# 1. Introduction

## 1.1 Overview

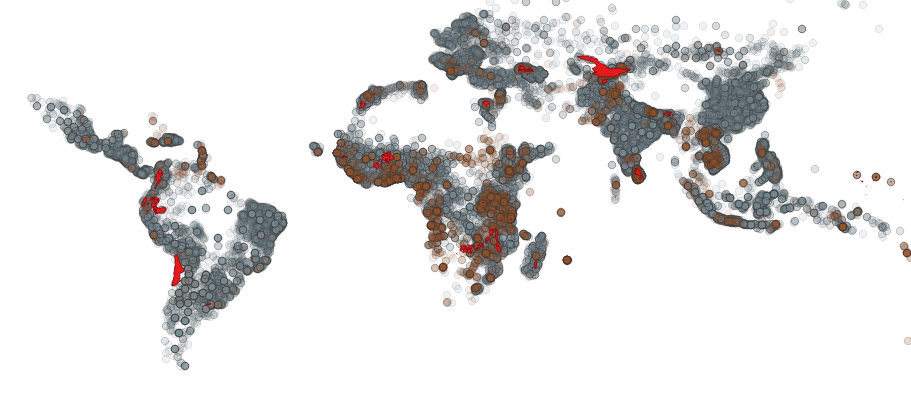
This session will focus on identifying hotspots of aid activity, based on information across multiple donors. Further, it will explore different approaches to estimating where aid projects have been allocated under conditions of spatial uncertainty (i.e., the exact location of aid is not recorded). Weighting based on nighttime lights information will be used as an example case.

For this lab, we will be using information on the full GCF portfolio, known locations of Chinese aid, and World Bank aid information. Because all sources of data have considerable geospatial uncertainty, we will further examine different strategies for use of this data.

## 1.2 Retrieving Data

First, we’ll be downloading a few datasets to facilitate each step of this lab. I have downloaded all of the below into a new folder I created called “0\_raw”. As we retrieve this data, we will also load it into QGIS.

(**NOTE**: If, for any reason, the interfaces and files below are not accessible, all files are provided in the accompanying data package for this session available online at <http://gcf.geoquery.org>).

1. The first data source has been provided – the full GCF portfolio. You can open this in QGIS by adding the shapefile “GCF\_Projects”.
2. Second, we will retrieve information on two other Donor’s projects – China and the World Bank.
   1. For **Chinese Aid:** First, go to <https://www.aiddata.org/data/geocoded-chinese-global-official-finance-dataset> and download the Global Chinese Official Finance dataset. Extract this file into your 0\_raw folder, and in QGIS add the “all\_flow\_classes.csv” data by using the CSV (delimited text) importer (denoted by  on the data loading menu). Ensure your geometry definition CRS is set to WGS 84 (this should be the default), then click Add.
   2. For **The World Bank**: First, go to <https://www.aiddata.org/data/world-bank-geocoded-research-release-level-1-v1-4-2> and download the World Bank Geocoded release. Extract this file into your 0\_raw folder, and in QGIS from the “data” folder add the “level\_1a.csv” file using the delimited text importer.
3. Briefly, we will visualize the data so we can see what’s going on. For both the level\_1a layer (World Bank projects) and the all\_flow\_classes layer (Chinese Aid), se their transparency to approximately 50%. Your resultant map should look something like this:  
   
4. In my case, GCF project locations are highlighted in red; Chinese in brown, and World Bank in grey. One key note is in the data representation – in this case, GCF projects have much more detailed representations than either World Bank or Chinese aid projects. Because – for example – Chinese Aid was geocoded using very limited information, only a single point approximating the rough location of where aid was allocated exists for each case. Contrasted to this, for the GCF a specific polygon represents where the Aid was allocated. A single point in the case of Chinese Aid can represent Aid that is known to – for example – have been allocated somewhere within the country the point is closest to!

## 1.3 Data Cleanup and Prep

1. Let’s filter the data so that (1) we are only analyzing high-precision projects, and (2) we’re only examining environmentally-focused aid projects.
2. First, right click on “all\_flow\_classes” (Chinese Aid) and click the Attribute Table. You’ll note there is a column – Precision Code – that has number in it ranging from 1 to 6. Precision Code’s of 1 indicate the point was placed near a named place (i.e., a capital city) in which Aid was known to be allocated. A Precision Code of 6 or greater would indicate it was placed somewhere near the correct country.
3. For our purposes, we’re only going to focus on aid that was both (a) known to have been allocated within or near a city (Precision code 1), and (b) for environmental purposes. To do this, we’re going to use the “select features based on an expression” option in the attribute table () – go ahead and click on it.
4. In this new window, you’ll see three columns. In the second column, expand the “Fields and Values” menu to see all of the different criteria you can use for searching.
5. In our case, we’re going to use the following expression, which we’ll type into the far left text box – note we’ll be intentionally casting a broad net, as it is rare that Chinese Aid is specifically allocated to environmental protection efforts (rather, it is built in to other types of projects, but left unspecified). As an analyst, you could instead choose to search only for explicit “Environmental” projects, but would likely only find 10 or less!:

"precision\_code" = 1

AND

(

regexp\_match( "ad\_sector\_names",'General Environmental Protection')

OR

regexp\_match( "ad\_sector\_names",'Other Multisector')

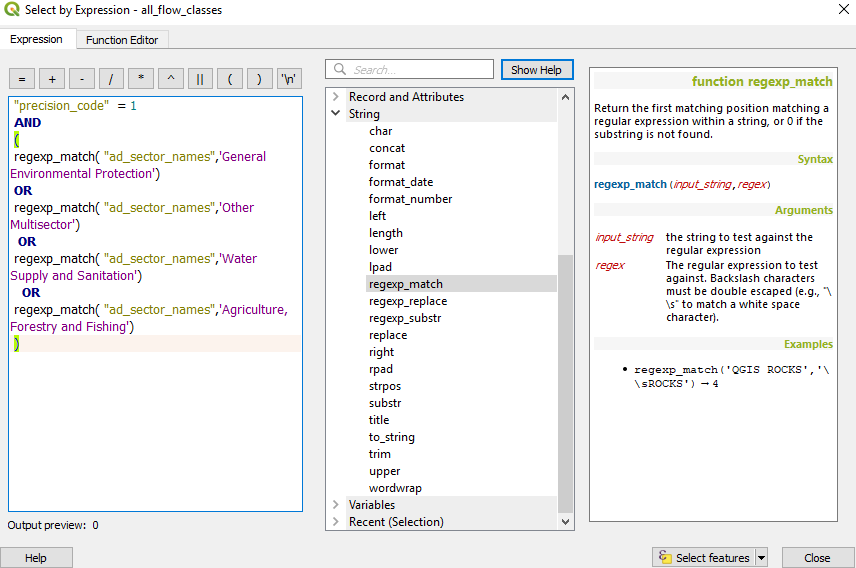
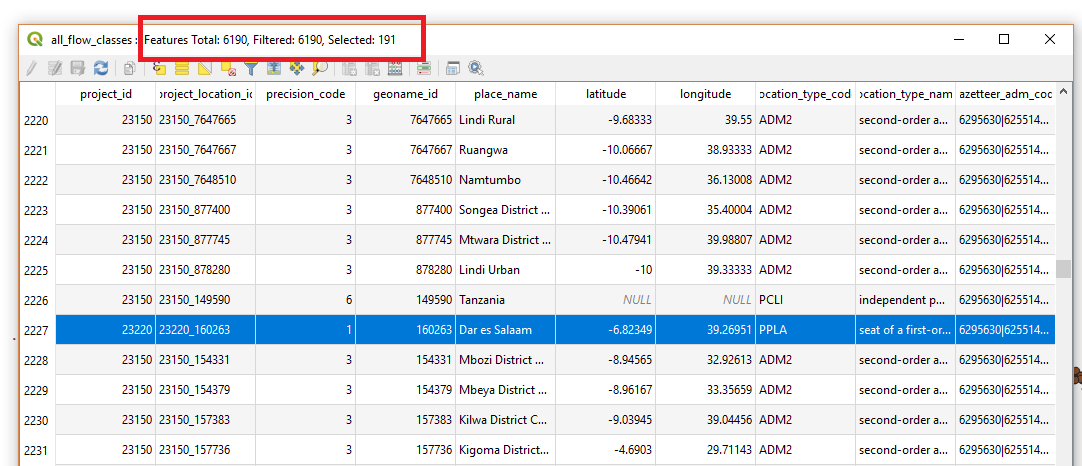
OR

regexp\_match( "ad\_sector\_names",'Water Supply and Sanitation')

OR

regexp\_match( "ad\_sector\_names",'Agriculture, Forestry and Fishing')

)

1. Go ahead and click “Select features” at the lower-right. Your window should look something like this when you do it:  
   
2. Close the select feature menu, and look back at the attribute table. At the very top (in the title) it will tell you how many features you have selected. In our case, we have identified 191 projects of all 6,190 that met our criteria – if you scroll up and down on the attribute table, you’ll see some of these are highlighted:  
   
3. Close the attribute table, and then right click on “all\_flow\_classes” in the layers. Click Export, then “Save As”. Choose a ESRI shapefile format, make sure to check “save only selected features”, and then name the file “Chinese\_highres\_environment.shp”. I saved my new file in a new folder, which I called “1\_envAid”. Go ahead and click OK, and then remove the old “all\_flow\_classes” map from your layers.
4. Now, let’s repeat the process for World Bank aid. Right click on level\_1a, and choose the attribute table. Click on the same “Select Features using an Expression” button () as before.
5. World Bank aid is slightly different from Chinese Aid, as it can be grouped into multiple categories (i.e., it can be “General Environmental Protection” AND “Water Supply and Sanitation” at the same time). Because of this, our expression from before – which sought to match exact words – won’t work. Instead, we need to search for specific words; further, because the World Bank always explicitly notes Environmental aid, we don’t need to search for other types of aid that is likely to contain it, so our expression can be simpler.
6. To do this, we’ll use the ILIKE operator – a tool that searches for matches. Specifically, our new code will look like:

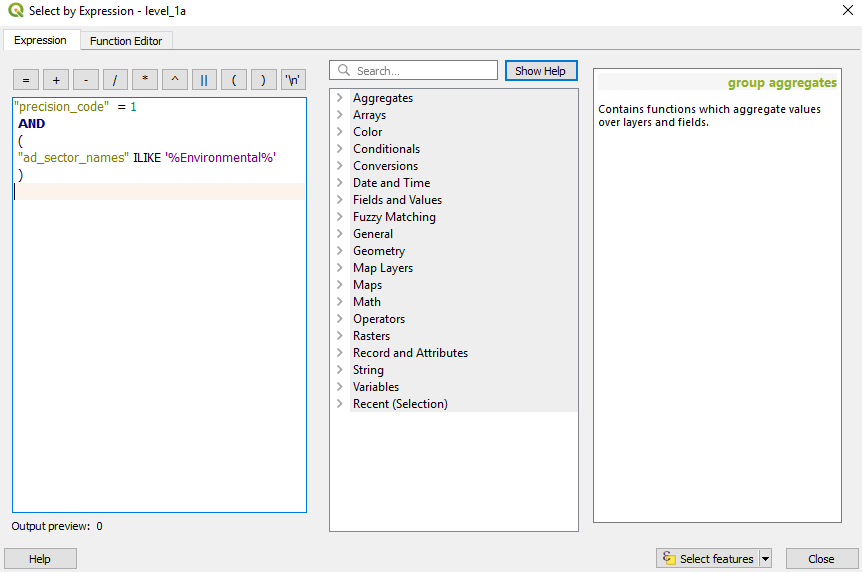
"precision\_code" = 1

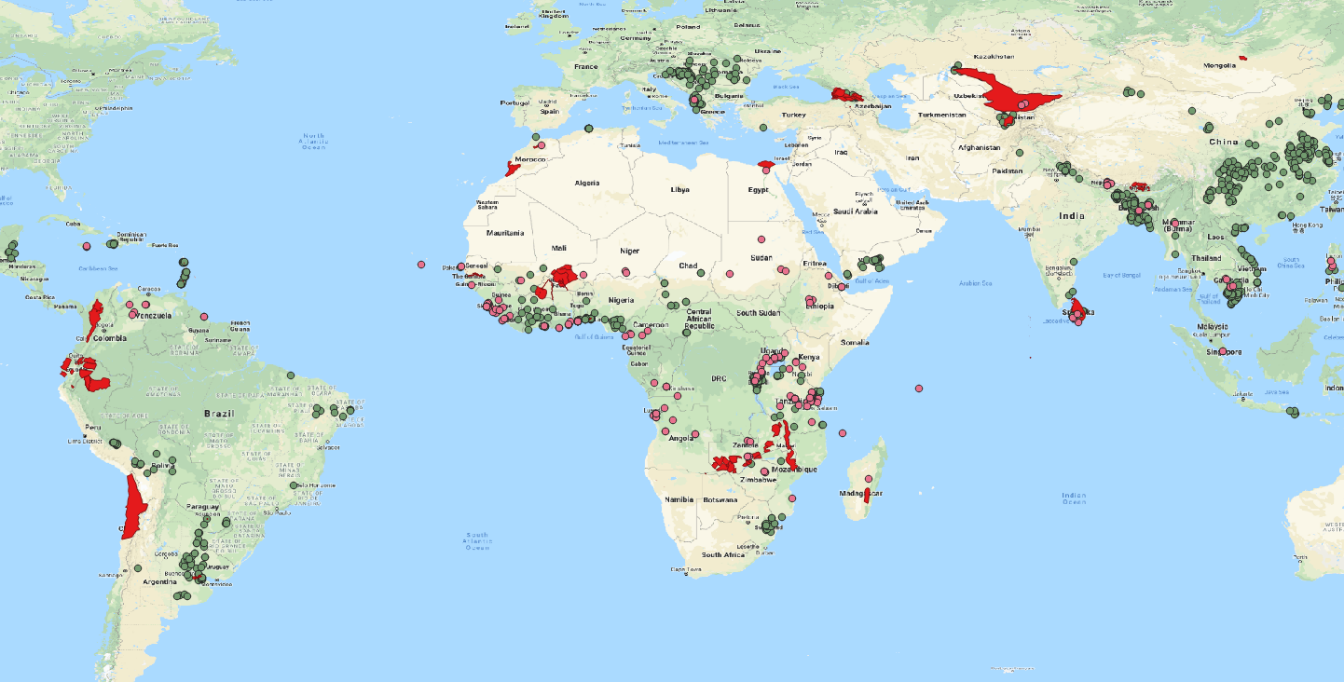
AND

(

"ad\_sector\_names" ILIKE '%Environmental%'

)

Go ahead and click “Select Features”, and then close out of the expression editor:  


1. Take a look at the resultant selection – you’ll see that 1,038 features of our original 61,243 are selected.
2. Close the attribute table, and then right click on “level\_1a” in the layers. Click Export, then “Save As”. Choose a ESRI shapefile format, make sure to check “save only selected features”, and then name the file “WorldBank\_highres\_environment.shp”. I saved my new file in a new folder, which I called “1\_envAid”. Go ahead and click OK, and then remove the old “level\_1a” map from your layers.
3. After all of this filtering, you should be left with a map of only Chinese and World Bank projects where we believe there is some form of environmental aid being allocated. After quickly adding a Google Maps background using the Quick Map Service we’ve used in previous sessions, my map looks like this (with World Bank projects as green dots, Chinese as pink, and GCF locations as red polygons):  
   

**1.4. Hotspot Analysis**

1. In this section, we’re going to explore what areas of the world receive the most attention from both (a) Chinese Investments, (b) World Bank Projects, and (c) Both China and the World Bank. We’ll then visually compare these to the locations of GCF projects.

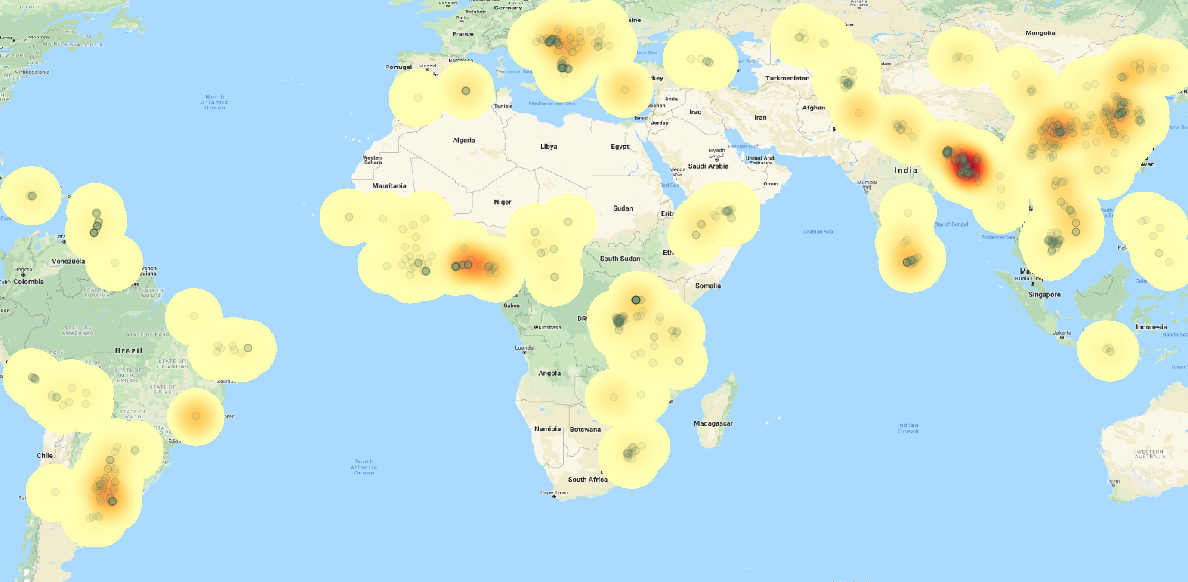
2. First, open the Processing Toolbox (At the top of QGIS, Processing -> Toolbox), and then at the top search for “Kernel Density Estimation” (KDE). KDE is the tool we’ll be using to create heatmaps of each donor’s activities.

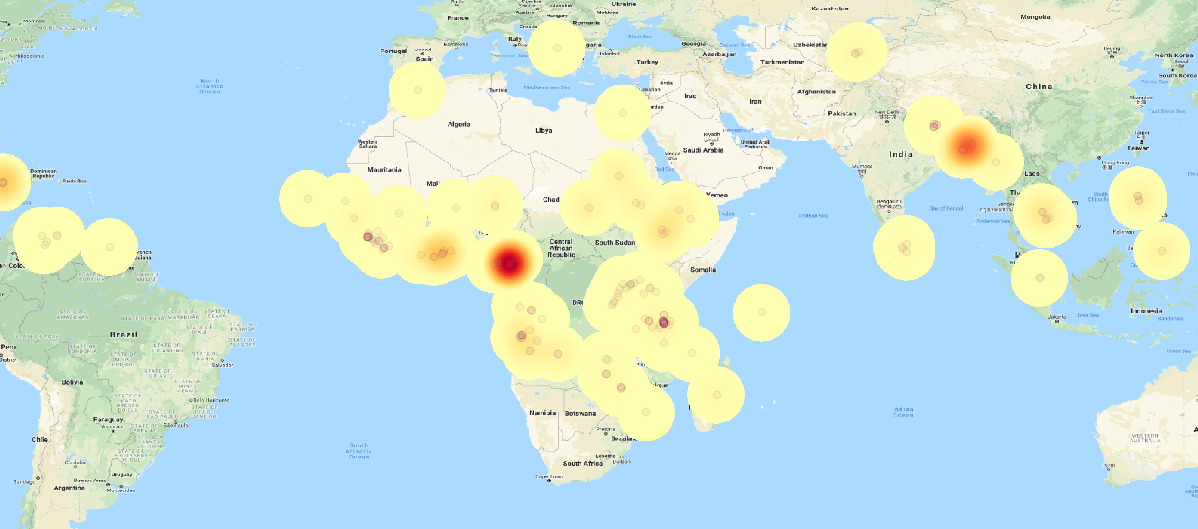
3. Double click on the KDE in the Toolbox.

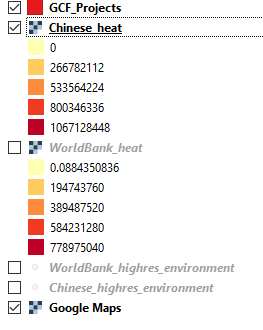
4. Once open, set the following options: (A) Choose the World Bank as your dataset; (B) Chance the Radius to “5.0”, and choose a name to save your new heatmap as (I saved mine into a new folder, 2\_heatmaps, and named my first one “WorldBank\_heat.tif). Importantly, we now want to weight our heatmap by dollars committed. Because not all Bank projects have the same dollar values, we need to change how we create our heat map to accommodate different funding levels. Click the “Advanced Parameters” option, and then choose “Weight from Field”. In this dropdown, choose “even\_split” – representing the estimated committed funds at each location. Finally, click Run. It may take a few minutes.

5. Add the new data layer you just produced to your map – note you may need to remove a temporary layer called “heatmap” depending on your operating system.

6. You should now see a heat map! Let’s make it a bit more interpretable – right click on WorldBank\_heat in your layers, go to the properties and symbology. Change the render type to singleband pseudocolor, and choose the YlOrRd (Yellow Orange Red) color ramp. Click classify and apply – your map should now look something like this, with darker red areas indicating higher levels of World Bank activity, and yellow colors indicating lower levels. No color indicates no nearby activity in the Environmental Sector:



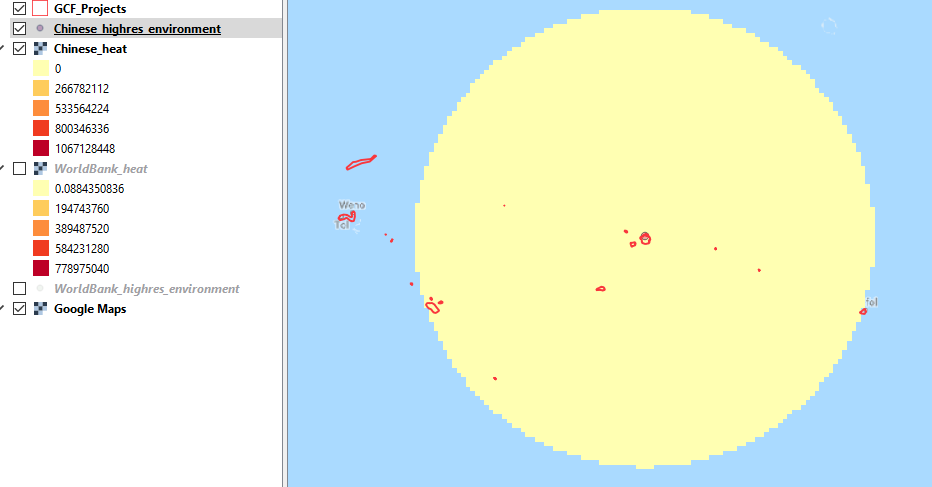
7. Repeat these steps for the Chinese Aid – remember to weight by the even\_split field in your KDE, and set the radius to 5! I named my new heatmap “Chinese\_heat.tif”. After loading and symbolizing my Chinese Data the same way as the World Bank data, my new map appears like this:  


8. Finally in this section, I want to visually compare the hotspots of Chinese and World Bank activity to the locations of GCF projects. To do so, I first drag the GCF\_Projects layer so that it is at the top of my layers menu – it should look something like this:  


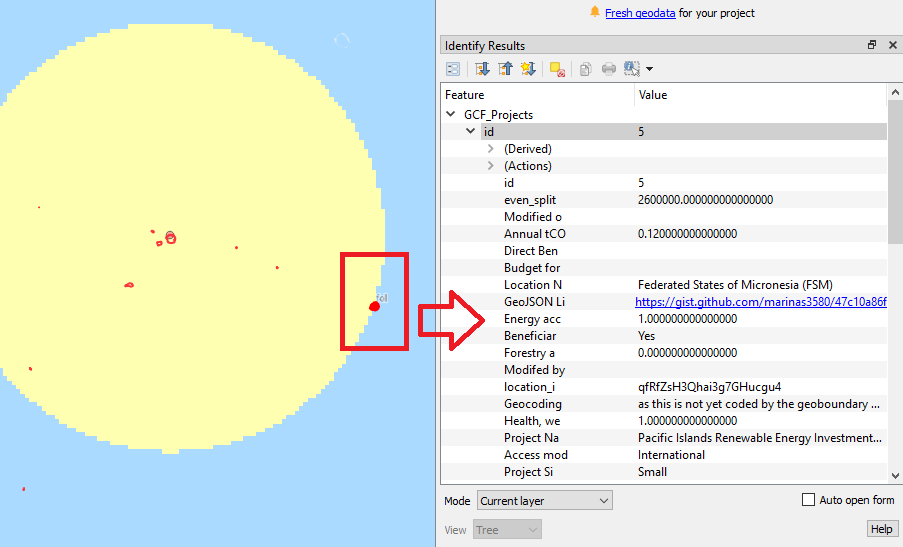
9. Now, I’m going to change the symbology on my GCF\_Projects layer – first, I click on the “Simple Fill” option at the top, and change the fill color to 0% Opacity. Second, I change the Stroke Width to 0.5, and click Apply. Now, each GCF location is represented by an outline.

10. I can now zoom to any GCF project location and see if it is overlapping other Chinese or World Bank projects, and if it’s in areas of high or low intensity, simply by turning the “Chinese\_heat” layer on in my layers (as seen in the above figure – Chinese\_heat is checked). At the aggregate layer, it is clear that the GCF has more environmental activities in South America and the country of Georgia than China, as well a sa number of island states and some activities in Mongolia.

11. Now, let’s say I want to look specifically at an area where there is substantial overlap. We’ll focus on activities in the Federated States of Micronesia.

12. First, zoom your map (using the zoom tool) to that location. Then, in the layers screen, turn Chinese\_highres\_environment on and drag it to the top layer. Your layers and map should look something like this:  


13. You’ll see in red the specific Islands that GCF projects have targeted; unlike the GCF, we only know that China targeted aid to this region, but no the specific islands (represented by the dot in the middle of the yellow circle). Let’s learn more about this aid – first, click on the “Chinese\_highres\_environment” layer in your layers, second click on the “info” tool at the top of QGIS (), and then finally click in the center of the large yellow circle on the single point of Chinese aid.

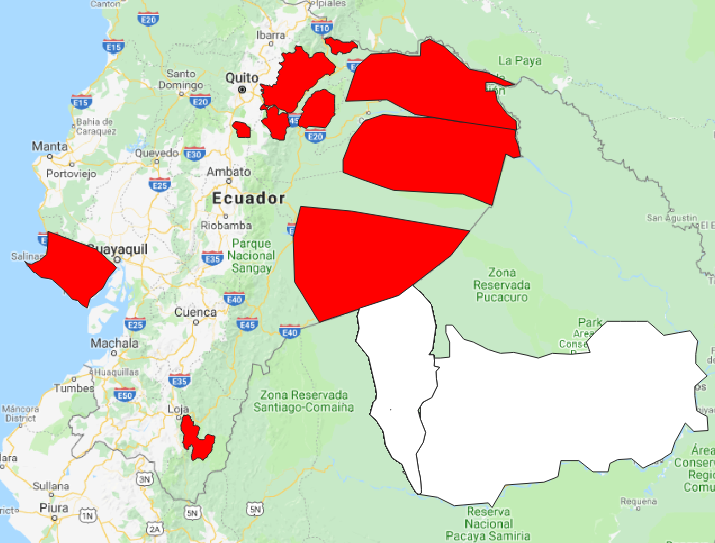
1. At the lower-right, a new “Identify Results” menu will appear. This will provide all of the information that we have on this particular Chinese project – i.e., the title “China constructs fisheries zone in Micronesia”; the end date (2007); and the most specific location available (Pohnpei).
2. You can do the same thing for GCF projects – first, select GCF\_Projects in the layers box, then use the info tool to click on any GCF project. For example, if you click on Project ID 5 you can see the Pacific Islands Renewable Energy Investment project, along with all metadata:  
   
3. You can repeat this exercise with the World Bank to identify overlap with WB environmental projects as well.

**1.5. Weighting Aid**

It is very common that the exact location of where aid is allocated is unknown – instead, we only know the geographic regions to which Aid was likely to have been allocated. In the case of the GCF, projects are known at highly variable levels – ranging from specific areas within natural reserves to entire reserves. To make better estimates of where aid was precisely allocated, it is common to use satellite imagery to generate more precise estimates. For example, if renewable energy investments are increased, it can be hoped that areas that received such investments would have an increased level of nighttime lights emissions that we can see from a satellite. Thus, we can choose to generate estimated regions of aid based on ancillary satellite sources.

For this example, we’ll be using a region covering Ecuador and Peru, encompassing a small set of example GCF projects – I have already prepared the relevant GCF data layer for you, and it can be loaded into QGIS from the folder 3\_PopWeight.

1. First, load the GCF\_EcuadorPeru.shp file from the folder 3\_PopWeight.

2. Next, let’s visualize the problem we’re trying to solve. Right click on GCF\_EcuadorPeru, and go to properties and Symbology. Change “Single Symbol” to “Graduated” at the top, and then for the Column choose the filed “Total Fina” (representing the total financing). Click apply and close the symbology menu – you should see a map that looks something like this:  


3. Common sense tells us that all areas in red did not receive the exact same amount of financing, nor did all of the areas in White. So, we want to make more refined estimates.

4. To do so, we’re going to assume that areas with high nighttime lights intensities received more aid – a process called spatial imputation. Researchers could easily choose to instead assume that areas with more vegetation received more aid (i.e., using NDVI as we did in an earlier session), or any other layer that may be appropriate for their case.

5. To do this, we’re going to retrieve a version of nighttime lights referred to as DMSP (you may also choose to use more contemporary styles of nighttime lights such as VIIRS, but these require much more computational power to process).

6. To retrieve DMSP data, you can go to <http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html> and choose a year you would like nighttime lights for; alternatively, I have provided a 2010 image in the folder 3\_PopWeight. The best year is generally up to the analyst, and it should be reflective of the data that she or he believes best represents where aid is most likely to have been distributed (i.e., if Aid in 2011 was given based on the economic conditions in 2010, then 2010 data may be best).

7. Add the DMSP tif data to QGIS (in this case, the file is named “F162010.fg\_15\_35\_43\_50.op\_blended.v3a.avg\_vis.tif”). Once added, you’ll just see a black box – let’s fix that by clipping to only our area of interest. Go to Raster 🡪 Extraction 🡪 Clip Raster by Extent. For your input layer, choose the DMSP tif; for clipping extent click the three dots beside the entry and choose “Use Layer Extent”. Choose GCF\_EcuadorPeru from the dropdown, and click OK. Finally, scroll down to the bottom and under Clipped, click the three dots to the left and save the file – I put mine in 3\_PopWeight, and called it “NTL\_GCF.tif”.

8. Add the new layer to your view, and zoom in to it (right click -> zoom to layer). Let’s do some quick visualization steps to see what we just did:

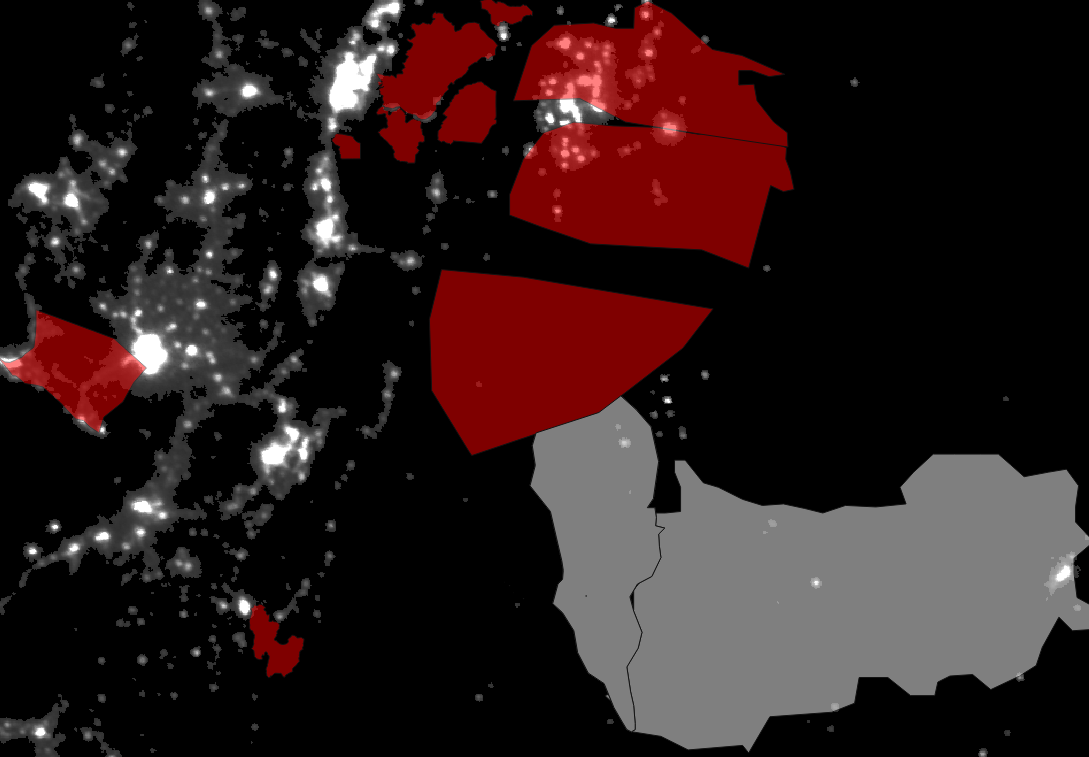
A) In your layers, drag GCF\_EcuadorPeru to the top.

B) Symbolize GCF\_EcuadorPeru so that it is around 50% transparent (layer rendering).

C) Right click on NTL\_GCF, and go to it’s symbology.

D) Set the “Max” to 15, and click apply – this will allow us to see smaller values. Click OK.

After doing this, your map should look something like:



9. Our basic goal is to distribute the aid that is currently in very large polygons (our known areas of GCF funding) to be more likely near areas where we know people live (areas with lights at night). As noted before, this is one of many ways geospatial uncertainty can be dealt with – i.e., one could instead use NDVI to improve estimates.

10. To do this, we need three total rasters: the first is the nighttime lights raster we already have. The second is a raster that – for each of our GCF projects – tells us what the total nighttime lights value is. Finally, we need a third raster that tells us – for each cell – what the estimated total amount of funding for a given project is. These three rasters will allow us to calculate:

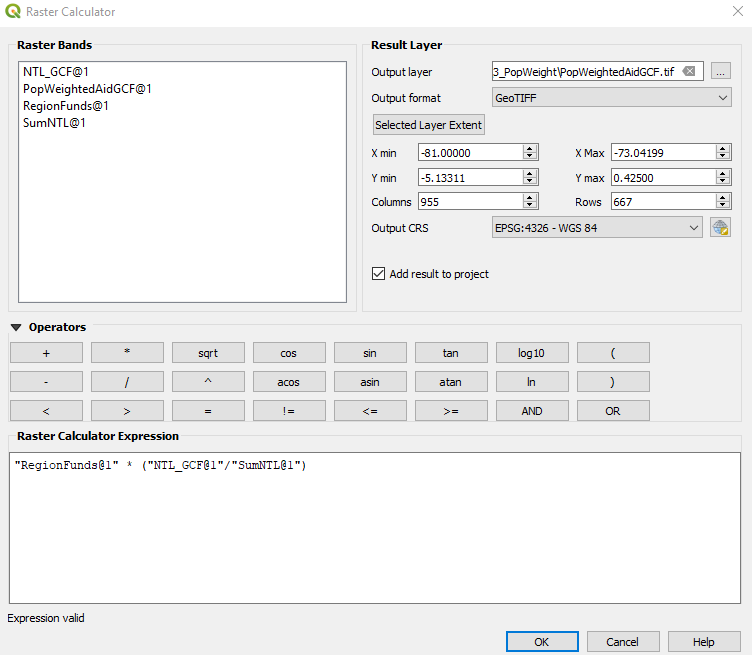
(Total Funds for Entire Project Region) \* (Nighttime Lights / Sum of Nighttime Lights)

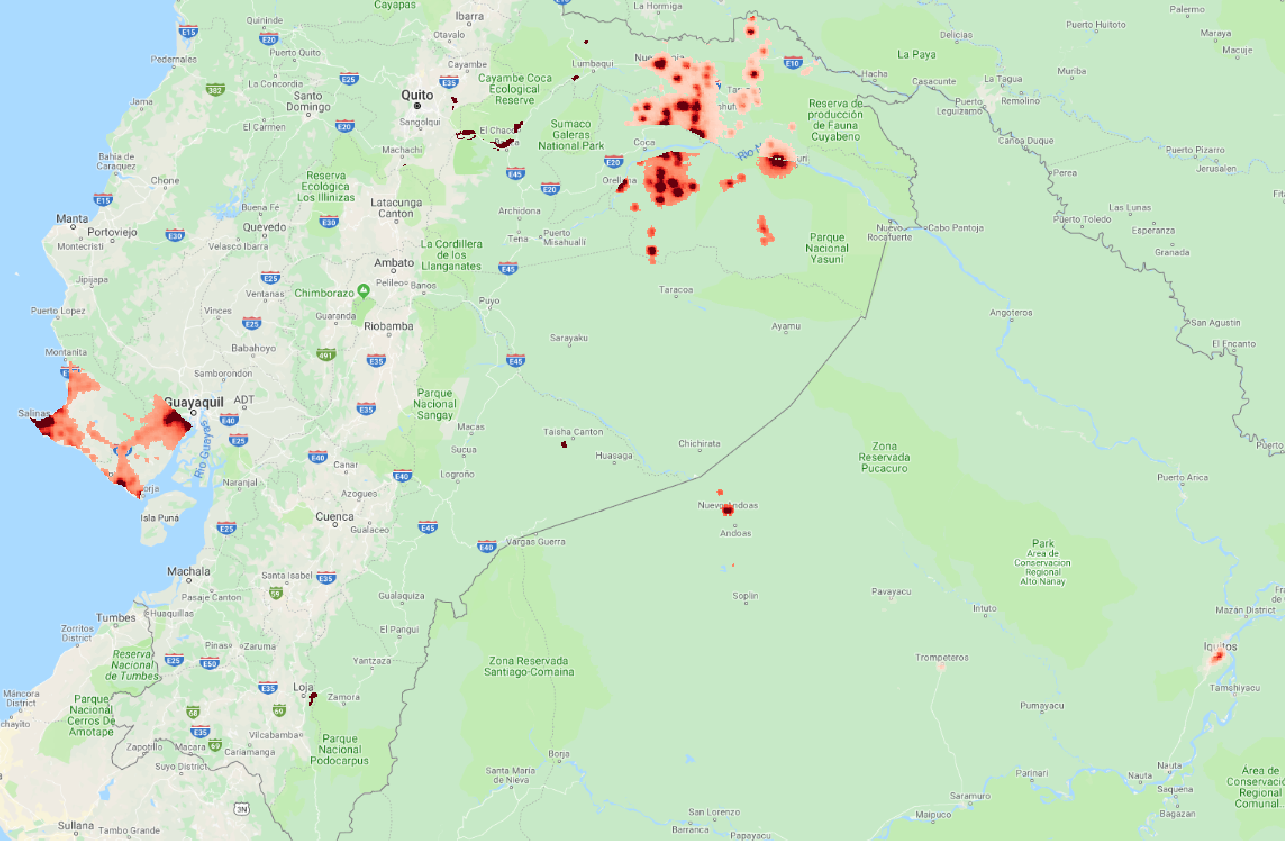
11. Let’s create the first raster (total funds), as this is the easiest. A the top menu for Q, go to Raster 🡪 Conversion 🡪 Rasterize (Vector to Raster).

12. For the input layer, choose the GCF\_EcuadorPeru.

13. Under “Field to use for a burn-in value”, choose “even\_split”.

14. For “Output raster size units”, choose “Georeferenced Units”.

1. For both “Width/Horizontal” and “Height/Vertical” resolutions, choose 0.008333 (the same as the Nighttime Lights layer).
2. For output extent, click the three dots to the right, and choose the NTL\_GCF.
3. At the bottom, under the word “Rasterized”, choose to save the file. I chose to save mine as “RegionFunds.tif” in the 3\_PopWeight folder.
4. Click “Run”, and then add your new raster to your legend. You may need to remove the “Rasterized” item, depending on your operating system. You’ll note the new raster is very simple – just big white triangles where projects are known to have existed (note you may need to change the symbology to a pseudocolor in order to see all projects).
5. Now, let’s create the third raster – the sum of nighttime lights for each area. This will be a two step process.
6. First, open the processing toolbox (Processing 🡪 Toolbox) and search for zonal statistics.
7. In the zonal statistics option box, for raster layer choose NTL\_GCF.
8. For the raster band, Band 1 (the default) can be left alone.
9. The vector layer containing the zones should be GCF\_EcuadorPeru
10. For Output column prefix, choose “NTL\_”.
11. Click the three dots next to “Statistics to calculate”, and uncheck everything but Sum.
12. Click run.
13. Right click on GCF\_EcuadorPeru in the layers, and scroll all the way to the right – you’ll now see a new column called “NTL\_sum”, populated with the highest value in each case.
14. We’re going to create a raster using that max – go back to Raster 🡪 Conversion 🡪 Rasterize, and repeat the process above, this time using the field “NTL\_sum” for burn-in. Everything else should be identical. I saved my new layer as “SumNTL.tif” in the 3\_PopWeight folder.
15. Finally, we’re going to use the raster calculator to build our refined estimates of aid. Go to Raster 🡪 Raster Calculator.
16. For your formula, repeat what we wrote above – your equation should look like:  
    
17. Finally, in the upper-right choose an output later. I saved mine as “PopWeightedAidGCF.tif”, and put it in the 03\_PopWeight folder.
18. Click OK, and then right click on your new layer in the Layers section and go to properties 🡪 Symbology.
19. Change the render type to Singleband pseudocolor, and the color ramp to Reds. I also changed the opacity of 0 values to 0%, so only locations that received aid show up.
20. Under Min/Max Value Settings, set the max to 5,000 (to allow us to see smaller estimates).
21. Finally, take a look at your final map – this is our best refined estimate as to how much aid was allocated, given the assumption of nighttime lights based distributions. My final map looks like:



1. Time permitting, try your own aid allocation assumption – i.e., distribute aid according to NDVI or another source layer.