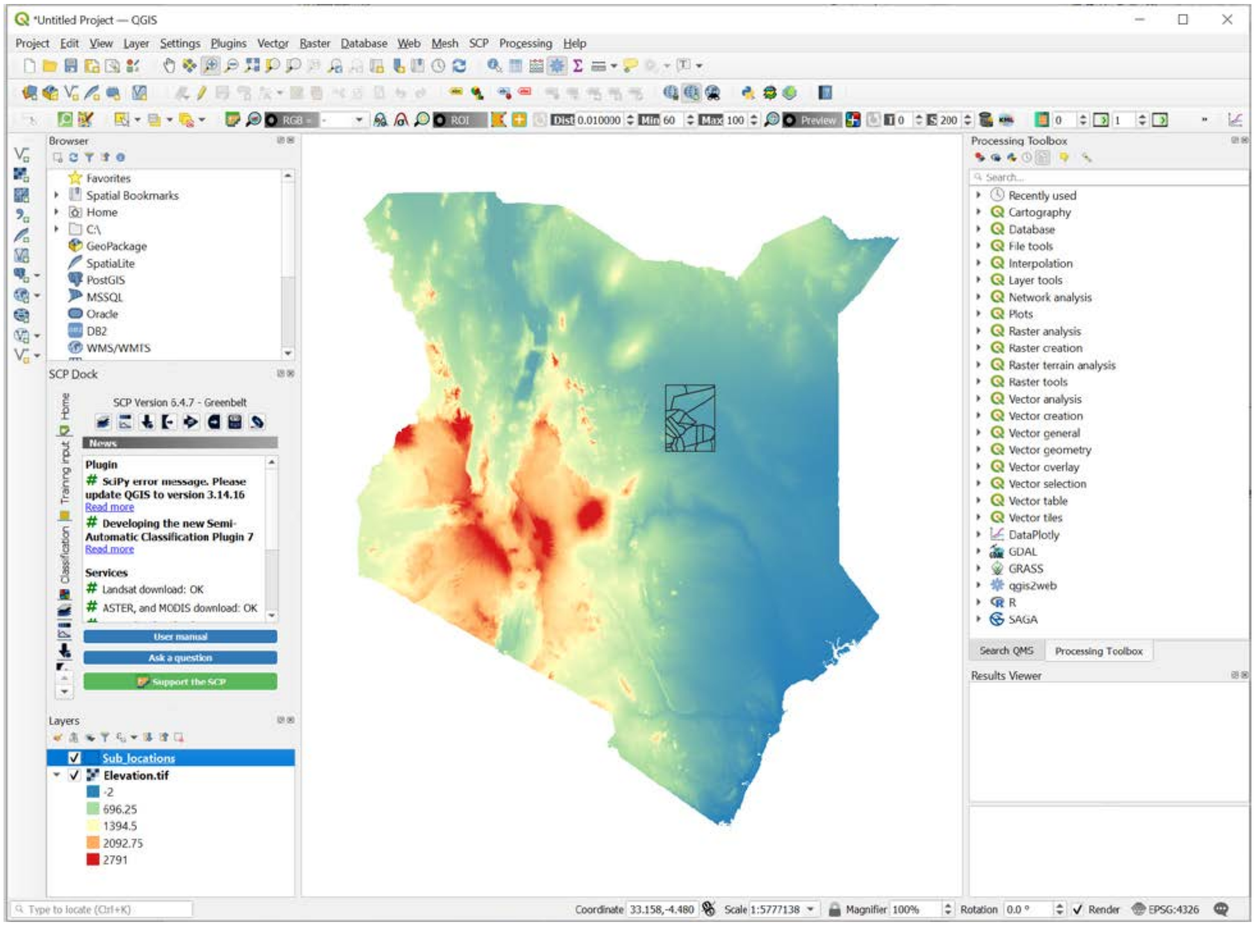
The red shades represent higher elevation points while blue and green areas represent low elevation regions and yellow representing medium elevation areas.

In the next section we will determine the average elevation for each sub location.

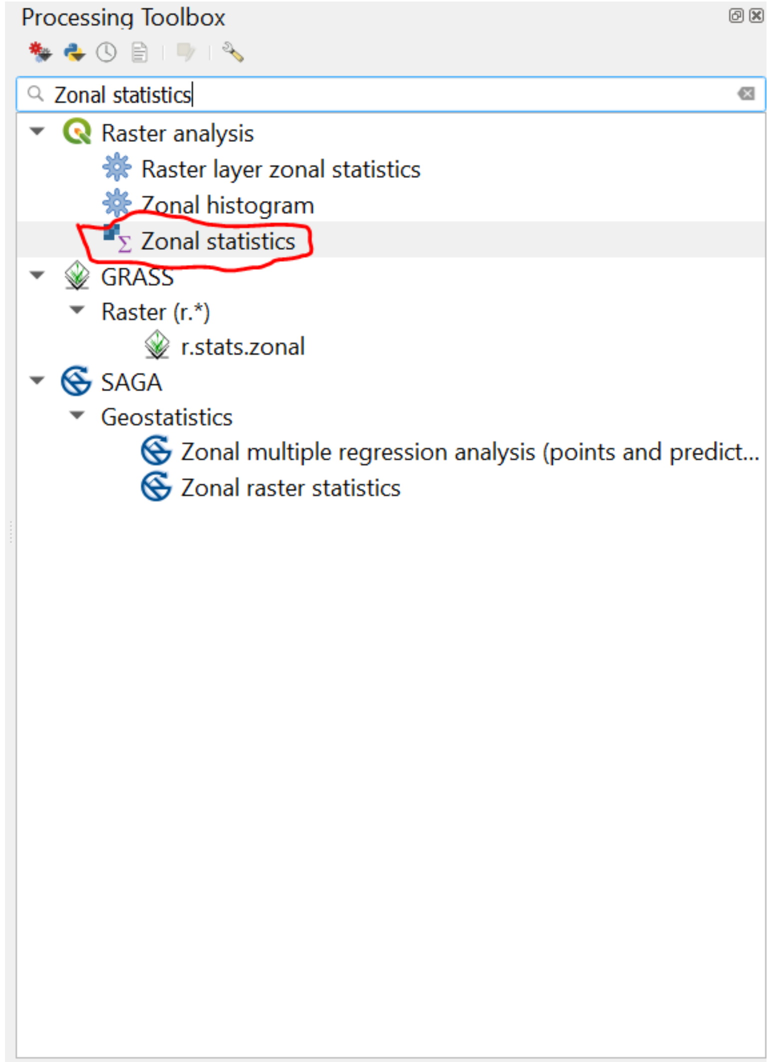
4.6 Average elevation in each sub-location

In this section, we are going to find the average elevation of each sub-location in our study area. This is just for demonstration of vector and raster data analysis. We will require; elevation data (which is the RASTER value data), and the sub-locations (Which is the VECTOR zone data. In QGIS, we will use a tool known as zonal statistics tool.

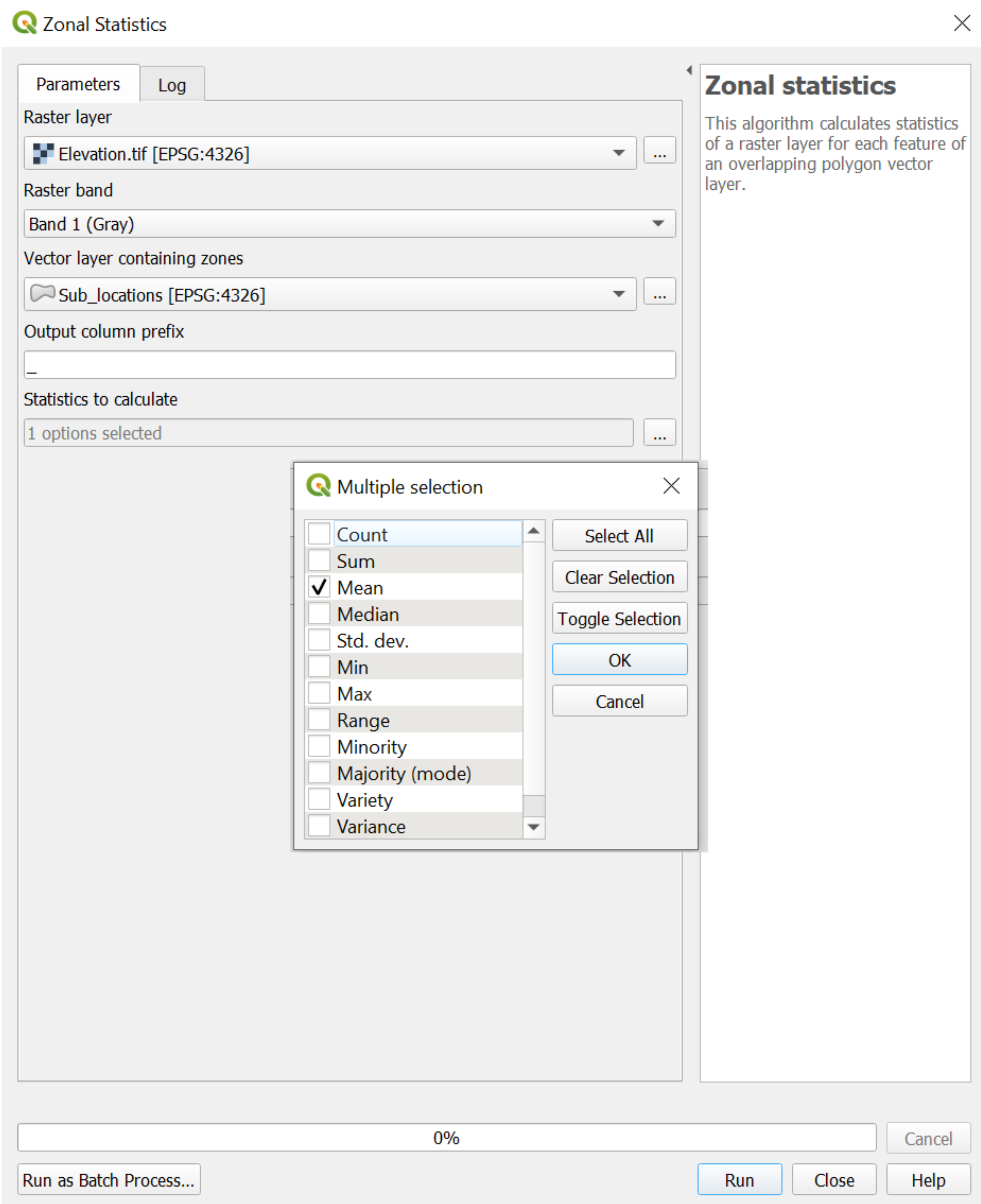
- To the current project, add area of study boundary,**If you don't remember the procedure, check the earlier steps on adding Vector data to QGIS**.
- Add the sub-location shapefile
- Once added be sure to change the color, as shown earlier to indicate only the areas boundary with no fill.
- It should be similar to this one



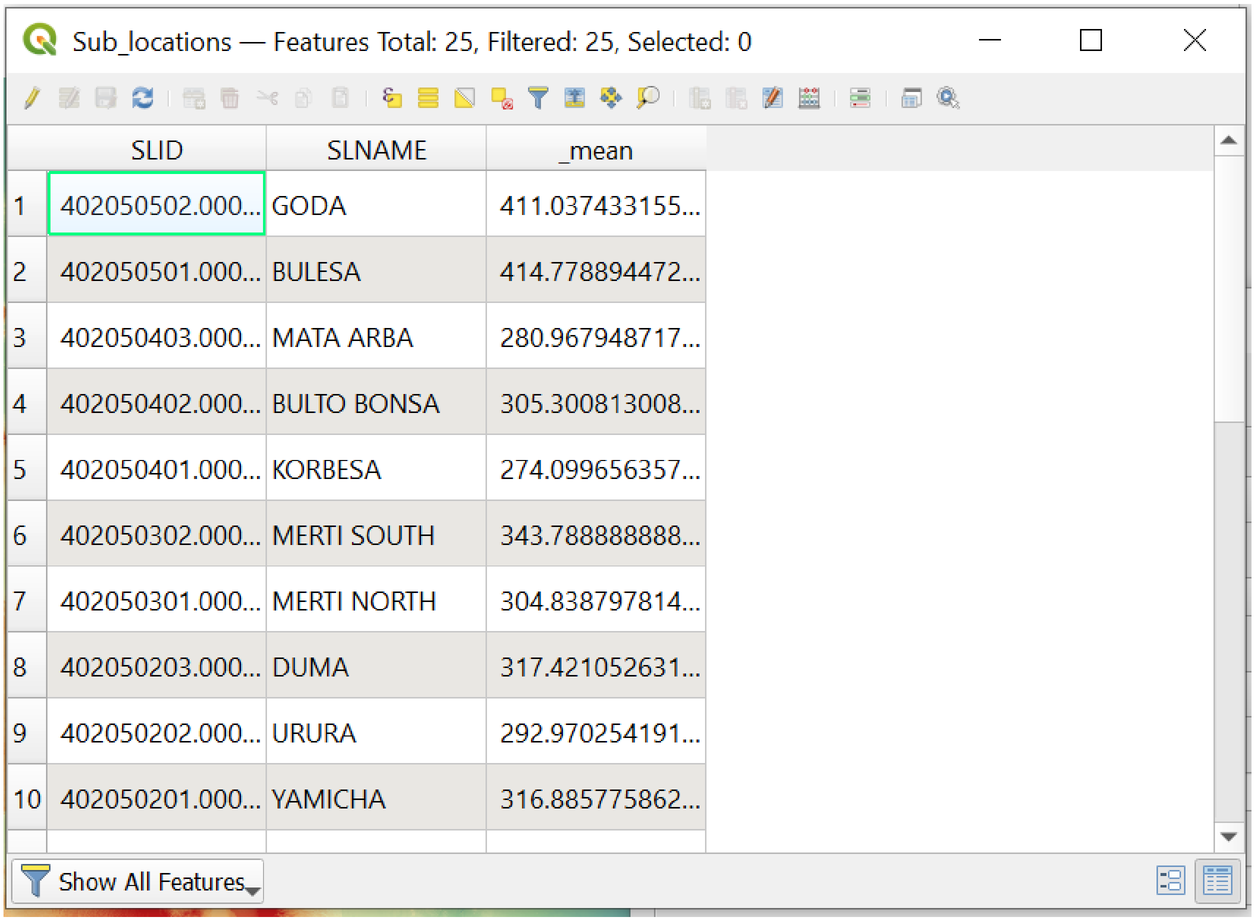
- To get the average elevation for each sub-location we will calculate the statistics for each sub-location using the Zonal statistics tool.
- The Zonal statistics tool outputs a table with statistics such as Mean, Standard Deviation, Median etc.
- In QGIS go to ***processing toolbox*** , in the dialogue window, on the **search box** type ***Zonal Statistics*** and press **Enter**.



- Click on the Zonal Statics tool: under **Raster layer** select ***Elevation.tif*** under **Vector Layer Containing zones** select ***Sub_location*** and finally click …. at **statistics to calculate** box and uncheck the other variables and remain with only ***Mean*** *checked.*



- Once done click ***OK*** and ***Run*** the tool.
- The tool adds the calculated elevation averages to the *Sub_locations.shp* attribute table.
- Right Click on ***Sub_locations→Attribute Table*** to view the averages. A column for ***_mean*** has been added which contains values for average elevation in each sub-location.


Sub_locations — Features Total: 25, Filtered: 25, Selected: 0

	SLID	SLNAME	_mean
1	402050502.000...	GODA	411.037433155...
2	402050501.000...	BULESA	414.778894472...
3	402050403.000...	MATA ARBA	280.967948717...
4	402050402.000...	BULTO BONSA	305.300813008...
5	402050401.000...	KORBESA	274.099656357...
6	402050302.000...	MERTI SOUTH	343.788888888...
7	402050301.000...	MERTI NORTH	304.838797814...
8	402050203.000...	DUMA	317.421052631...
9	402050202.000...	URURA	292.970254191...
10	402050201.000...	YAMICHA	316.885775862...

Show All Features

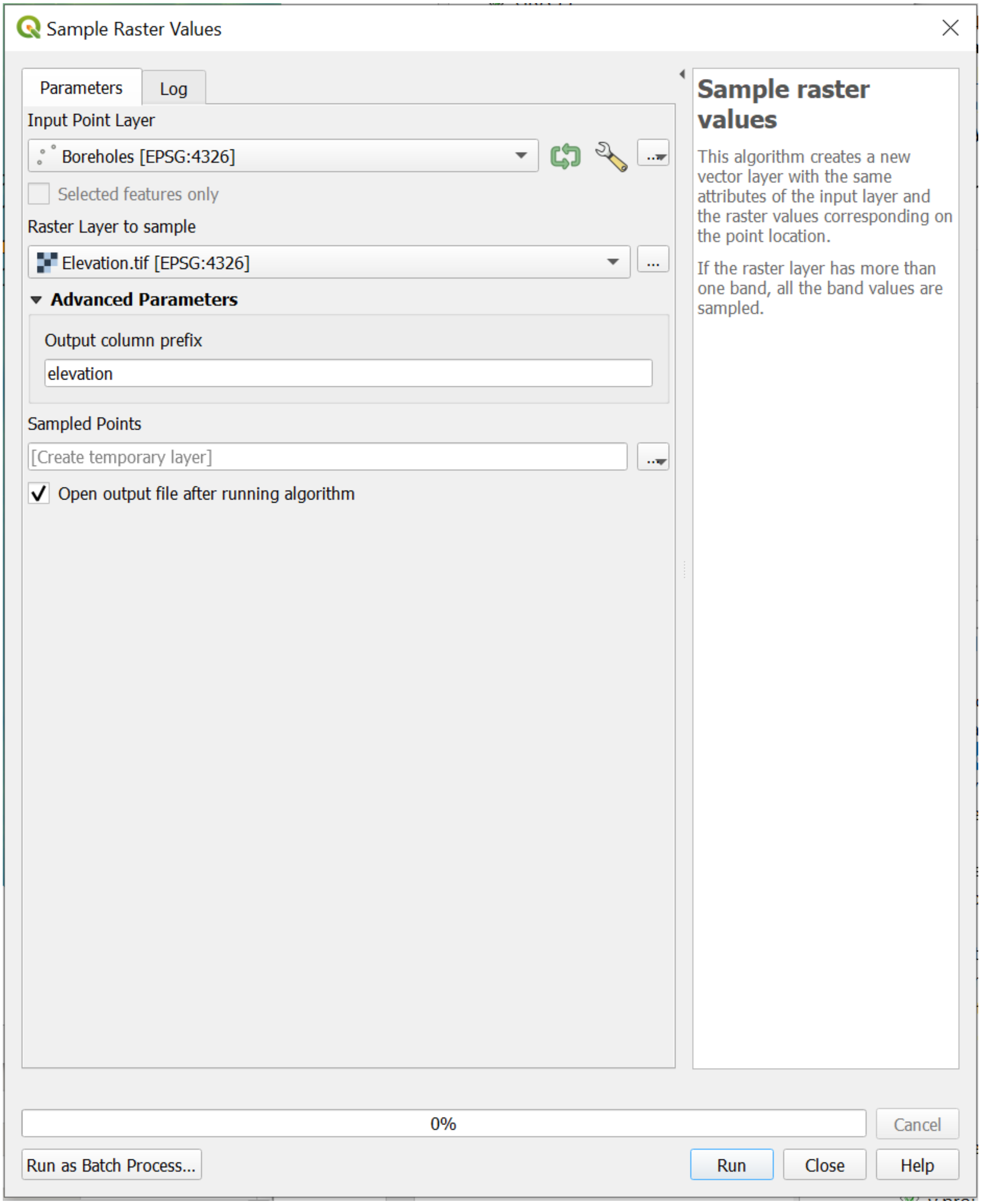
Close the attribute table and save the project.

We have extracted the average elevation for each sub location and next we want to extract the elevation of each borehole

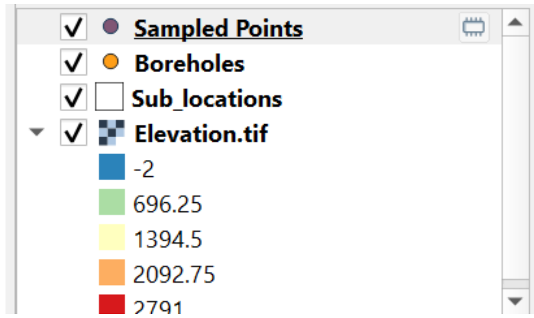
4.7 Extract raster value (elevation) to each point (borehole)

We are going to extract the elevation of each location in which the boreholes are found. To achieve this goal we are going to use a spatial analysis tool called Sample Raster analysis.

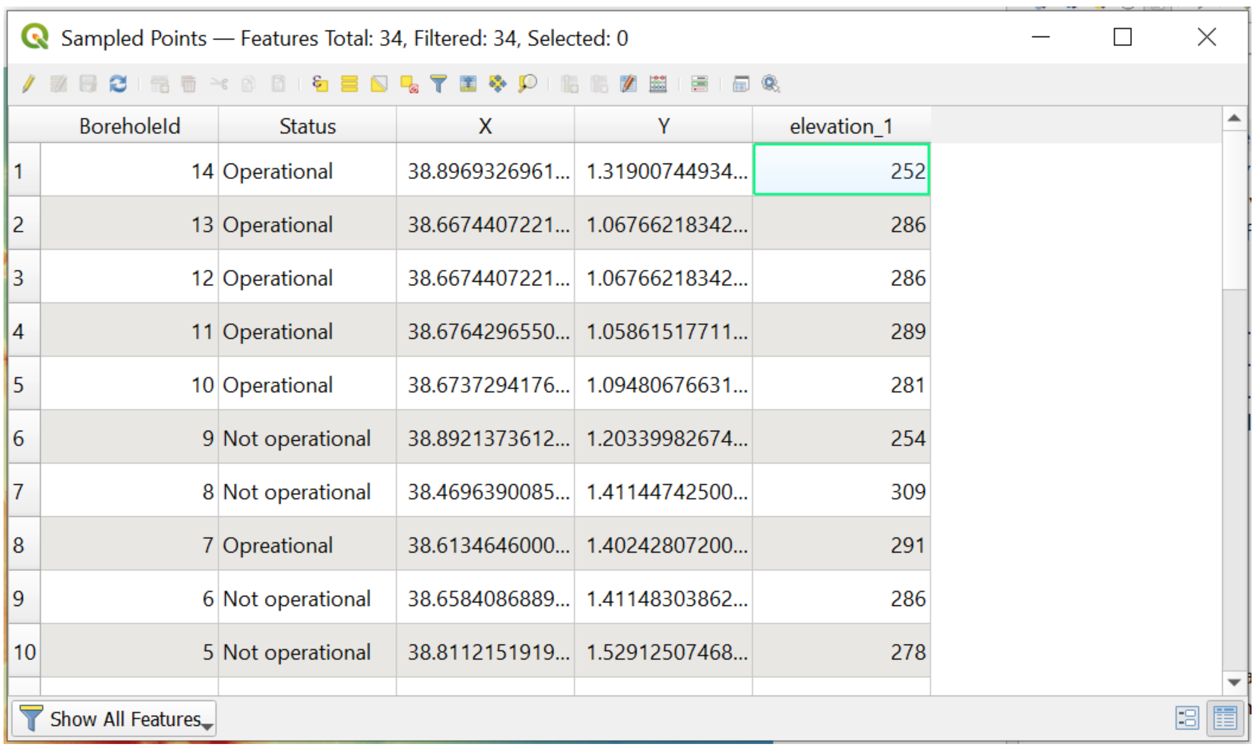
- Add ***Boreholes.shp*** as shown in the previous sections.
- Click on the Processing toolbox
- Search for Sample Raster Analysis
- Double Click on the tool. Under **Input point layer** select ***Boreholes [EPSG:4326]*** for **Raster Layer to Sample** input ***Elevation [EPSG:4326]*** , for **output column prefix** type ***elevation.***



- Click ***Run.***
- After the tool runs a new layer named ***Sampled Points*** is generated.



- The new layer is a layer with all the boreholes, with an addition of the elevation column.
- Open the attribute table to view the elevation height of each borehole.



	BoreholeId	Status	X	Y	elevation_1
1	14	Operational	38.8969326961...	1.31900744934...	252
2	13	Operational	38.6674407221...	1.06766218342...	286
3	12	Operational	38.6674407221...	1.06766218342...	286
4	11	Operational	38.6764296550...	1.05861517711...	289
5	10	Operational	38.6737294176...	1.09480676631...	281
6	9	Not operational	38.8921373612...	1.20339982674...	254
7	8	Not operational	38.4696390085...	1.41144742500...	309
8	7	Opreational	38.6134646000...	1.40242807200...	291
9	6	Not operational	38.6584086889...	1.41148303862...	286
10	5	Not operational	38.8112151919...	1.52912507468...	278

From the Attribute table we can see the last column has the elevation (elevation_1) of each borehole.

- Using steps that we learnt earlier in the tutorial, export the new layer as a shapefile and give it a name of your choice.
- Close the attribute table and save the project

We have learnt the third method to conduct spatial analysis to extract raster values for known locations for instance elevation of boreholes in our study area.

4.8 Estimating population of people accessing specific boreholes

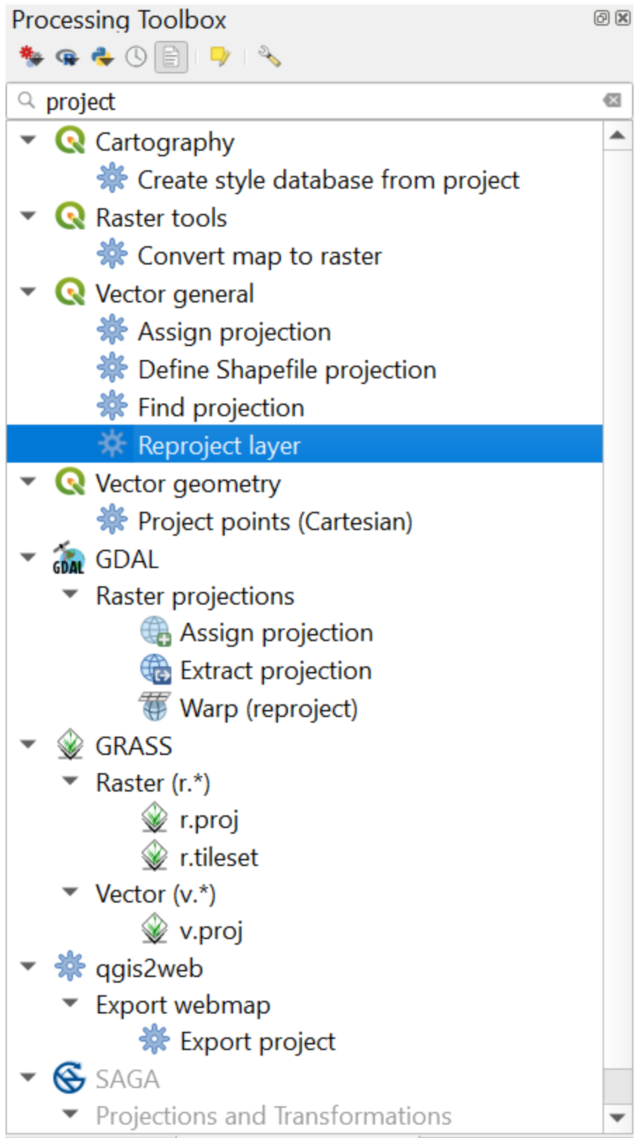
The next aspect of spatial analysis that we shall conduct will be to determine the number of population using each borehole. This can be people within a 1km radius of the borehole. For this task we shall require population (raster) data and boreholes (point) data.

4.8.1 Changing projection of vector data

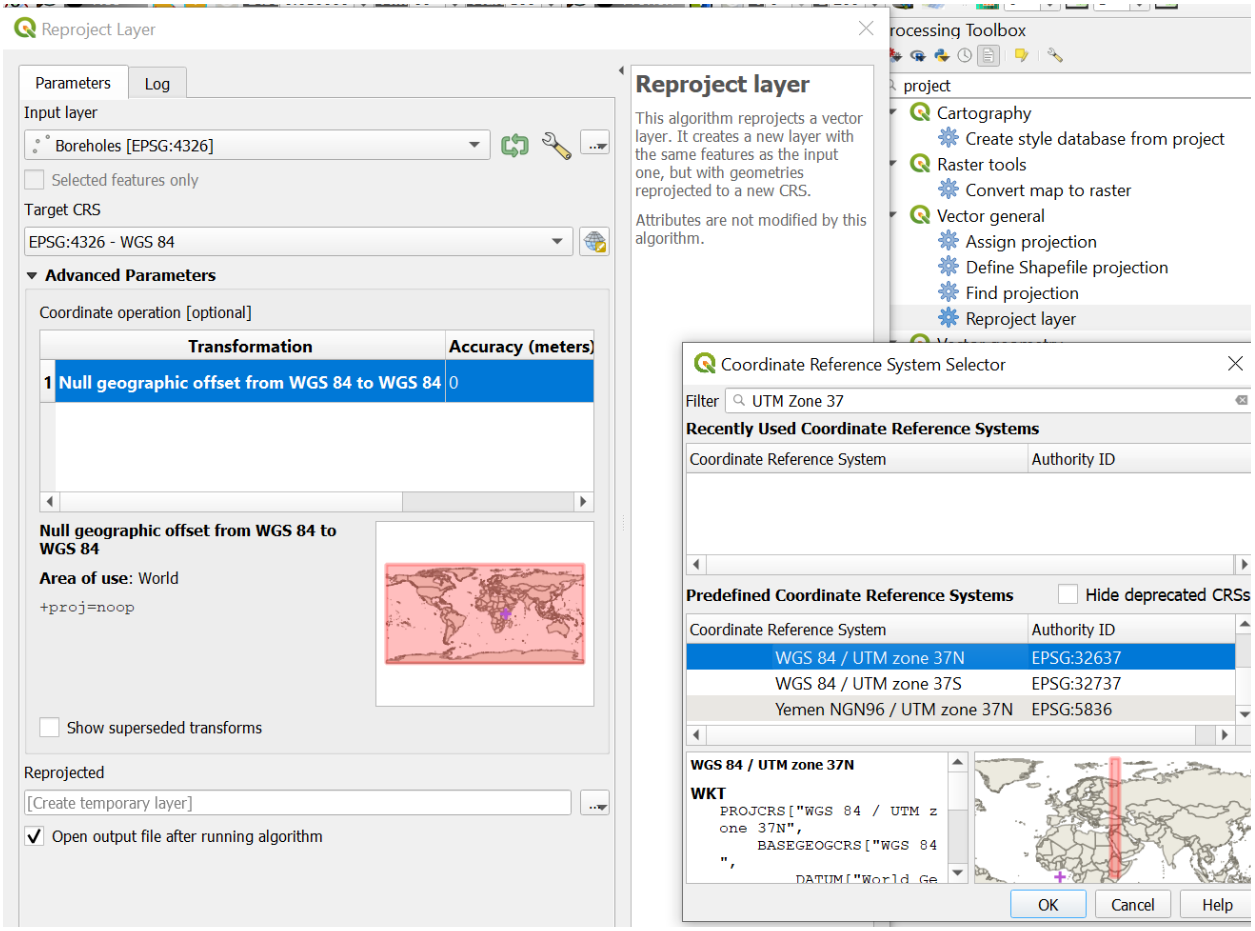
Spatial analysis requiring distance measurements require Layers to be converted to a projected coordinate system such as Universal Transverse Mercator (UTM). This is because projected coordinate systems to estimate accurate ground distances in metric units.

We will convert **Boreholes** shapefiles, which is in geographic coordinate system (World Geodetic System – 1984) to a desired UTM projected coordinate system.

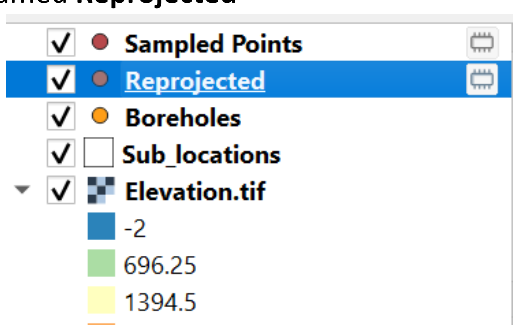
- In Processing toolbox type ***Project*** and **enter**
- Select **Reproject Layers**



- Once the tool is Open, under **Input layer** select ***Boreholes***
- For Target CRS click on the  button and type ***WGS 84 / UTM zone 37N***



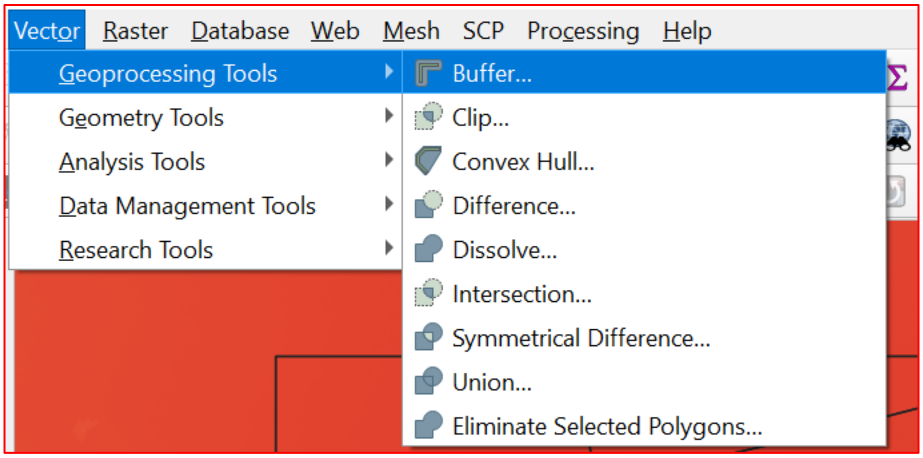
- Click ***OK*** and then **Run**
- A new layer is created named **Reprojected**



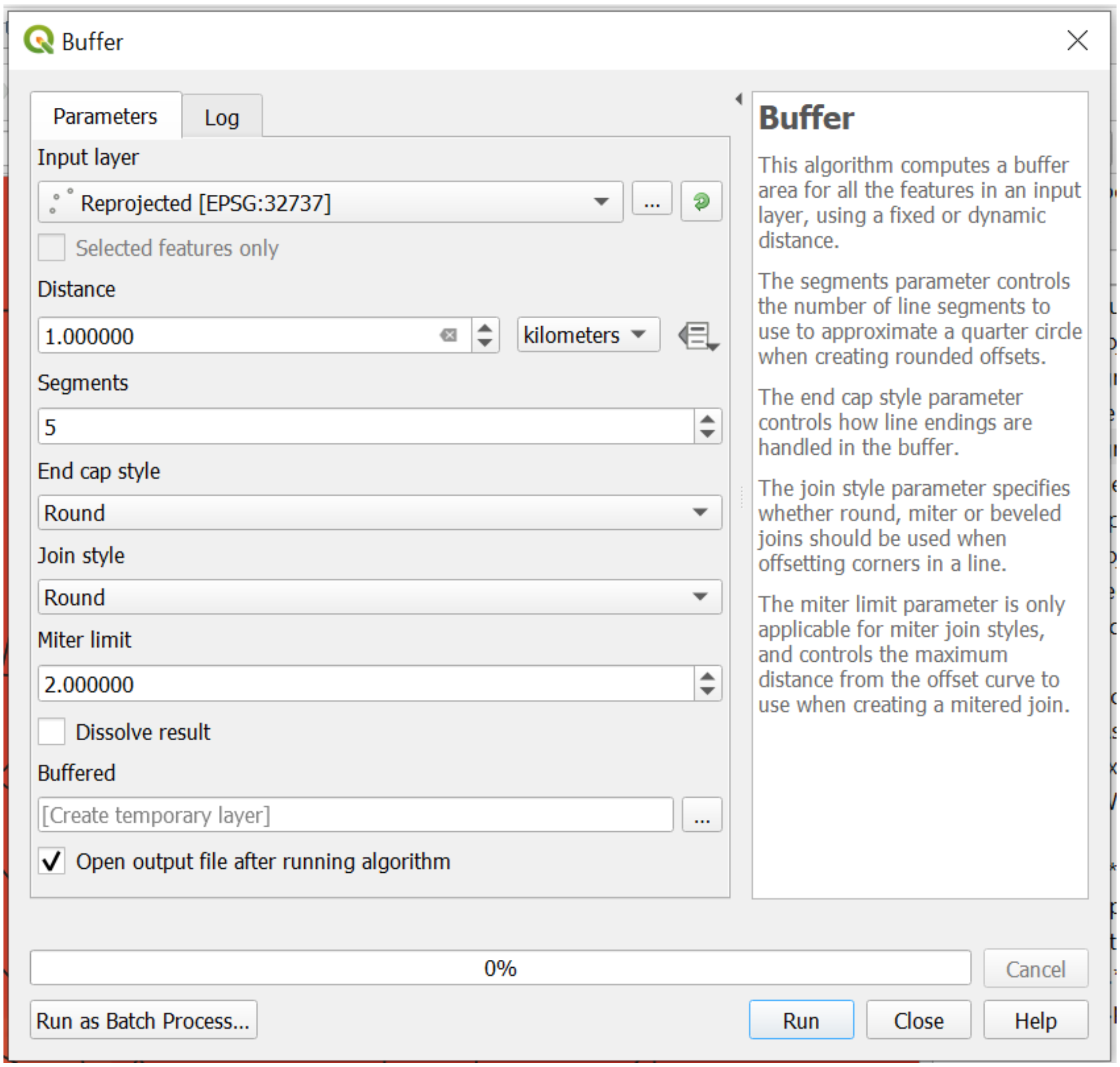
4.8.2 Buffer analysis on vector data

To create a polygon of uniform distance around vector features, we create a buffer. This is done through a tool that is referred to as buffer analysis.

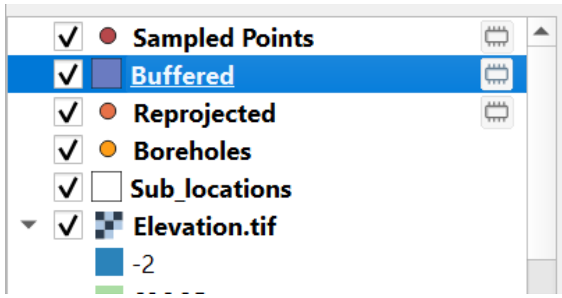
- We are going to use this layer to create a buffer of 1 km around each borehole.
- On the topmost tab, click on ***Vector→ Geoprocessing Tools→ Buffer***



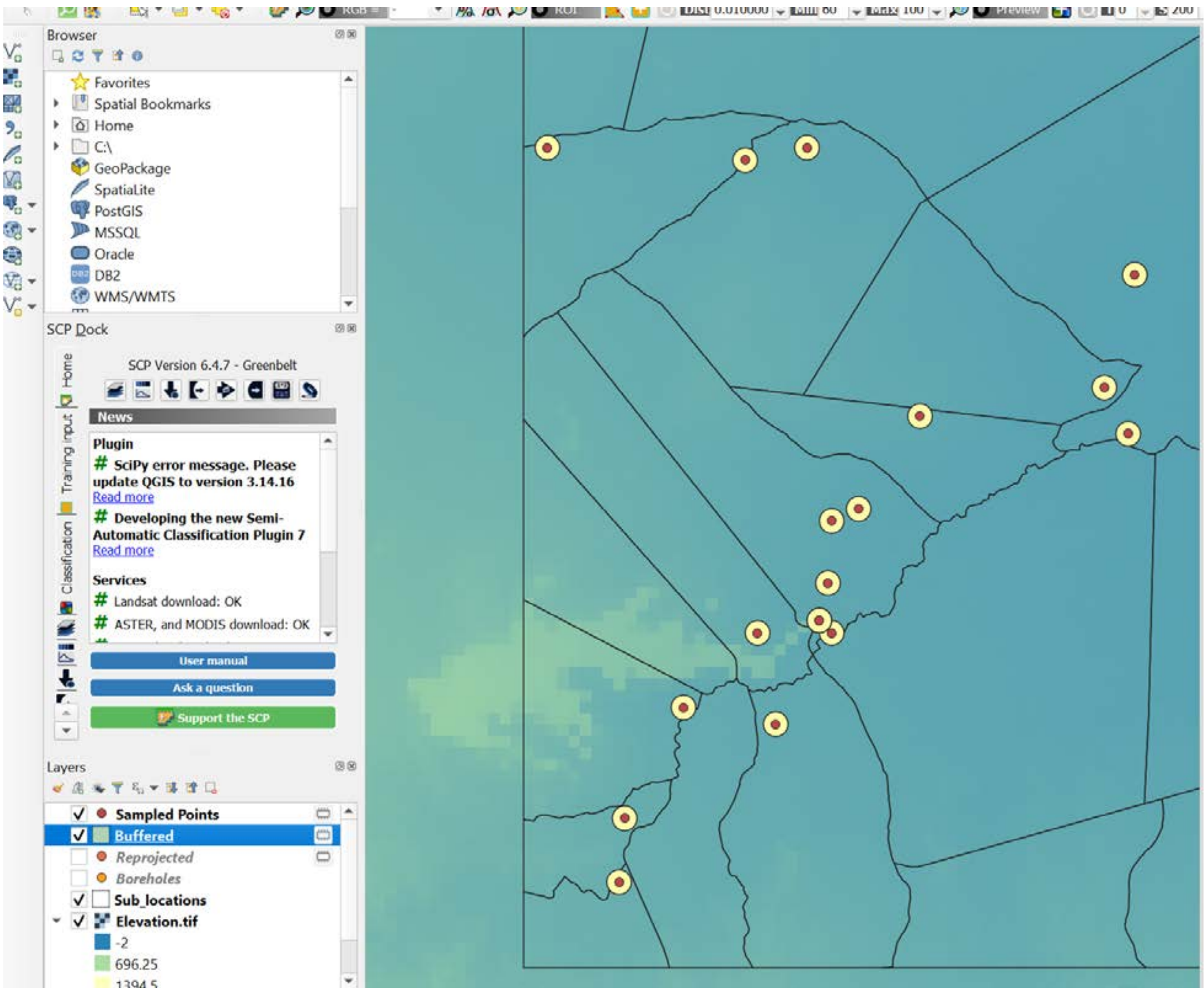
- On the new dialogue window, under **Input layer** select ***Reprojected [EPSG:32737]*** For **distance** Input ***1.0*** and select ***Kilometers.***



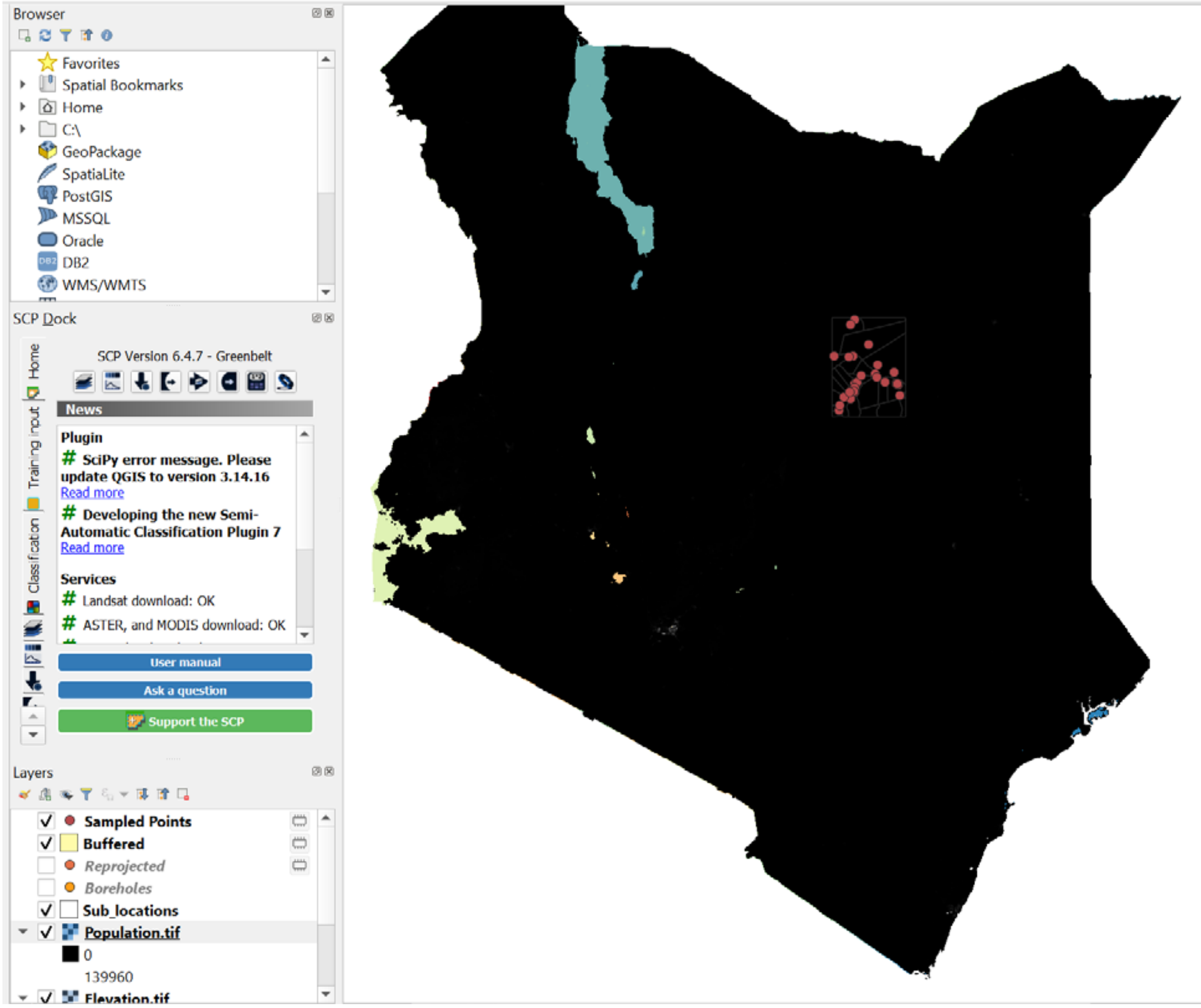
- Click ***Run***
- A new layer named ***Buffered*** is created and automatically added to the map.



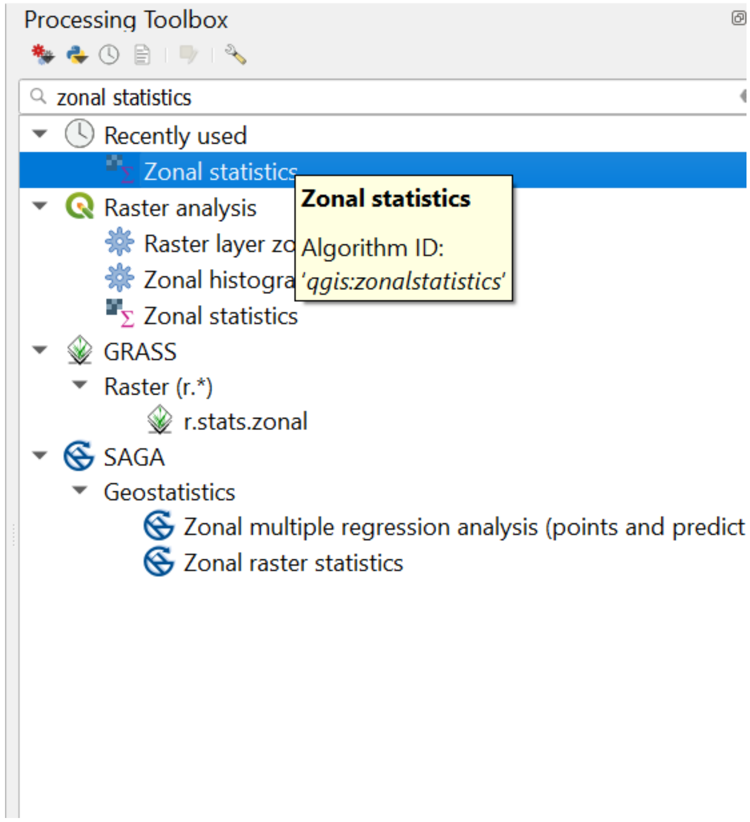
- The new layer ***Buffered*** indicates areas of radius 1km created around each borehole.
- We can see that the layer is indicated as circles around boreholes.



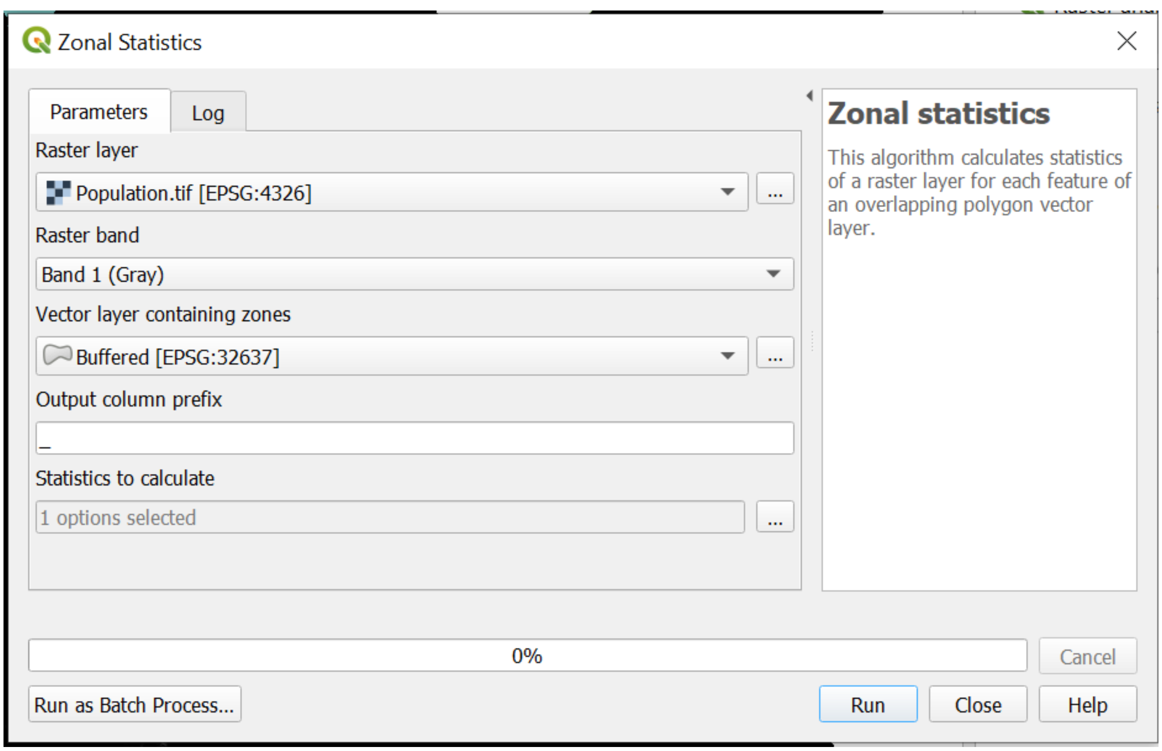
- We are again going to use Zonal statistics to determine the number of People within this region for each borehole.
- Now click the population data which is in Raster format, in the Raster folder, indicated as ***Population*** add it to the map.
- Once added the map should appear similar to this one



- Open the **Processing toolbox**  and search for **Zonal Statistics**



- Double click on ***Zonal Statistics***
- For **Raster Layer** select ***Population[EPSG:4326,*** for **Vector Layer Containing Zones** select ***Buffered[EPSG:32737]***
- Click ***Run***



- After running the tool, a new column is added to the attribute tables of the Buffered layer for the SUM of people in the 1km buffer.

Buffered — Features Total: 34, Filtered: 34, Selected: 0

	BoreholeId	Status	X	Y	_sum
1	13	Operational	38.6674407221...	1.06766218342...	1628.22103881...
2	12	Operational	38.6674407221...	1.06766218342...	1628.22103881...
3	11	Operational	38.6764296550...	1.05861517711...	665.516052246...
4	3	Not operational	38.7406797545...	1.21626489120...	109.645910263...
5	10	Operational	38.6737294176...	1.09480676631...	84.6546649932...
6	17	Not operational	39.1157744960...	1.02265475037...	84.6208820343...
7	33	Not operational	38.6225019939...	1.05860904084...	81.6983222961...
8	32	Not operational	38.5685819794...	1.00431479743...	47.9866037368...
9	26	Not operational	39.0882386956...	1.14114056569...	43.6885547637...
10	25	Not operational	39.0882386956...	1.14114056569...	43.6885547637...

Show All Features

- The values appear with decimal places since the population raster was encoded using the floating point data type. These can be converted easily to integers by creating a new field and assigning integer values of the _sum. Using the statistics tool, the sum of population in the buffer was approximately 4880.

Buffered — Features Total: 34, Filtered: 34, Selected: 0

	BoreholeId	Status	X	Y	_sum	Population
1	10	Operational	38.6737294176...	1.09480676631...	84.6546649932...	85
2	17	Not operational	39.1157744960...	1.02265475037...	84.6208820343...	85
3	33	Not operational	38.6225019939...	1.05860904084...	81.6983222961...	82
4	23	Not operational	38.9709315652...	1.14903173253...	8.23799121379...	8
5	6	Not operational	38.6584086889...	1.41148303862...	7.65894460678...	8
6	11	Operational	38.6764296550...	1.05861517711...	665.516052246...	666
7	32	Not operational	38.5685819794...	1.00431479743...	47.9866037368...	48
8	26	Not operational	39.0882386956...	1.14114056569...	43.6885547637...	44
9	25	Not operational	39.0882386956...	1.14114056569...	43.6885547637...	44
10	29	Operational	39.0972267821...	1.13331366821...	43.5979924201...	44

Show All Features

-

In this section we have learnt how to extract the number of people in a particular polygon from a population raster. For instance we have created a 1 Km buffer around the boreholes and determined the number of people living within the 1 Km radius.

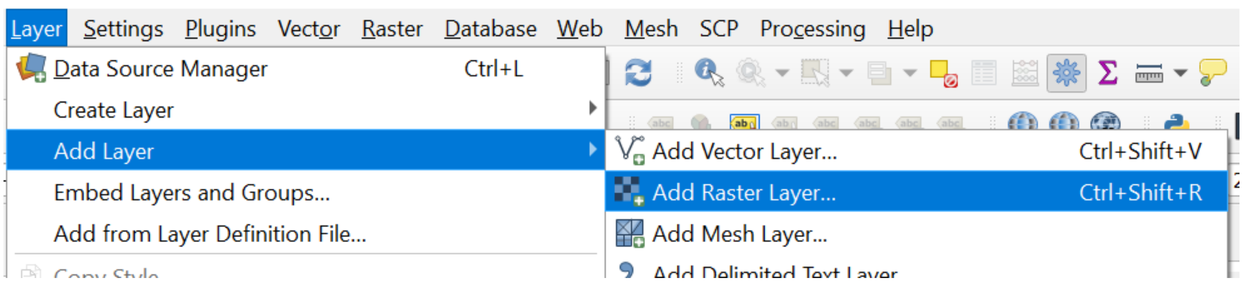
- Close the attribute table, save the project
- Start a new project

4.9 Land Use/ Land Cover

We are going to add a land cover map and clip using our area of study map so that we can view the land cover of the study area.

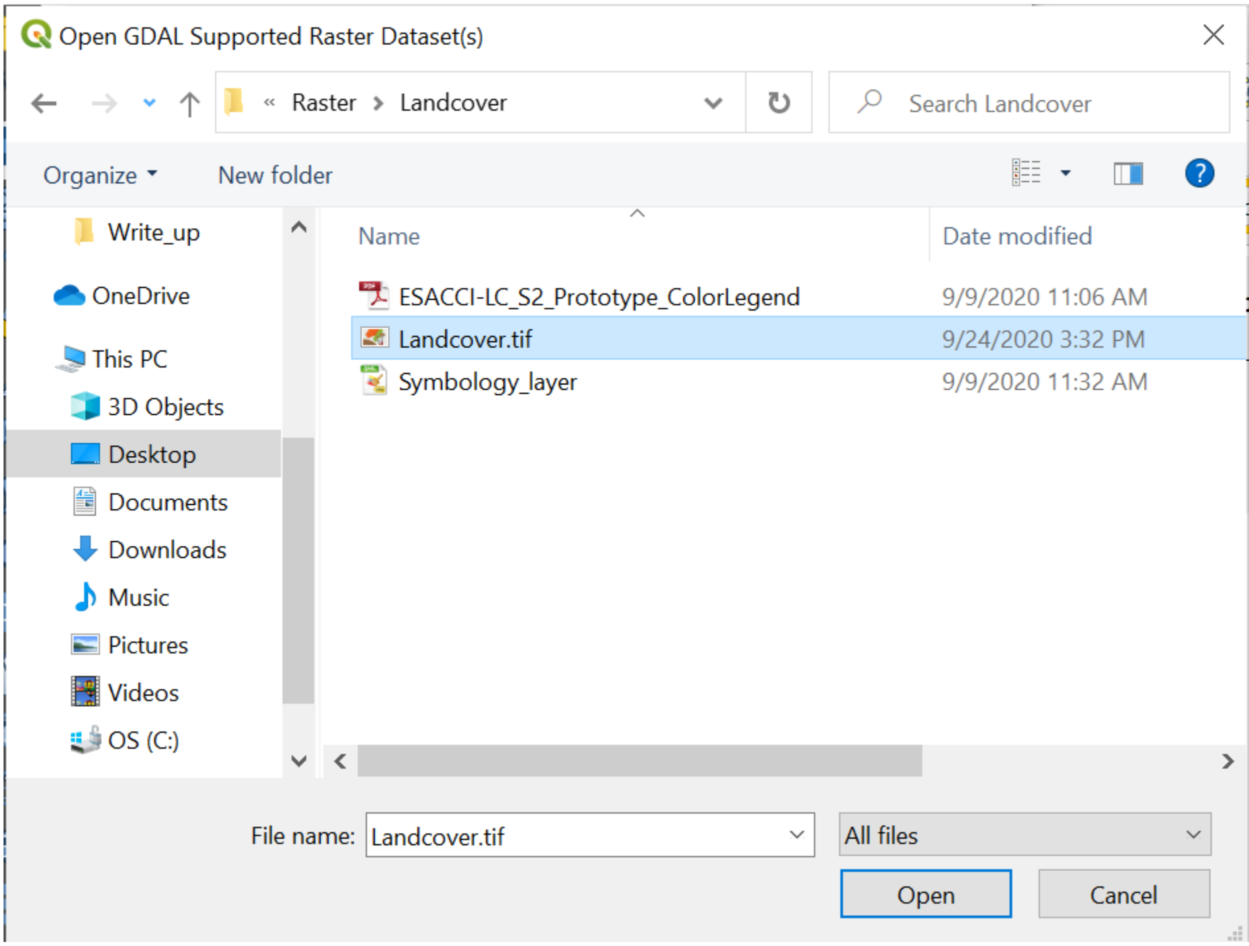
Land cover data files are usually in Raster format. When adding land cover data in raster format use the procedure for adding Raster files that we learnt earlier on.

- Click on ***Layer→Add Layers→Add Raster Layers…***

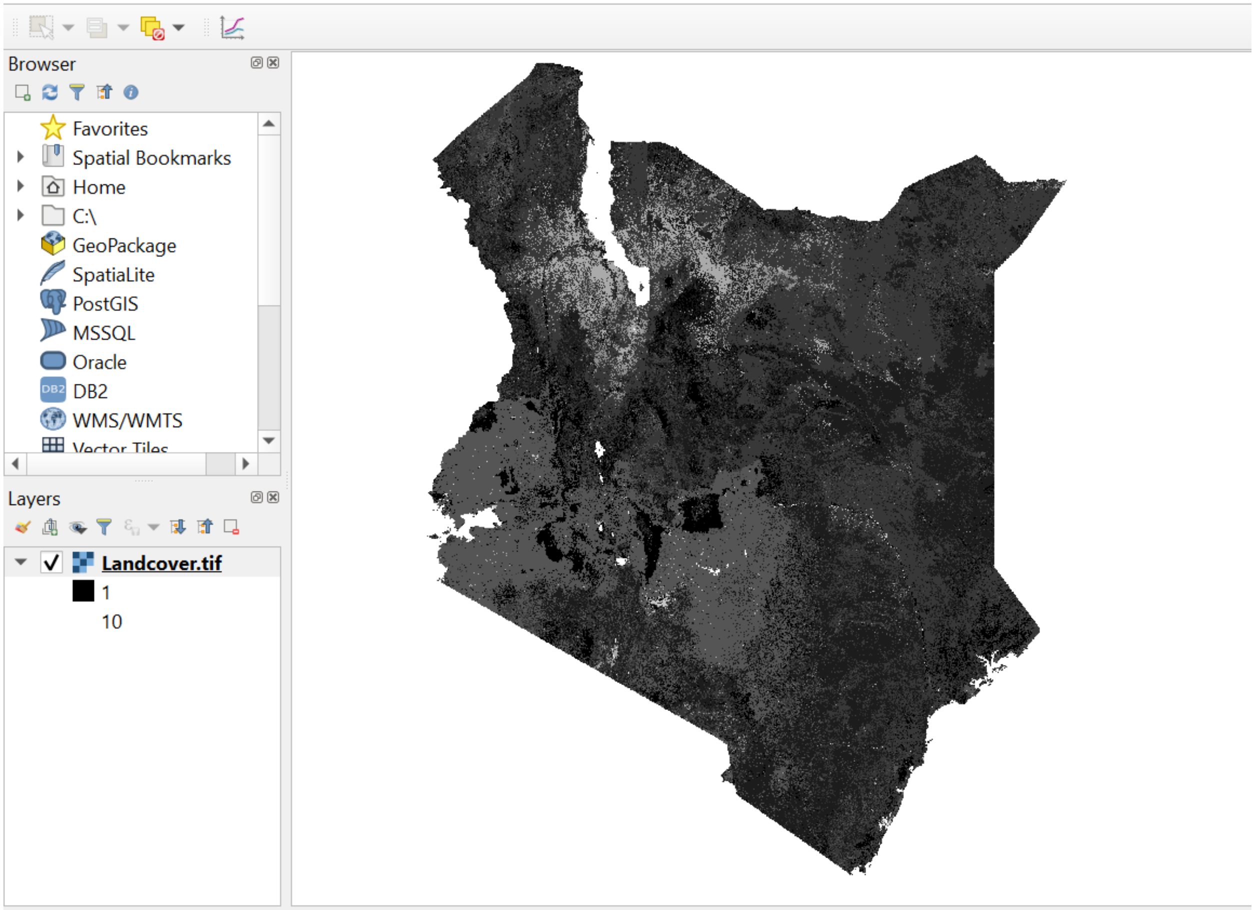


- On the new dialogue use the  button to navigate to the folder with ***Landcover*** and click open

- Once open select ***Landcover.tif*** Click Open, on the new dialogue window click ***Add*** then ***Close*** the dialogue window

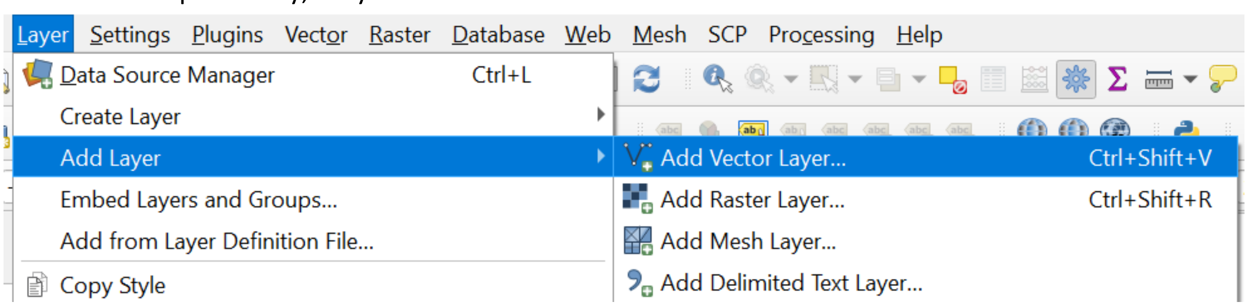


- The layer is added to the map canvas. It should appear similar to the one below.

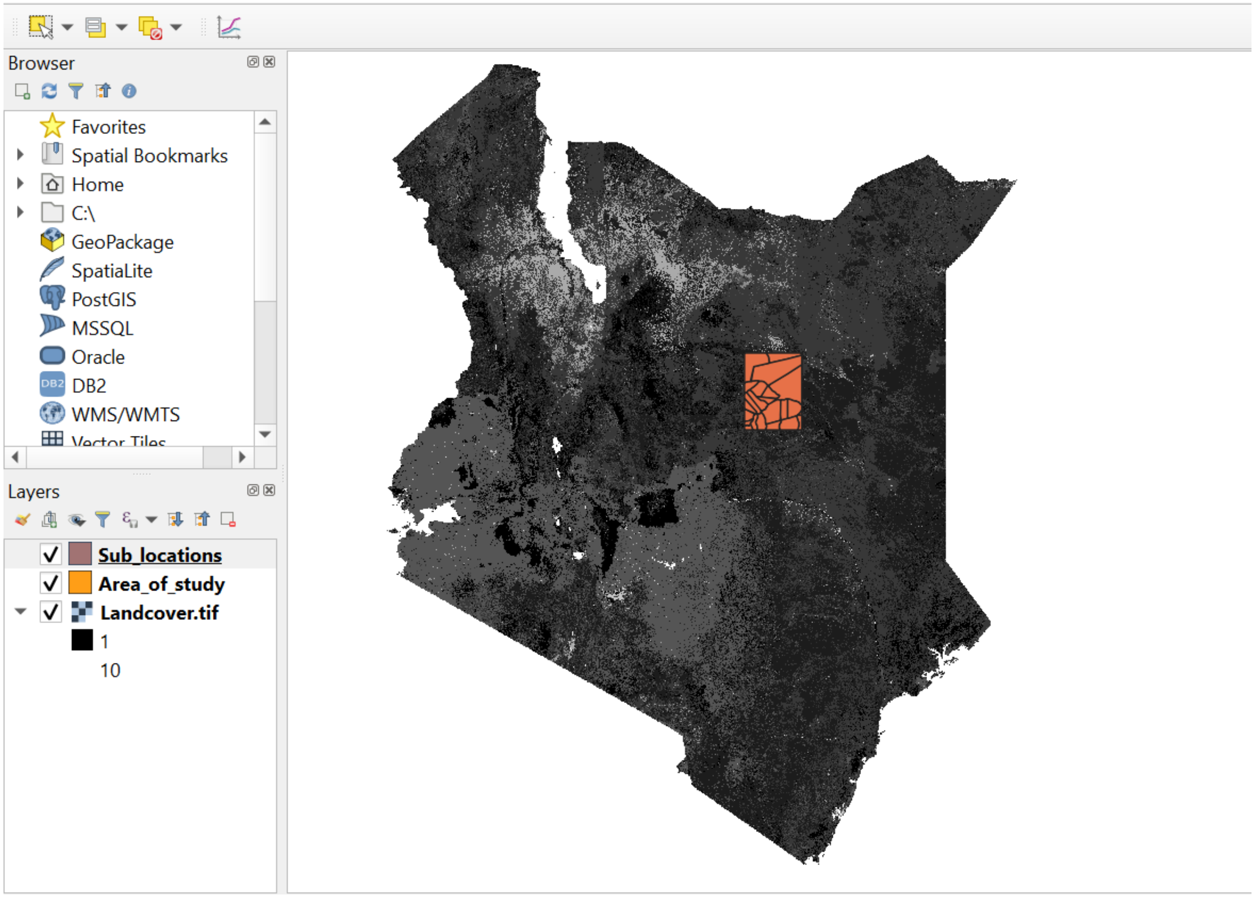


Now we are going to clipping the layer using our study Area;

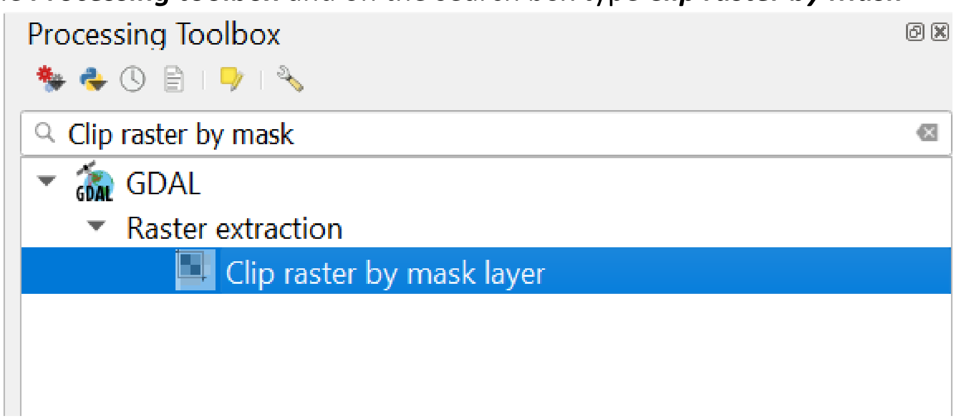
- Add the following layers to the canvas ***Area_of_study.shp, and Sub_locations.shp.*** Remember we used them previously, they are both Vector data files.



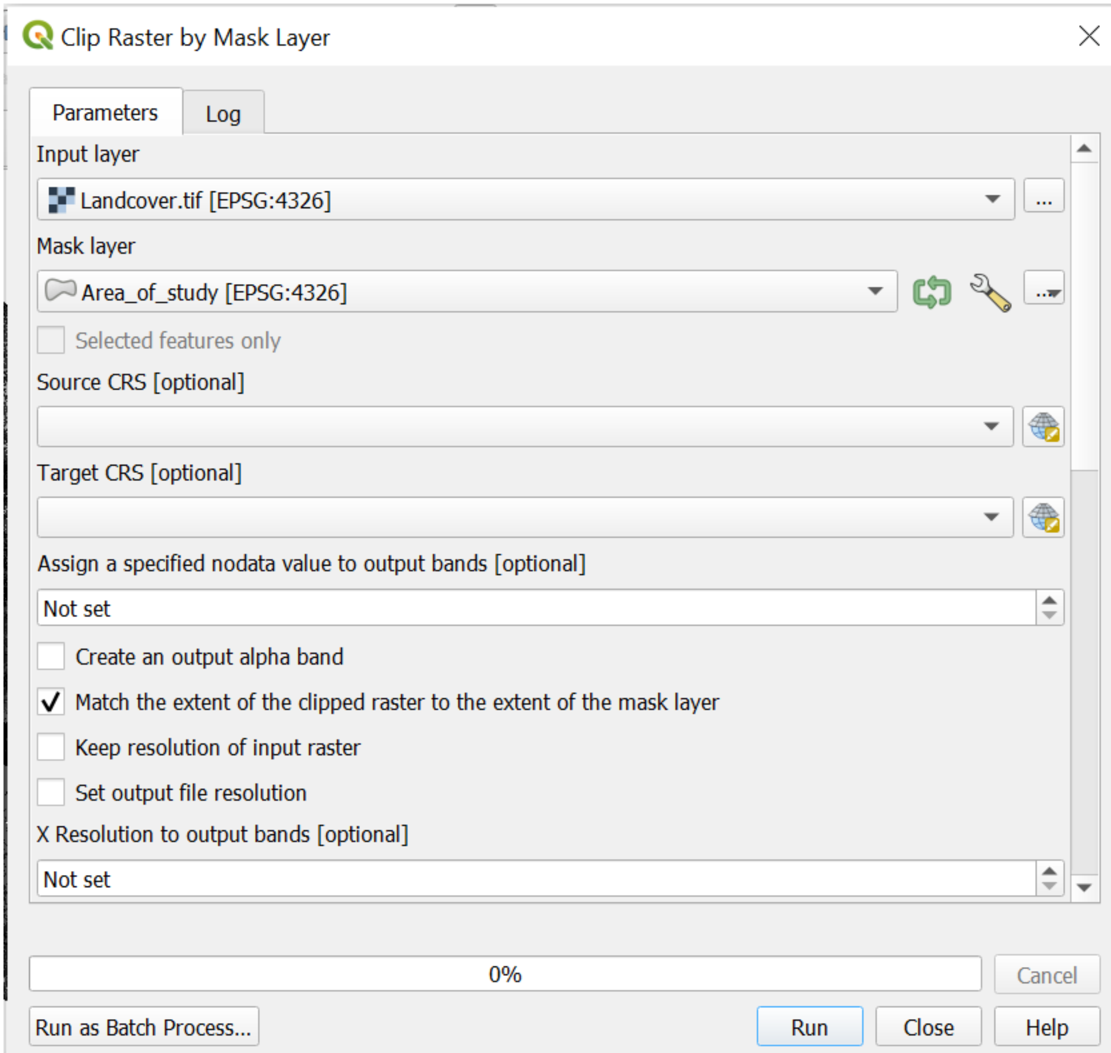
- Once added they should appear as below



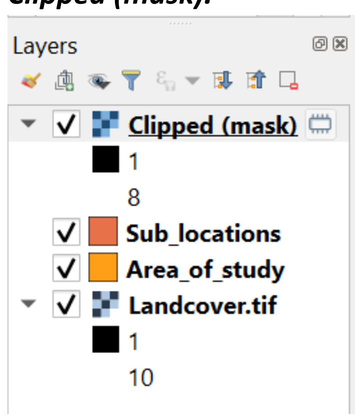
- The next step is extracting the land cover within the Area of study.
- Open the ***Processing toolbox*** and on the search box type ***Clip raster by Mask***



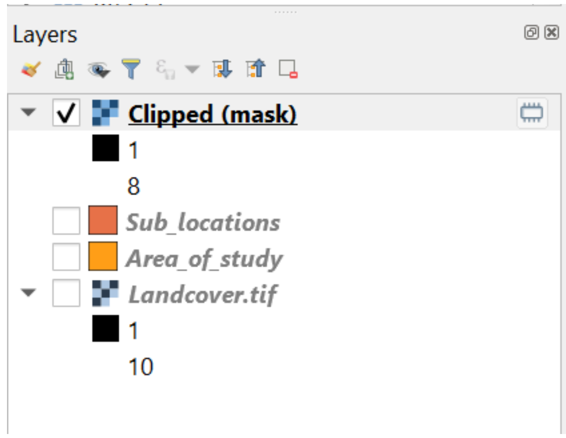
- Click to Open the tool for Input layer select ***Landcover[EPSG:4326]*** for Mask layer select ***Area_of_study[EPSG:4326]*** then click ***Run***



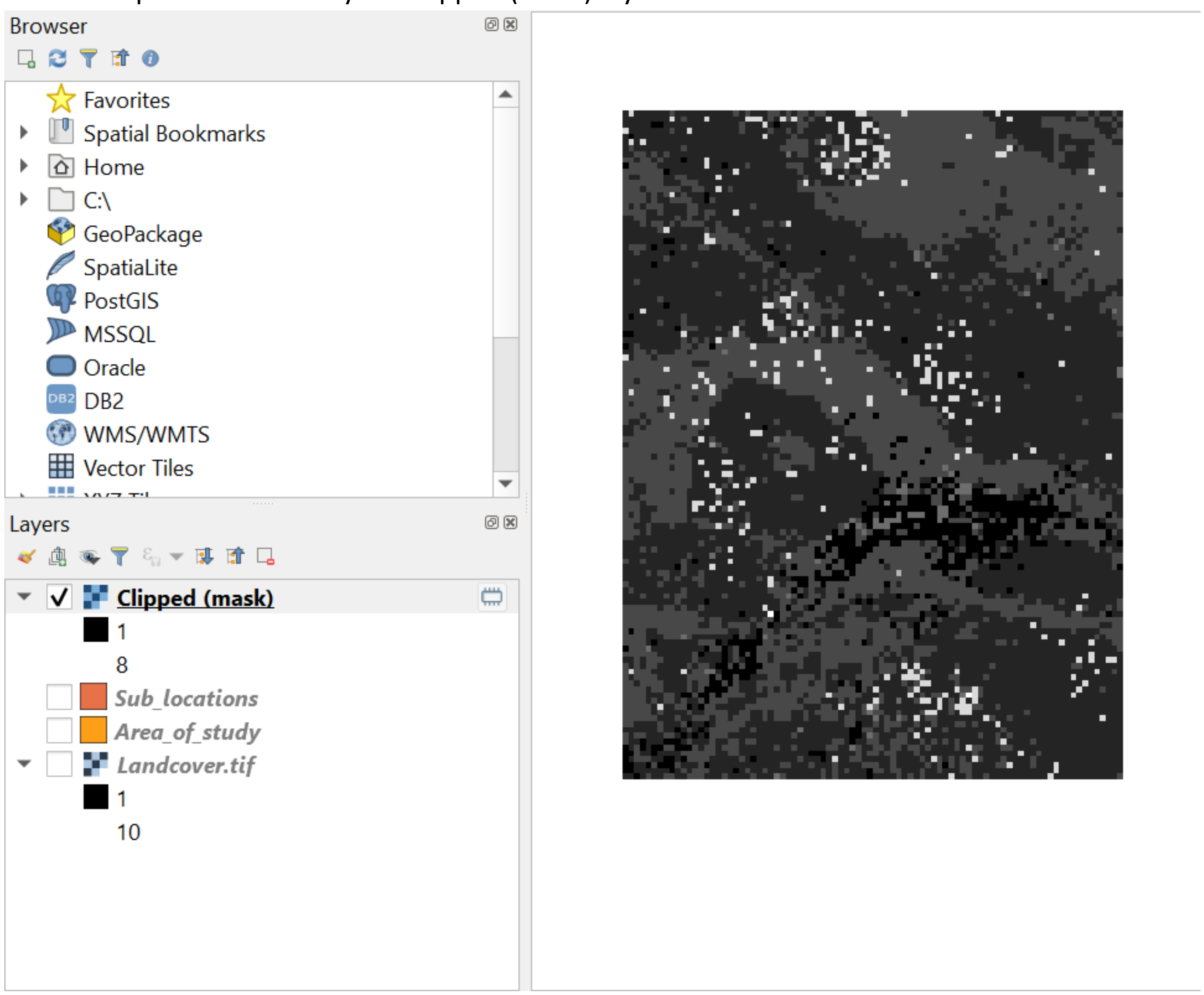
- The land cover layer is clipped and a new land cover layer clipped to the study area is added to the map. It is named ***Clipped (mask).***



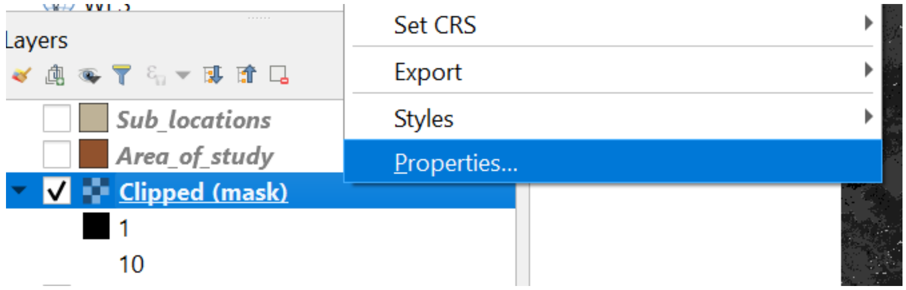
- Before we apply a different symbology we need to hide all (turning off) the other layers except ***Clipped (mask).*** To do this uncheck the box before each layer.



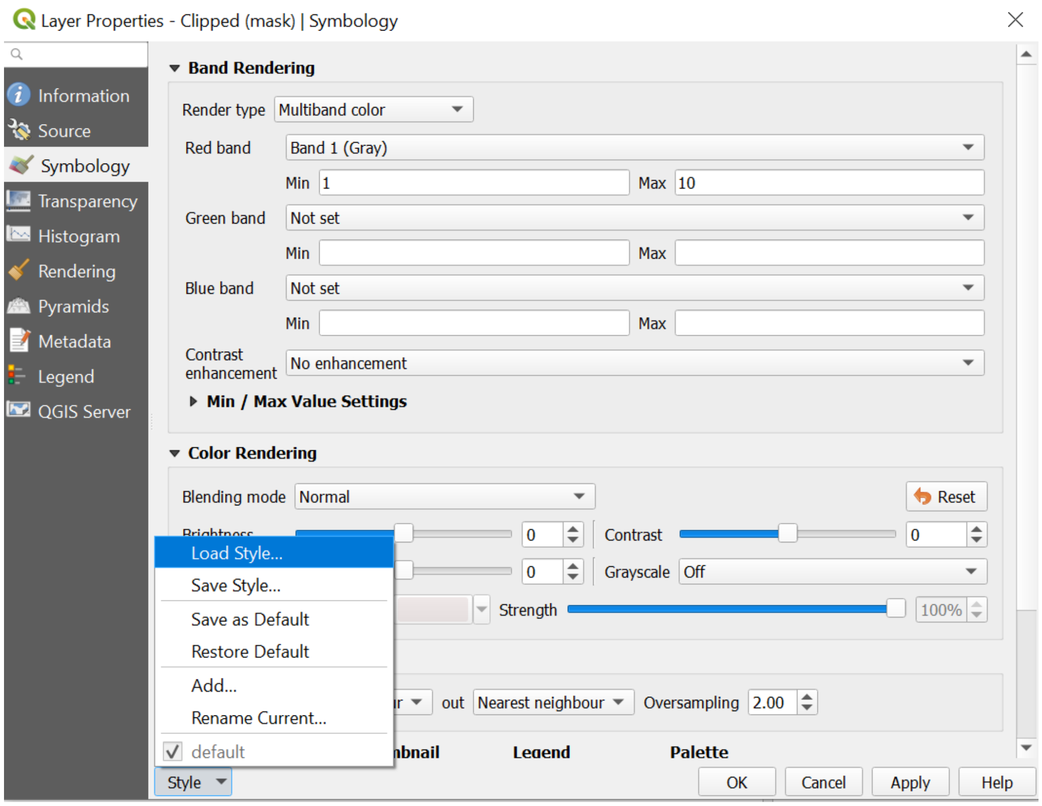
- On the map we can see only the clipped (mask) layer is visible



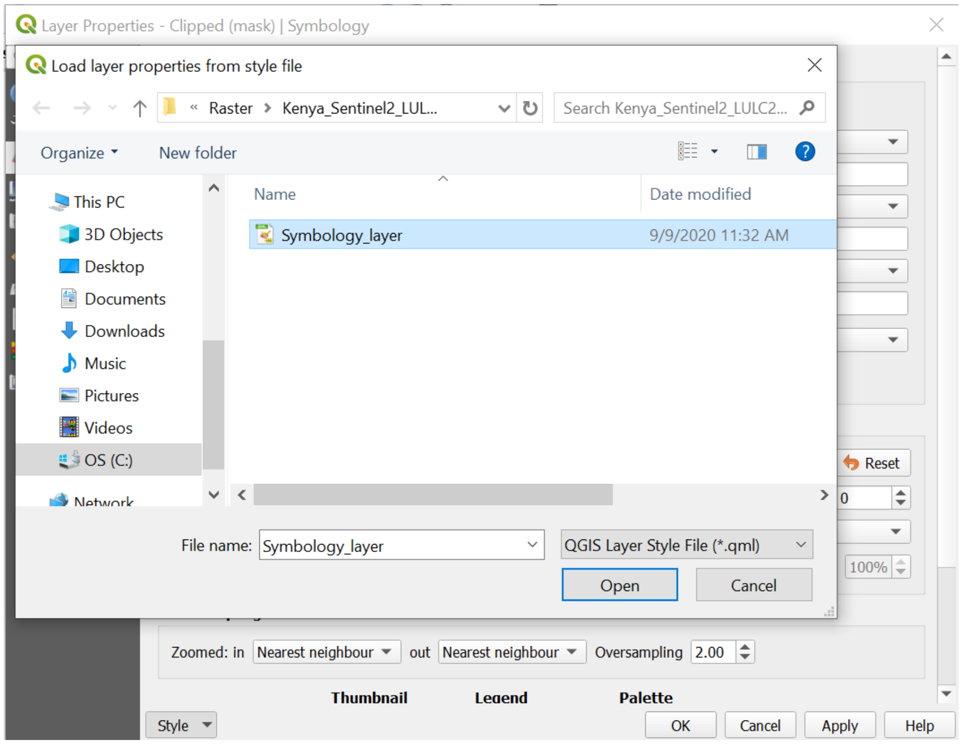
- The next step is to change the symbology of the layer in order to show the land cover classes better.
- Double Click on ***Clipped(Mask)*** layer→ ***Properties***



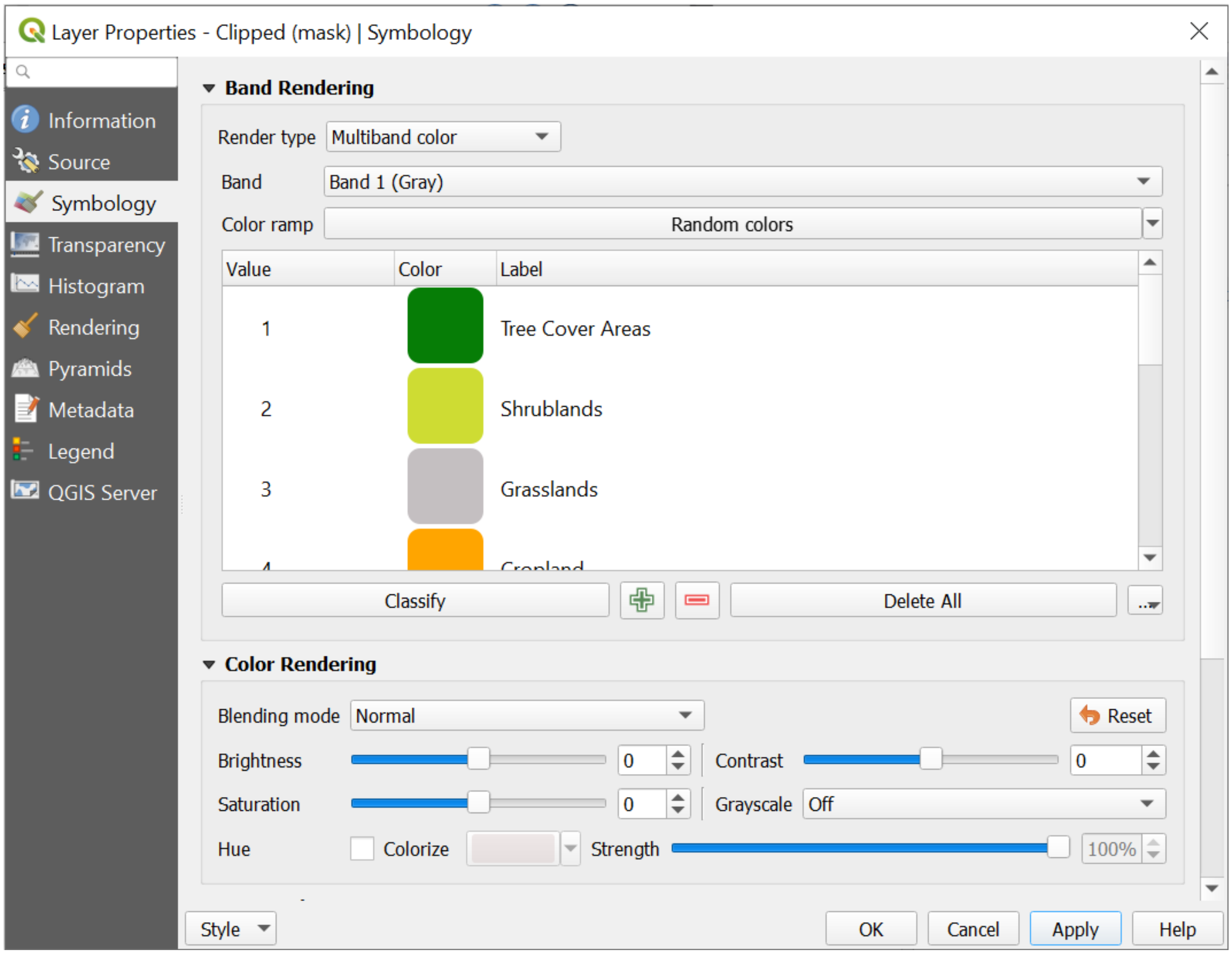
- Go to ***symbology,*** on the lower left corner click on ***Style, Load style…***



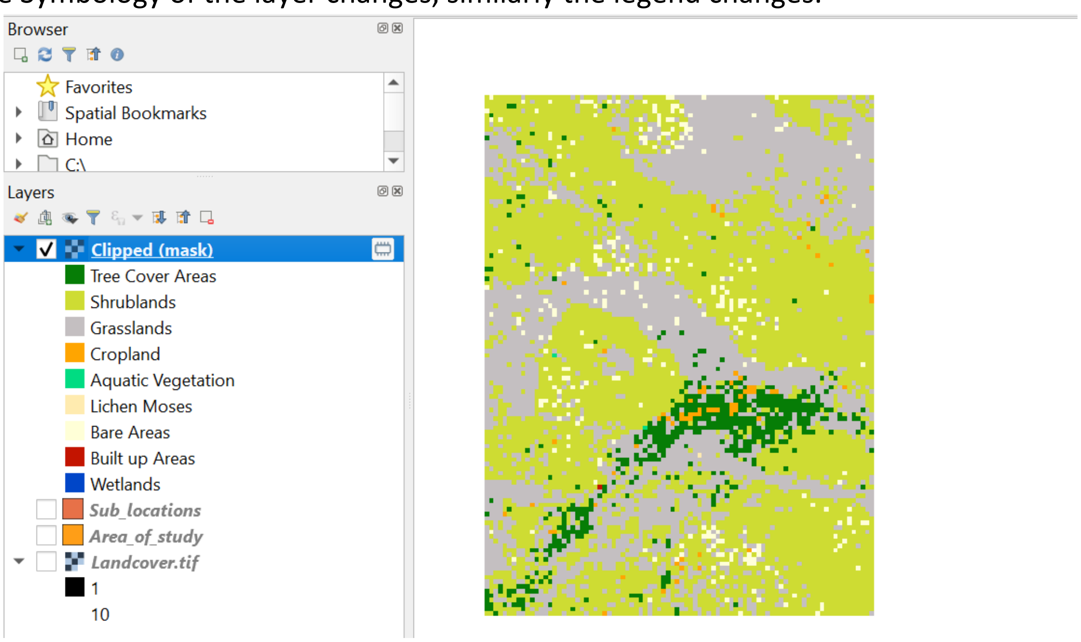
- On the new dialogue window that appears select ***Symbology_layer*** from the raster folder then click ***Open***



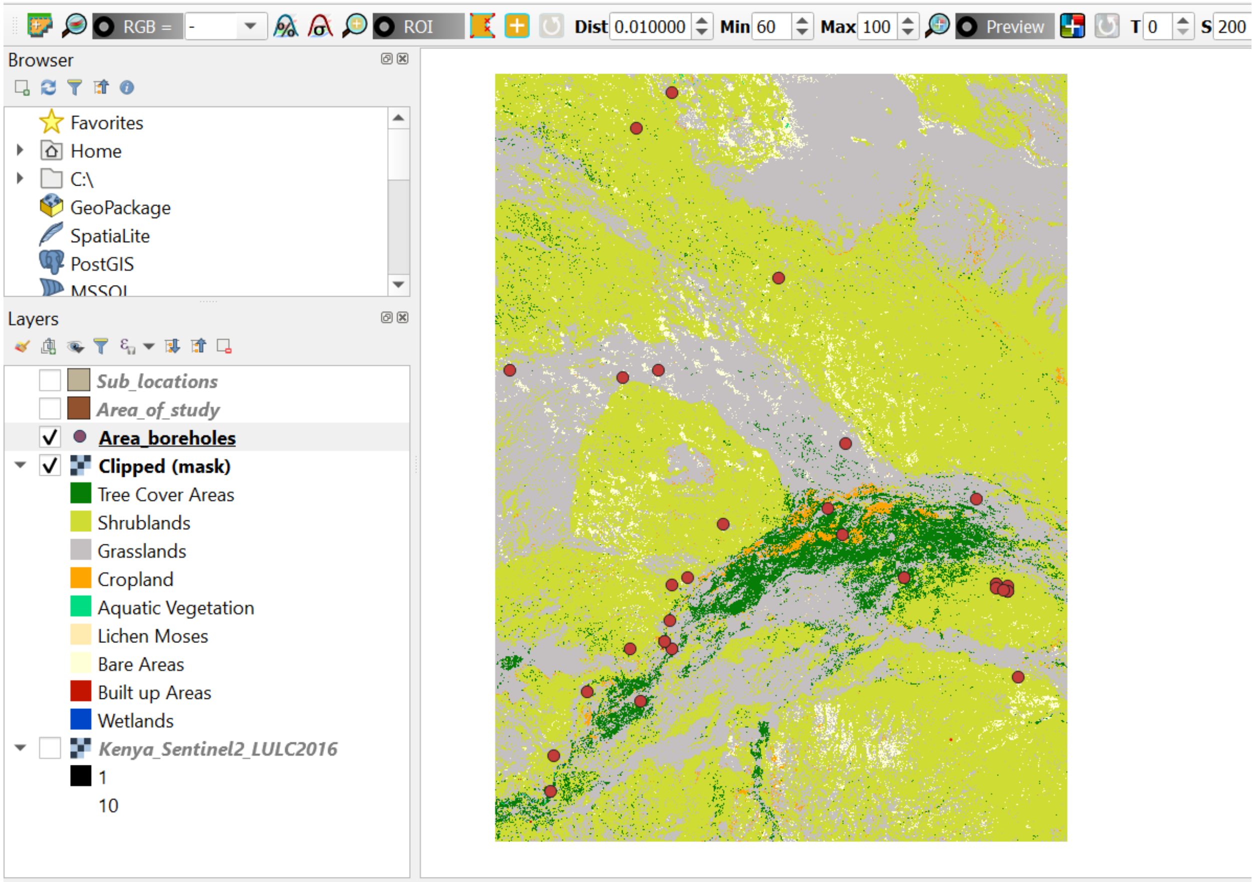
- Click ***Apply*** to render/Apply the symbology to the ***Clipped (Mask)*** Layer.



- Click ***OK***
- The Symbology of the layer changes, similarly the legend changes.



- To explore more, add the boreholes layer to understand their distribution within the land cover classes



We can see that majority of the boreholes are within the tree cover Areas.

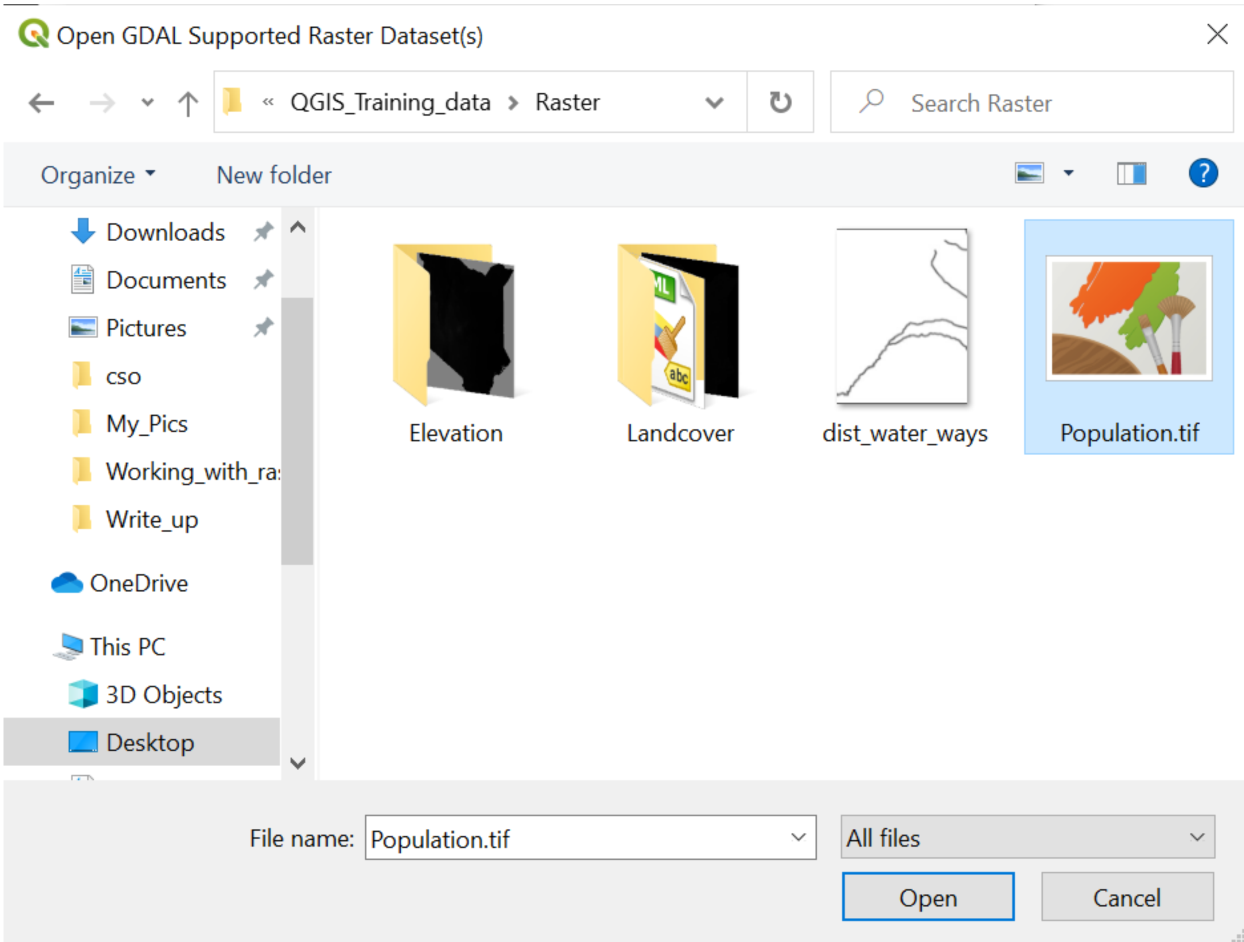
In this section we were able to extract part of a raster using a mask which is a vector, applied different symbology to classify land cover classes, created an overlay of Land cover and boreholes enabling us to identify their distribution in the study area.

Kudos, you have made a land cover map of the study area.

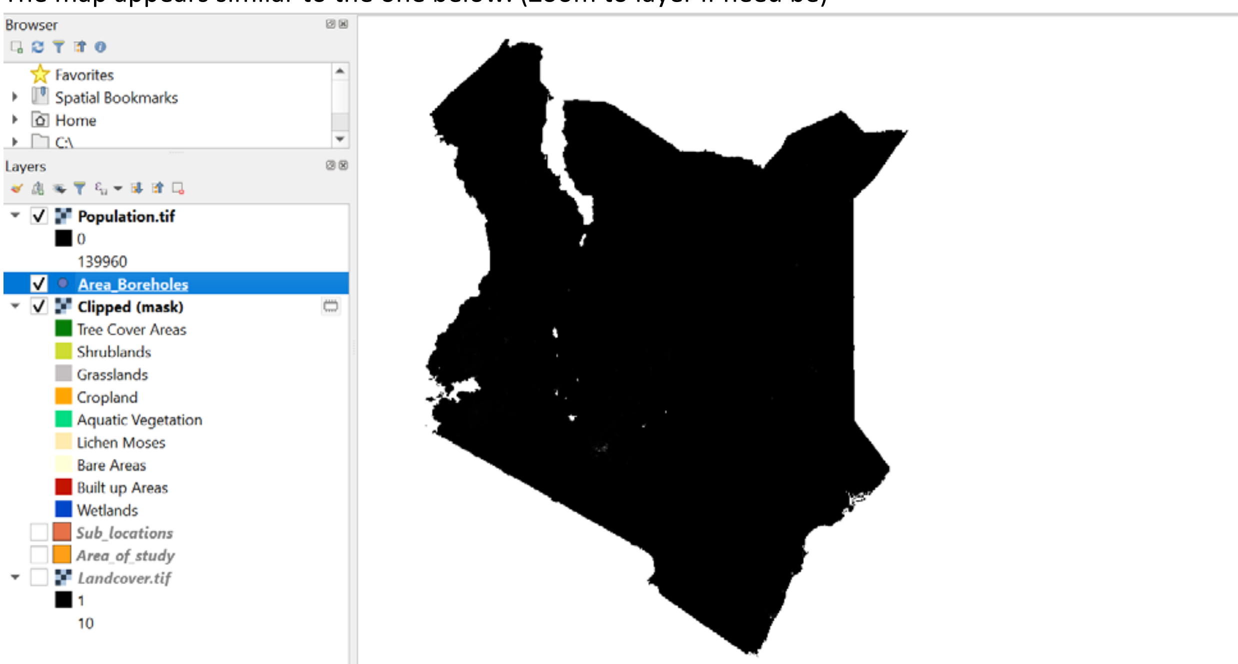
4.10 Population per land cover

In this section we will determine how many people are based within each land cover class. To achieve this we will Use the population Raster and extract the sum value for each Land cover.

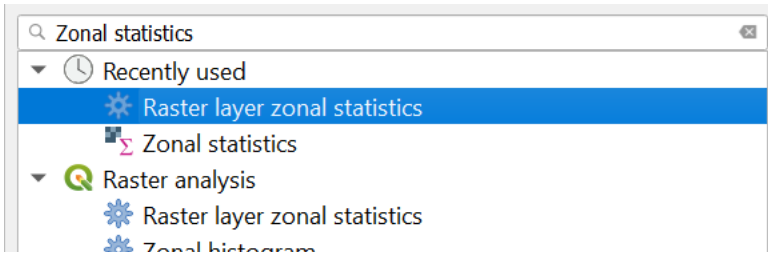
- Add Populations raster to the map, the raster is named ***Population.tif***



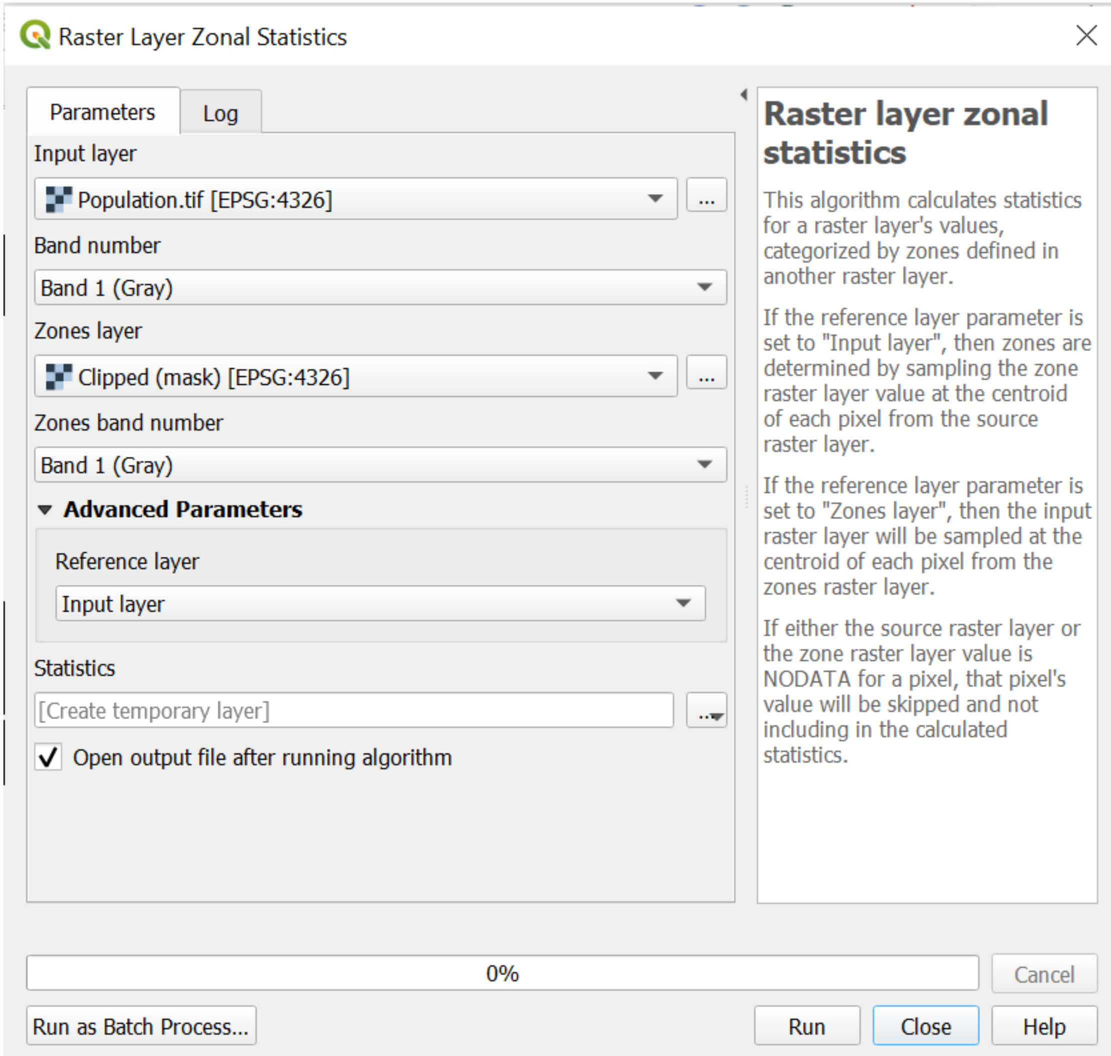
- The map appears similar to the one below. (Zoom to layer if need be)



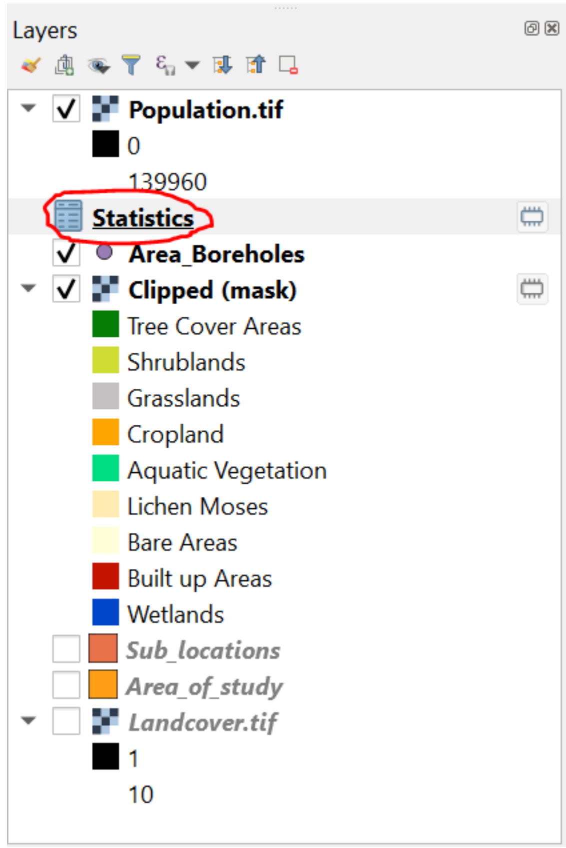
- The next step is to calculate the sum of people per land cover class using Zonal statistics
- Open the processing tool, search and open ***Raster Layer Zonal Statistics.***



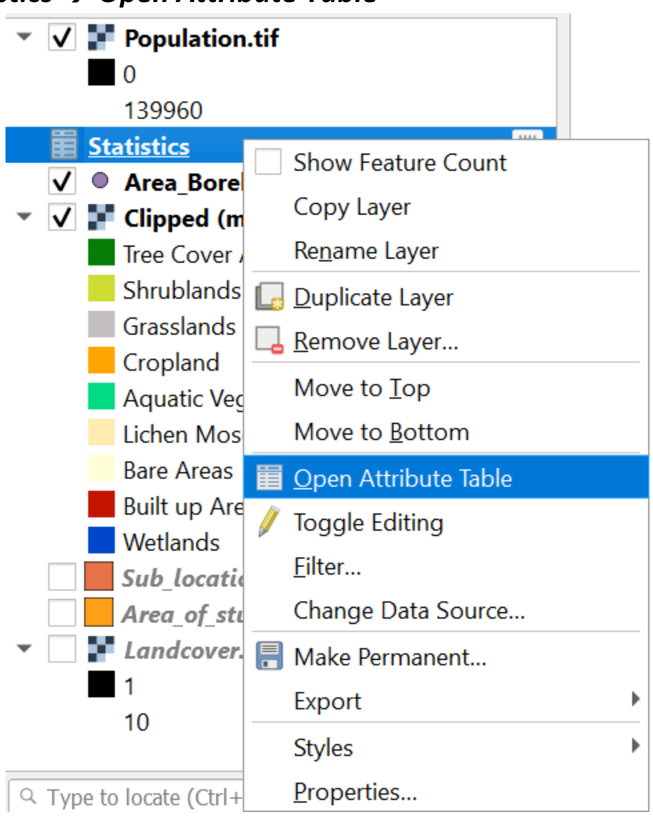
- Click and open the tool, input the parameters as they appear in the following snip.



- Click ***Run***
- The output of this tool is a sheet names ***Statistics*** with counts, sum and more statistics on the populations in the land cover classes.



- Right Click on **Statistics → *Open Attribute Table***



- The attribute table has a column for sum indicating the sum of people per land cover classes. **The name of the land cover class is not included in the table, however the table is arranged in the order of item occurrence in the legend**.

Statistics :: Features Total: 9, Filtered: 9, Selected: 0

	zone	deg^2	sum	count	min	max	mean
1	7.00000000	0.02040873	15914.25029927	595118	0.00063166	6.79236603	0.02674134
2	1.00000000	0.03630112	69155.53024771	1058540	0.00062308	6.40263510	0.06533105
3	8.00000000	0.00006341	7681.65114956	1849	0.08927021	6.82571268	4.15448953
4	3.00000000	0.25705965	239998.627950...	7495854	0.00062308	6.82571268	0.03201752
5	4.00000000	0.00614730	13223.36428316	179255	0.00067762	6.79236603	0.07376845
6	10.00000000	0.00001814	21.47037877	529	0.00130322	0.35725757	0.04058673
7	6.00000000	0.00007150	94.85786161	2085	0.00244024	4.12750292	0.04549538
8	5.00000000	0.00018659	218.09038540	5441	0.00087661	6.17552471	0.04008278
9	2.00000000	0.38454311	361641.952426...	11213269	0.00062708	6.74600649	0.03225125

After viewing close the attribute table and save the project.

We have successfully calculated the sum of people living on each land cover class in the study area. The difference of this statistics from the previous is that we have used raster for input and Zones layer, unlike earlier where we used a vector layer for zones input.

Summary

In this section we have worked with raster datasets. We have learnt how to add raster data to our map canvas, how to change or symbolize raster layers as well as done some basic spatial analysis using raster layers. We have used Zonal statistics tool to extract values such as average elevation of sub location, borehole elevation and also population per land cover class. Additionally, we have learnt how to clip a raster using the extract by mask tool.

Next we are going to style and output a printable map document.

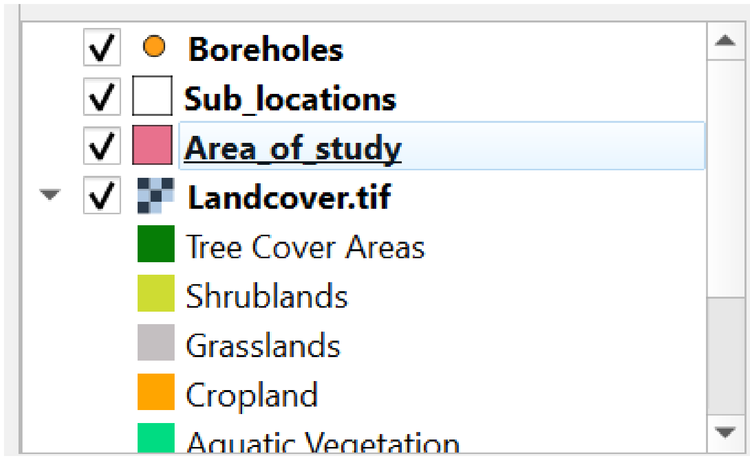
5. Styling and map design

Presentation of the results of any spatial analysis is as important as the analysis itself. Mostly, Outputs from a spatial analysis include Maps followed by charts. In this section we are going to style and design a simple map. The features that need to appear clear in the map include; Boreholes, Land cover and Sub locations.

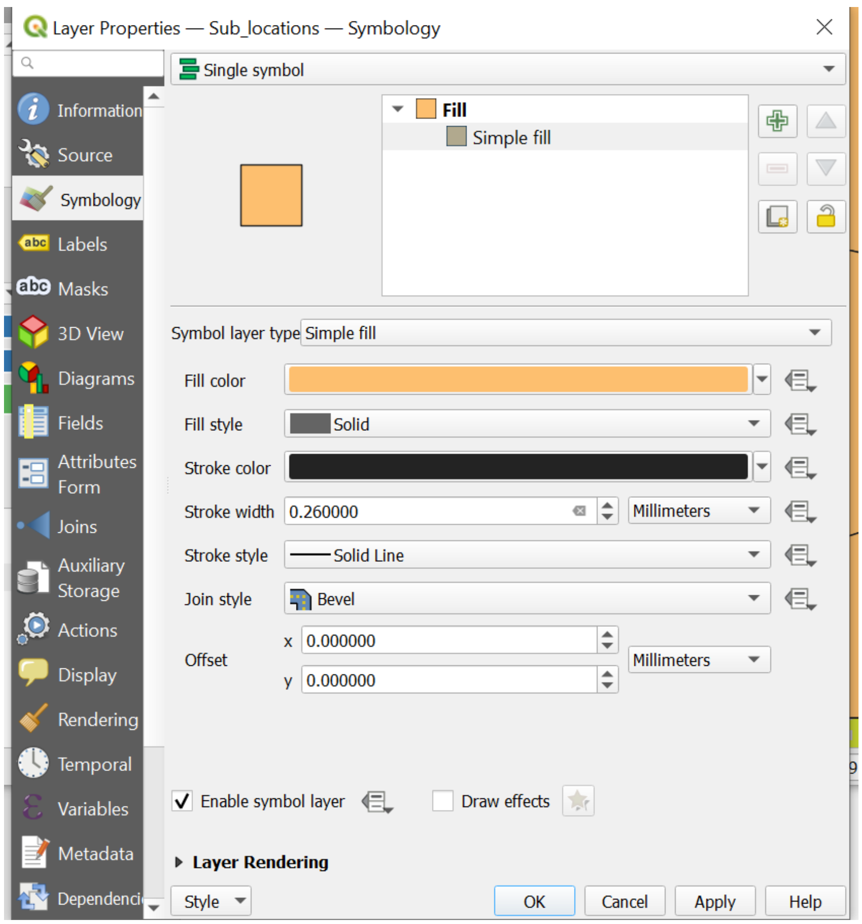
5.1 Preparing data for mapping

Before designing a map of interest, it is important to choose the right symbology and names for your data.

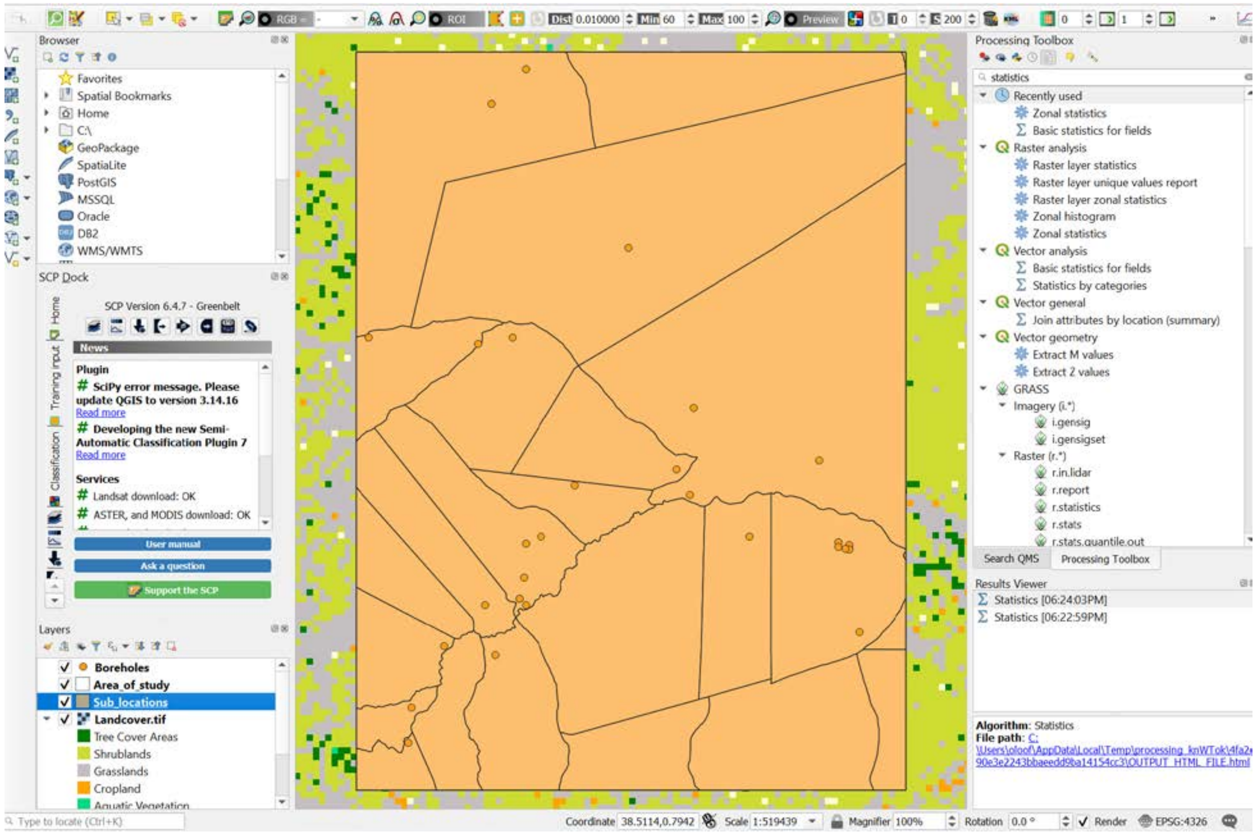
- In the map, turn off or remove all layers except the layers shown in the snippet below



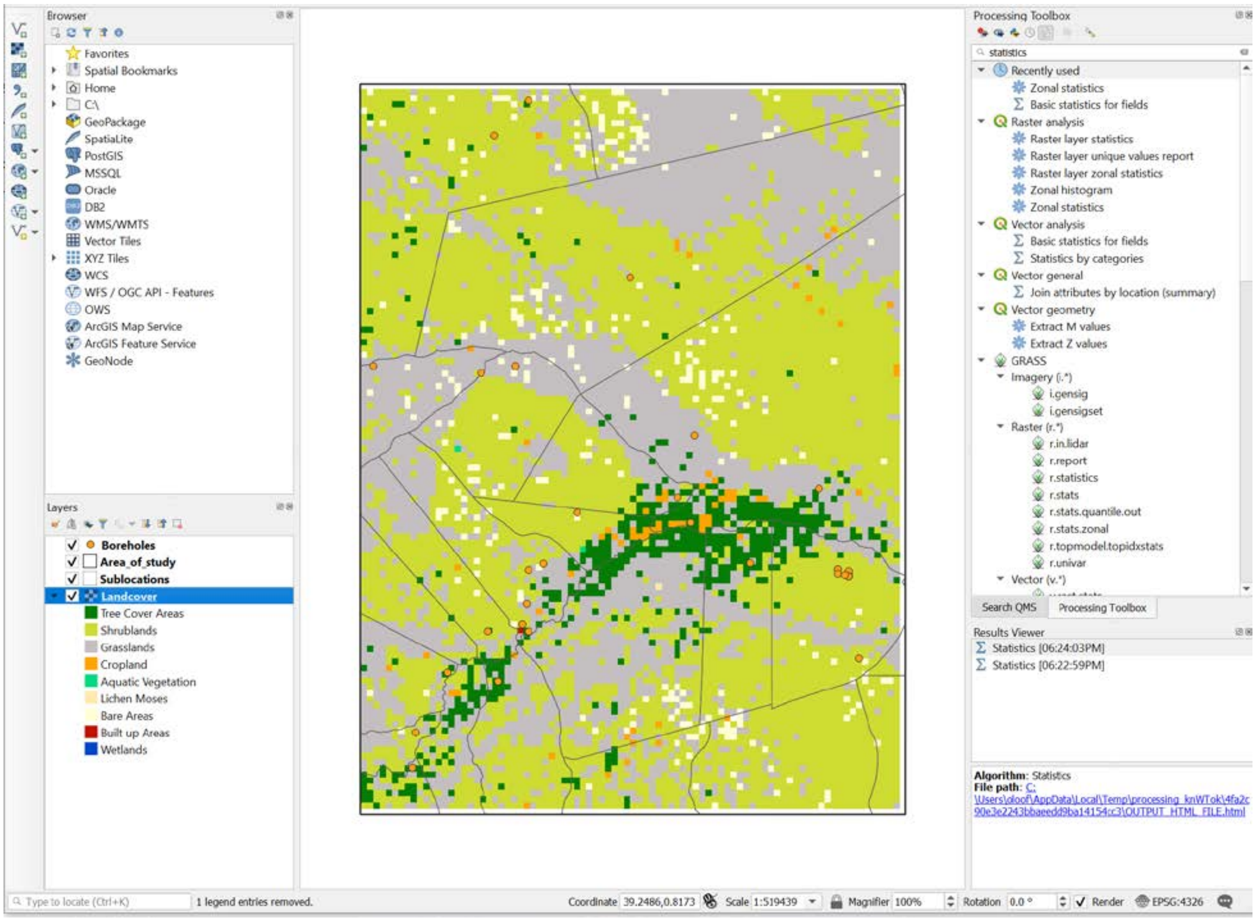
- Right click on ***Sub_locations→ Properties→ Symbology***
- Click on ***Simple Fill→ Fill style*** set to ***No brush.*** Click ***Apply*** then ***Okay***



- This removes the fill color and only boundaries are left



- We will repeat the same procedure for the Area_of_study layer
- Click on ***Area_of_study →Properties→ Symbology***
- Click on ***Simple fill*→Fill style** select ***No Brush***
- Click ***Apply*** then ***OK***
- The canvas should appear similar to the one below
- You can also use Properties → Source to change the name of each data layer, for instance Sub_locations to Sub-location and and Landcover.tif to Landcover

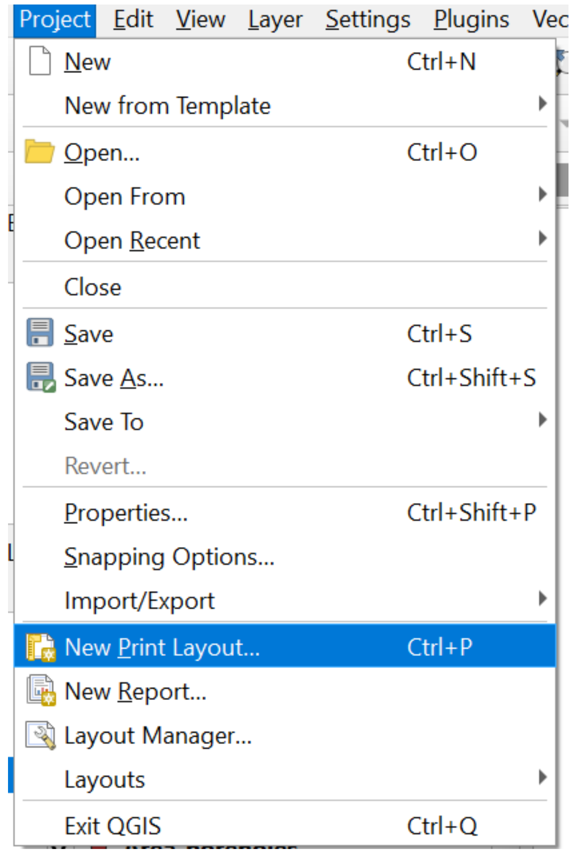


- Label the sub locations layer, **Using the procedure illustrated earlier in the tutorial**.

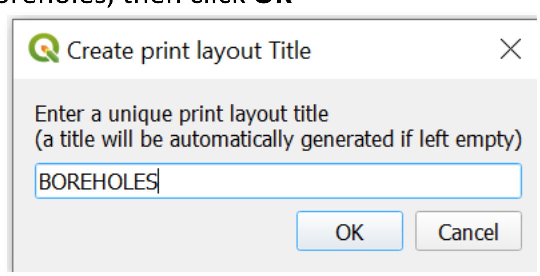
5.2 Prepare a map layout

A map layout is where you design your map by placing all the important map elements. The elements include the map data, legends, north arrow etc.

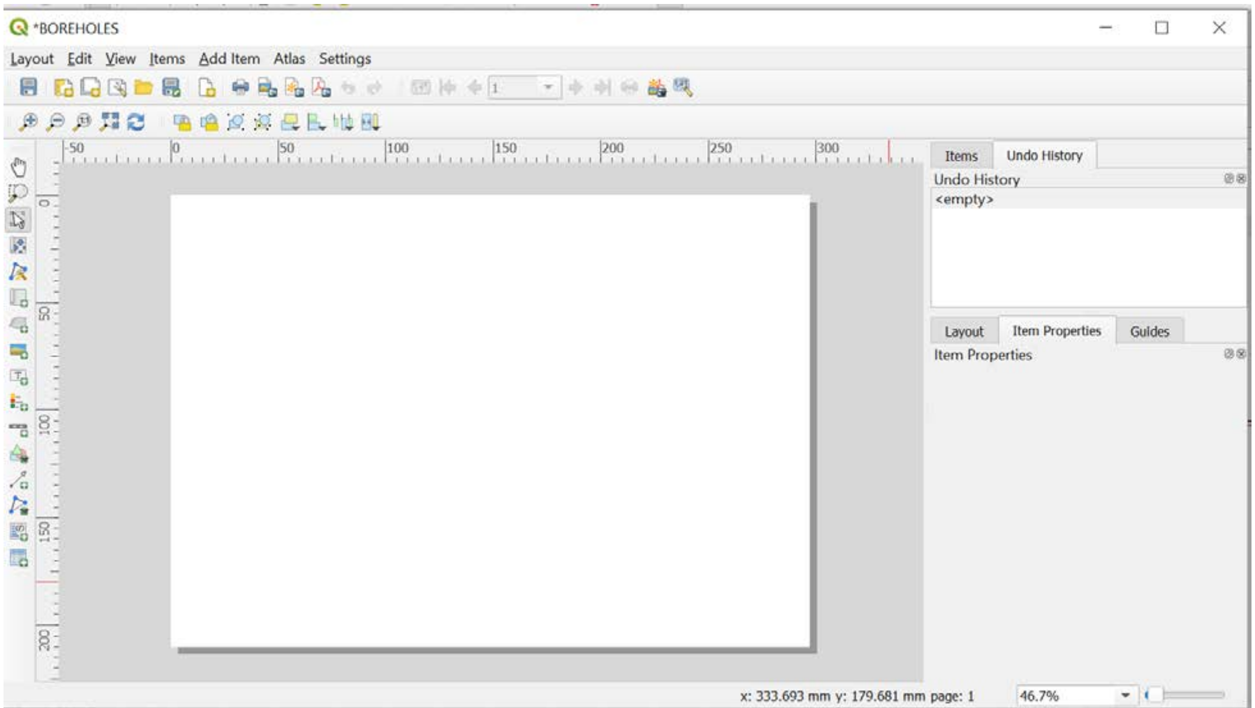
- To create a printable map we need to create a new print layout.
- Click on ***Projects → New Print Layout…***



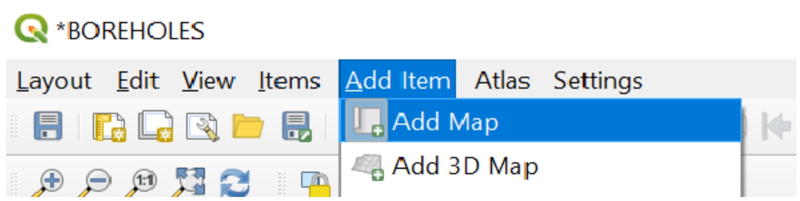
- Choose a name, here we've typed Boreholes, then click **OK**



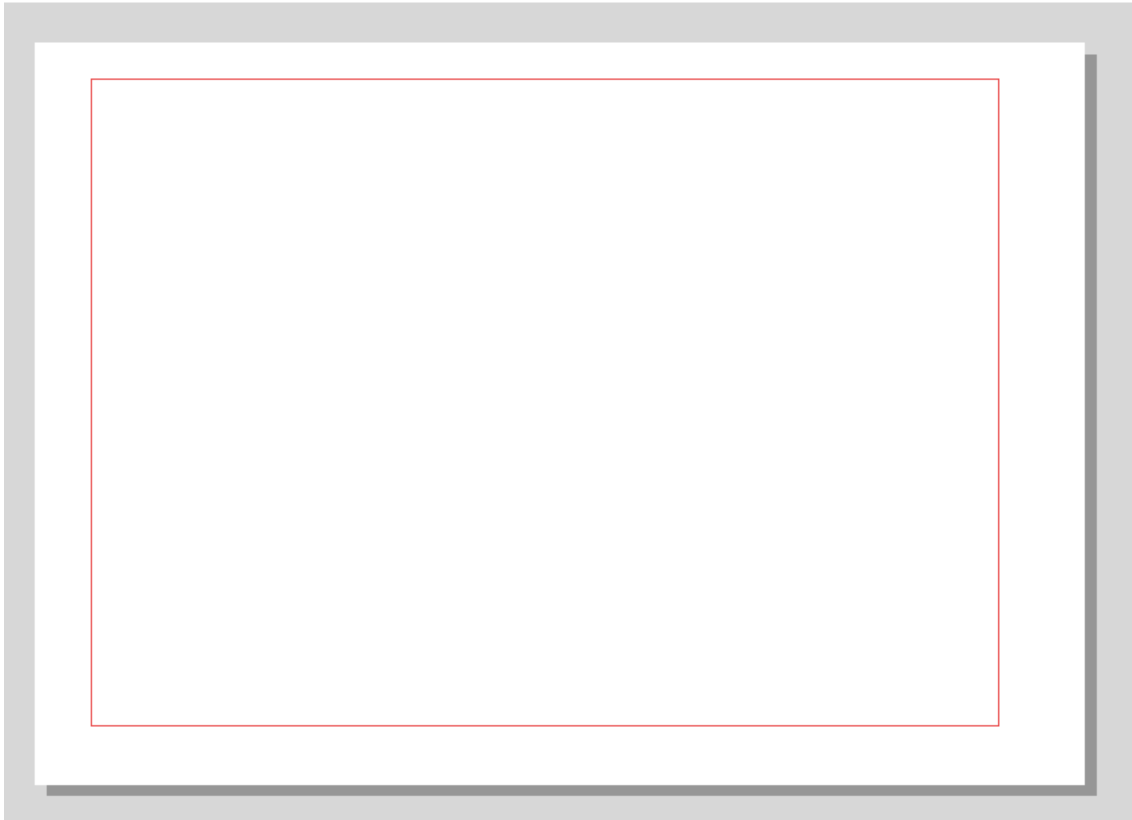
- A new print layout opens.



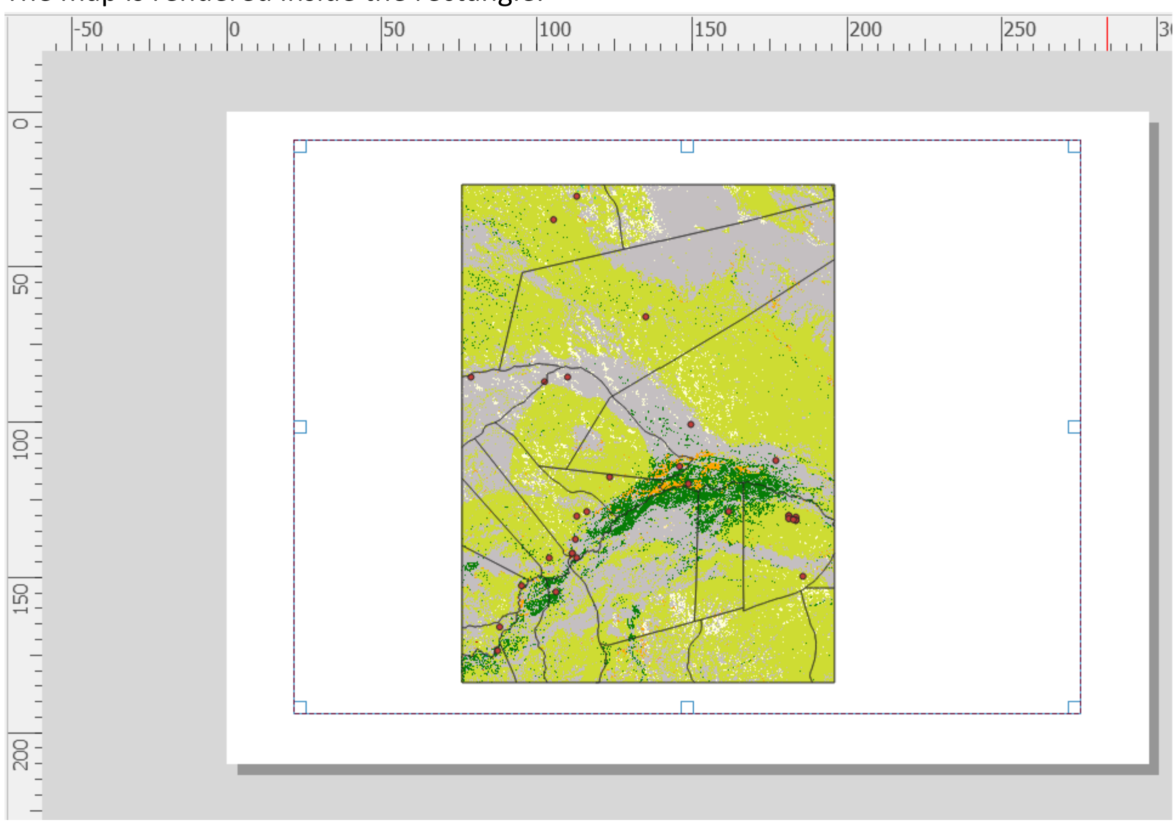
- Click on ***Add Item →Add Map***



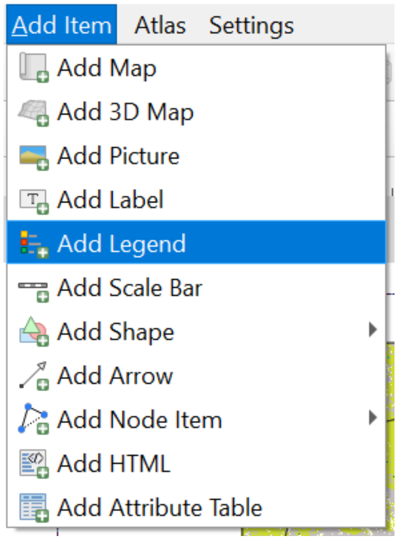
- Draw a rectangle on the canvas of the layout, where the map is rendered.



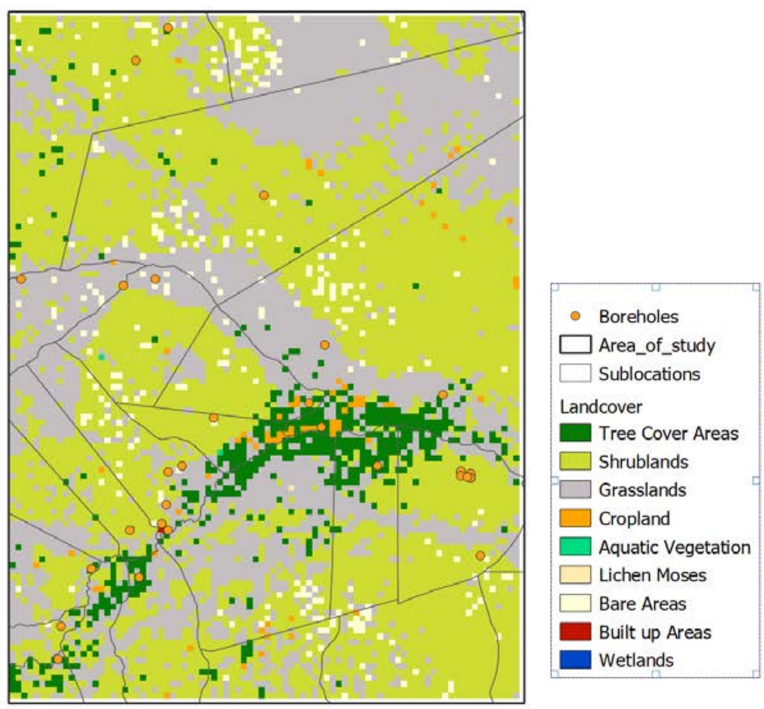
- The map is rendered inside the rectangle.



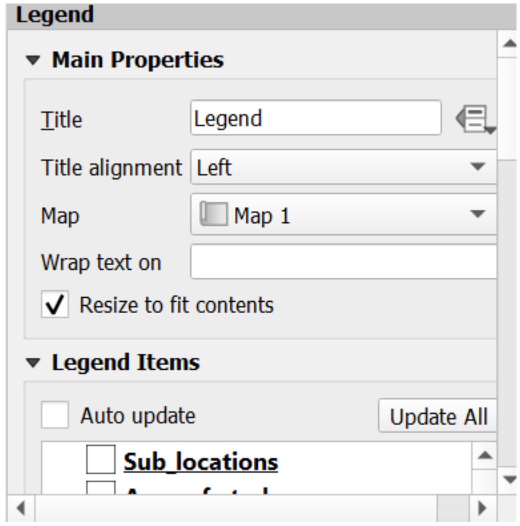
- Next we add the legend, click **Add Item, Add Legend**



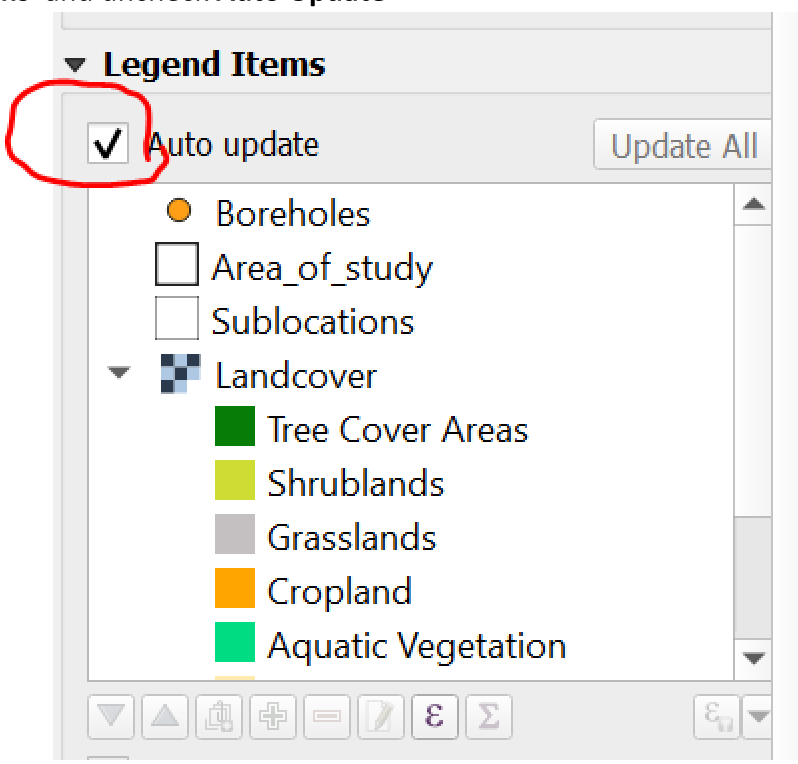
- In the map, draw a box where a legend should be added



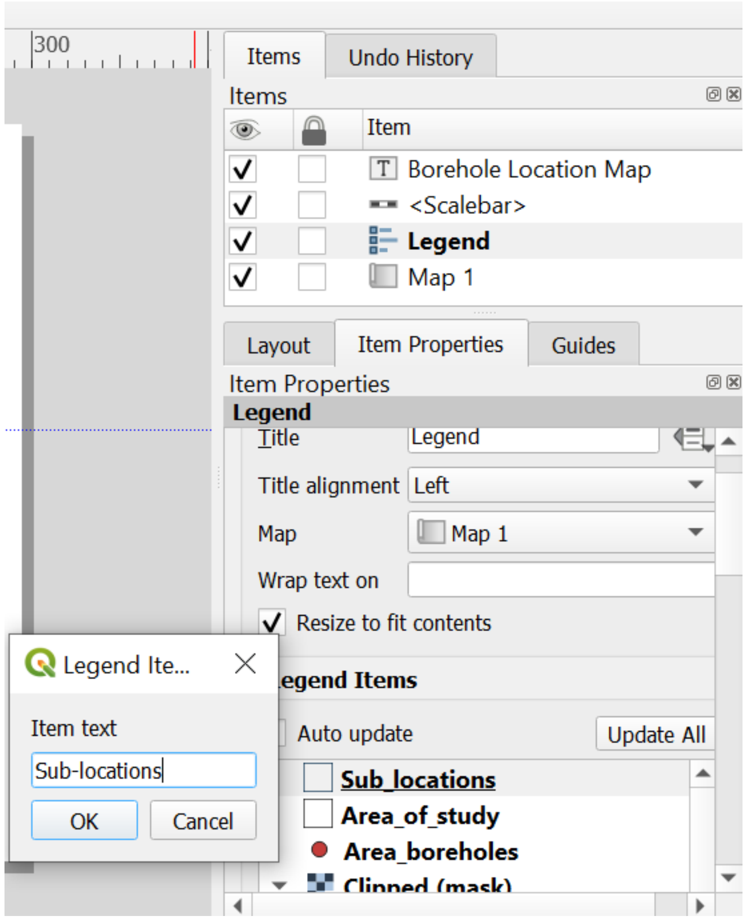
- On the tab before Legend Items Input the title as Legend



- The legend has some unnecessary layer, to remove them navigate to the right **Legend Items** and uncheck **Auto Update**

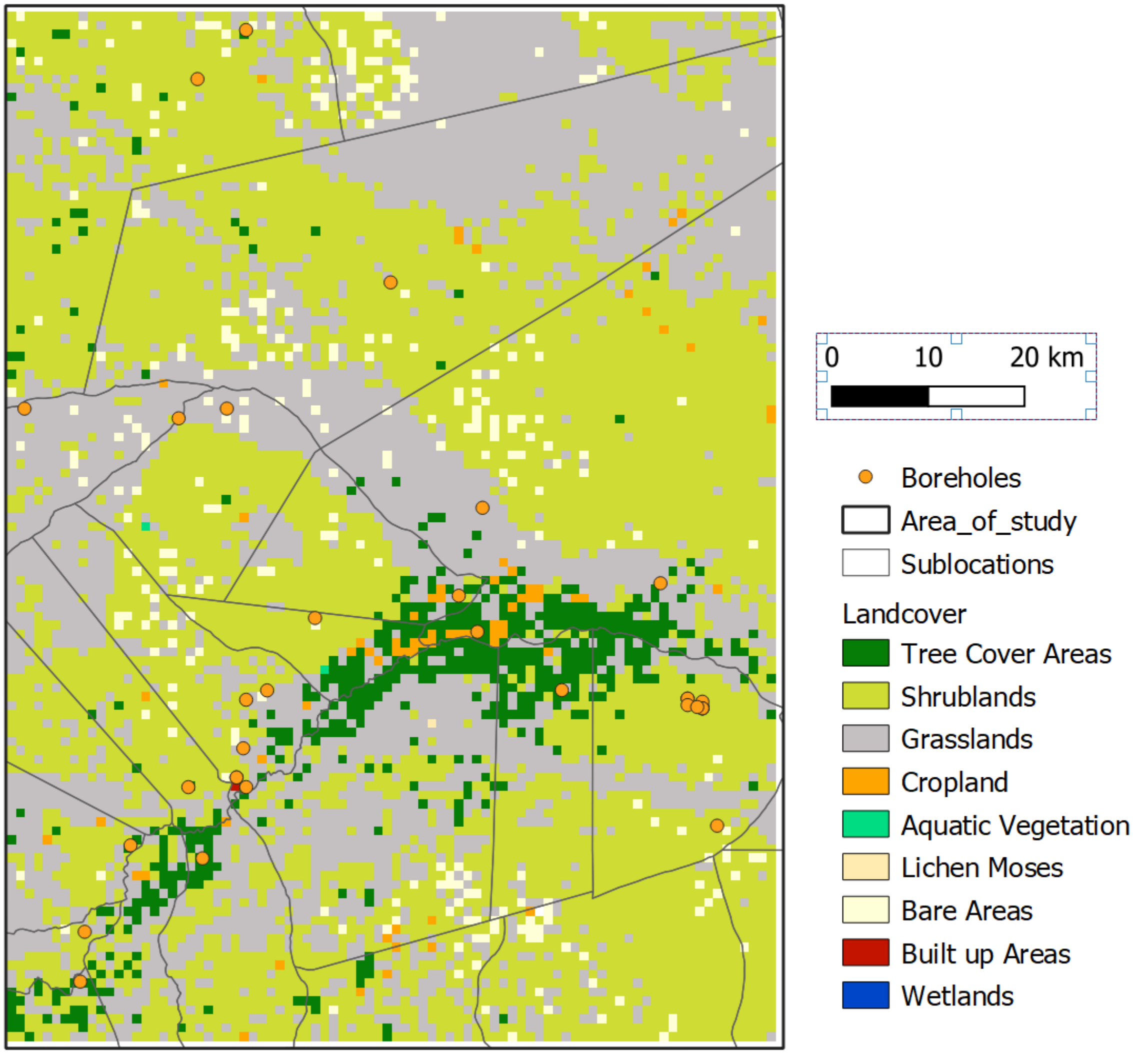


- To edit the names of the layers such as Sub_locations to Sub-locations double click on the layer name, in the dialogue box that appears input the desired name.

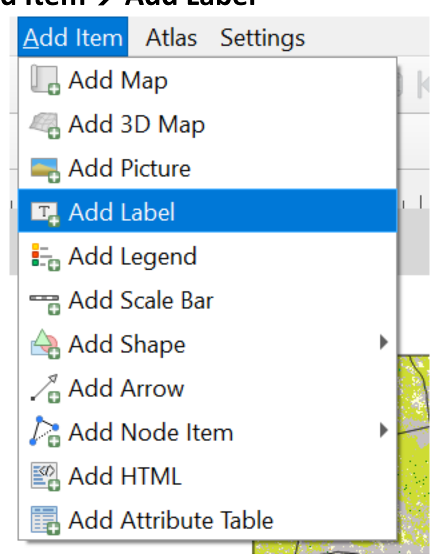


- Click **OK** to save the name. Repeat the procedure for all names which require to be changed

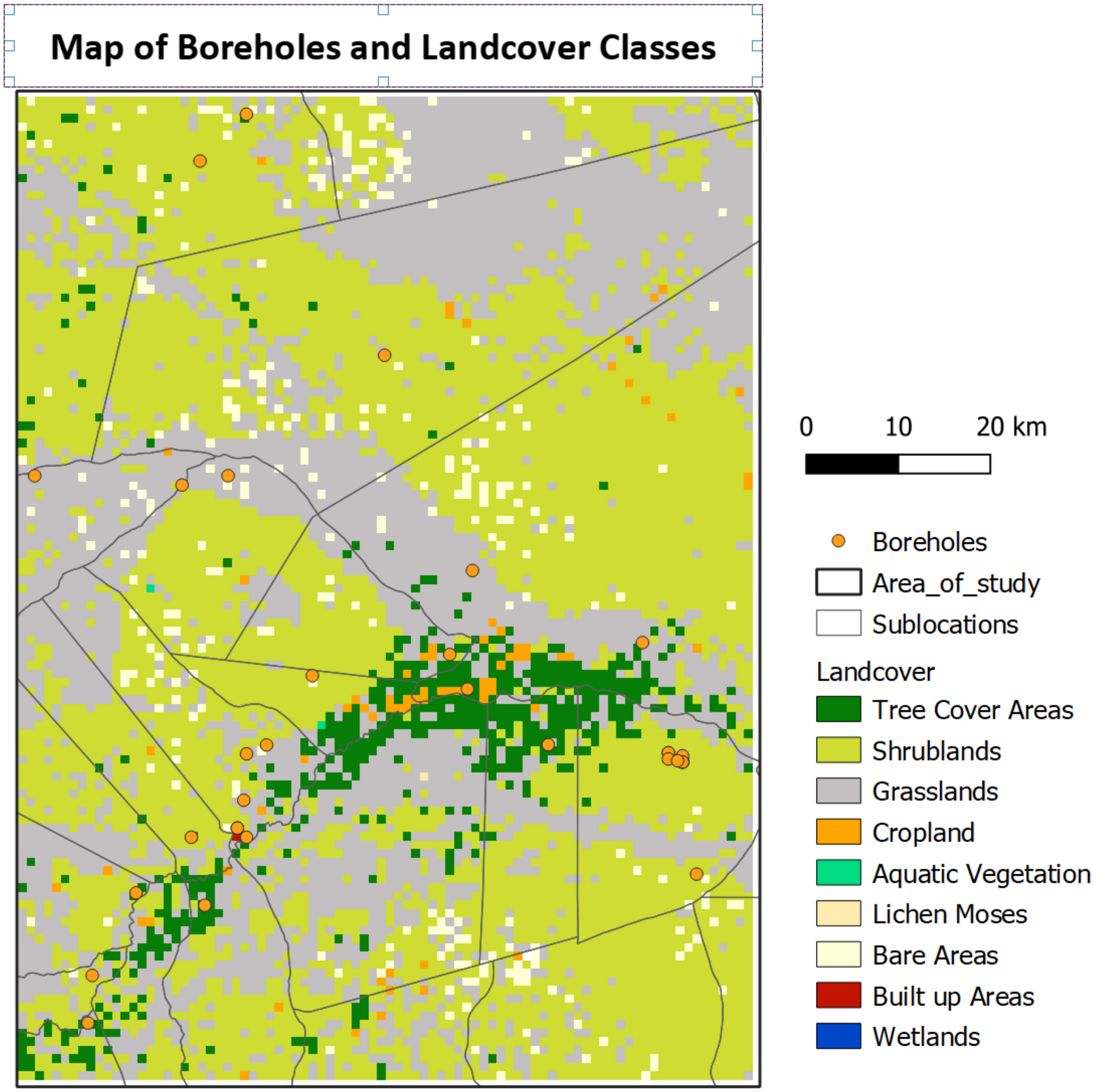
- The next step is to add scale and Title
- Click on Add Item → Scale bar
- On the Canvas, draw a box above the legend where the scale should be rendered



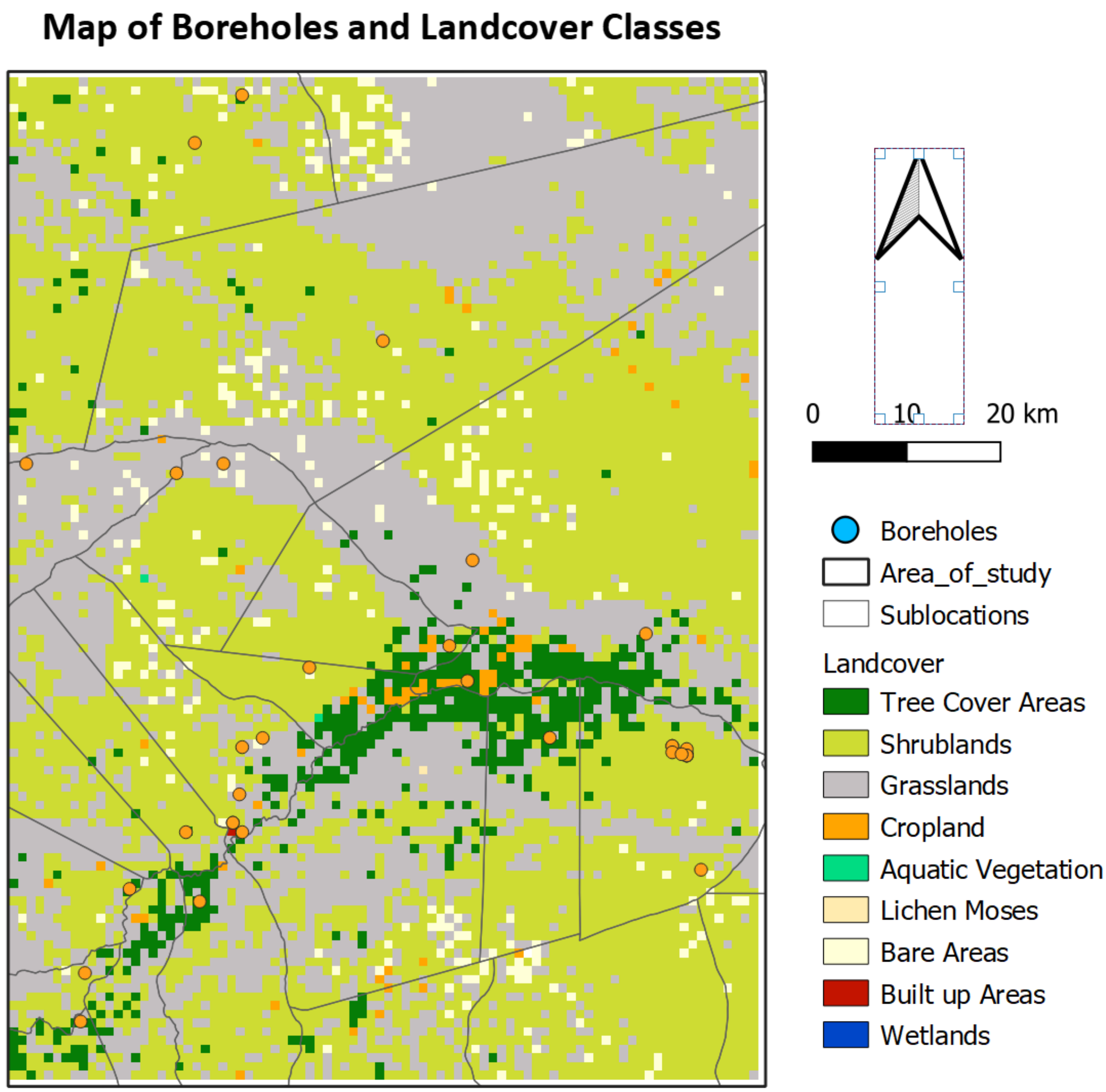
- Now the scale bar has been added to the map
- To add Map Title, Select **Add Item→ Add Label**



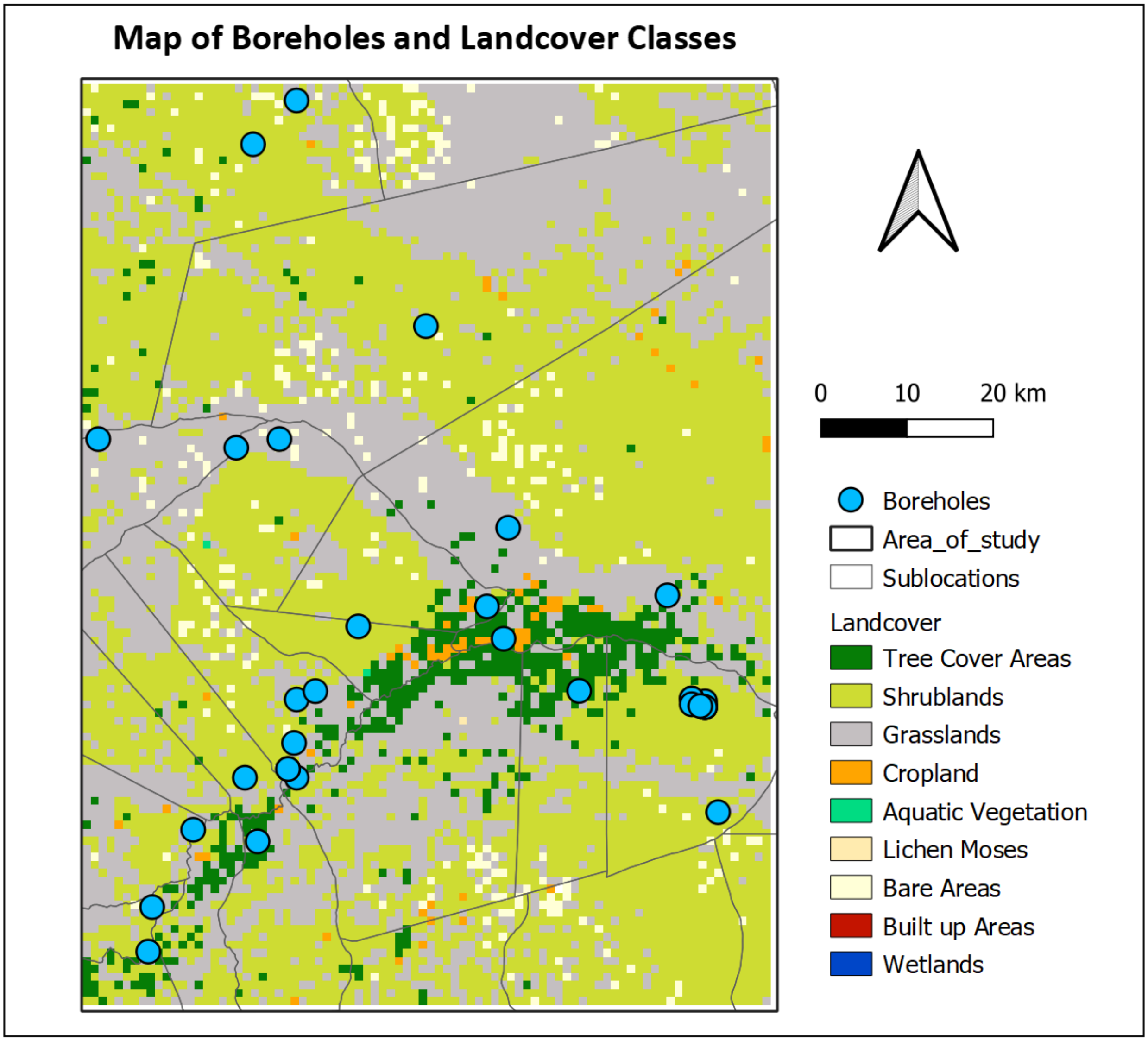
- Draw a box above the Map, on the far left inside the Tab labeled ***Main Properties***, delete Loren Ipsum and Type ***Borehole Location Map, scroll down and in Horizontal alignment check the circle labeled center. For Vertical Alignment check the circle next to Bottom.*** Now the Title is Centered



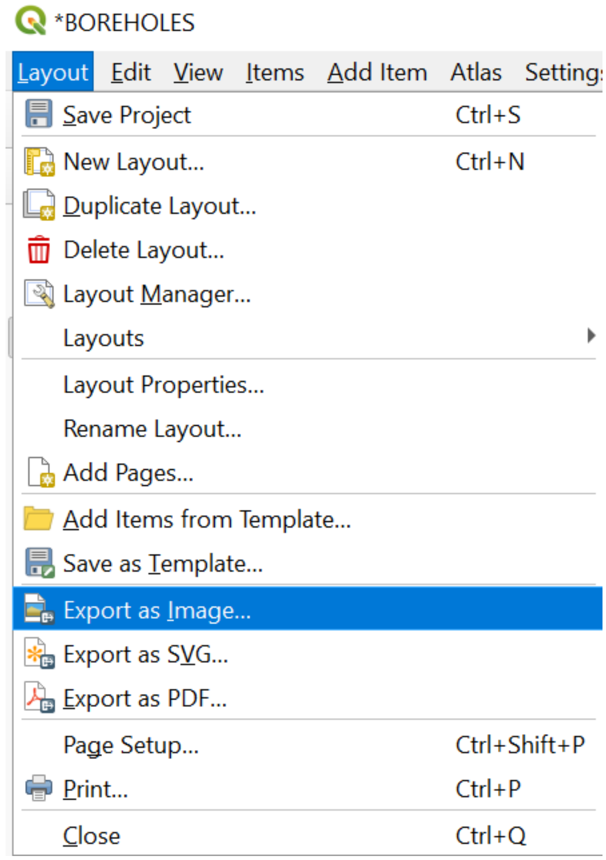
- We can also add a North Arrow using Add Item > North Arrow


Map of Boreholes and Landcover Classes
0
10
20 km
Boreholes
Area_of_study
Sublocations
Landcover
Tree Cover Areas
Shrublands
Grasslands
Cropland
Aquatic Vegetation
Lichen Moses
Bare Areas
Built up Areas
Wetlands

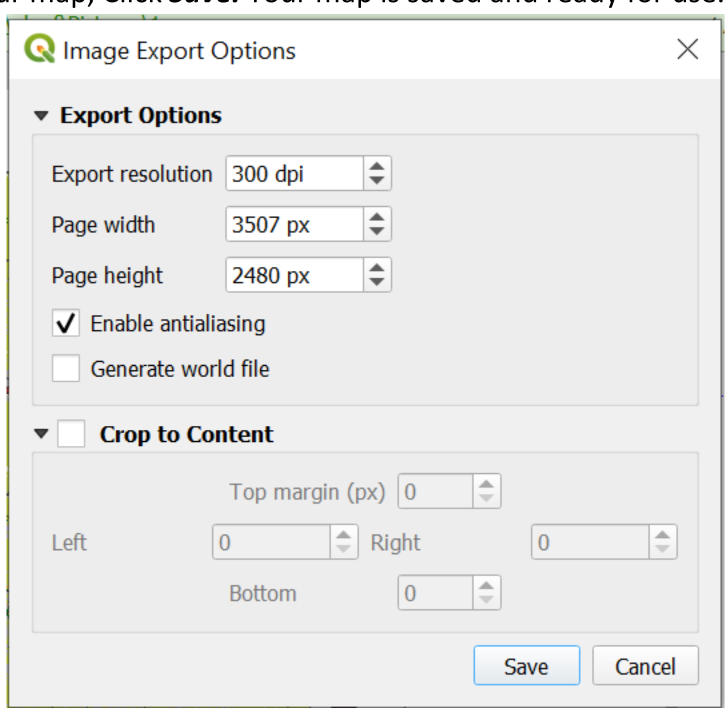
5.3 Export the map



- At this point the map is ready for export.
- Click on **Layout** there are options for export, as Image SVG or PDF.



- Click ***Export as Image...*** In the dialogue box that appears select a folder where you want to save your map, Click ***Save.*** Your map is saved and ready for use.



In the saved map remove the outer white spaces by Clipping.

6. Conclusion

In this exercise, you performed various GIS steps to assess access to water in the northern areas of Isiolo County of Kenya. This exercise was meant to introduce the participant to vector and raster data models which are the common data structures in GIS. In this task, you used both Vector (boundary of the study areas, Sub locations, and borehole) and Raster (Elevation, Population, Distance to Water and Land cover) data. We managed to add both vector and raster data and use them for analysis in QGIS. You also managed to execute different GIS operations on the vector and raster data. Specific operations included clipping, querying attribute tables, spatial join, buffering and point in polygon analysis. For raster data, you were able to add, extracting layer by mask, apply different layer symbology and implement zonal statistics. Finally, you were able to style and design a map from the data.

In the hypothetical example of access to water resources, you were able to determine the average distance to water ways from each borehole. You were also able to estimate the average population and elevation in each sub-location. Here, we used QGIS which is an open source and free GIS software, the standard procedures in this system are to a large extent similar to those in other tools. It is our hope that this basic level module allowed the participant to learn basic GIS concepts and their application in a typical GIS tool.