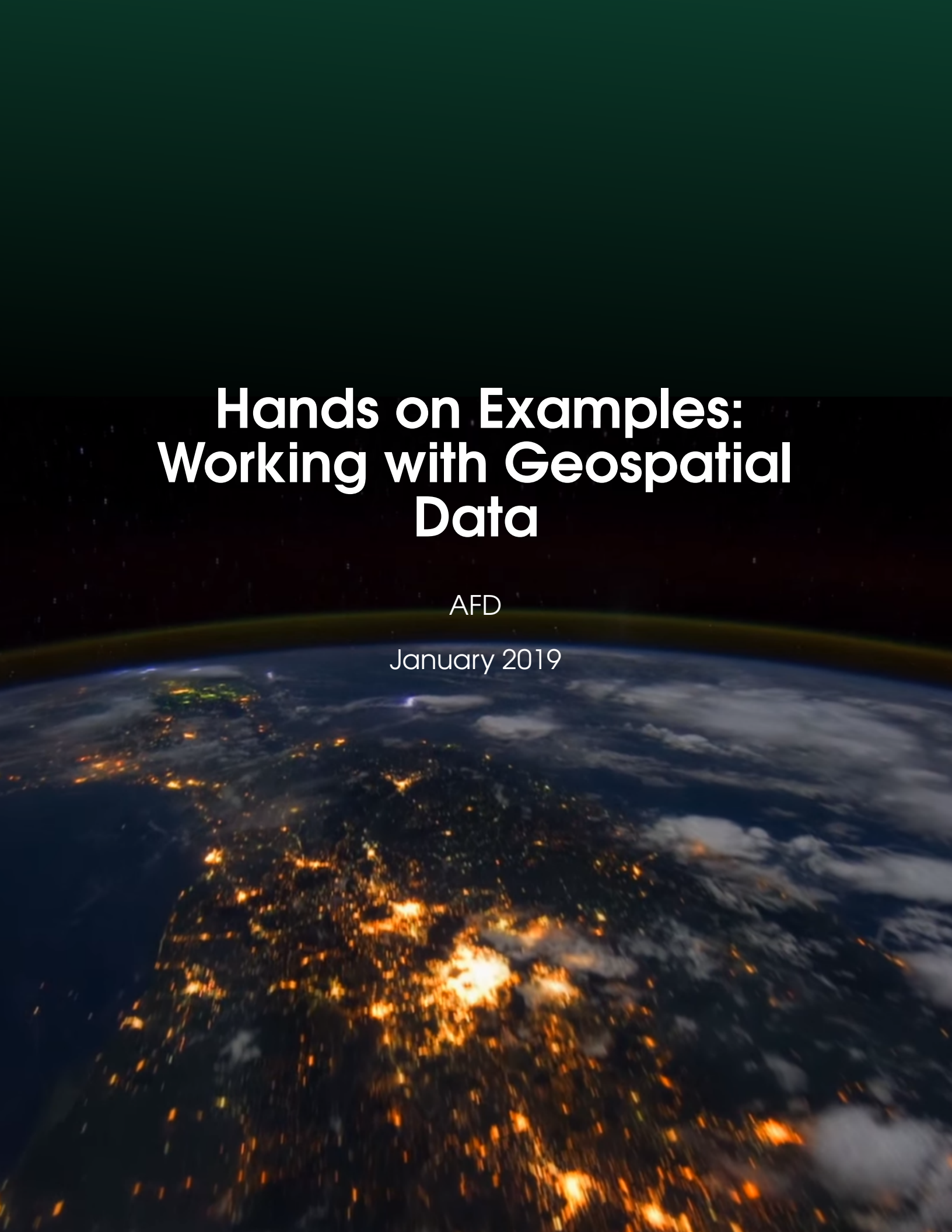
Hands on Examples: Working with Geospatial Data

AFD

January 2019



Hands on Examples: Working with Geospatial Data
GEOQUERY.ORG | GEO.AIDDATA.ORG

THIS WORK IS AUTHORED BY DAN RUNFOLA, SETH GOODMAN, AND ZHONGHUI LV, AS WELL AS MANY COLLABORATORS WITHIN THE AIDDATA RESEARCH LAB AT THE COLLEGE OF WILLIAM AND MARY. GEOQUERY IS MADE POSSIBLE BY THE SUPPORT OF USAID, KFW, HUMANITY UNITED, THE WORLD BANK, THE GLOBAL ENVIRONMENT FACILITY, THE MACARTHUR FOUNDATION, THE CLOUDERA FOUNDATION, AND THE COLLEGE OF WILLIAM AND MARY. THIS WORK WAS PERFORMED IN PART USING COMPUTATIONAL FACILITIES AT THE COLLEGE OF WILLIAM AND MARY WHICH WERE PROVIDED WITH THE ASSISTANCE OF THE NATIONAL SCIENCE FOUNDATION, THE VIRGINIA PORT AUTHORITY, VIRGINIA'S COMMONWEALTH TECHNOLOGY RESEARCH FUND AND THE OFFICE OF NAVAL RESEARCH

AidData is a research lab at the College William & Mary that works with governments and international organizations to improve the ways in which sustainable development investments are targeted, monitored, and evaluated. Its team of 35 program evaluators, policy analysts, and media and communication professionals have specialized expertise in GIS, remote sensing, machine learning, and causal inference. AidData's work is made possible through funding from and partnerships with USAID, the World Bank, the Asian Development Bank, the African Development Bank, the Islamic Development Bank, the Open Aid Partnership, DFATD, the Hewlett Foundation, the Gates Foundation, Humanity United, and 20+ finance and planning ministries in Asia, Africa, and Latin America.

As a research institution, we ask that users of GeoQuery cite:

Goodman, S., BenYishay, A., Lv, Z., & Runfola, D. (2019). GeoQuery: Integrating HPC systems and public web-based geospatial data tools. Computers & Geosciences, 122, 103-112.

Contents

1. Intro to Spatial Data and GIS

1.1 Setup and Introduction

Spatial data can be incredibly powerful and useful for evaluations, but requires understanding key spatial concepts as well as the tools and methods needed to work with spatial data. In this training we will explore how to use QuantumGIS (QGIS), a free and open source GIS platform, to visualize and perform basic analysis of several types of data (geocoded aid data, administrative boundaries, and satellite imagery). After introducing the basics, we will show how AidData's GeoQuery tool can be utilized to make working with spatial data drastically easier.

1.1.1 Training Environment

To get started, we will access Paperspace. Paperspace is a virtual desktop that has been pre-loaded with all the software and data you will need for this training.

To log in to Paperspace:

1. Go to `https://www.paperspace.com`
2. Click **SIGN IN** in the top right
3. For your email, enter **geo@aiddata.org**
4. Enter the password provided by the instructors
5. Click the **SIGN IN** button

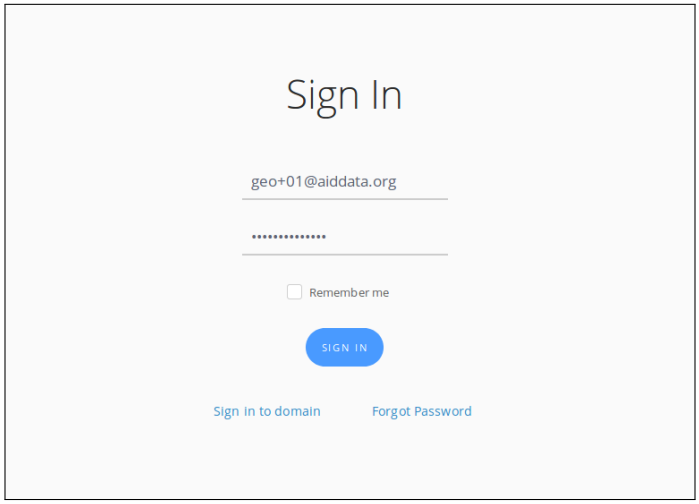


Figure 1.1: Signing in to Paperspace

To access your virtual desktop:

1. Once you have signed in, you should see a set of "machines" available for your Paperspace account
2. Select the machine with the number assigned to you (e.g., User 01 would be **Training 01**.
3. (Note: If you were assigned a single digit number, make sure to add a zero in front. For example, user **1** would use **Training 01**)

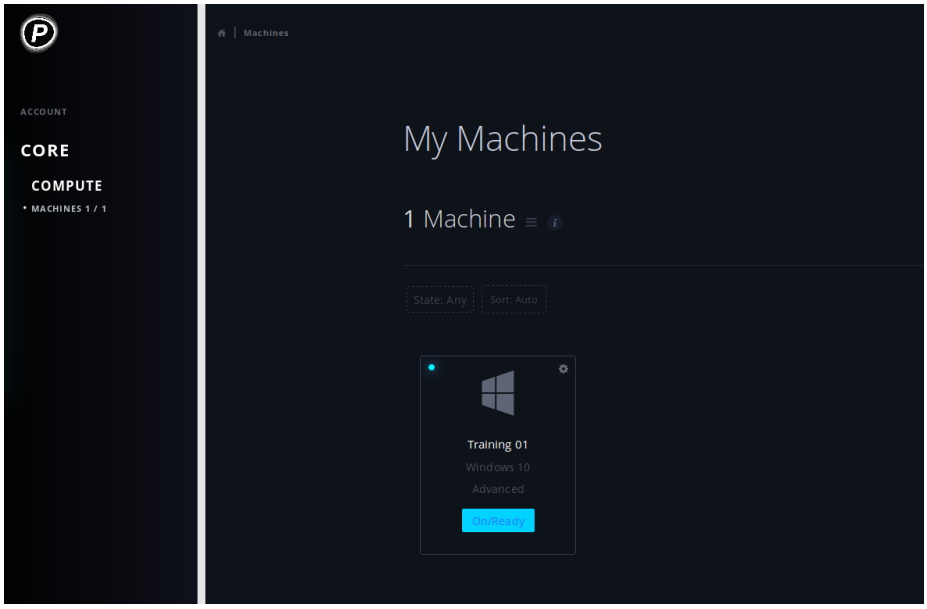


Figure 1.2: Virtual desktop machine selection

Inside your virtual desktop:

- Your virtual desktop may take a moment to connect and load
- Once ready, you should see a standard Windows 10 environment within your browser window (as seen below)
- On your desktop, you should see four main items:
    1. **Google Chrome**, with useful bookmarks preset
    2. **QGIS** Desktop 3.4
    3. A **data** folder which contains all the material used in this training
    4. **Training PDF** that has step by step instructions for everything we will cover during the training

Figure 1.3: Paperspace virtual desktop

Next we will open QGIS and explore its interface.

1.1.2 QGIS

To open QGIS, double click on the QGIS icon in your Paperspace desktop.

Once QGIS is launched you should see a window similar to the one below.

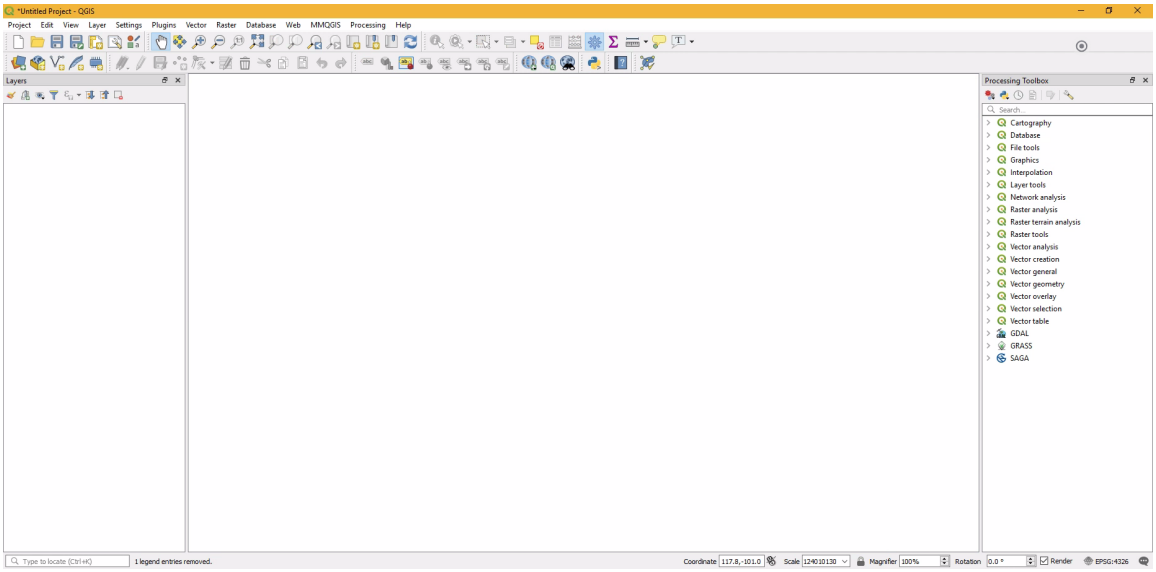


Figure 1.4: Screenshot of QGIS interface

There are several main elements to the QGIS interface:

1. The **toolbars** along the top include the most often used tools for loading and working with spatial data. If the tools are not visible in a toolbar, you can find them from the menus.
2. The **layer list** on the left side will populate with your spatial data layers as you add them.
3. The **processing toolbox** on the right shows the list of all available algorithms grouped in different blocks.
4. The **status bar** along the bottom shows you information about the current map view.
5. The **map canvas** in the center is where your spatial data will be visualized.

There are a wide variety of geospatial tools, functions, and analysis procedures available through the core QGIS library, however, some GIS processes have yet to be integrated into QGIS. Fortunately, there is an increasingly large population of developers and users who are contributing to QGIS's functionality by developing plugins (searchable using the *Plugins* menu). All plugins used in this tutorial have been pre-installed and activated for you.

More on QGIS can be found at `https://qgis.org/`.

1.2 Loading and Exploring Data

In this section we will explore several datasets which represent the types of data you might want to include in a Geospatial Impact Evaluation (GIE), and basic processing steps needed to prepare data for a GIE. Nigeria's State boundaries will represent the unit of analysis for the GIE, the commitment dollar values of Global Environment Facility (GEF) projects can be used as a treatment indicator, and NDVI (a satellite based measure of the "greenness" of vegetation) could be used as an outcome measure.

These datasets are all available in the **training_data** folder included on your Desktop, and are summarized in the list below.

1. **Nigeria State Boundaries**
   - Description - Spatial features defining the boundaries of first order administrative zones (states) in Nigeria
   - Source - GeoBoundaries database of administrative zones[1]
   - File Type - Vector (GeoJSON, .geojson)
2. **GEF Geocoded Aid Projects**
   - Description - Geocoded data for 724 GEF projects spanning 4978 locations. Includes projects identified as land degradation, biodiversity, multifocal area, and programmatic. The file *level1a.csv* contains project, location, and financial data merged together using unique project and location ids. Project level commitments have been disaggregated evenly across all locations associated with a project. For more detail about this dataset, please refer to the readme PDF file provided with the data.
   - Source - AidData
   - File Type - Comma Separated Values (.csv)
3. **NDVI (Normalized Difference Vegetation Index) Satellite Imagery**
   - Description - NDVI is a satellite based measure of the "greenness" of vegetation. This is a yearly aggregate of NDVI for 2010.
   - Source - NASA LTDR AVHRR v4[2]
   - File Type - Raster (GeoTiff, .tif)

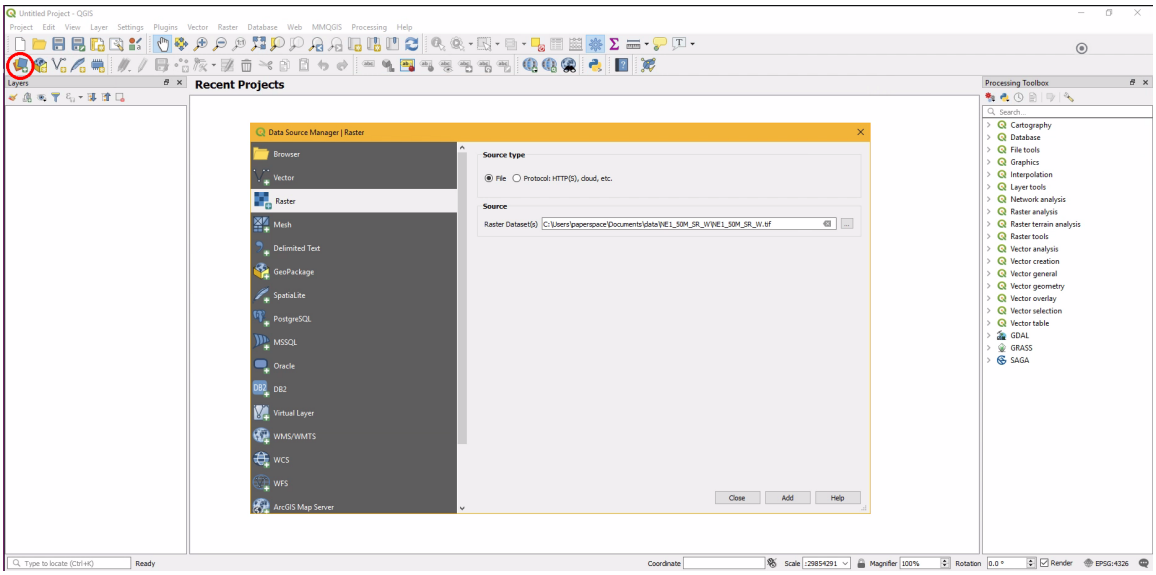
1.2.1 Base Layer

To get started, let's add a background layer to QGIS so we have some spatial context when we add our other data.

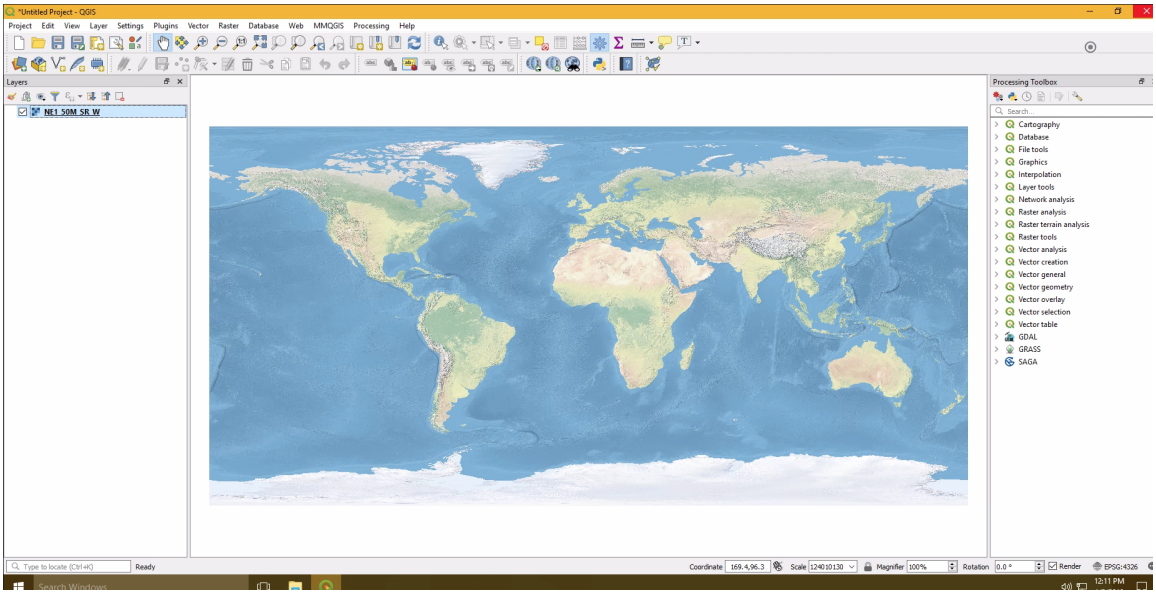
1. Click "Open Data Source Manager" button on the far left of the second row of the top toolbar (shortcut: Ctrl+L).
2. (Note: If you are ever unsure about which button is which, just hover over the button with your cursor to see a tooltip.)
3. Go to the "Raster" tab and then use the browse button (...) for the "Source" field to navigate to and select the **.../Desktop/training_data/NE1_50M_SR_W/NE1_50M_SR_W.tif** file
4. (Note: some files, such as TIF, may not show the ".tif" extension in your file browser. Simply look for the matching base file name or file the "TIF File" file type)

[1] http://www.geoboundaries.org

[2] https://ltdr.modaps.eosdis.nasa.gov/cgi-bin/ltdr/ltdrPage.cgi



5. Click "Add" to load the base layer, and then click "Close". You should see a map similar to below.



1.2.2 Saving Your Work

Before we get too far, let's make sure we save our work.

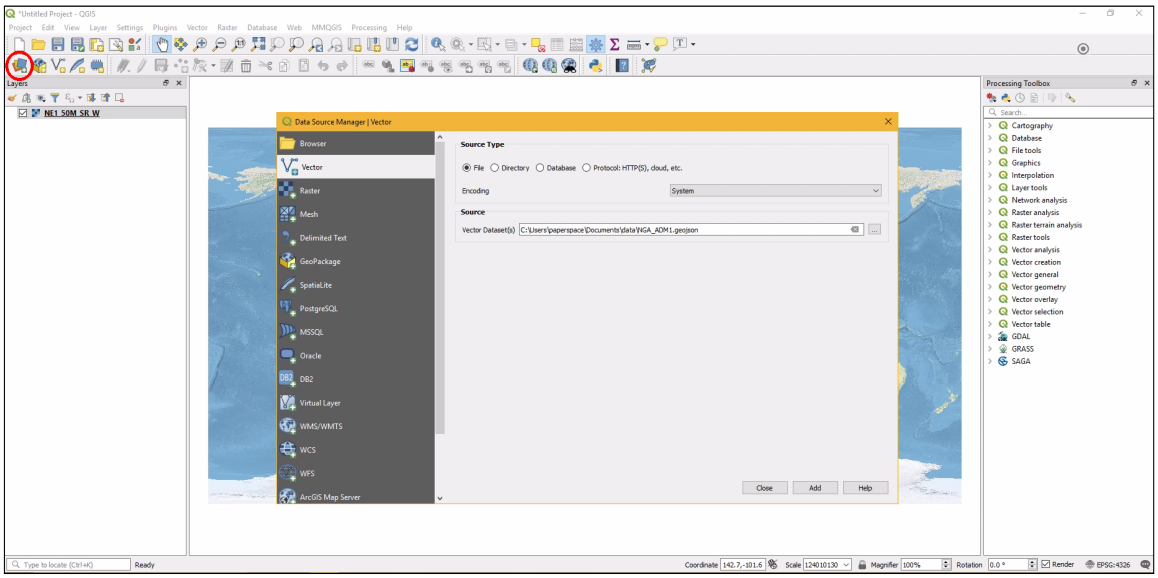
1. Go to the "Project" menu in the top left, then click "Save as"
2. Save your project to the "workspace" folder on your Desktop as a "qgz" file.

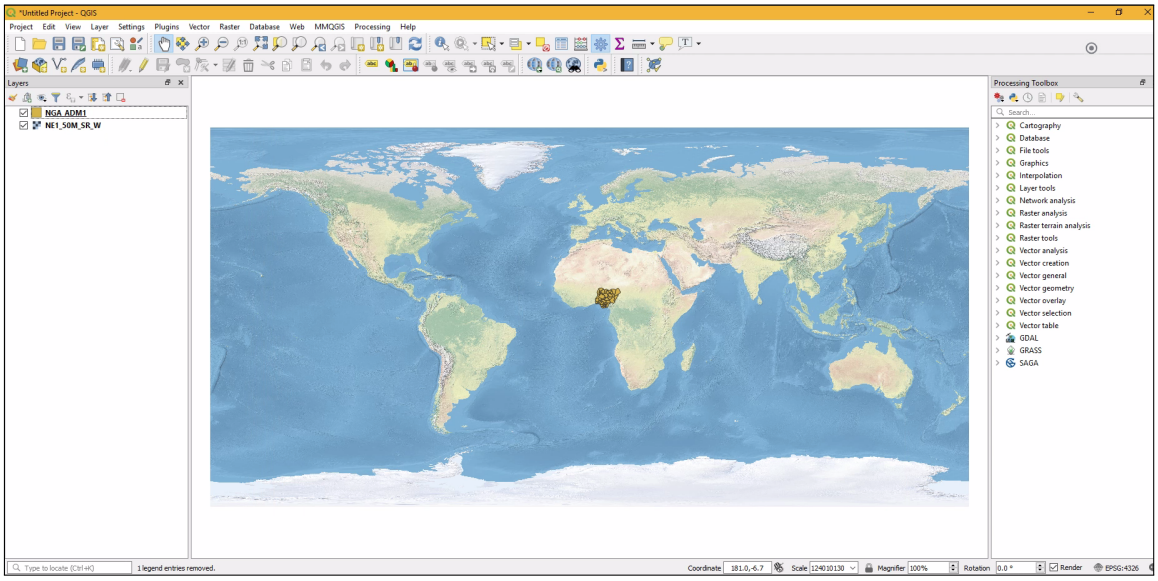
This saves your entire workspace as a project, which is essentially a list of references to file locations, layer styling choices, and other settings used. You will be able to open this single project file in the future instead of having to load each layer again.

1.2.3 Nigeria State Boundaries

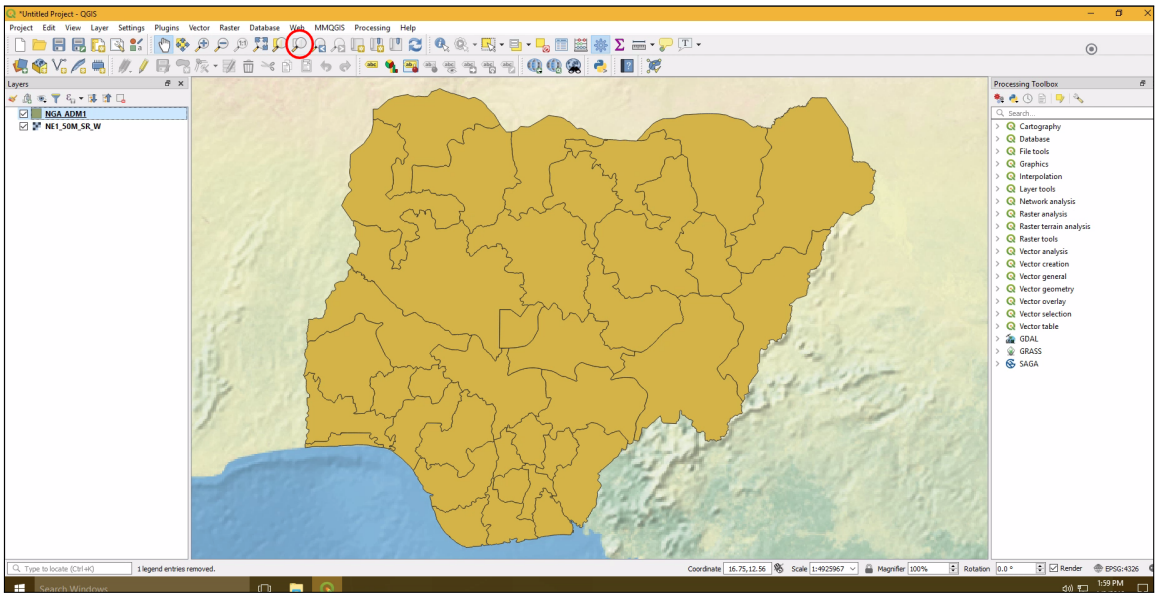
Now we can add in our Nigeria state boundaries.

1. Open the Data Source Manager again, but select the "Vector" tab this time.
2. Find and select the **.../Desktop/training_data/NGA_ADM1.geojson** file.
3. Click "Add" to add this layer.





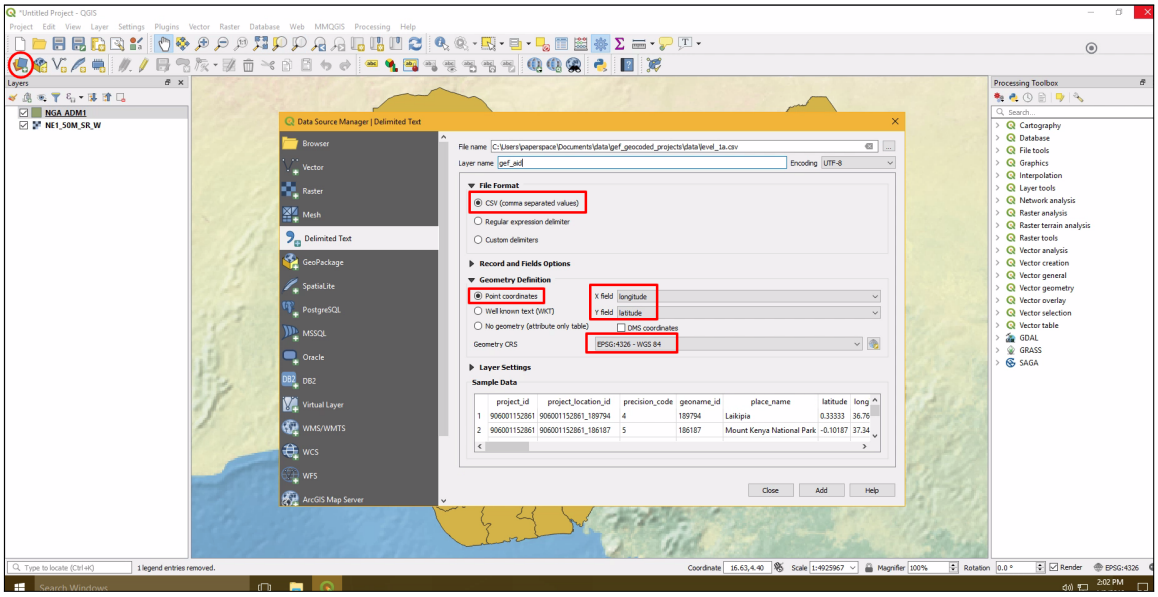
4. You can then right click on the layer in the layer list on the left side of the screen, and select "Zoom to Layer" to get a better view
5. Alternatively, you can use the "Zoom to Layer" button in the top toolbar.



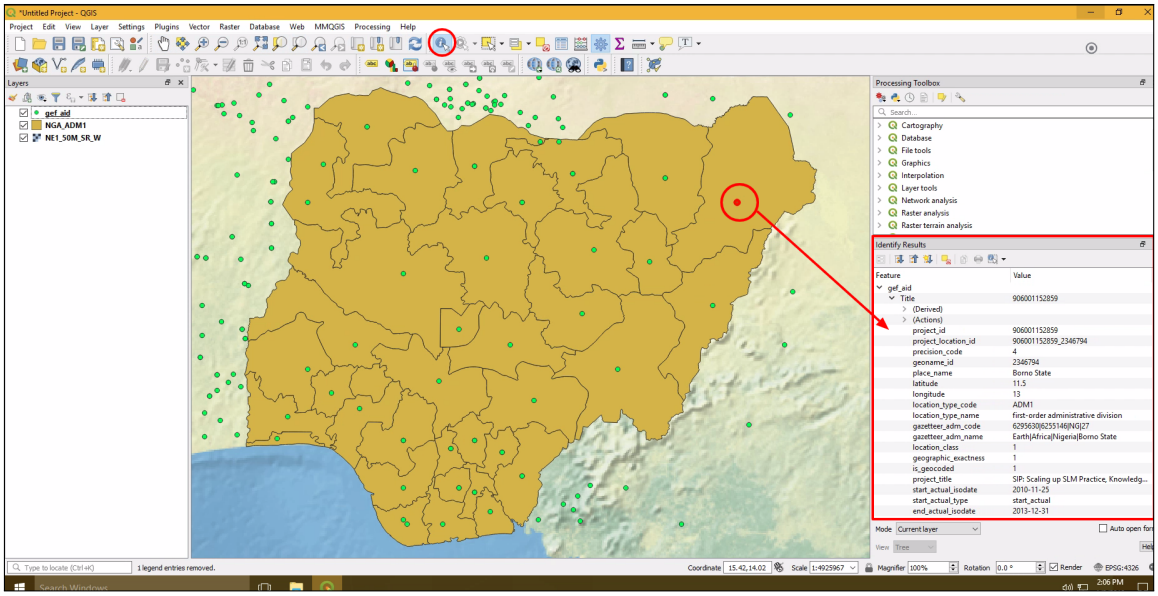
1.2.4 GEF Geocoded Aid Projects

Next, we will add our aid projects.

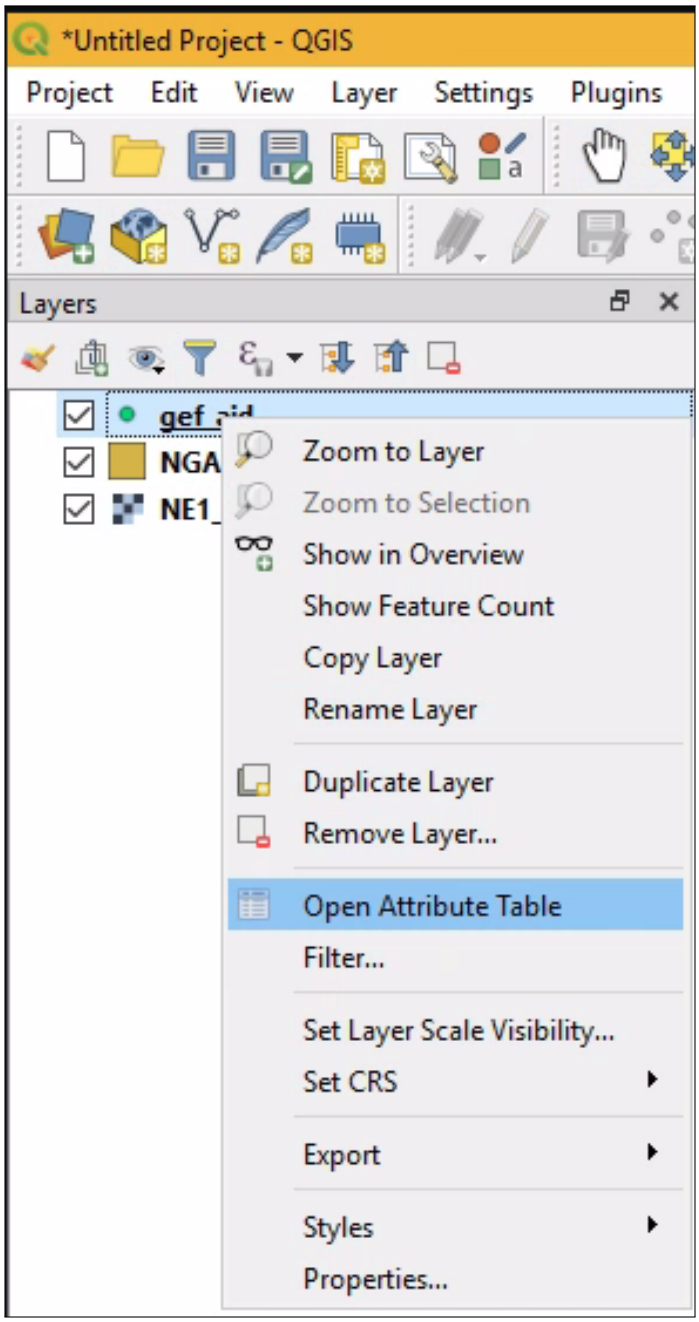
1. Open the Data Source Manager again, and select the "Delimited Text" tab this time.
2. For "File name", select the **.../Desktop/training_data/gef_geocoded_projects/data/level_1a.csv** file.
3. You can set the "Layer name" to "gef_aid"
4. QGIS should automatically load the selected .csv file to the window and detect the longitude and latitude fields.
5. If it does not load automatically, make sure the "File Format" is set to "CSV" and "Geometry Definition" is set to "Point coordinates"
6. Set "Geometry CRS" to "EPSG:4326 - WGS 84"
7. Click **Add** to load the selected .csv file, then "Close"

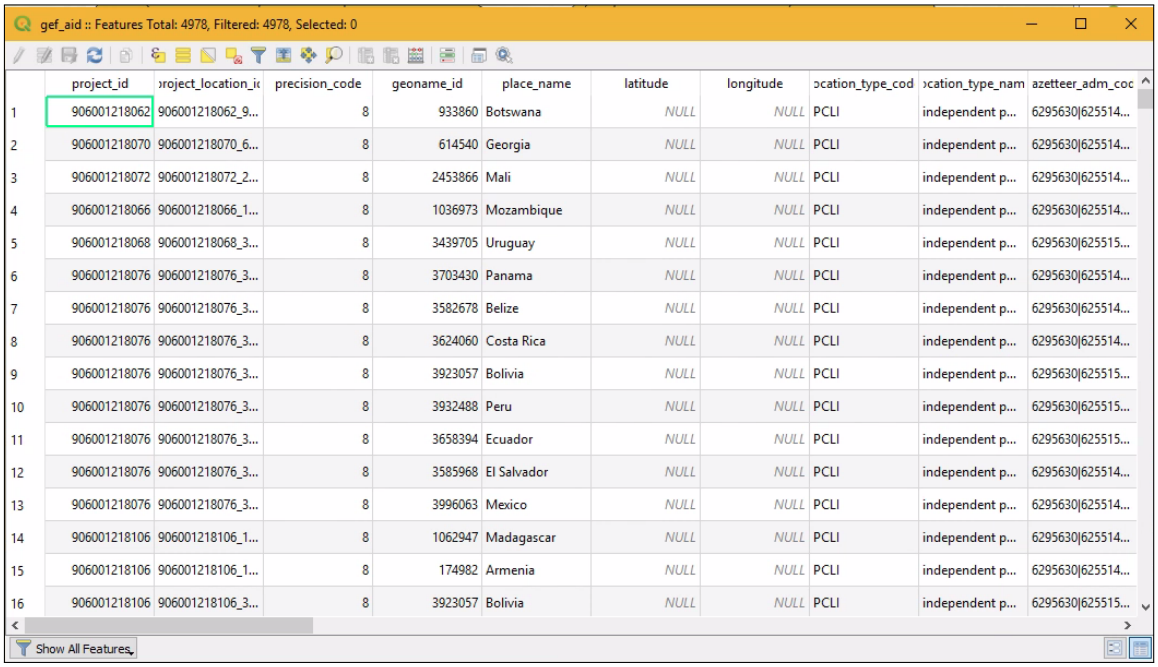


8. You can use the "Identify Features" button (top toolbar, shortcut: Ctrl+Shift+I) to click on any of the points added to the map. This will display details about the point in the "Identify Results" window on the right side.



9. To further explore this data, right click the **gef_aid** layer and click **Open Attribute Table**. The attribute table displays information associated with each features, or project location, in the layer. Each row represents a geocoded project location, and each column contains specific information about the project location such as the project title, status, start year, or commitments.


*Untitled Project - QGIS
Project
Edit
View
Layer
Settings
Plugins
Layers
NGA
NE1_
Zoom to Layer
Zoom to Selection
Show in Overview
Show Feature Count
Copy Layer
Rename Layer
Duplicate Layer
Remove Layer...
Open Attribute Table
Filter...
Set Layer Scale Visibility...
Set CRS
Export
Styles
Properties...



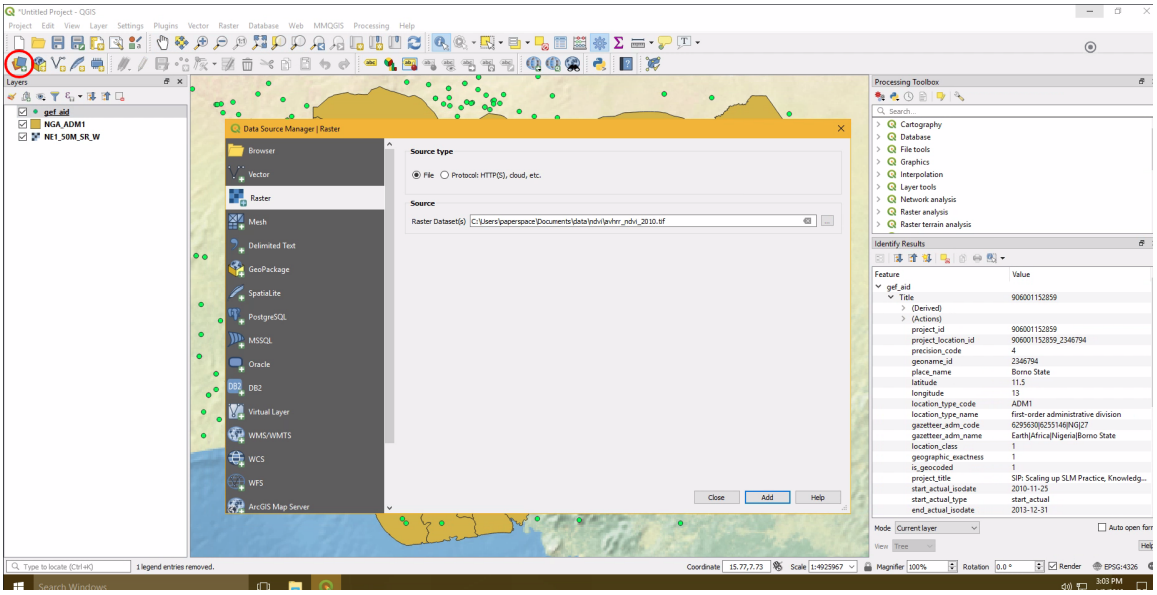
	project_id	project_location_ic	precision_code	geoname_id	place_name	latitude	longitude	cation_type_cod	cation_type_nam	azetteer_adm_coc
1	906001218062	906001218062_9...	8	933860	Botswana	NULL	NULL	PCLI	independent p...	6295630|625514...
2	906001218070	906001218070_6...	8	614540	Georgia	NULL	NULL	PCLI	independent p...	6295630|625514...
3	906001218072	906001218072_2...	8	2453866	Mali	NULL	NULL	PCLI	independent p...	6295630|625514...
4	906001218066	906001218066_1...	8	1036973	Mozambique	NULL	NULL	PCLI	independent p...	6295630|625514...
5	906001218068	906001218068_3...	8	3439705	Uruguay	NULL	NULL	PCLI	independent p...	6295630|625515...
6	906001218076	906001218076_3...	8	3703430	Panama	NULL	NULL	PCLI	independent p...	6295630|625514...
7	906001218076	906001218076_3...	8	3582678	Belize	NULL	NULL	PCLI	independent p...	6295630|625514...
8	906001218076	906001218076_3...	8	3624060	Costa Rica	NULL	NULL	PCLI	independent p...	6295630|625514...
9	906001218076	906001218076_3...	8	3923057	Bolivia	NULL	NULL	PCLI	independent p...	6295630|625515...
10	906001218076	906001218076_3...	8	3932488	Peru	NULL	NULL	PCLI	independent p...	6295630|625515...
11	906001218076	906001218076_3...	8	3658394	Ecuador	NULL	NULL	PCLI	independent p...	6295630|625515...
12	906001218076	906001218076_3...	8	3585968	El Salvador	NULL	NULL	PCLI	independent p...	6295630|625514...
13	906001218076	906001218076_3...	8	3996063	Mexico	NULL	NULL	PCLI	independent p...	6295630|625514...
14	906001218106	906001218106_1...	8	1062947	Madagascar	NULL	NULL	PCLI	independent p...	6295630|625514...
15	906001218106	906001218106_1...	8	174982	Armenia	NULL	NULL	PCLI	independent p...	6295630|625514...
16	906001218106	906001218106_3...	8	3923057	Bolivia	NULL	NULL	PCLI	independent p...	6295630|625515...

10. The Attribute Table also allows you to select and filter locations, create new fields using the field calculator, and zoom to selected fields. For example, you may wish to filter based on the precision with which project location were able to be geocoded (Note: we can take a moment to discuss geocoding methodology and precision here if there is additional interest.)

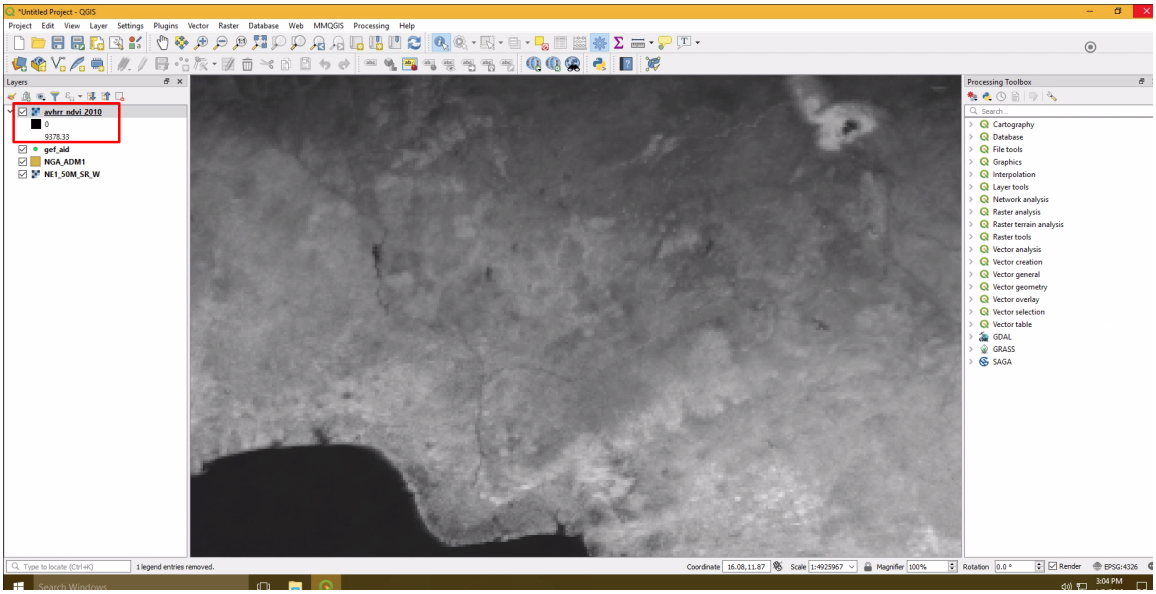
1.2.5 NDVI Satellite Imagery

Finally, we will add in our NDVI raster.

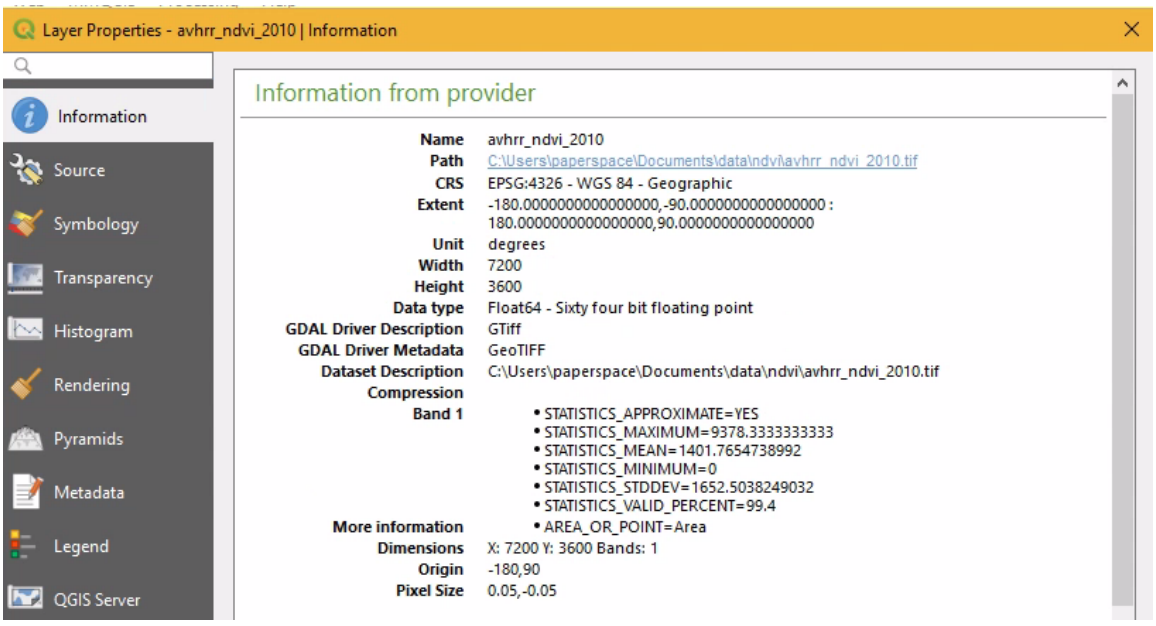
1. Using the Data Source Manager, go to the Raster tab and select the **.../Desktop/training_data/ndvi/avhrr_ndvi_2010.tif** file, then Add and Close



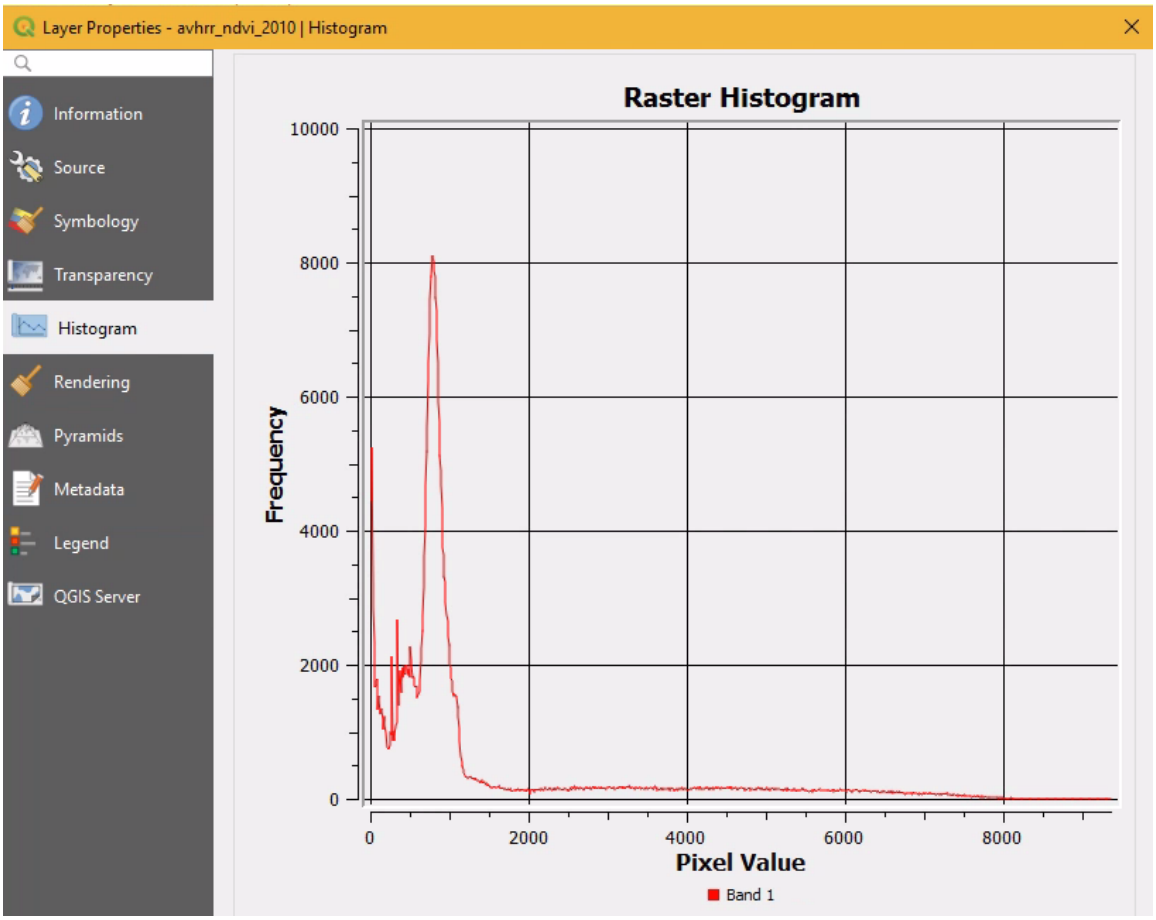
2. The NDVI layer defaults to a grey scale display. To change this, and explore other information about this layer, right click on the layer in the Layers window and select "Properties"



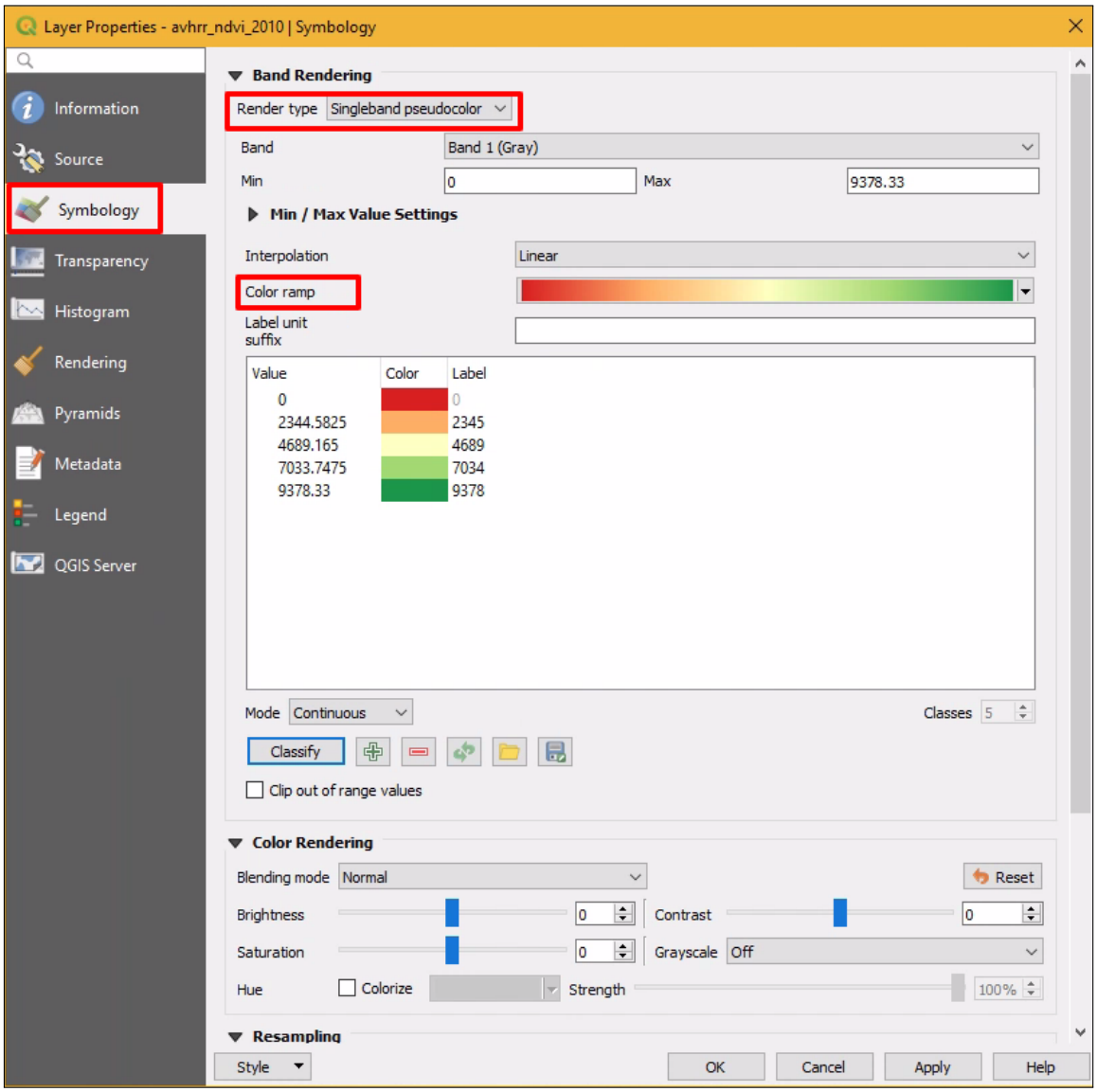
3. The "Information" tab will show basic statistics and metadata for this layer



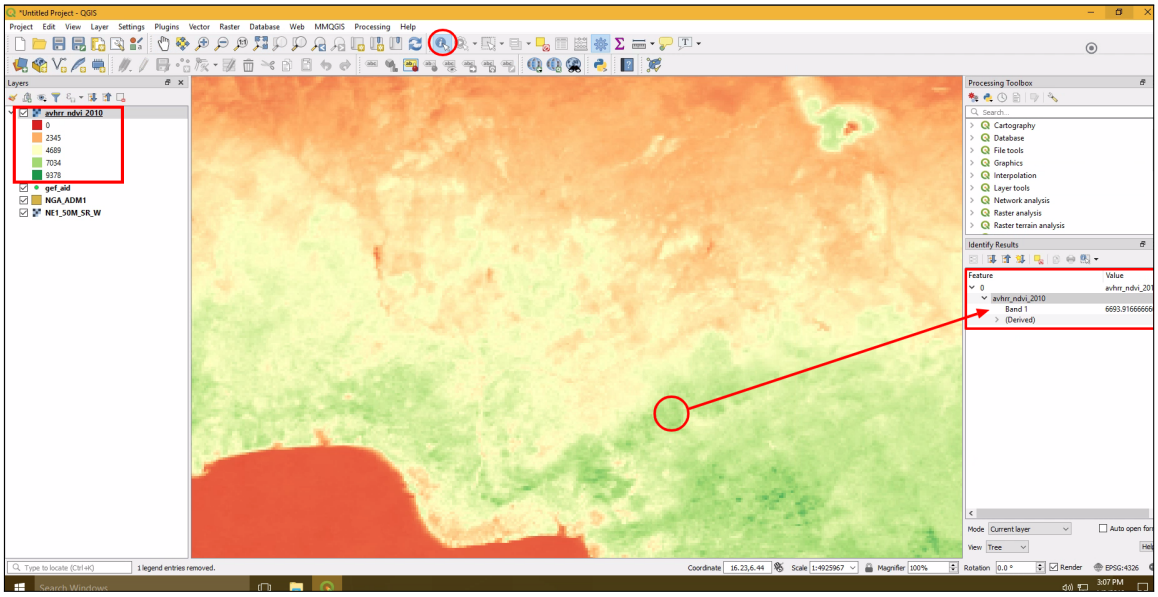
4. The "Histogram" tab will display the distribution of pixel values (corresponding to NDVI measurements) across the entire raster layer. You may need to click the "Compute Histogram" button first. For more details on NDVI and how it is calculated, check out `https://gisgeography.com/ndvi-normalized-difference-vegetation-index/`



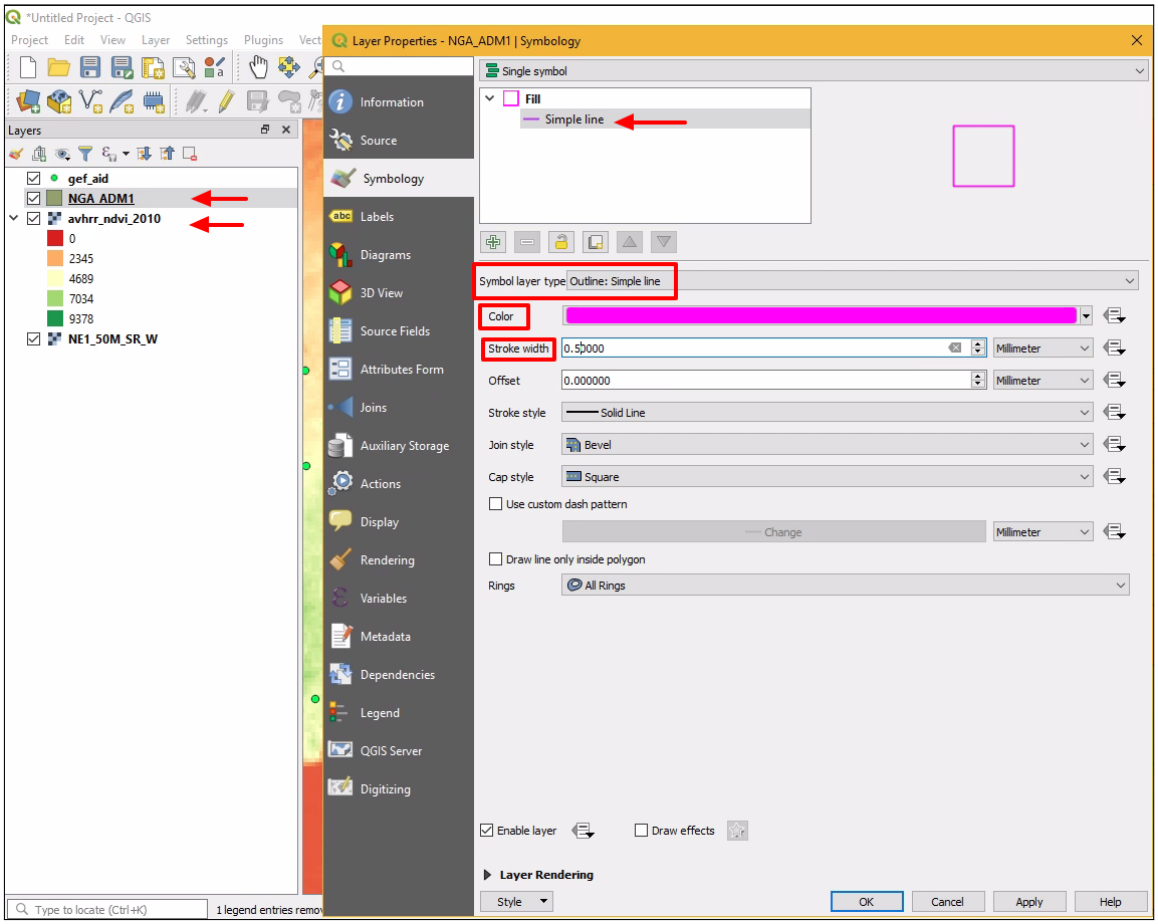
5. Go to the "Symbology" tab to change the style or coloring of this layer to be more intuitive for NDVI. We will use the Red-Yellow-Green color ramp in order to indicate low NDVI with and high NDVI with green.
    - Change "Render Type" to "Singleband Pseudocolor"
    - Click on the dropdown arrow for "Color ramp" and then go to the "All Color Ramps" option and select "RdYlGn"
    - Next click the "Classify" button, then "Apply" and "OK"



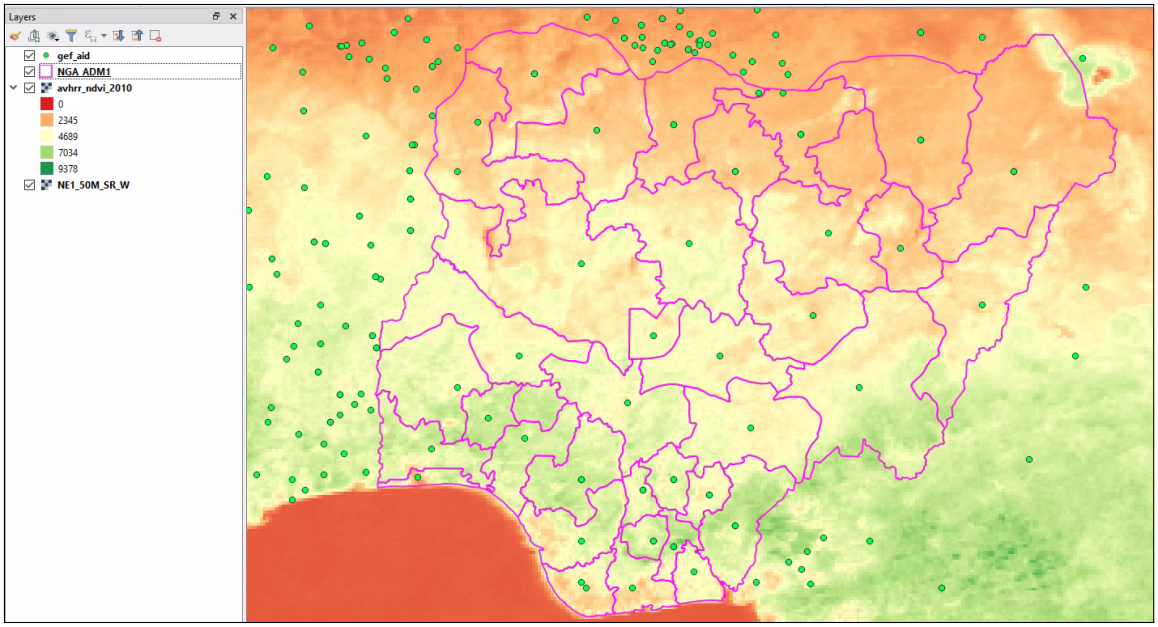
6. The new raster layer style should now be applied, like below



7. You can use the same "Identify Features" tool from before to look at individual pixel values by clicking at different points on the map. The values will appear in the "Identify Results" window on the right.
8. To visualize our Nigeria state boundaries and aid projects on top of the NDVI layer, we can adjust the order of layers in the "Layers" window by simply dragging layers.
9. Move the "NGA_ADM1" and "gef_aid" layers above the "avhrr_ndvi_2010" layer
10. Then, we can change the boundaries to display as only an outline instead of a solid fill.
11. Right click on "NGA_ADM1", go to "Properties" and then "Symbology"
12. Click on "Simple fill" (below "Fill" in the white box)
13. Then change "Symbol layer type" to "Outline: Simple line"
14. Change the "Color" and increase the "Stroke width" to improve visibility



15. Then click "Apply" and "OK".



2. Preparing Data for Analysis

2.1 Manual Processing

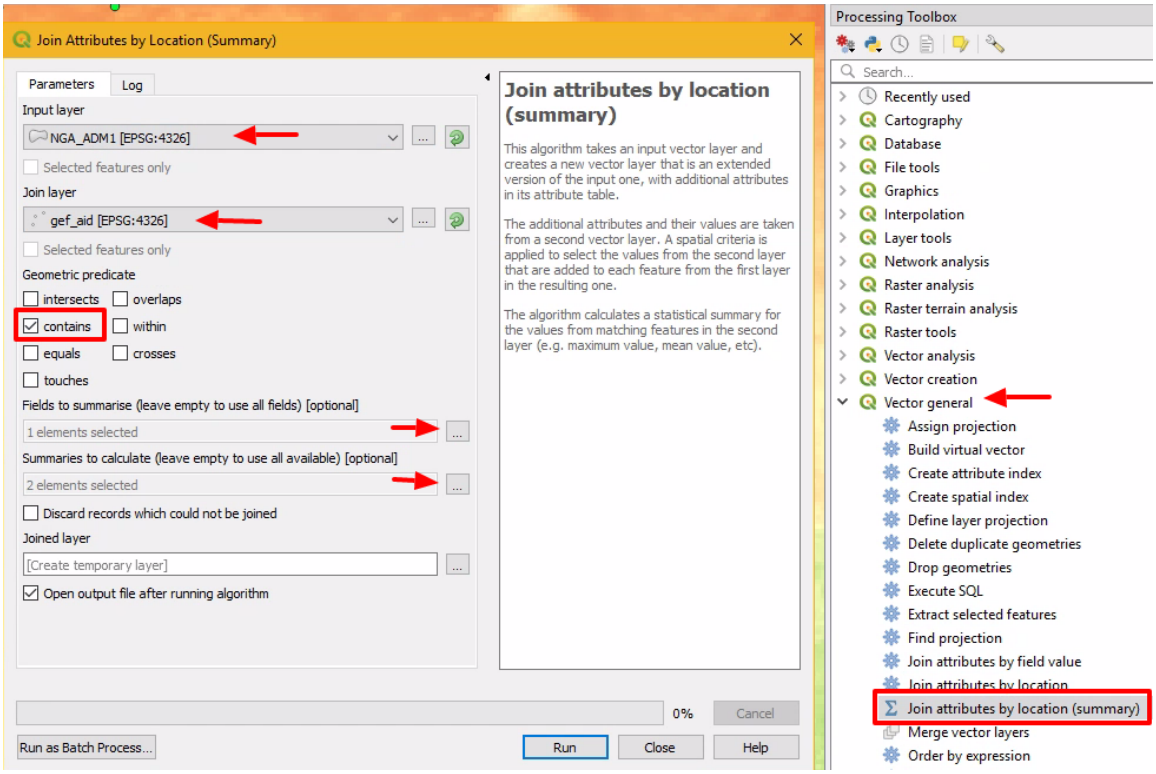
Using states in Nigeria as our unit of analysis, we will now explore how to aggregate both our GEF Geocoded Aid Project (point data) as well as our NDVI satellite imagery (raster data) to these units. This will provide us with a sum of project commitments for each state, as well as a mean NDVI value for each state.

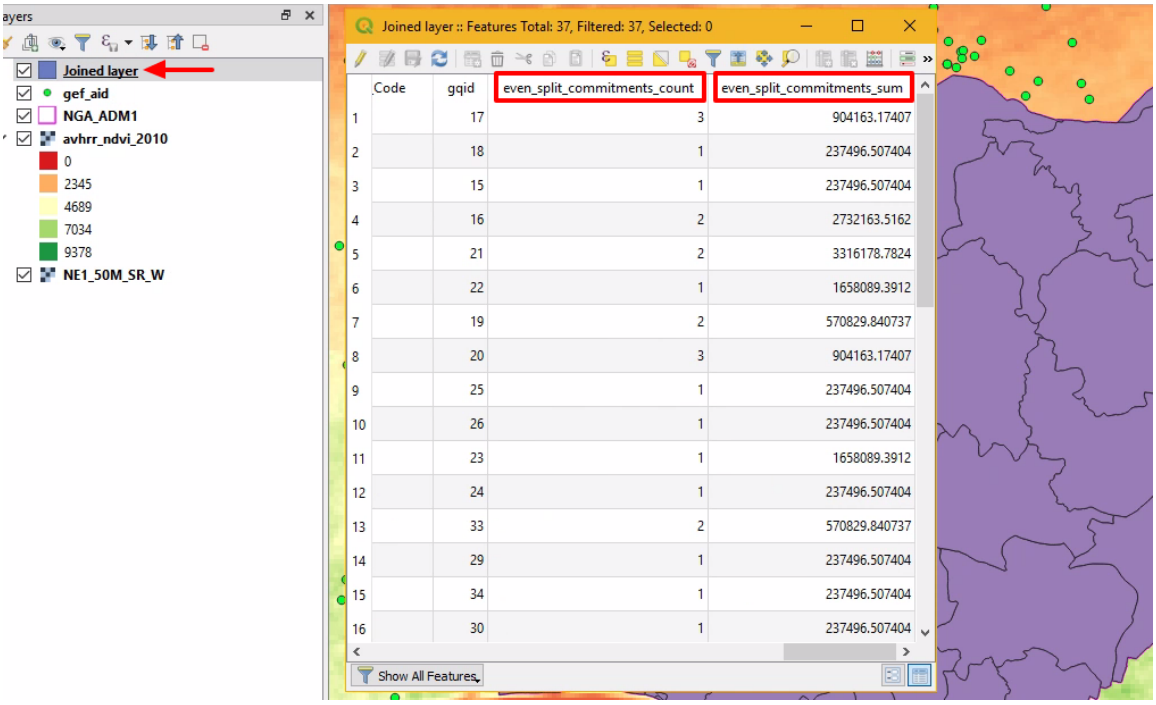
In this section we will first explore the tools and methods used to perform these aggregations manually in QGIS. In the next section we will see how the GeoQuery web platform can be used to make this process drastically faster and easier.

2.1.1 Point Data - Aid Projects

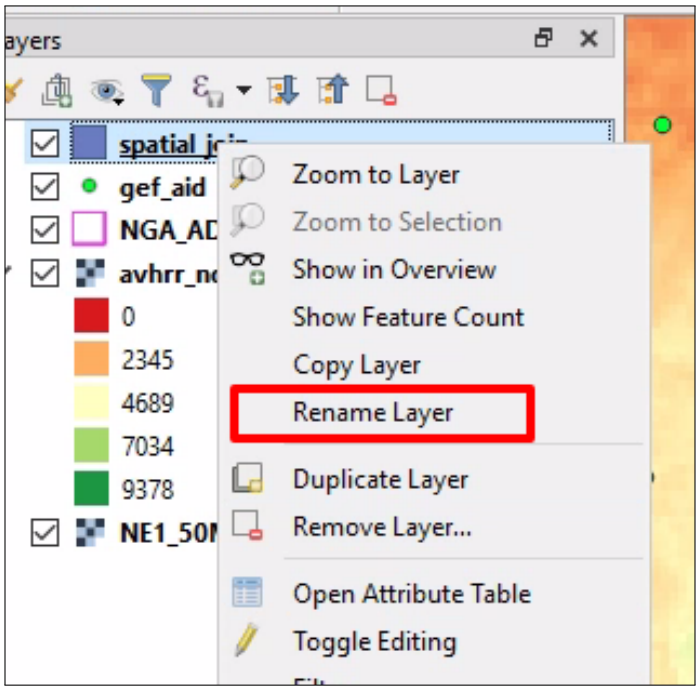
In this section, we are going to utilize a "Spatial Join" tool to calculate the total amount of project commitments going to each state. The spatial join is an operation that aggregates the attribute data from one layer (GEF geocoded aid) to another (Nigeria ADM1) based on a spatial association (e.g., overlapping) and an aggregation function (summation).

1. In the "Processing Toolbox" (right side of window) expand the "Vector general" section and then double click "Join attributes by location (summary)"
2. Set "NGA_ADM1" as the Input layer (this is the layer we want data aggregated to)
3. Set "gef_aid" as the Join layer (layer which has data we want to aggregate)
4. Select "Contains" as the spatial operation in "Geometric predicate".
5. Select "even_split_commitments" from "Fields to summarise"
6. Select "Count" and "Sum" from "Summaries to calculate"
7. Then click **Run**, and "Close" once it finishes (this may take a minute to complete, and will add a new layer named "Joined layer" to you map when it is done)

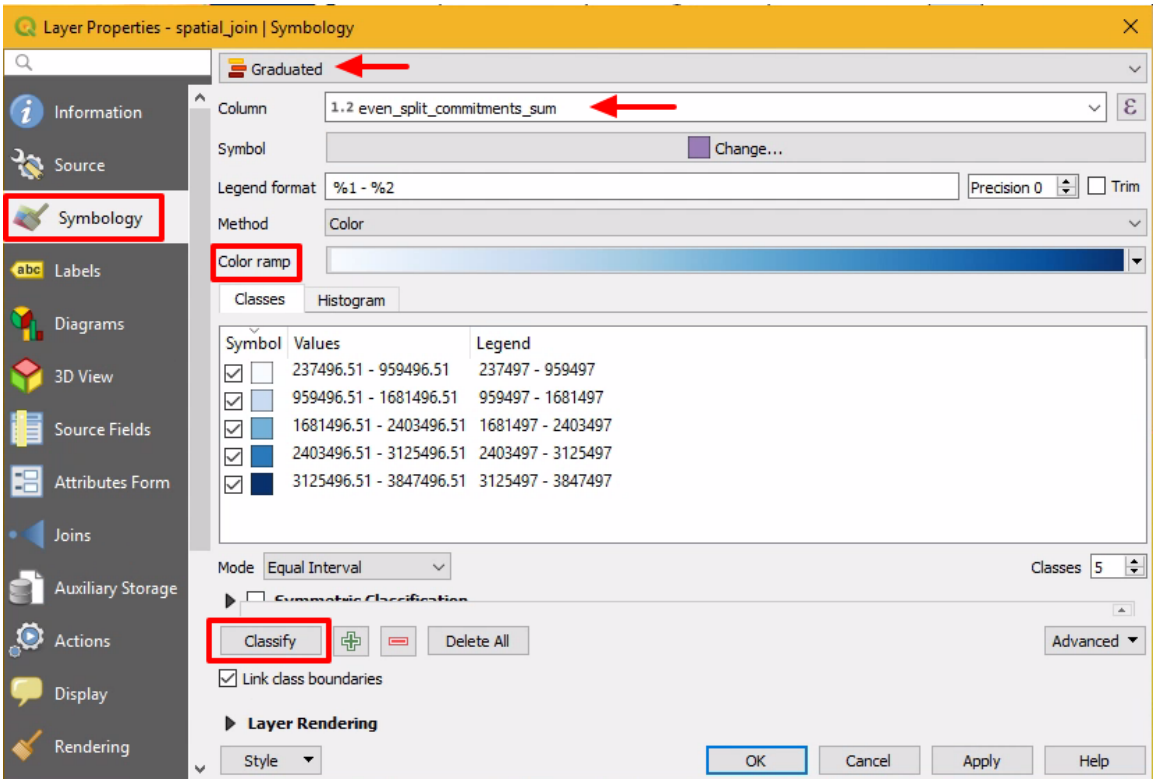


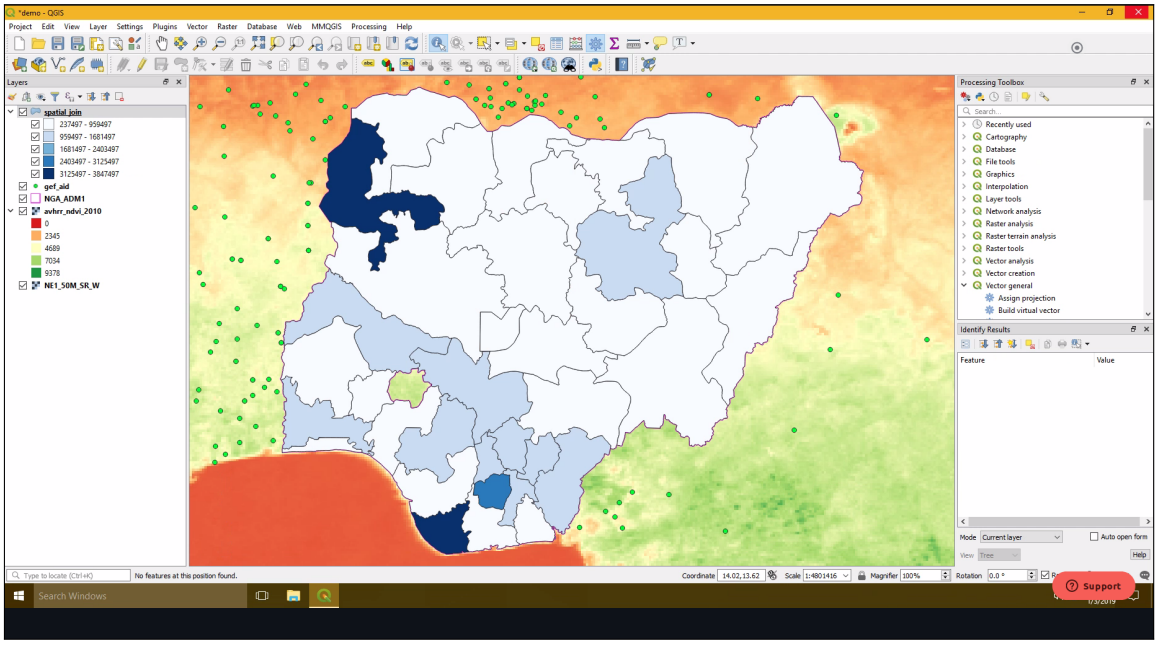


8. Check the attributes by right clicking the new layer and open the attribute table. The "even_split_commitments_count" and "even_split_commitments_sum" columns are the output. The count is the number of projects in each state. The sum is the sum of project commitments for each state.
9. Right click on this layer and select "Rename Layer" to change the layer name to "spatial_join_tmp"
10. To save this temporary layer to a file, right click on the "spatial_join_tmp" layer and go to "Export" and select "Save Features As..."
11. From the resulting window, select "GeoPackage" for Format and set the File name to **.../Desktop/workspace/spatial_join"**, set layer name to "spatial_join", then click "OK"
12. You can now remove the temporary "spatial_join_tmp" layer by right clicking on it and selecting "Remove"



13. Now let's visualize the distribution of aid across states in Nigeria
14. Right click on the **spatial_join** layer and click "Properties", then go to the "Symbology" tab
15. Change the "Single Symbol" drop down at the top to "Graduated" then select the "even split" column from the drop down directly below
16. There are various data classification mode provided by QGIS, such as "Equal Interval", "Equal Count", "Natural Break (Jenks)", "Standard Deviation", "Pretty Break". These classification modes use different statistical algorithms to break down the data into separate categories.
17. Click the "Color ramp" dropdown and select "Blues"
18. Click the "Classify" button below
19. (Note: do not worry if the values of your classification do not match the screenshots exactly)
20. Finally, click "OK"

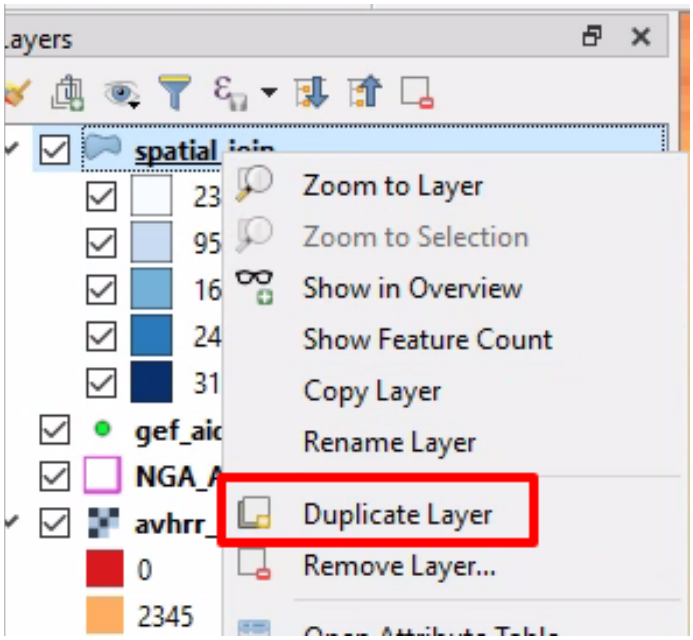

Layer Properties - spatial_join | Symbology
Information
Source
Symbology
Labels
Diagrams
3D View
Source Fields
Attributes Form
Joins
Auxiliary Storage
Actions
Display
Rendering
Graduated
Column
1.2 even_split_commitments_sum
Symbol
Change...
Legend format
%1 - %2
Precision 0
Trim
Method
Color
Color ramp
Classes
Histogram
Symbol
Values
Legend
237496.51 - 959496.51
237497 - 959497
959496.51 - 1681496.51
959497 - 1681497
1681496.51 - 2403496.51
1681497 - 2403497
2403496.51 - 3125496.51
2403497 - 3125497
3125496.51 - 3847496.51
3125497 - 3847497
Mode
Equal Interval
Classes
5
Classify
Delete All
Advanced
Link class boundaries
Layer Rendering
Style
OK
Cancel
Apply
Help



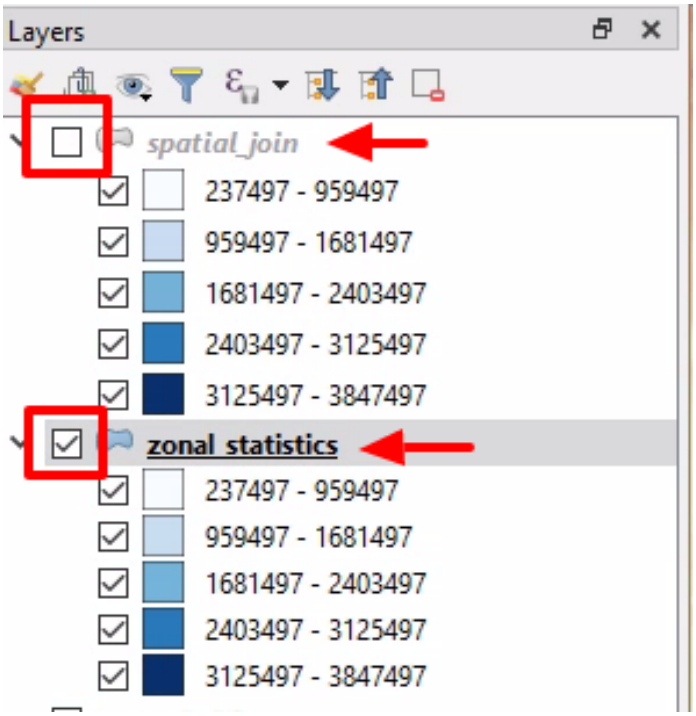
2.1.2 Raster Data - NDVI

With the zonal statistics tool, a summary statistic is calculated for each of the administrative states, based on overlapping pixel values from the input raster dataset. In this case, summary statistics of NDVI values for each administrative state will be calculated.

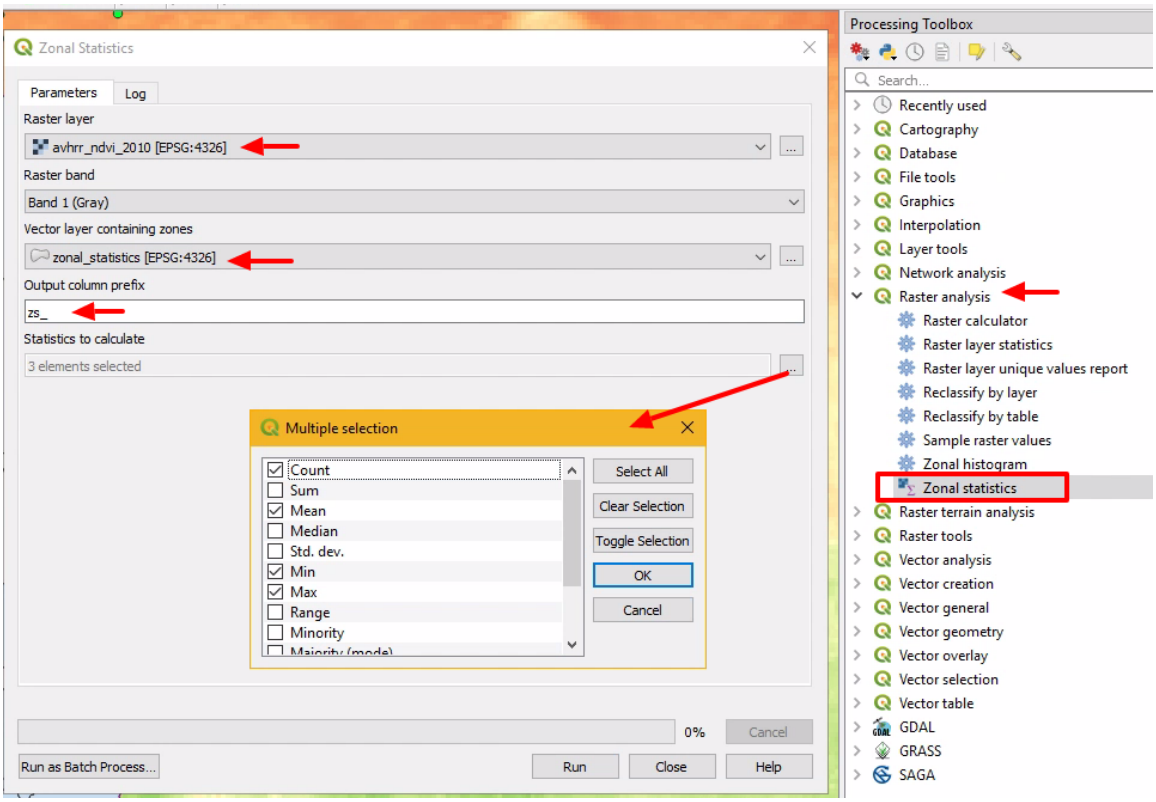
1. First, create a copy of the "spatial_join" layer by right clicking and selecting **Duplicate**
2. Then rename the new layer ("spatial_join copy") to "zonal_statistics" by right clicking and selecting **Rename**



3. You can adjust the visibility of the "spatial_join" layer and new "zonal_statistics" layer by clicking the checkboxes next to each layer. Let's turn off the "spatial_join" and turn on the "zonal_statistics" layer



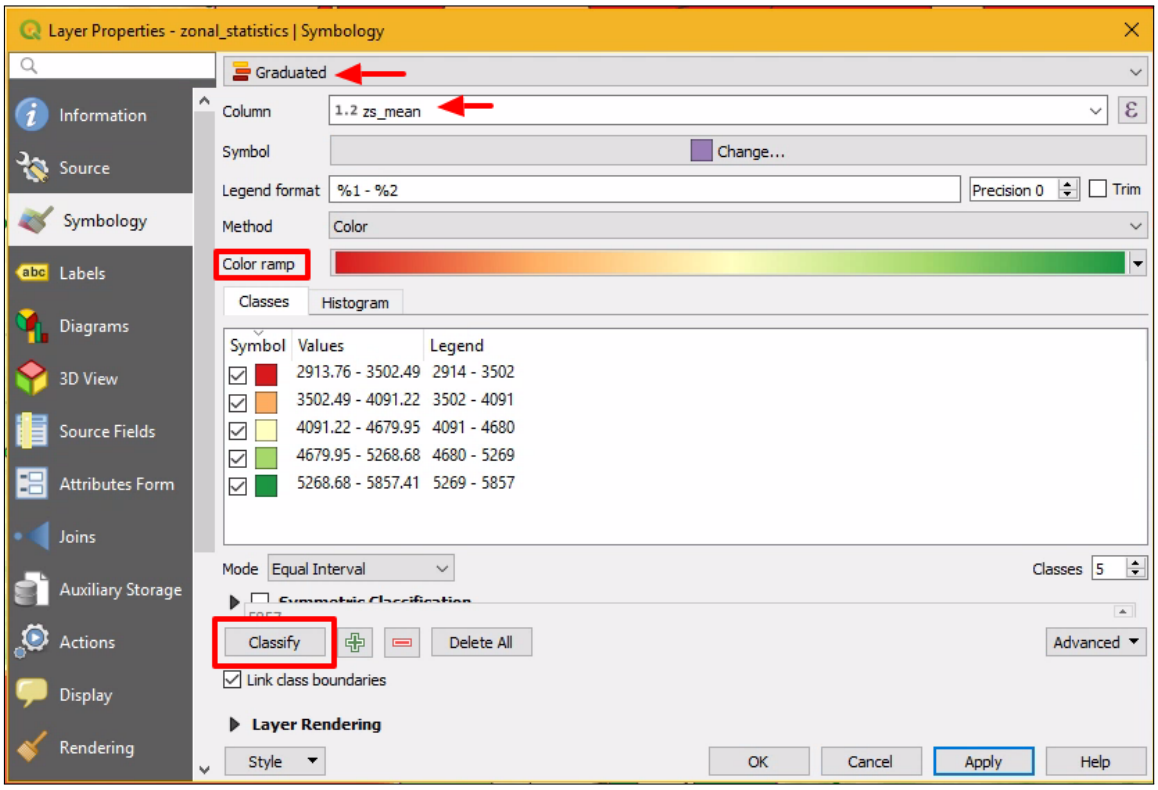
4. Open “Zonal Statistics” by going to the "Raster analysis" menu in the "Processing Toolbox" on the right, and then selecting "Zonal statistics"
5. Set “avhrr_ndvi_2010” layer to be your Raster layer.
6. Set "zonal_statistics" to be the vector layer.
7. Set the “Output column prefix” to “zs_”.
8. Select "Count", "Mean", "Min", and "Max" from the "Statistics to calculate" options
9. Click "Run" and then "Close" when it finishes. This may take a moment to aggregate the raster NDVI value to your Nigeria state boundaries

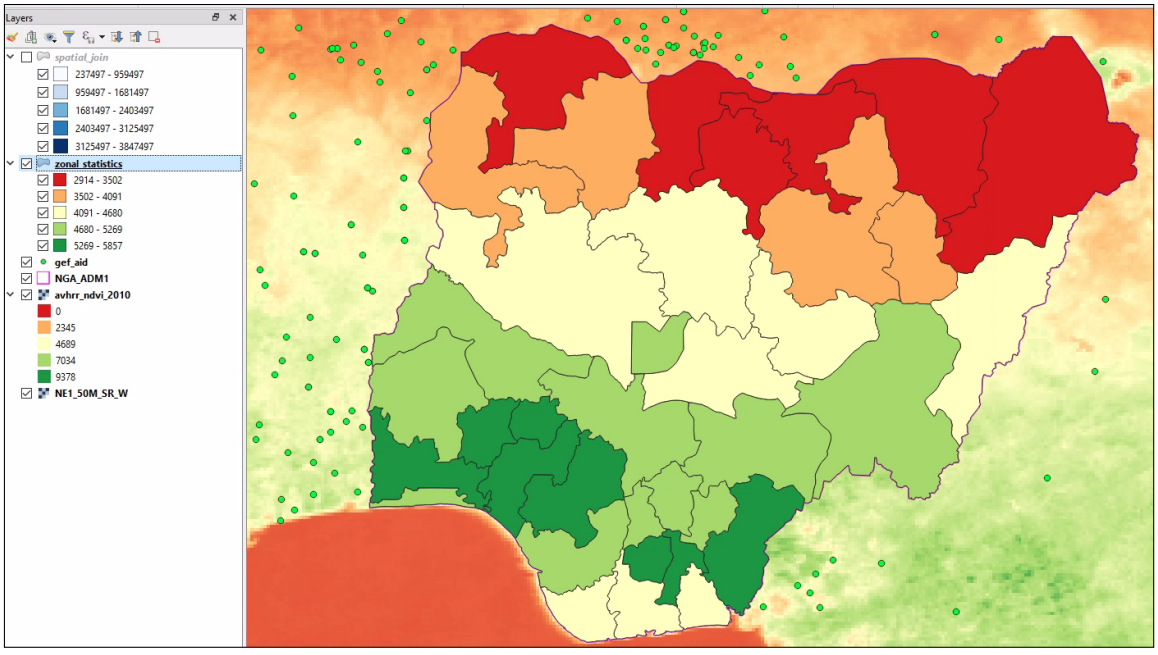


10. Next, right click "zonal_statistics" layer and click “Open the Attribute Table” to check the new fields added by zonal stats
11. The new field representing the average NDVI value in each state can be found in the "zs_mean" column. Additional "zs_" columns have been generated representing the other aggregation methods (count of pixels, min NDVI value, max NDVI values)

	itments_sum	zs_count	zs_mean	zs_min	zs_max
1	904163.17407	779	3065.275780915...	2436.083333333...	4599.916666666...
2	237496.507404	1463	4485.34455456824	3298.333333333...	5553.083333333...
3	237496.507404	573	3569.268179173...	1946.833333333...	4874.166666666...
4	2732163.5162	167	5678.874750499...	4541.583333333...	7128.75
5	3316178.7824	1194	3584.454285315...	1928.583333333...	4935.833333333...
6	1658089.3912	948	5053.961849507...	3904.666666666...	6496
7	570829.840737	674	3407.112141444...	2498.333333333...	5107.833333333...
8	904163.17407	790	3125.815084388...	2385.083333333...	4298.333333333...
9	237496.507404	877	4629.813473964...	3445.75	5489
10	237496.507404	2378	4356.884636949...	1509.083333333...	5753.583333333...
11	1658089.3912	1109	4789.035392245...	3321.333333333...	6092.583333333...
12	237496.507404	123	4752.901084010...	1703.416666666...	6390.833333333...
13	570829.840737	1072	3003.029695273...	1777.916666666...	4823.166666666...
14	237496.507404	277	5857.413357400...	4844.416666666...	6897.166666666...
15	237496.507404	1938	5052.861154110...	3172.25	7020.25
16	237496.507404	909	5148.187660432...	3513.25	6791.583333333...

12. Now let's visualize these average NDVI value for each state. Right click the "zonal_statistics" layer and click "Properties" then go to the "Symbology" tab
13. Select "Graduated" from the menu at the top of the style window
14. Select "zs_mean" from the list of columns
15. Change the Color ramp to "RdYlGn" by clicking on the dropdown arrow to the right of the color ramp. Then hover over "All Color Ramps" to see a second dropdown with RdYlGn
16. Then click "Classify", and "OK"


Layer Properties - zonal_statistics | Symbology
Information
Source
Symbology
Labels
Diagrams
3D View
Source Fields
Attributes Form
Joins
Auxiliary Storage
Actions
Display
Rendering
Graduated
Column
1.2 zs_mean
Symbol
Change...
Legend format
%1 - %2
Precision 0
Trim
Method
Color
Color ramp
Classes
Histogram
Symbol
Values
Legend
2913.76 - 3502.49
2914 - 3502
3502.49 - 4091.22
3502 - 4091
4091.22 - 4679.95
4091 - 4680
4679.95 - 5268.68
4680 - 5269
5268.68 - 5857.41
5269 - 5857
Mode
Equal Interval
Classes
5
Classify
Delete All
Advanced
Link class boundaries
Layer Rendering
Style
OK
Cancel
Apply
Help


Layers
spatial_join
237497 - 959497
959497 - 1681497
1681497 - 2403497
2403497 - 3125497
3125497 - 3847497
zonal statistics
2914 - 3502
3502 - 4091
4091 - 4680
4680 - 5269
5269 - 5857
gef_aid
NGA_ADM1
avhrr_ndvi_2010
0
2345
4689
7034
9378
NE1_50M_SR_W

2.2 Using GeoQuery

Now that we have explored how to manually prepare data using QGIS, you can imagine what it would be like if you had 20 different datasets you need to aggregate to 10 different sets of administrative boundaries.

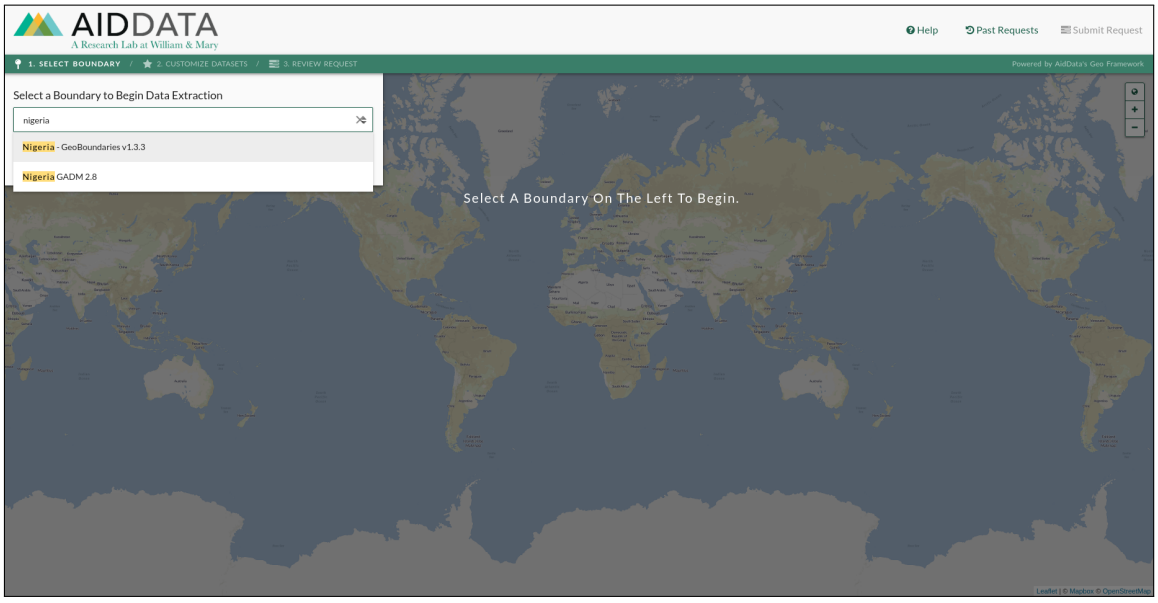
To help reduce technical barriers and processing burdens often associated with incorporating spatial data into research, AidData developed GeoQuery to do the work for you. GeoQuery offers over 45 datasets on aid, conflict, the environment, population, and more that can be aggregated to administrative boundaries for every country in the world.

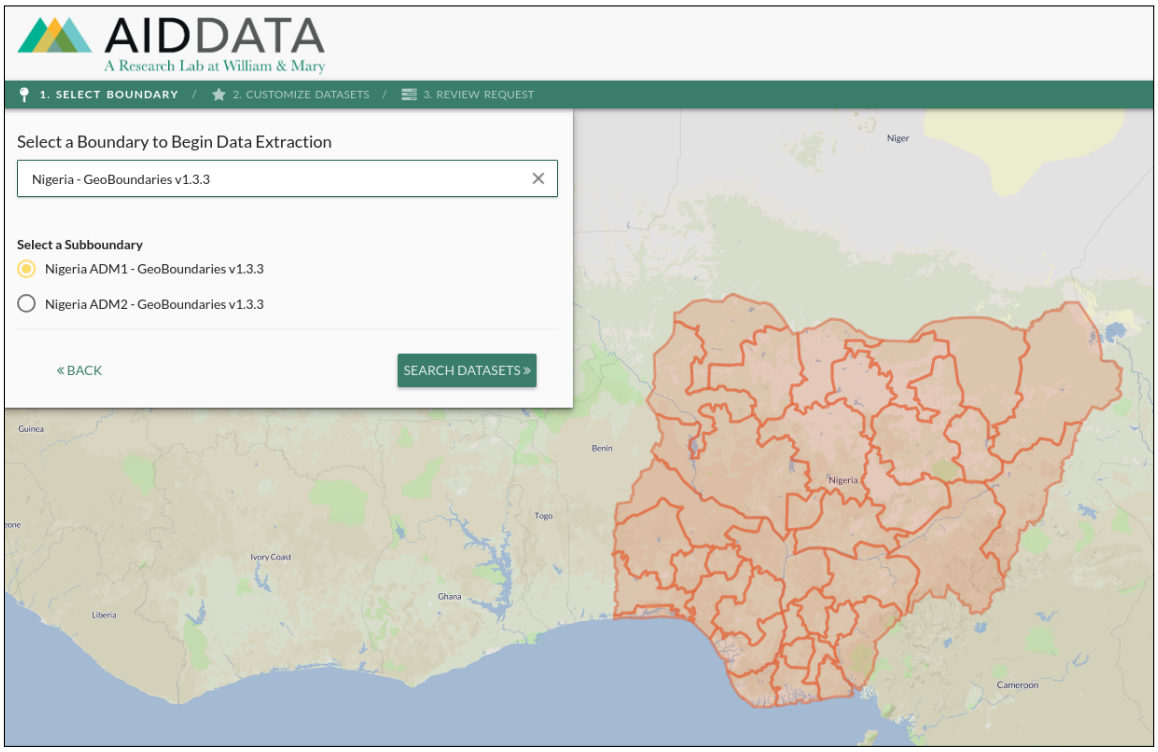
In this section we will look at how you can replicate the work we have just done in a fraction of the time by using GeoQuery.

2.2.1 Web Interface

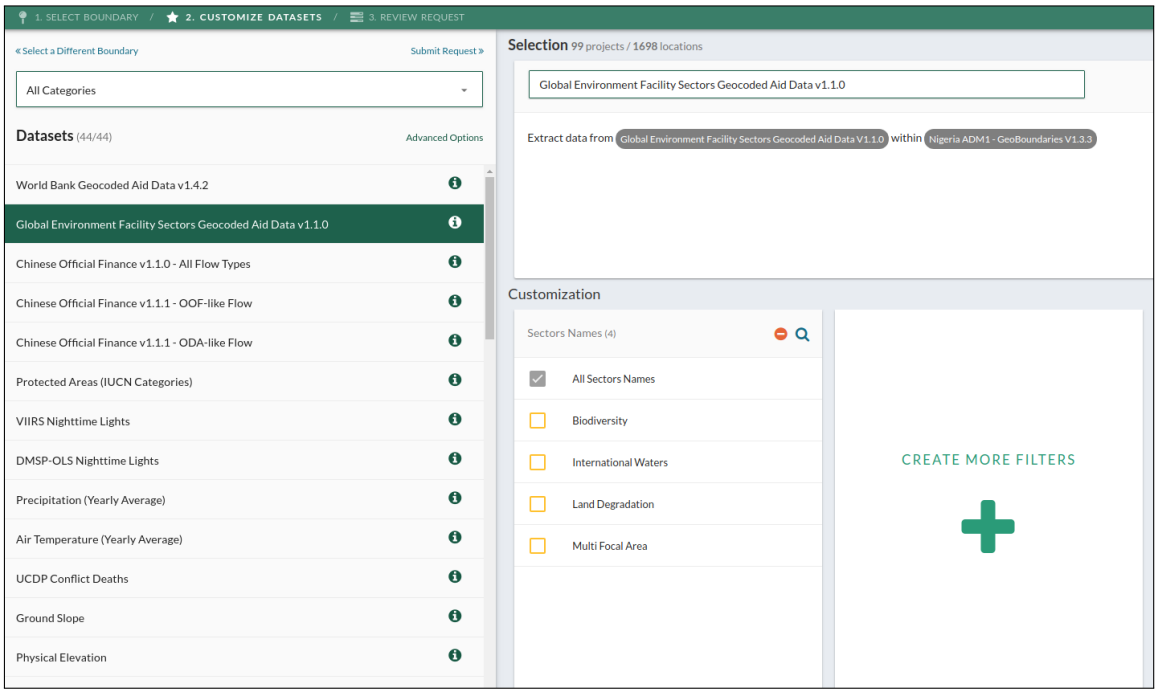
To get started with GeoQuery, go to `geo.aiddata.org` and click the **Get Data** button.

1. The first page of GeoQuery is where you can select the unit of analysis - or the geographic boundaries you are interested in aggregating data to. Let's search for "Nigeria"
2. Select ADM1 or state option for Nigeria (**Nigeria ADM1 Boundary - GADM 2.8**)

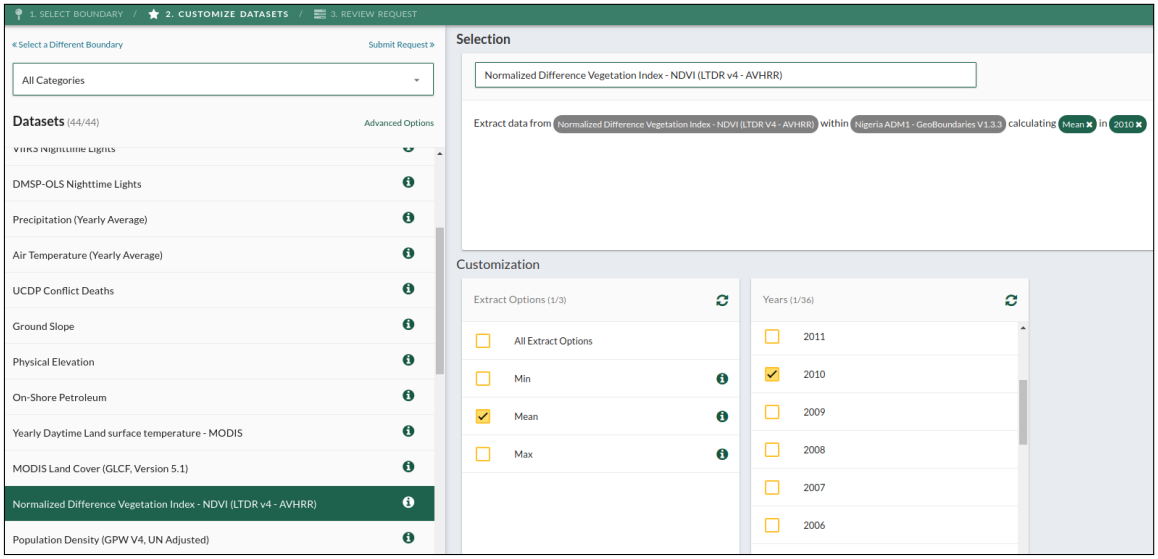




3. Click the **Search Datasets** button to load all datasets in GeoQuery with data for Nigeria
4. Select **Global Environment Facility Geocoded Aid Data v1.1.0** on the left of the data selection screen. For this example, we will leave the default "All Sectors" filter and not select any additional filters. Other filters available include: project title, project status, commitment amount, start and end year, as well as location type and other geocoding details.
5. Click the **Add to Request** button.



6. Next click the **Normalized Difference Vegetation Index - NDVI (LTDR v4 - AVHRR)** dataset from the menu on the left
7. Choose **Mean** as the Extract Options, and **2010** under Years.
8. Click the **Add to Request** button.

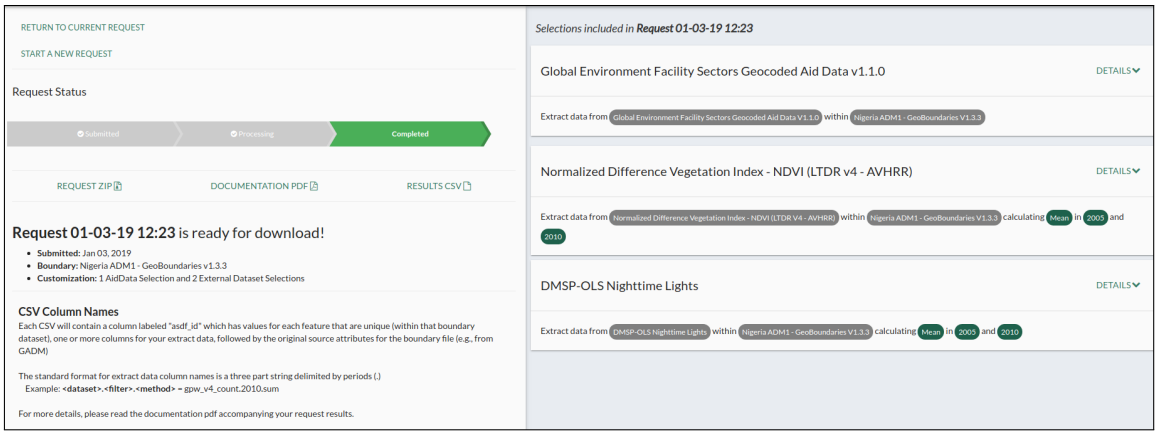


9. Click Submit Request in the upper-right, then Review Request.
10. Enter the email address you would like the results sent to on the right, and click submit

With the above steps completed, your request will have triggered a job on the SciClone high performance computing cluster at the College of William and Mary. All jobs are processed dynamically,

so depending on the complexity of your request you will likely receive a completion email within 5 minutes to approximately 4 hours for extremely large or complex jobs.[1]

When your job is complete, you will receive an email directing you to the permanent download page for that data. You can also always retrieve your history of data requests and downloads by clicking "Past Requests" at the upper-right of GeoQuery and entering your email address.



If for some reason you cannot submit a new job or it is taking more than a few minutes to complete, you can view the results page for an existing request we created for the same data:

`http://geo.aiddata.org/query/#!/status/5c2e4511c15e002940b7ff8f`

2.2.2 Understanding Results

Once your request is completed, you can download three different files: a PDF containing descriptions of your data, a CSV containing your data itself, and a zip file containing everything generated during your request (including CSV and documentation). If you download the CSV and open it in a common program such a Libre Office or Excel, you will see an output very similar to that seen below.

	A	B	C
1	asdf_id	globalenvironmentfacility_geocodedresearchrelease_level1_v1_1_0.2f7c253.sum	ltdr_avhrr_ndvi_v4_yearly.2010.mean
2	0	13879.0323585	4068.06387235
3	1	1979661.38747	2503.54696242
4	2	2282784.5174	2644.79167597
5	3	2507967.70174	7207.18815481
6	4	4927579.805	6878.50744427
7	5	7424211.90277	3532.9963398

Each row represents a geographic boundary, and each column represents data associated with that boundary. In the PDF automatically prepared for you, you will find a description of exactly what each column means, how it was generated, and how to cite the underlying information. To identify what a column name means, the quickest way to find it is to open your documentation PDF and

[1] The average request time is about 2 minutes but in some cases jobs may take 24 hours (or longer) depending on the number of users on the system (i.e. if many users concurrently put in extremely complex jobs). If you have put in a request for a job and it has been longer than 24 hours, you can reach out to the GeoQuery support team at geo@aiddata.wm.edu.

search through it (i.e., ctrl+f) for the name of the column. Units and other key information are provided in the documentation as well.

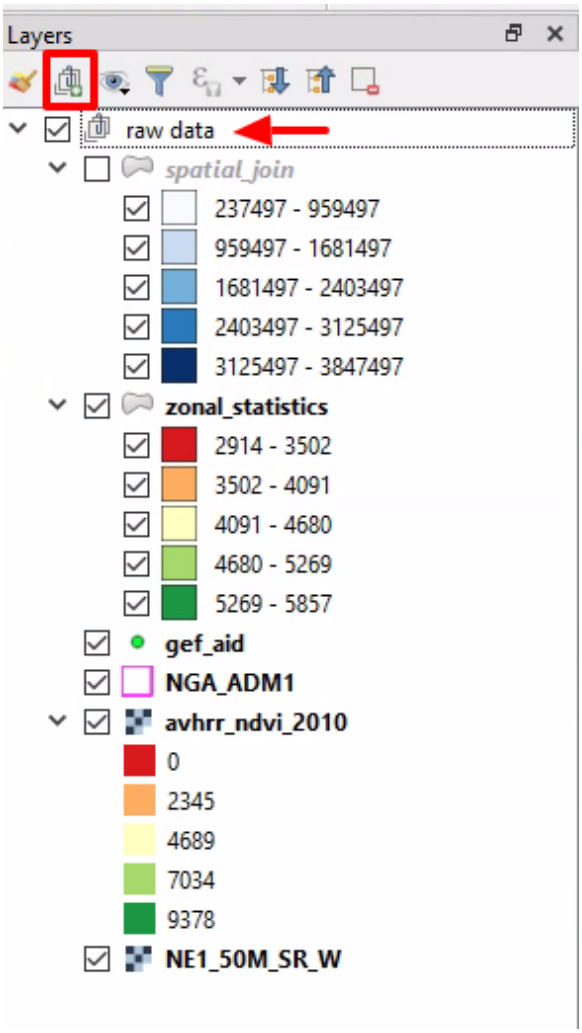
Selection 2 - Normalized Difference Vegetation Index - NDVI (LTDR v4 - AVHRR)

Title	Normalized Difference Vegetation Index - NDVI (LTDR v4 - AVHRR)
Name	ltdr_avhrr_ndvi_v4_yearly
Version	4
Column Names	Format: "ltdr_avhrr_ndvi_v4_yearly.<temporal>.<method>" for all combinations of <temporal> and <method> which can be found in the "Temporal Selection" and "Extract Types Selected" fields below (1 columns total)
Temporal Selection	2010
Extract Types Selected	mean (average NDVI value per unit of analysis)
Description	Yearly value for Normalized Difference Vegetation Index (NDVI). Created using the NASA Long Term Data Record (v4) AVHRR data.
Details	Created by aggregating daily data to monthly by taking the maximum value, then averaging the monthly data to get yearly values. All negative NDVI values were truncated to 0 and saturated pixels were adjusted to the max of the normal NDVI range (10000).
Bounding Box	[[[-180.0, 90.0], [-180.0, -90.0], [180.0, -90.0], [180.0, 90.0], [-180.0, 90.0]]]
Date Added	2017-07-25
Date Updated	2017-07-25
Source Name	NASA/Goddard Space Flight Center
Source Link	http://ltdr.nascom.nasa.gov/ltdr/ltdr.html
Citation	Pedelty JA, Devadiga S, Masuoka E et al. (2007) Generating a Long-term Land Data Record from the AVHRR and MODIS Instruments. Proceedings of IGARRS 2007, pp. 1021–1025. Institute of Electrical and Electronics Engineers, NY, USA.
Variable Description	positive NDVI values 0:10000
Resolution	0.05
Factor	10000.0

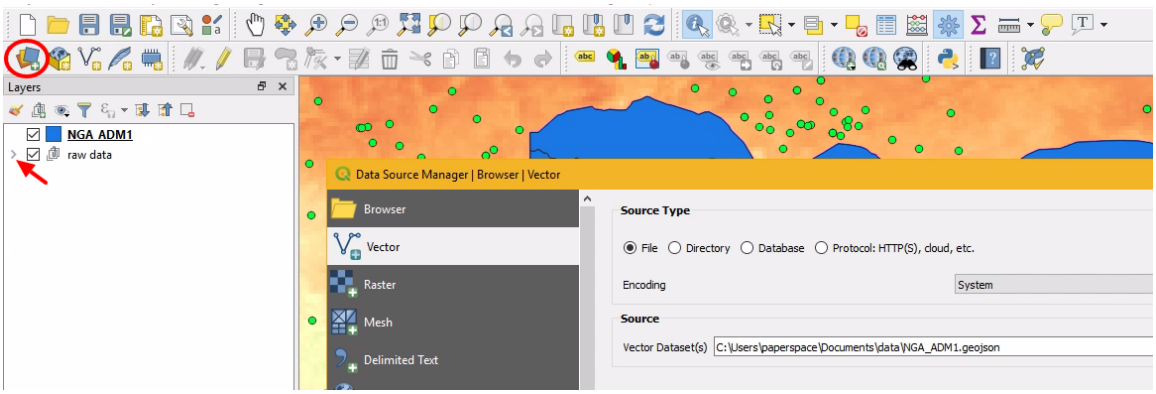
2.2.3 Visualizing Results in QGIS

In this section, we are going to visualize the data that you request from the last section. To save time, we included the request results in the *data* folder for you.

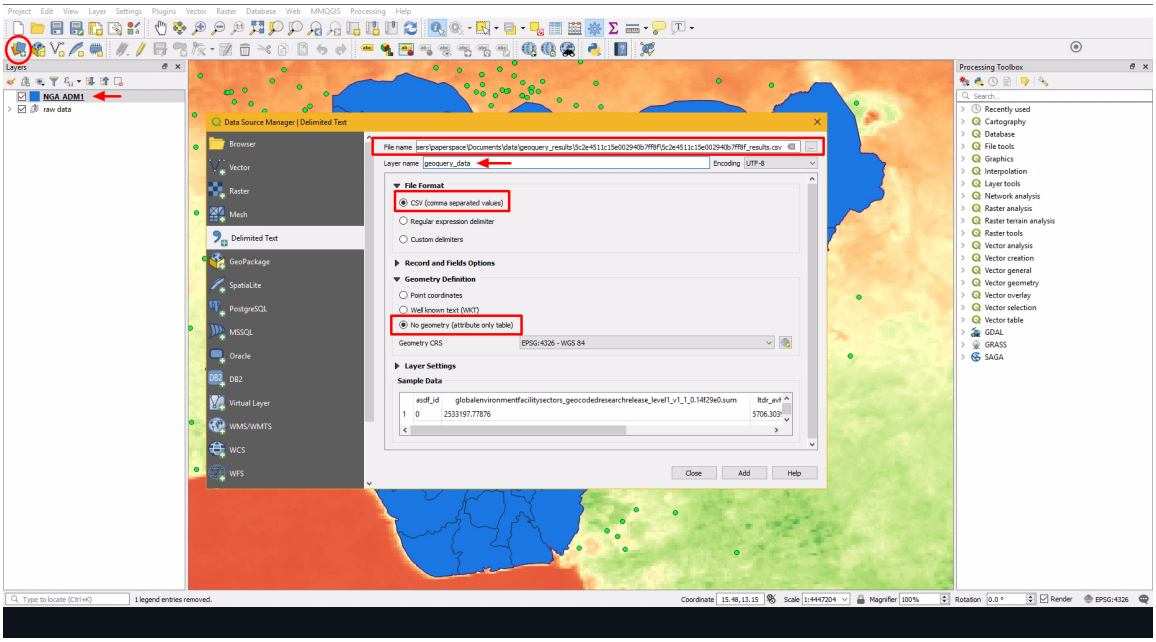
1. First, let's move all of our existing layers into a group so that we can manage our data more easily.
   - Use the "Add Group" button in the "Layers" window to create a new group called "raw data"
   - Then select all the existing layers and drag them into this group
   - You can now toggle the visibility of this group or collapse the group's contents to make room



2. Next, let's load in a fresh copy of our Nigeria state boundaries

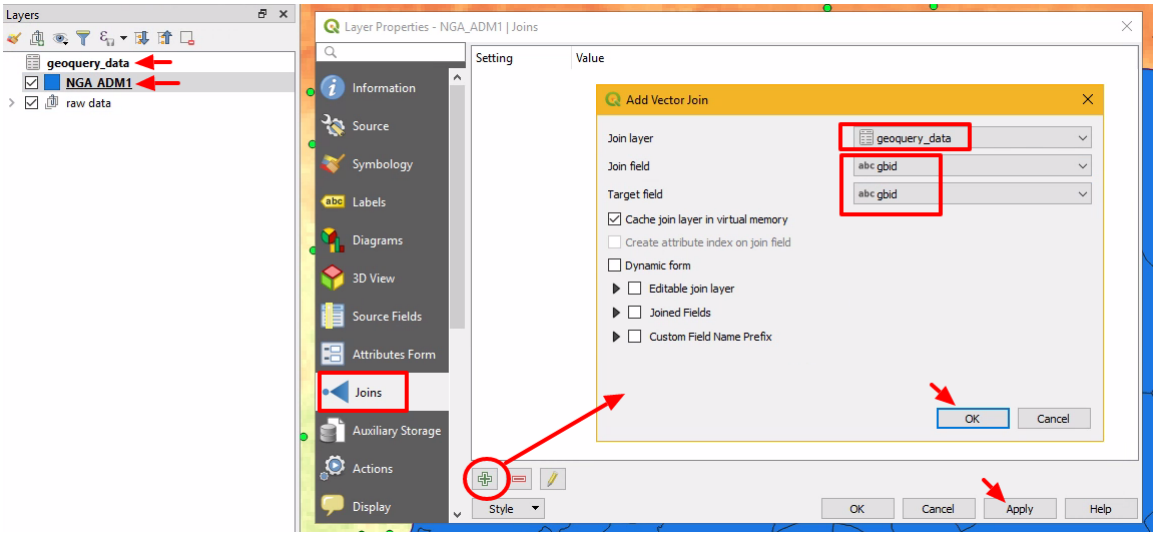


3. Now we will add the GeoQuery results to QGIS as a data only layer. We will join this to our Nigeria State boundaries after.
4. Add the .csv file by opening the Data Source Manager and using the "Delimited Text" tab to add the **.../Desktop/training_data/geoquery_results/5c2e4511c15e002940b7ff8f/5c2e4511c15e002940b7ff8f_results.csv** file
5. Name the layer "geoquery_data"
6. Since this .csv is an attribute only file, make sure to select "No geometry" in the Geometry definition.
7. Click "Add" and then "Close"
8. (Note: this may add this layer to the "raw data" group. Simply drag the "geoquery_data" layer outside the group if needed)

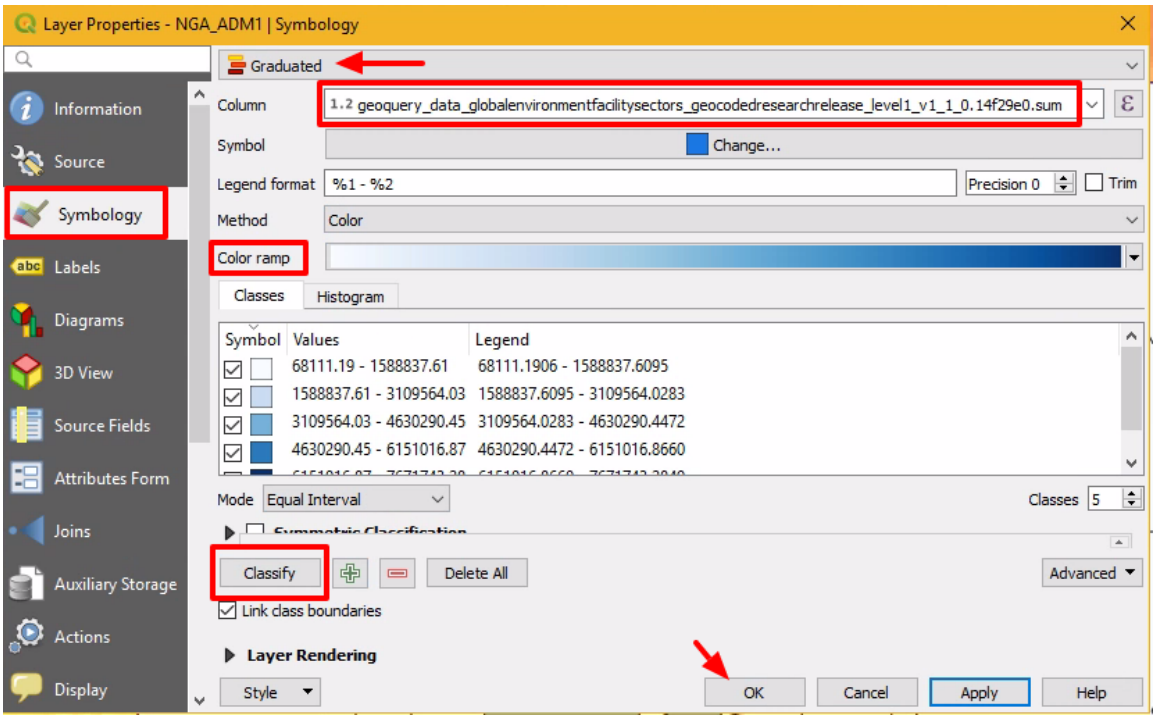


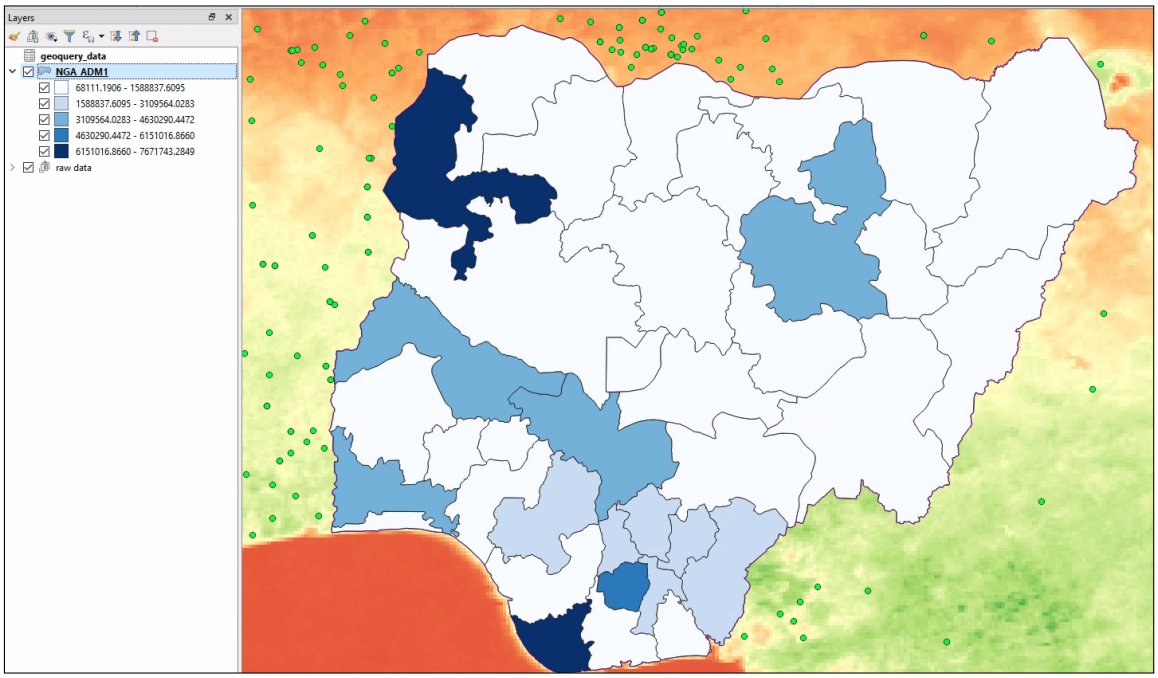
9. Now let's join the GeoQuery data to our boundaries by right clicking on the new "NGA_ADM1" layer and going to "Properties" and then the "Joins" tab
10. In the "Joins" tab, click "+" sign on the bottom.

11. Choose "geoquery_data" as "Join Layer".
12. Select "gbid" as both the "Join field" and "Target field"
13. Click "OK" and then "Apply"

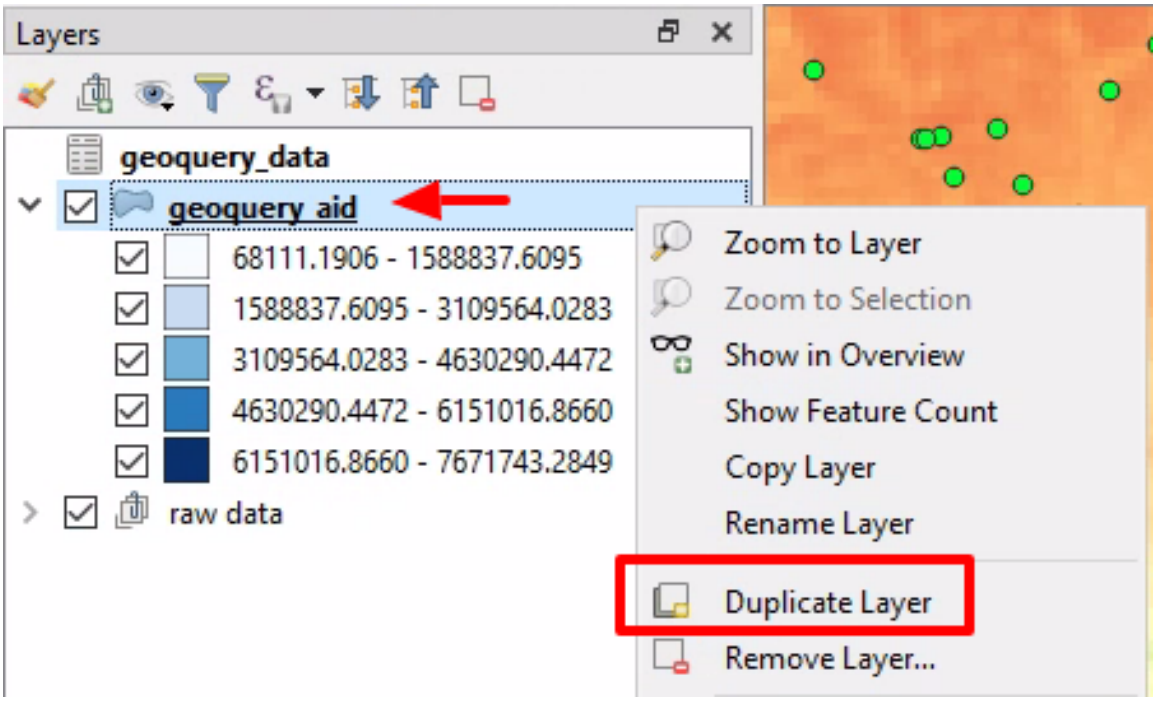


14. Without closing the Properties window, go to the Symbology tab to style our layer with the newly joined data
15. Change the "Single Symbol" drop down at the top to "Graduated".
16. Select "geoquery_data_globalenvironmentfacility..." as the column.
17. Set the color ramp to "Blues" by clicking on the dropdown arrow to the right of the color ramp.
18. Click "Classify" and the "OK"

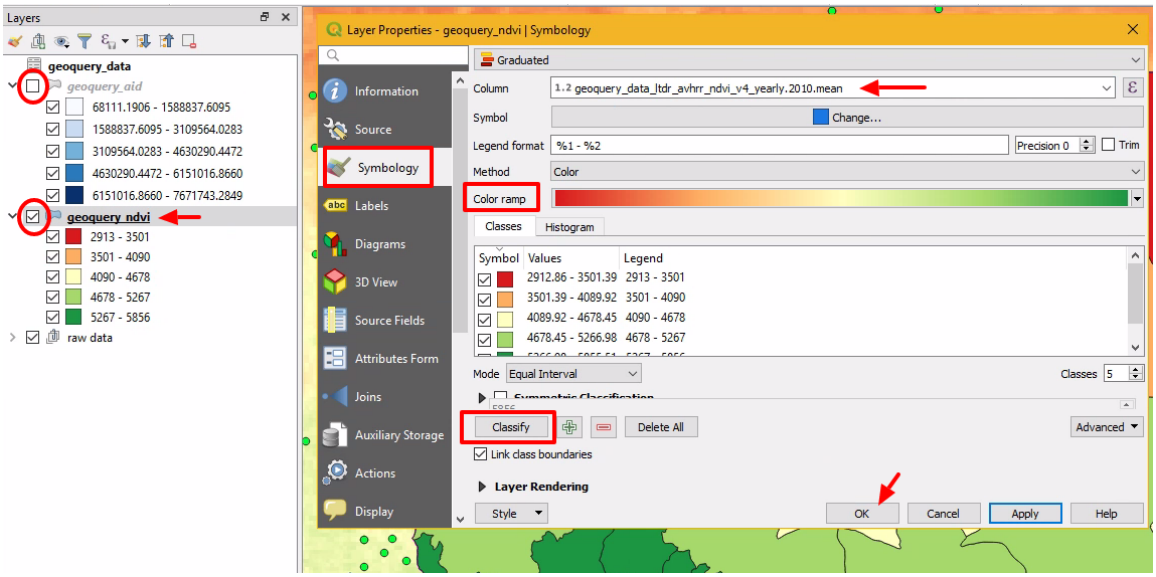


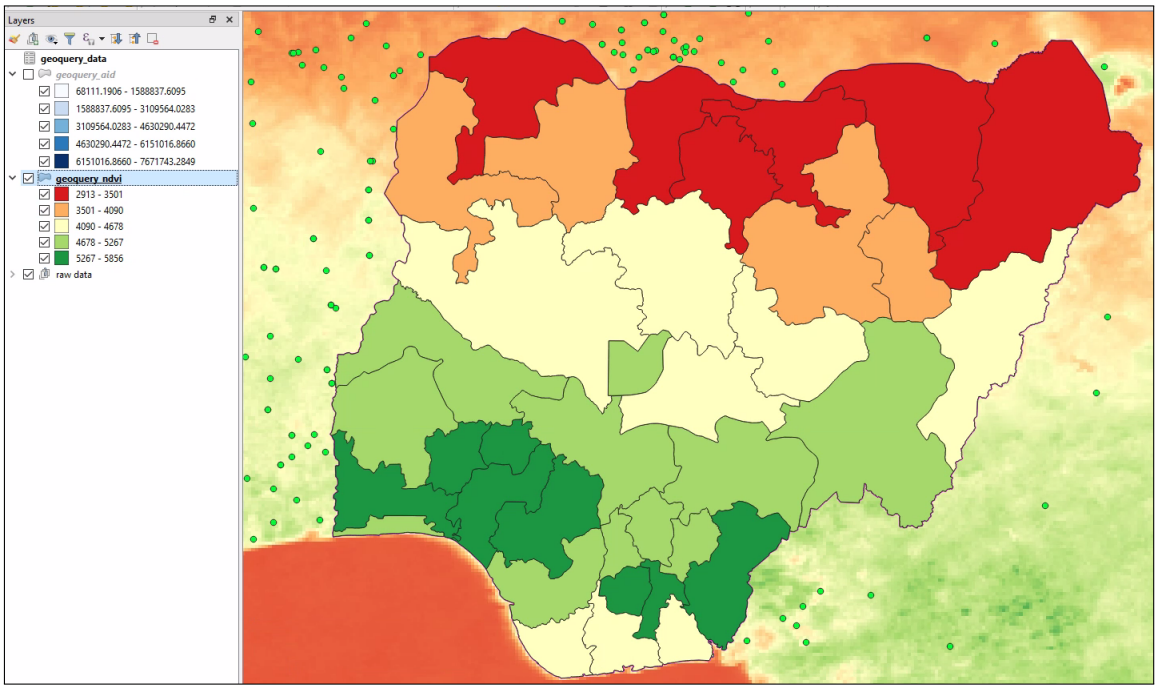


19. To visualize the NDVI data from GeoQuery, we will rename our layer to "geoquery_aid", create a duplicate of the layer, and rename the duplicate to "geoquery_ndvi"



20. Open the layer properties for the "geoquery_ndvi" layer and go to the "Symbology" tab.
21. Change the column to "geoquery_data_ltdr_avhrr_ndvi_v4_yearly.2010.mean"
22. Change the Color ramp to "RdYlGn" by clicking on the dropdown arrow to the right of the color ramp. Then hover over "All Color Ramps" to see a second dropdown with RdYlGn
23. Click "Classify" and then "OK".





For general raster data, GeoQuery should usually provide similar results to the zonal statistics method used in QGIS, but improvements in underlying algorithms often result in at least some difference. Differences may be more noticeable in some cases for a variety of reasons including raster measurement data resolution and the size of boundary features being used.

The methods for aggregating aid data used in GeoQuery are substantially different than the point based method we explored in QGIS. GeoQuery uses codes assigned during the geocoding process to associate each project location with more realistic polygon based geometries. Aid is then dispersed across these polygons rather than being assigned to a single point location.

3. Aid Targeting Exercise

3.0.1 Targeting Aid Using GeoQuery

Here we will show a very basic example of how data from GeoQuery can be manipulated in common spreadsheet format, for targeting or other statistical or analytical purposes, and then loaded back into QGIS to spatially visualize the results.

For this example, instead of examining the raw aggregated values of aid and NDVI we receieved from GeoQuery, let's generate rankings to perform a simple targeting based on what states receive the least amount of relative aid and have the lowest relative NDVI values.

1. Open your GeoQuery CSV in Excel or Libre Calc (**.../Desktop/training_data/5ae89a5bc15e00da98ff0f87/5ae89a5bc15e00da98ff0f87_results.csv**)
2. create two new columns called "aid_rank" and "ndvi_rank"
3. use the **RANK** function to rank the aid aggregations in column B and the ndvi aggregations in column C (e.g., =RANK(B2, B$2:B$33)

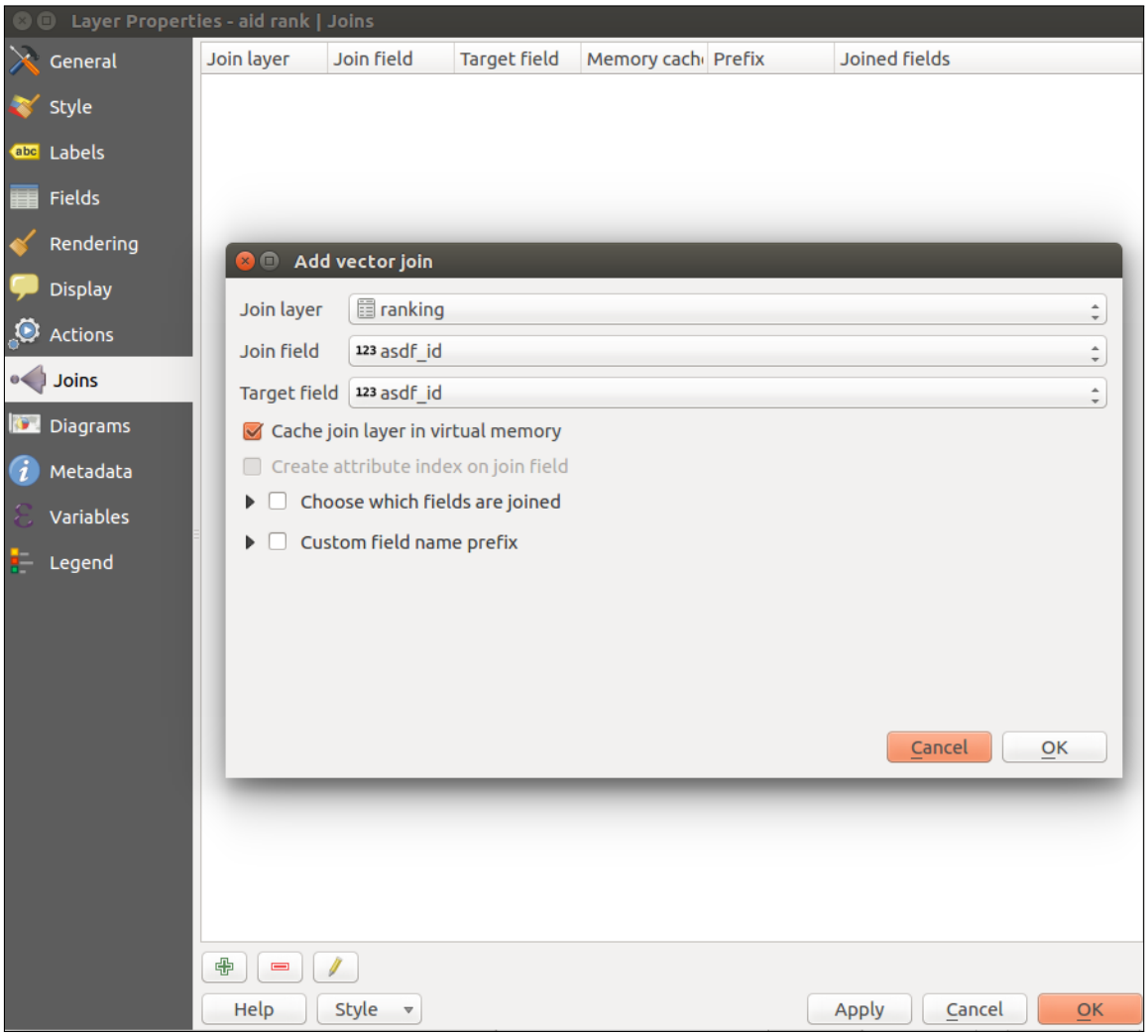
R2 =RANK(B2,B$2:B$33)

	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	CCN_1	NAME_0	Shape_Area	VARNAME_1	ENGTYPE_1	NAME_1	ID_0	HASC_1	CCA_1	ID_1	TYPE_1	Shape_Length	NL_NAME_1	ISO	aid_rank	ndvi_rank
2		Mexico	0.455085		State	Aguascalientes	145	MX.AG		1	Estado	2.805223		MEX	30	26
3		Mexico	6.425989		State	Baja California Sur	145	MX.BS		2	Estado	56.642878		MEX	10	32
4		Mexico	6.935157		State	Baja California	145	MX.BN		3	Estado	30.144752		MEX	9	31
5		Mexico	4.9281		State	Campeche	145	MX.CM		4	Estado	19.770088		MEX	8	1
6		Mexico	6.232582		State	Chiapas	145	MX.CP		5	Estado	18.455846		MEX	5	4
7		Mexico	22.179338		State	Chihuahua	145	MX.CH		6	Estado	24.718163		MEX	2	30
8		Mexico	14.083734		State	Coahuila	145	MX.CA		7	Estado	21.914328		MEX	14	29
9		Mexico	0.479665		State	Colima	145	MX.CL		8	Estado	4.582133		MEX	28	7
10		Mexico	0.137604		Federal District	Distrito Federal	145	MX.DF		9	Distrito Federal	1.6242		MEX	32	25
11		Mexico	11.092258		State	Durango	145	MX.DU		10	Estado	19.119391		MEX	12	23
12		Mexico	2.677437		State	Guanajuato	145	MX.GJ		11	Estado	9.371481		MEX	26	21
13		Mexico	5.532332		State	Guerrero	145	MX.GR		12	Estado	16.703545		MEX	4	10
14		Mexico	1.942748		State	Hidalgo	145	MX.HI		13	Estado	9.344134		MEX	19	12
15		Mexico	6.595676		State	Jalisco	145	MX.JA		14	Estado	24.655214		MEX	11	13
16		Mexico	1.993861		State	M├®xico	145	MX.MX		15	Estado	9.501047		MEX	15	19
17		Mexico	5.082846		State	Michoac├ín	145	MX.MC		16	Estado	16.117549		MEX	7	11
18		Mexico	0.435073		State	Morelos	145	MX.MR		17	Estado	3.081499		MEX	29	16
19		Mexico	2.393142		State	Nayarit	145	MX.NA		18	Estado	15.818823		MEX	22	8
20		Mexico	5.841792		State	Nuevo Le├│n	145	MX.NL		19	Estado	16.610488		MEX	24	20
21		Mexico	7.707076		State	Oaxaca	145	MX.OA		20	Estado	23.746206		MEX	1	9
22		Mexico	2.888997		State	Puebla	145	MX.PU		21	Estado	13.837959		MEX	20	18
23		Mexico	0.999768		State	Quer├®taro	145	MX.QE		22	Estado	5.807059		MEX	27	17
24		Mexico	3.764375		State	Quintana Roo	145	MX.QR		23	Estado	29.526269		MEX	6	3
25		Mexico	5.527162		State	San Luis Potos├¡	145	MX.SL		24	Estado	14.811564		MEX	21	22
26		Mexico	4.780722		State	Sinaloa	145	MX.SI		25	Estado	38.314801		MEX	25	15
27		Mexico	17.256229		State	Sonora	145	MX.SO		26	Estado	38.578136		MEX	17	28
28		Mexico	2.030078		State	Tabasco	145	MX.TB		27	Estado	13.90615		MEX	13	5
29		Mexico	7.013091		State	Tamaulipas	145	MX.TM		28	Estado	49.011123		MEX	23	14
30		Mexico	0.358126		State	Tlaxcala	145	MX.TL		29	Estado	2.938221		MEX	31	24
31		Mexico	5.931054		State	Veracruz	145	MX.VE		30	Estado	32.873701		MEX	3	6
32		Mexico	3.295681		State	Yucat├ín	145	MX.YU		31	Estado	13.469015		MEX	16	2
33		Mexico	6.690325		State	Zacatecas	145	MX.ZA		32	Estado	21.147914		MEX	18	27
34																

4. Save this as a CSV file (NOT an .xls file) in your workspace as "ranking.csv"

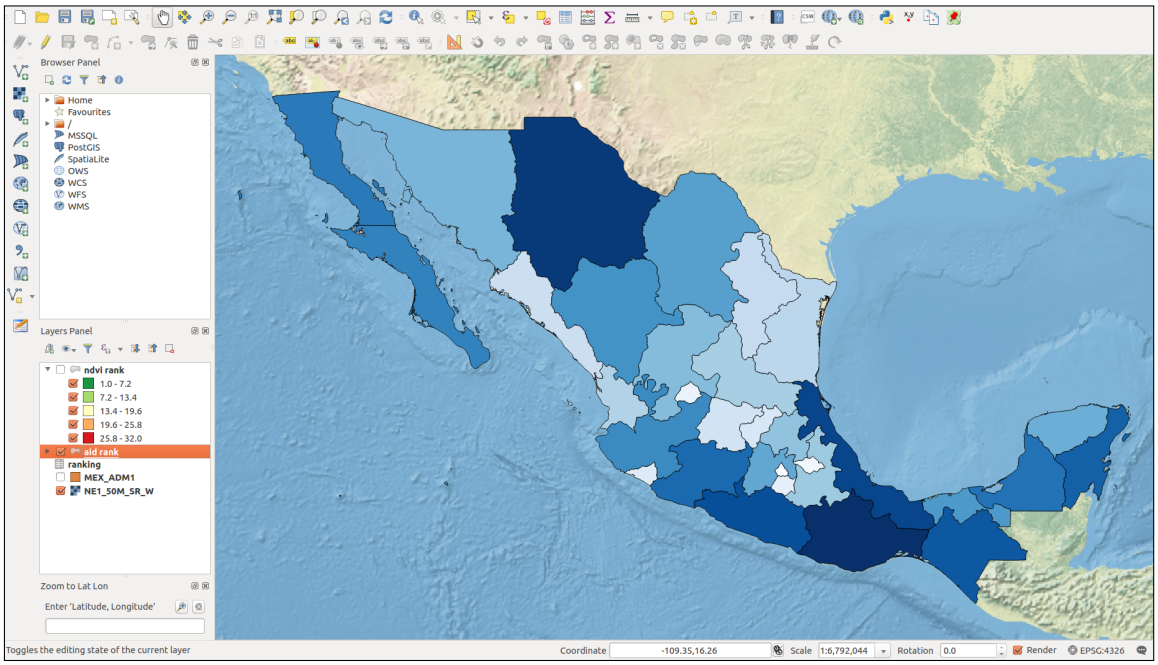
Now that we have our ranking data, we will join it back to our states like before. Then we can generate a map for each set of rankings which we can visually compare.

1. Add the .csv file by click "Add Delimited Text Layer" on the left toolbar.
2. Click "Browse" to add the **.../Desktop/training_data/workspace/ranking.csv**
3. Since this .csv is an attribute only file, make sure to select "No geometry" in the Geometry definition.
4. Click "OK" to add this layer.
5. Create another copy of our "NGA_ADM1" layer and rename it to "aid rank"
6. Go to Properties > Join and click "+" sign on the bottom.
7. Choose "ranking" as "Join Layer".
8. Choose "asdf_id" for "Join Field" and "Target Field".
9. Click OK



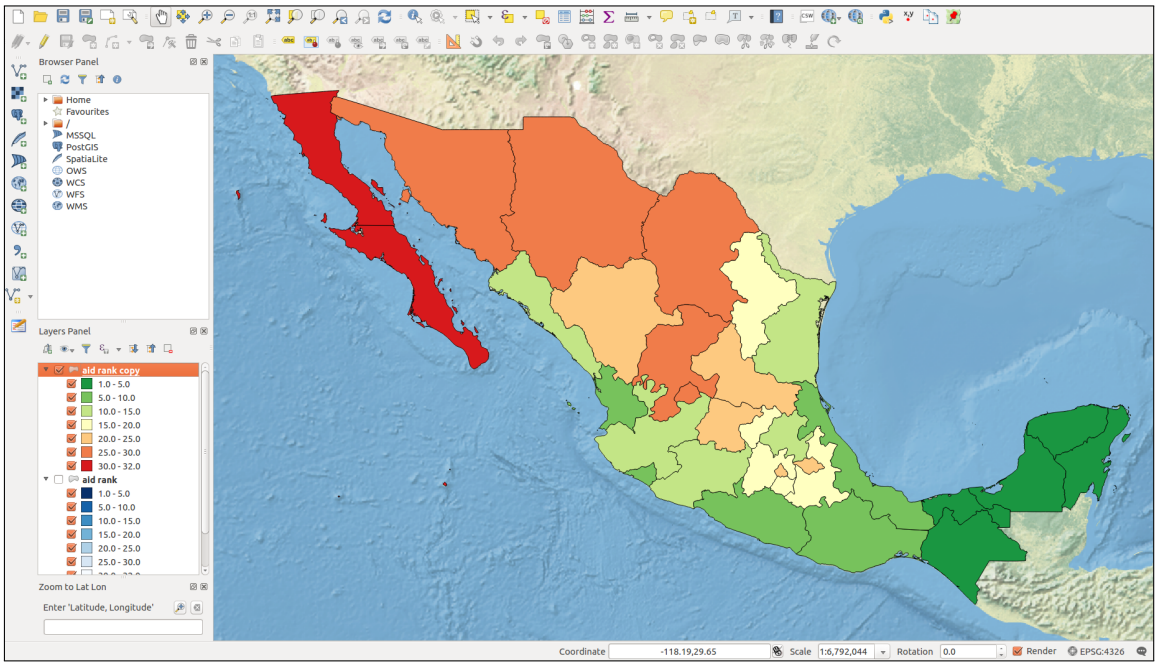
To visualize aid ranks:

1. Open the "aid rank" layer properties.
2. Navigate to Style tab, change the top option to "Graduated".
3. Select "ranking_aid_rank" as the column.
4. Select the "invert" option next to color ramp since lower ranks indicate more aid
5. Choose "Pretty Breaks" in classification mode.
6. Change the number of classes to 10
7. Click "Classify".
8. Click "OK".



And to visualize NDVI ranks:

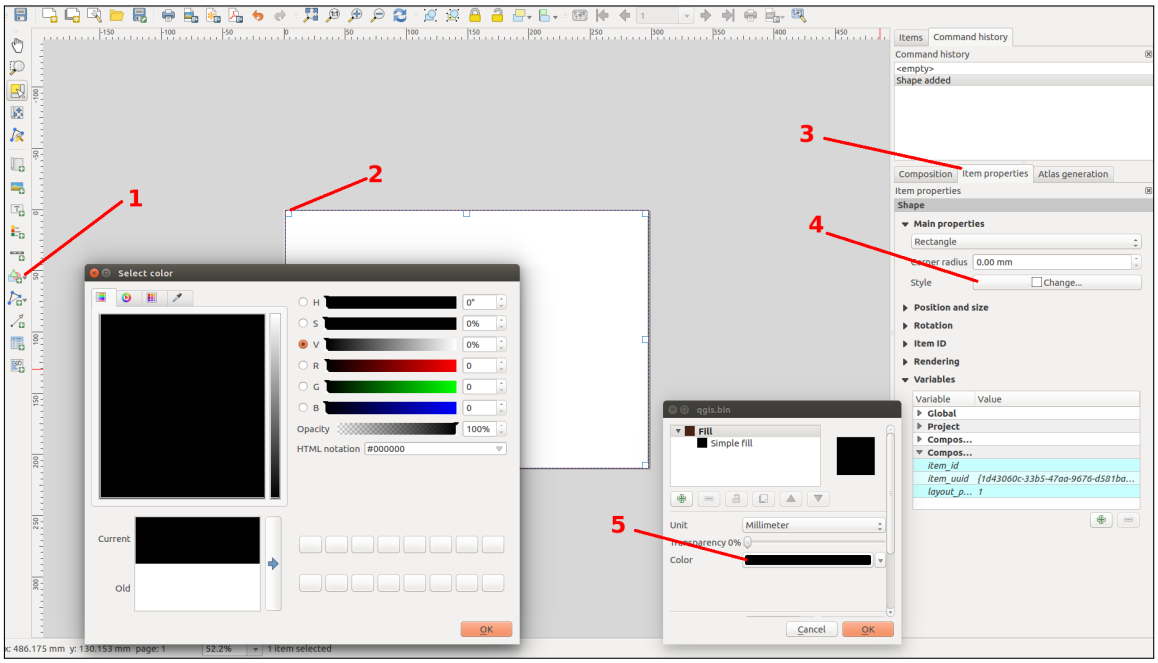
1. Make a copy of the "aid rank" layer
2. Rename this layer to "ndvi rank"
3. Open the layer properties.
4. Navigate to Style tab, change the top dropdown to "Graduated".
5. Select "ranking_ndvi_rank" as the column.
6. Set the color ramp to "RdYlGn"
7. Select the "invert" option next to color ramp since lower ranks indicate large NDVI values
8. Choose "Pretty Breaks" in classification mode.
9. Change the number of classes to 10
10. Click "Classify".
11. Click "OK".



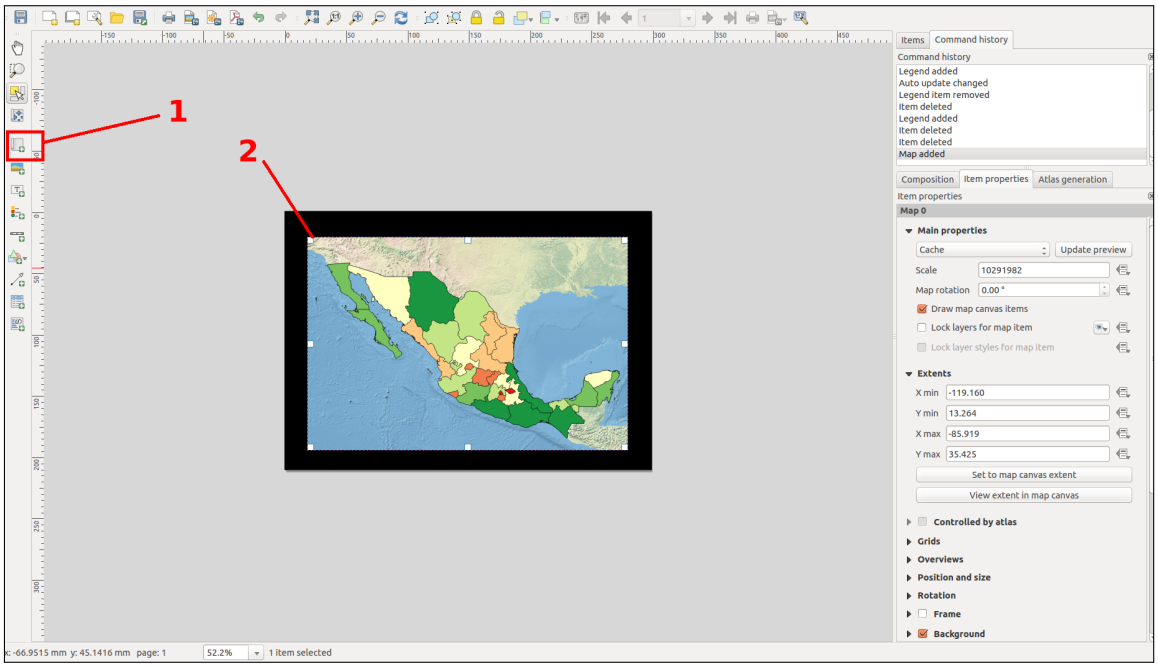
3.0.2 Exporting Maps

As a final step, we can add a legend to these maps and export them to a PDF using the QGIS Print Composer tool. Let's start with the NDVI rank map. Before opening up the Print Composer, make sure the current QGIS window displays the spatial data you want to print out, as the Print Composer captures what is currently visible.

- Click the Project menu on the main toolbar along the top, and select "New Print Composer"
- Assign a title - "ndvi rank map" - and click OK
- First, let's add a border for our map
  - Step 1- Click the **Add Shape** tool on the left
  - Step 2- Draw a rectangle over the print area that we want our border to cover
  - Step 3- Click on the **Item properties** menu on the right
  - Step 4- Click **Change** for the Style options
  - Step 5- Click on the Color field and change to black
  - Step 6- Click OK on the color and style windows

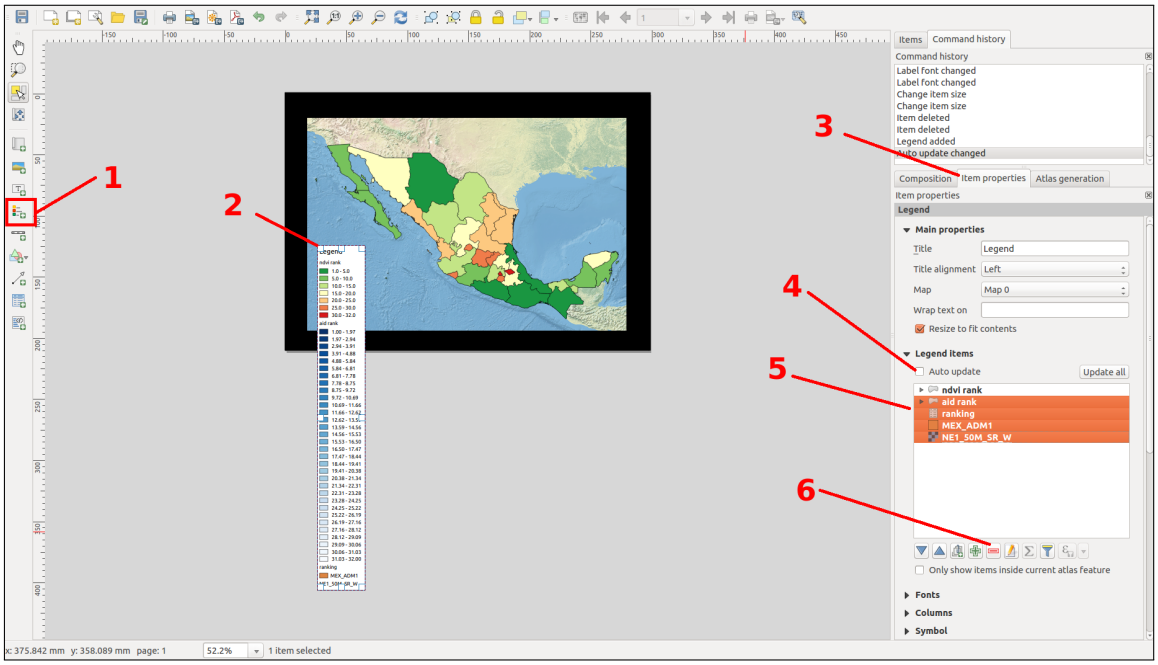


- Next, we will add our map
  - Step 1- Click the **Add Map** tool on the left
  - Step 2- Draw a rectangle over the area you want the map to cover

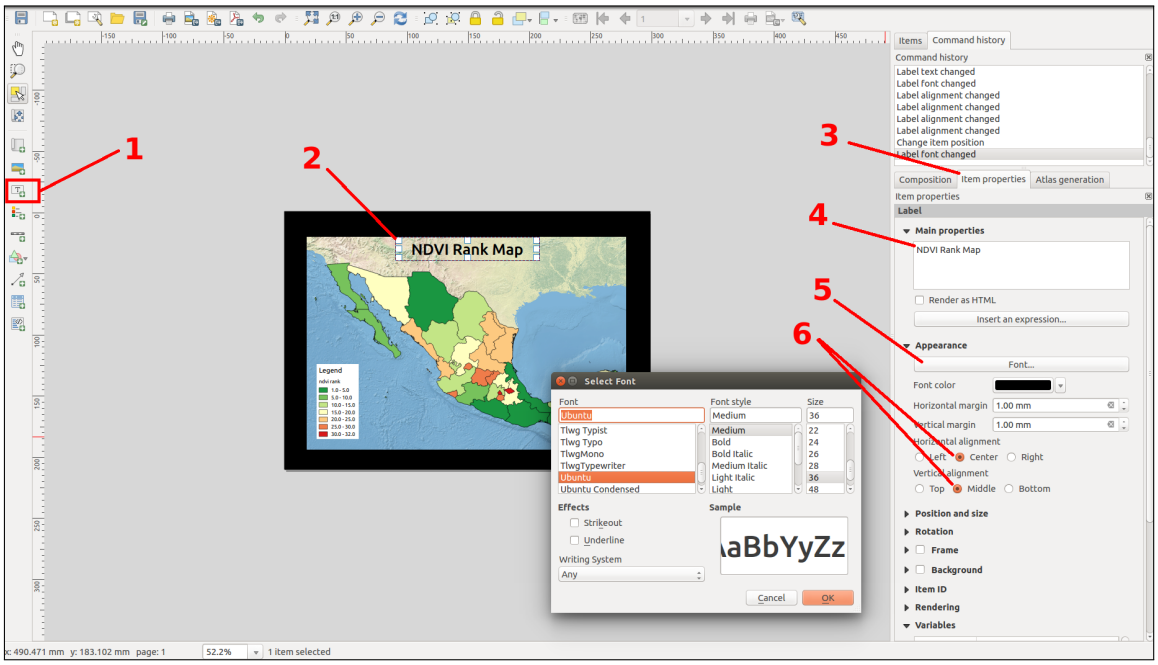


- Let's also add a legend
  - Step 1- Click the **Add Legend** tool on the left
  - Step 2- Draw a rectangle over the area you want the legend to cover
  - Step 3- Select the **Item properties** menu

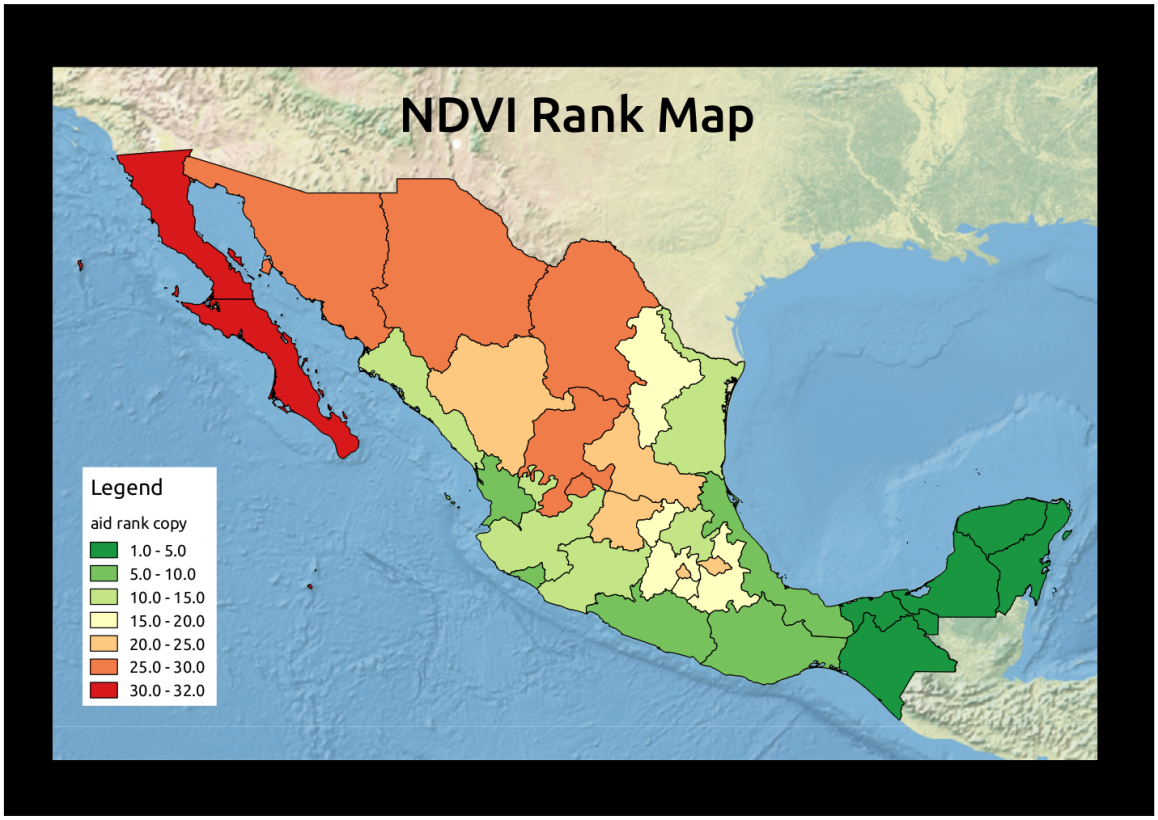
- Step 4- Turn off **Auto update**
- Step 5- Select all the legend entries except for **ndvi rank**
- Step 6- Click the red minus arrow to remove the unwanted legend entries



- Finally, let's add a title
  - Step 1- Click the **Add Label** tool on the left
  - Step 2- Draw a rectangle over the area you want the title to cover
  - Step 3- Select the **Item properties** menu
  - Step 4- Add a title to the text field
  - Step 5- Click **Font** to adjust the font type and size
  - Step 6- Vertically and horizontall center the text

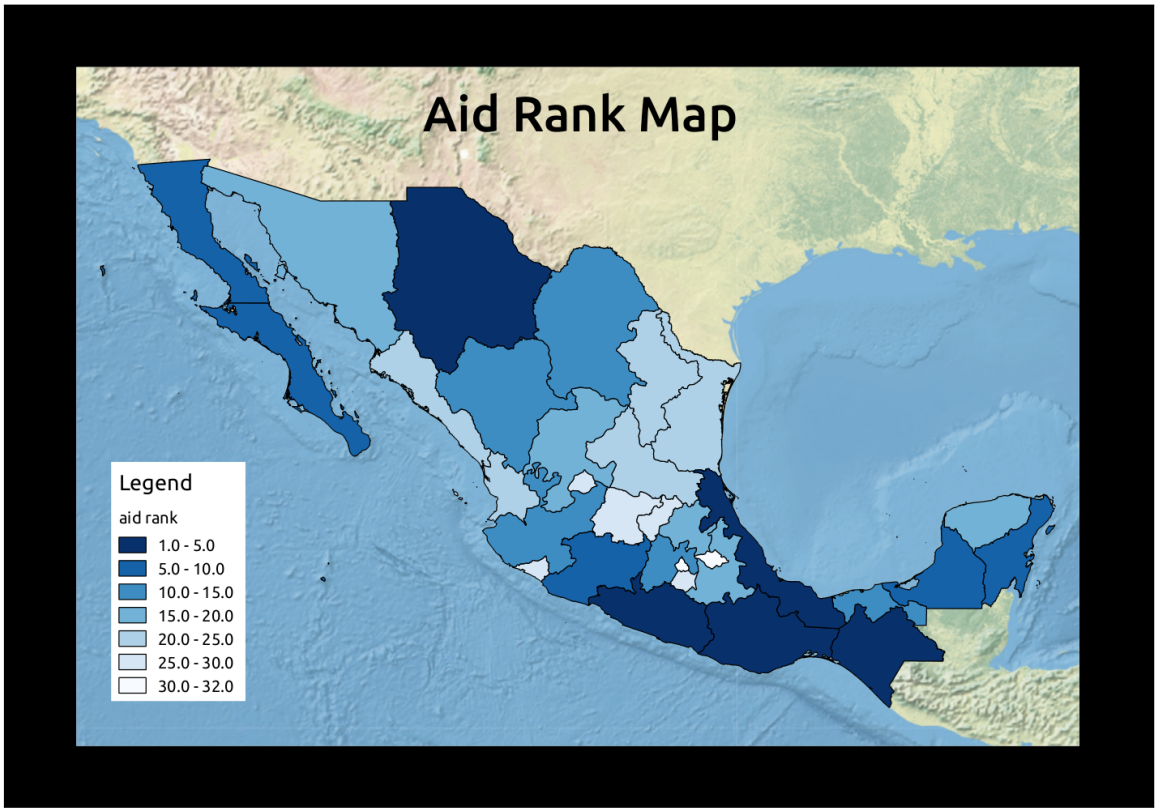


- To print the map to a PDF, use the **Export as PDF** button on the upper horizontal toolbar and select a location to save to.



Now you can repeat these steps for the aid rank layer. First, make sure to go back into the main

QGIS window and toggle off the "ndvi rank" layer and toggle on the "aid rank" layer - remember the Print Composer uses what is visible on the map.



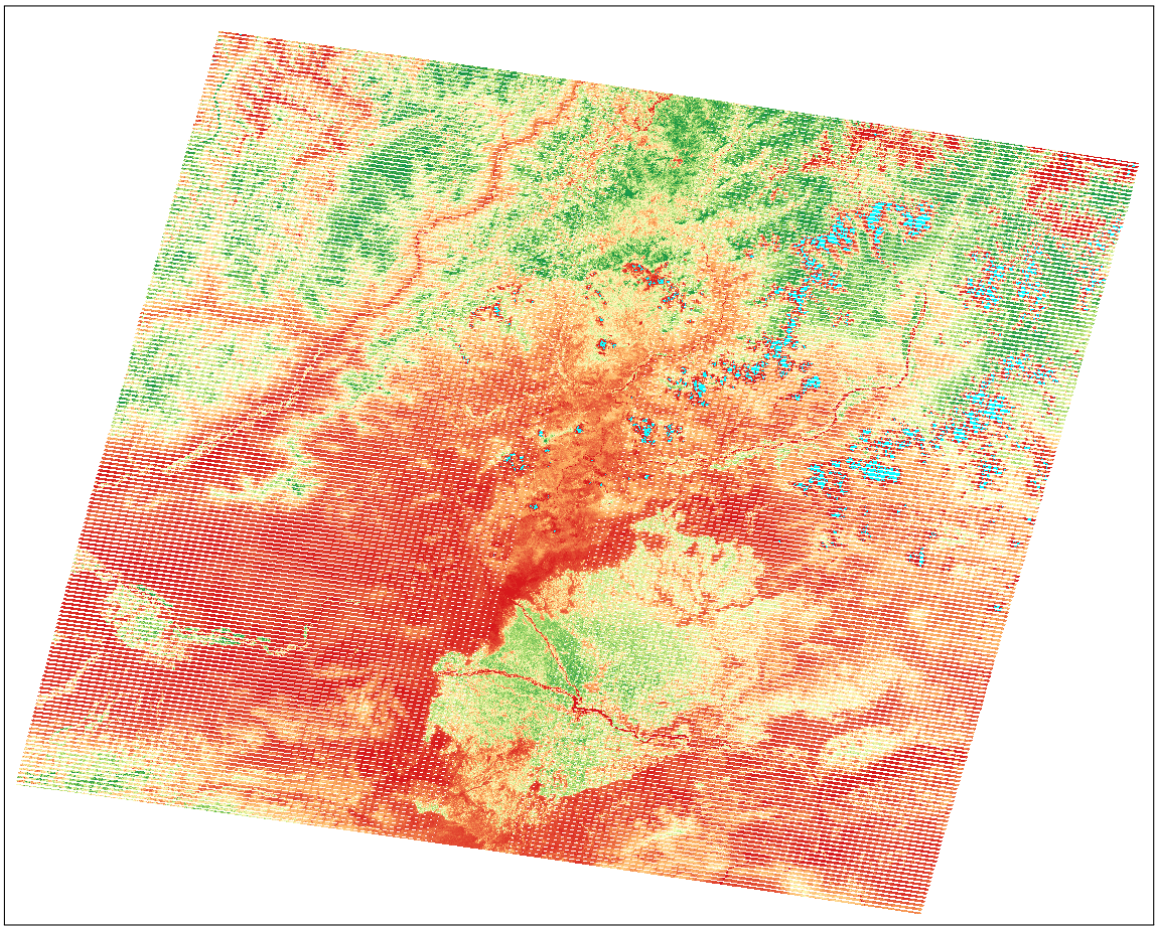
4. Exploring Landsat 7 Satellite Data

4.1 NDVI

Here is an example of an NDVI scene derived from Landsat 7, acquired from the Bulk Download and Processing tool which accompanies Earth Explorer

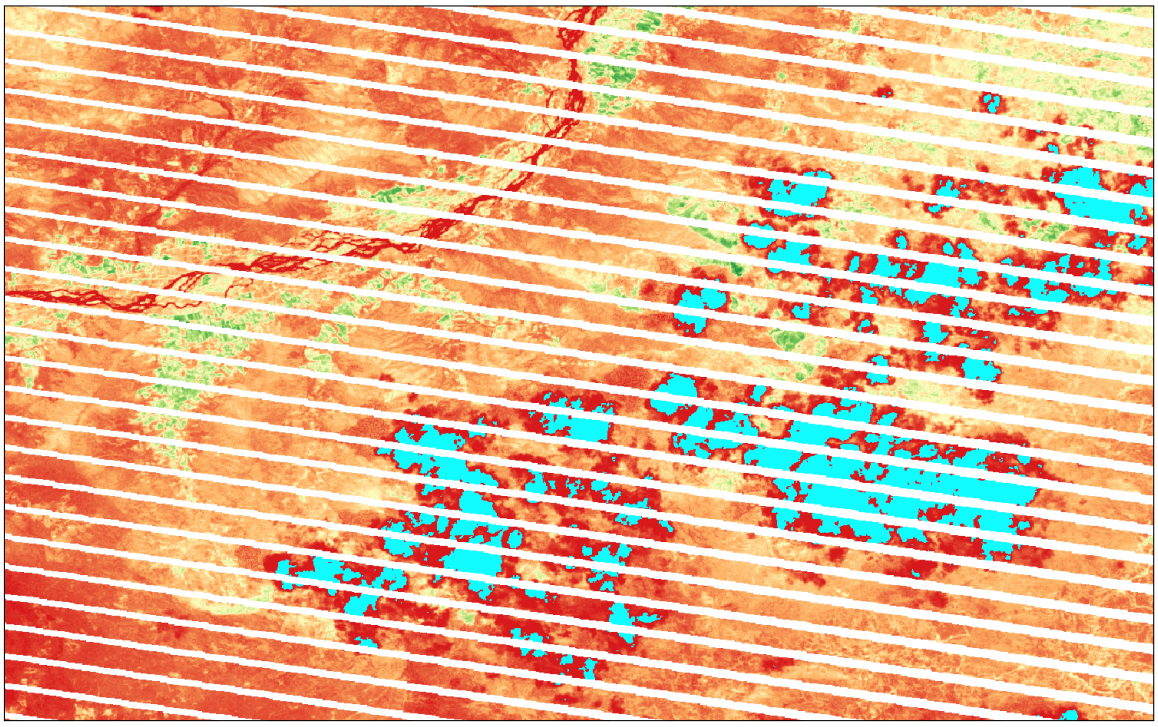
1. Go back to QGIS
2. Click the "Add Raster Layer" tool
3. Select **.../Desktop/training_data/ndvi_data/ndvi_example/ndvi.tif**
4. Right click on the new layer in the "Layer Panel" and use "Zoom to Layer"

In this scene of NDVI for an area of Afghanistan, we use a Red Yellow Green color ramp to indicate low > high NDVI values.



If you zoom in and explore this scene, you can notice two colors not on the Red Yellow Green scale, which represent types of problems we encounter with satellite imagery.

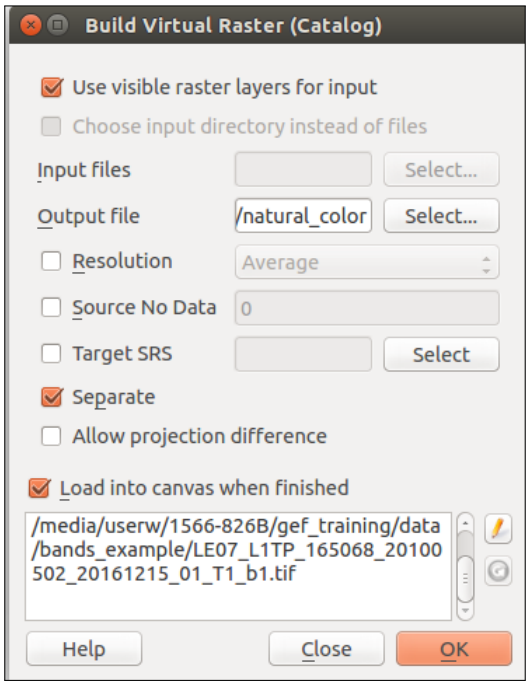
1. White - this represents no data. In this case there a bands of no data across the entire image. This is due to a failure in the Scan Line Corrector (SLC) on-board Landsat 7 which impacts all images after the failure (2003). While this specific issue is not common, it is indicative of various issues and quirks that need to be accounted for and a quite common in satellite imagery.
2. Light Blue - this represent areas which were covered by clouds. Cloud cover is an ever present issue when using satellite imagery. Solutions include dropping scenes from analysis, ignoring or masking cloud pixels, and aggregating multiple scenes.



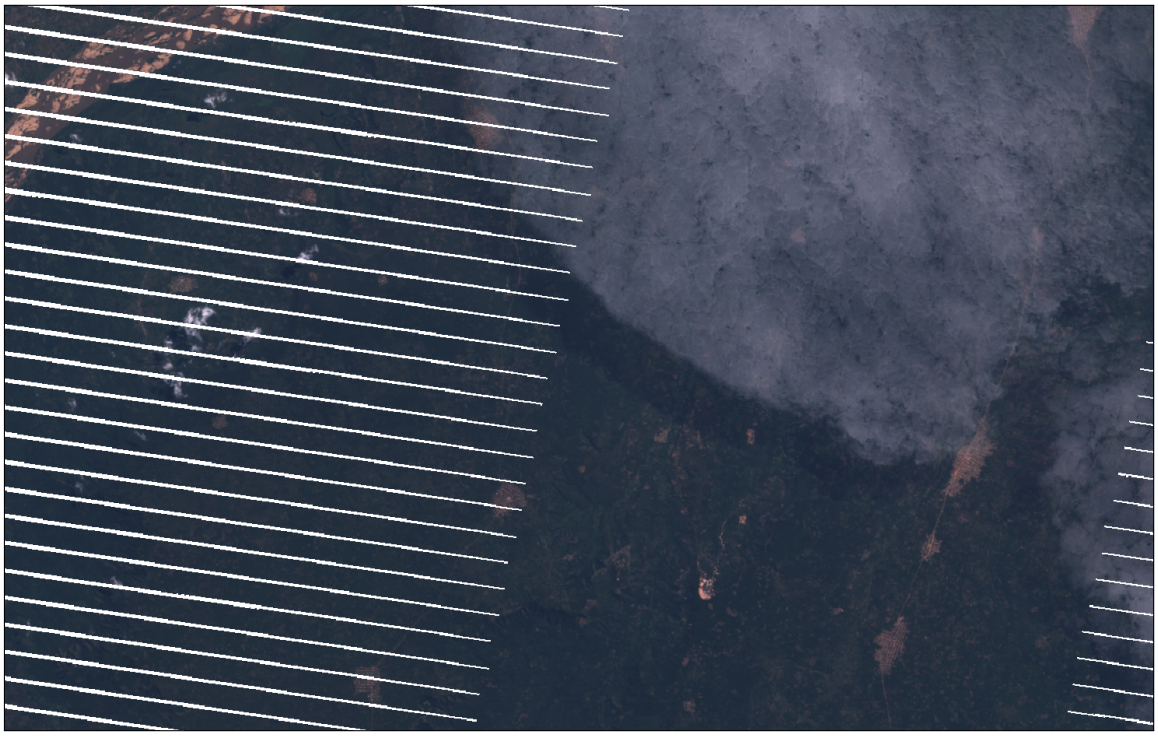
4.2 RGB Bands

Next we will look at some individual bands for a scene of Landsat imagery, and use QGIS to composite them into a natural color (Red Green Blue) image.

1. Click the "Add Raster Layer" tool
2. Select the 3 layers ending in **b1, b2, and b3** within **.../Desktop/training_data/bands_example** to load individual bands for a scene
3. Turn off the visibility of any layers in QGIS except for these 3 layers
4. Go to "Raster > Miscellaneous > Build Virtual Raster"
5. Select "Use visible raster layers for input"
6. Select a path for "Output file" (e.g., **.../Desktop/training_data/workspace/natural_color.vrt**
7. Select the "Separate" option (this will put each input file into a corresponding RGB band, otherwise they will be merged into a single band)
8. Select "Load into canvas when finished"
9. Click "OK"
10. Right click on the new layer in the "Layer Panel" and use "Zoom to Layer"



You will notice the same SLC issue, and will also be able to see numerous clouds across the image.



References for this example:

1. `https://landsat.usgs.gov/how-do-landsat-8-band-combinations-differ-landsat-7-or-landsat-5-satellite-data`

2. https://gis.stackexchange.com/questions/185064/how-to-make-band-composite-image-in-qgis?utm_medium=organic&utm_source=google_rich_qa&utm_campaign=google_rich_qa

5. Data Sources

5.1 Boundary Data

- Global Administrative Boundaries
  - https://gadm.org/
  - http://geoquery.org/geoboundaries/
- Global Protected Areas
  - https://www.protectedplanet.net/
- US Census Boundaries
  - https://www.census.gov/geo/maps-data/data/tiger-cart-boundary.html
- Humanitarian Data Exchange (HDX) - Mali Example - offers broad range of development oriented data
  - https://data.humdata.org/dataset/administrative-boundaries-cod-mli
- World Food Programme - range of spatial data available
  - https://geonode.wfp.org/

5.2 Satellite and Measurement Data

- USGS Earth Explorer
  - Excellent tool for acquiring Landsat scene as well as a range of other satellite data from government programs
  - Bulk download tool also available to use with scene lists acquired from Earth Explorer
  - https://earthexplorer.usgs.gov/
- ESA Land Cover
  - Example of data viz portal along with manual download of zip containing all files for dataset
  - http://maps.elie.ucl.ac.be/CCI/viewer/index.php
- Nighttime lights
  - Example of satellite collection a type of data changing
  - Example of standardized, public, web scrapable links
  - https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html
  - https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html
- SEDAC - Socioeconomic Data and Applications Center
  - Example of curated data portal
  - Login required
  - Manual downloads
  - http://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-adjusted-to-2015-unwpp-country-totals-rev10
- LTDR NDVI
  - Example of updated data changing download format
  - https://ltdr.modaps.eosdis.nasa.gov/cgi-bin/ltdr/ltdrPage.cgi?fileName=products
  - https://ladsweb.modaps.eosdis.nasa.gov/archive/

6. Appendix

6.1 Key Terms and Definitions

- **Vector** - Vector data are comprised of vertices and paths (coordinate points and the lines connecting them. The three basic types of vector data are points, lines and polygons (areas).[1]
- **Raster** - Raster data is made up of pixels (also referred to as grid cells). They are usually regularly-spaced and square but they don't have to be. Rasters often look pixelated because each pixel has its own value or class.[2]
- **NDVI** - Normalized Difference Vegetation Index quantifies vegetation by measuring the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs). Strong positive values are considered an indicator of "greenness" of vegatation, or vegatative density.[3]
- **Histogram** - A histogram is a plot that lets you discover, and show, the underlying frequency distribution (shape) of a set of continuous data.
- **Administrative boundaries** - These are government based boundaries such as country borders, states, and districts. Administrative boundaries, or ADMs, are often labeled as ADM0 (country), ADM1 (next finest boundary, e.g., state), ADM2 (e.g., district), and so on (based on the country).
- **Geocoding** - This is the process of assigning a geospatial reference (i.e., a vector point, line, polygon) to a placename or description of an area (e.g., the hospital building 100 meters south of the river, next to the park)
- **Zonal Statistics** - This is the process of aggregating the values from raster data (pixels) to vector data. For example, a city represent by a circle (vector data) may overlap/contain 100 pixels representing temperature (raster data). Taking the average value of those pixels would produce a mean estimate of temperature in the city. [4]

[1] https://gisgeography.com/spatial-data-types-vector-raster/
[2] https://gisgeography.com/spatial-data-types-vector-raster/
[3] https://gisgeography.com/ndvi-normalized-difference-vegetation-index/
[4] http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/h-how-zonal-statistics-works.htm