Lab 2. The Roots of Human History

Lab in Brief:
The Purpose of this assignment is to introduce you to thinking about finding solutions using spatial data – i.e., "making fancy maps." In this case, you'll be learning how to process and use satellite imagery to attempt and identify locations for your brand new pacific island villa, and using that data to defend your choices. Further, you will learn how to produce a new type of data visualization from scratch using datasets: a "map". Believe it or not, geographers still make those things!

Data and Materials:
This time around, we'll be using the open source GIS program "Quantum GIS" (Q for short), along with satellite information drawn from a number of different satellite and space shuttle missions (including the Shuttle Radar Topography Mission [SRTM], MODIS Satellite Series, and a variety of weather and climate satellites).
Data can be downloaded at: http://labs.aiddata.wm.edu/COLL100/

Objectives:
In this lab, you will (a) answer a series of questions regarding the relative suitability of different locations within French Polynesia for human habitation (i.e., "where would you want to build your house if you got to live on a pacific island"), taking into account a variety of environmental concerns (including sea level rise), and (b) select two sites – one that is a strong choice, and one that Is a weak choice. Finally, you will produce a mapped visualization that defends your choice.

Deliverables:
By Friday, September 25th at 11:59PM EST, you will submit the lab deliverables described on the final page of the lab to blackboard.

Grading:
This lab will be graded according to the same rubric used for Lab 1, which can be found on blackboard. The four key criteria are Creativity, Clarity, your ability to Define the Problem and use pertinent information to defend your site selections. **It is important that you spend time working on your map so that it is visually appealing – you may even consider mixing strategies from your infographic creation with the map to prove your points.** What interesting statistics can you pull from your data to make a better product for "System 1"?

Part 1: Getting Setup

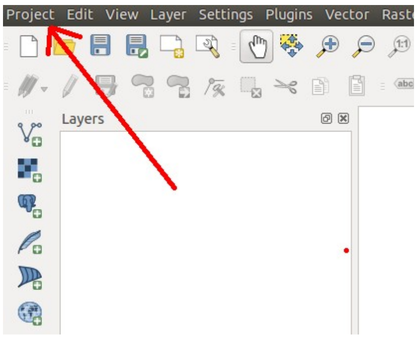
1. First, create a new folder on your H drive, and then download all of the data from black-board and put it into that folder. You'll have quite a lot in there, including vectors (i.e., "shapefiles") and raster (i.e., "tif" files) data.

Fun ("for Dan") Geography Fact: A vector is generally used to define boundaries, for example a state, on a map. Rasters store information that is collected over space, for example satellite data or temperature.

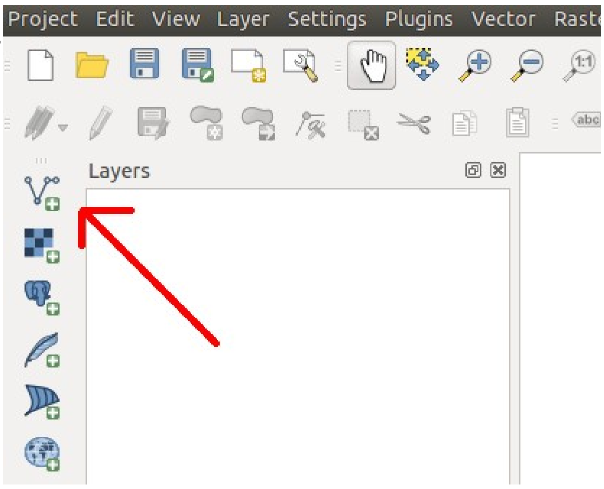
2. Go to the Start button and type in "Qgis", or simply "Q". You'll see Quantum GIS Desktop appear as an application – go ahead and launch it. Note it will take a little while to start the first time (2-3 minutes), but should be much faster after this.

*Fun ("for You") Fact: Quantum GIS is **free**, and you can install it on your computer and do this entire lab at home (if you really want!). The URL is: http://www.qgis.org/*

3. Once Q (we'll call it Q from now on) opens, the firs thing you'll want to do is create a new "project" (this is what you'll be saving into). To do that, click the "Project" button at the upper-left, then click "Save As". Pick your H drive folder and create a name you'll remember!

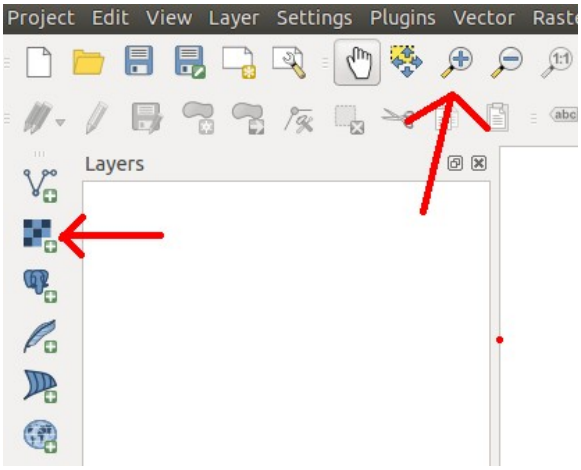


4. Now, we're going to add in our first bit of GIS data. Ultimately, we're attempting to identify the best sites across French Polynesia for your island villa (or, if we're being intellectuals, human habitation writ large), so we want to load a boundary that tells the computer where French Polynesia is. Because it is a boundary file, it's saved as a vector, so click on the "Add Vector Layer" button, then navigate to your H:\ drive and find the file "PYF_adm0.shp", and add it to your map.

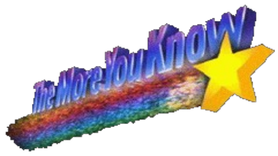


Fact number 3(ish): you can download boundaries for any country in the world from http://www.gadm.org !

5. You're probably thinking “well, that doesn't look very impressive”. All you've loaded are the national boundaries for French Polynesia, but it's awfully hard to tell what's really going on without something called a “basemap” - i.e., data that shows where you are in the world. Let's add one of those. The basemap is our first *raster* file, so you're going to click “Add Raster Layer.” Now, navigate to your H drive and find the file “NE2_HR_LC_SR_W_DR.tif”, and add that to your map.

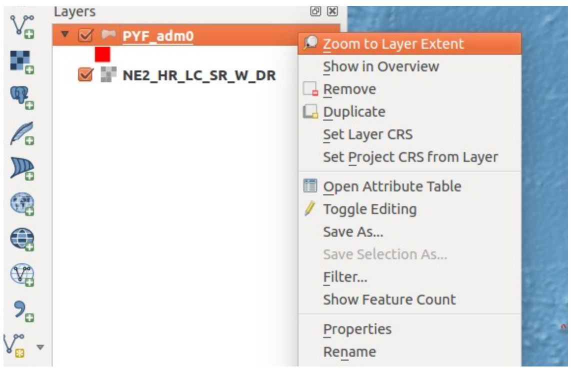


That long file name stands for “Natural Earth 2 – High Resolution – Land Cover – Shaded Relief – Water and Drainages. You can have many different types of base maps with a variety of different features included. Take a look at http://www.naturalearthdata.com/ for more!

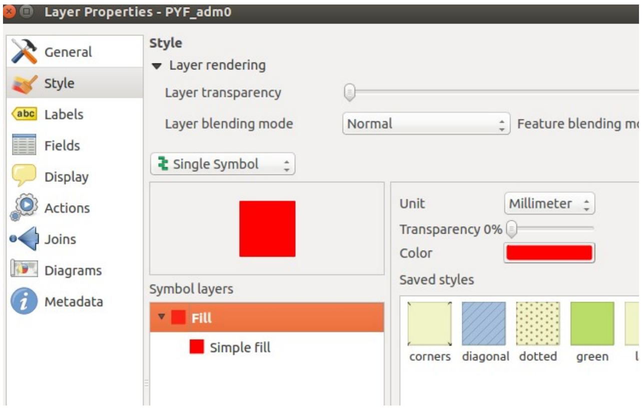


6. To understand what you just did, you'll want to do two things. **First**, use the magnifying glass tool to zoom in on one of your islands. You'll see it looks kind of fuzzy – that's caused by what's called “spatial resolution” - i.e., satellites can only see as well as the camera that powers them. Better satellites have finer spatial resolution, which makes things less fuzzy. **Second**, use the zoom-out magnifying glass (with the “minus” inside it) to zoom out and see the entire world (you may need to click a few times; you can also use your mouse wheel). You can zoom in to any area to see how it's represented with the base map. Hundreds of basemaps exist – for example, if you use maps.google.com and type in “French Polynesia”, you'll see a basemap with much finer spatial resolution. However, the file size for such images is far larger than systems like blackboard (or even your computers!) could hold.

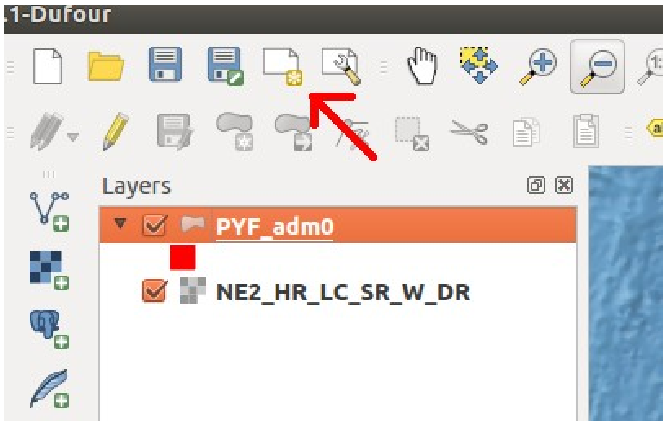
7. You may notice you “lost” your original islands when you added the base map. This is because the base map was effectively put “on top” of the polygons. To fix this, drag the PYF_adm0 to the top of the list on the left (this list is called the “Table of Contents”). Once you've done that, let's head back to French Polynesia by right clicking on “PYF_adm0” in our Table of Contents and choose “Zoom to Layer Extent”.



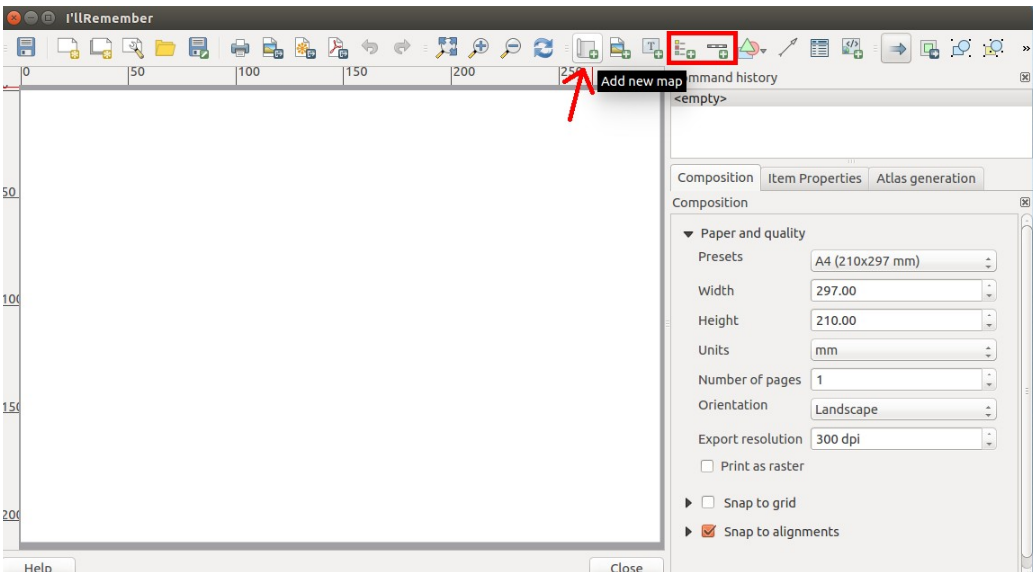
8. Even now, those islands and atolls are difficult to see, so let's make them more obvious. Right click on “PYF_adm0” in the table of contents, and choose “Properties”. Change the color (under “Style” on the left) to something easier to see – I chose bright red, but you can do whatever is easiest for you. You can also play with things like transparency here to make your map easier to understand and more visually appealing.



9. Once you're happy with your coloring, close that window (you can always go back later). We're going to use what we just produced to make our first – very basic – map. Click on the “Map Composer” in the upper left, which will open a new window. Now, name your map something you'll remember later.



10. The window that just opened is the “Map Composer”. All we're going to do this time is add the map you just produced, a legend (i.e., something that explains what symbols on the map mean), and scale bar (a representation of how many miles an inch is), but you'll want to do much more for the map you turn in for this lab so make yourself comfortable with this tool. Let's get started: click on the Add new map button and then drag a square in the white canvas. Adjust you r map so you can see it, then try the add legend and scalebar options.



11. While you won't need to save this test map, in the future you can export it to a PDF document by clicking the “Export to PDF” button at the top of the composer. To save your map, simply close this window – you can pick up where you left off by clicking the button next to the map composer on the main window (the one with a wrench), then choosing your map name.

What you should be comfortable doing at this point (ask if you're not!):

→ *Loading Vector Data into Q*

→ *Loading Raster Data into Q*

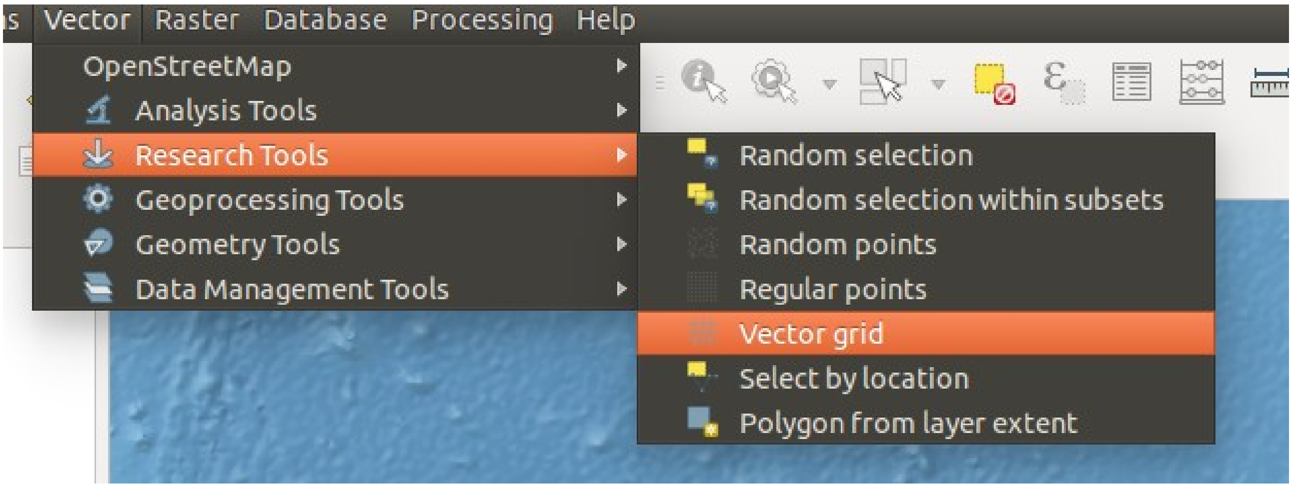
→ *Making a map using Q*

→ *Interacting with the Q Table of Contents and main Map window (zooming around!)*

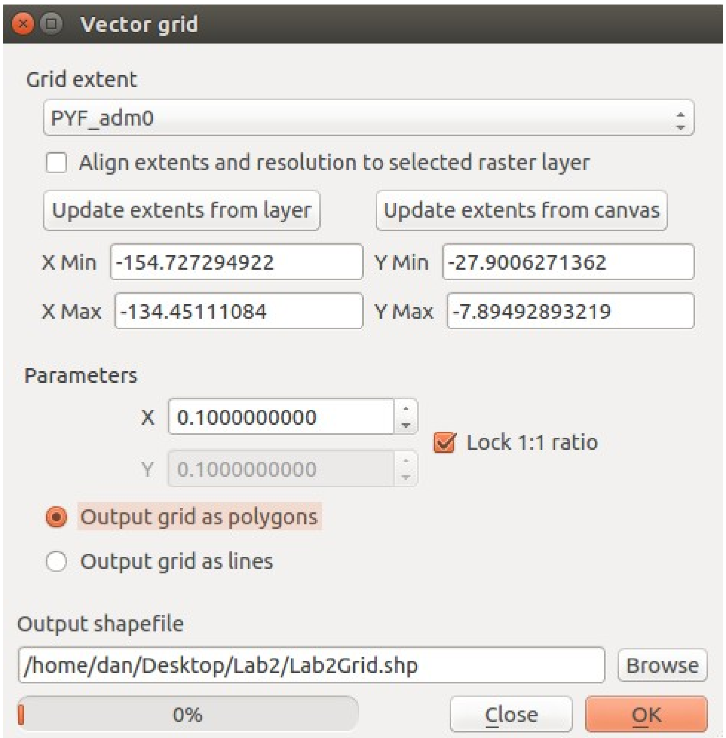
Part 2: Data Preparation

In this part of the lab we're going to prepare a variety of datasets to identify the most suitable areas for human habitability using an approach called "index analysis". First, we're going to define our areas of interest – in this case, a grid of areas you could potentially live in. Second, we're going to collect data for each of those grid "cells".

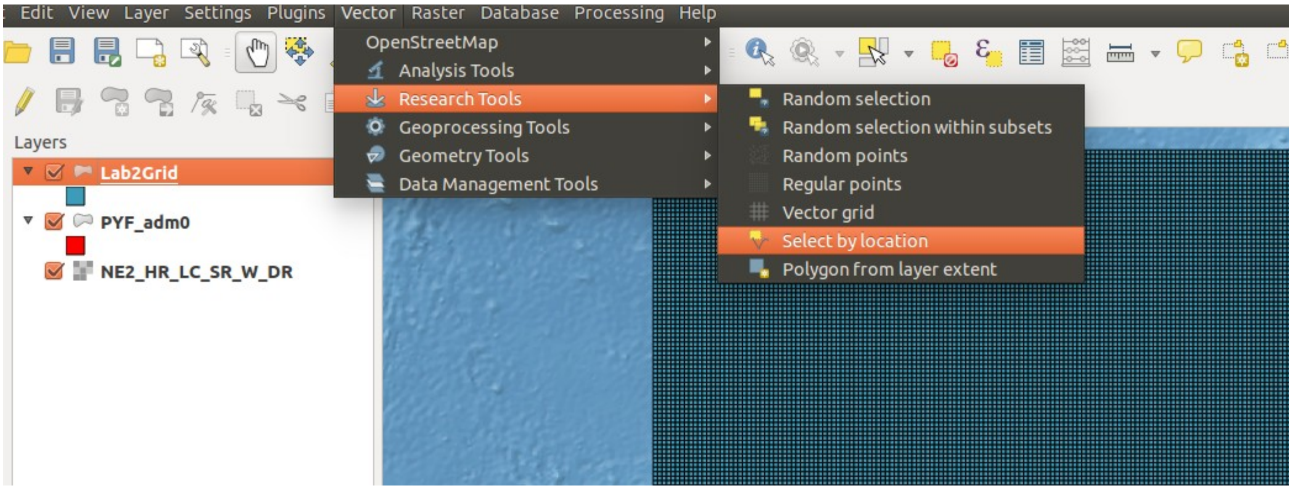
1. First, we're going to make a set of boundaries that covers our study area to use as the basis for our analysis. To do this, go to "Vector" at the top, choose research tools, then choose "Vector Grid".



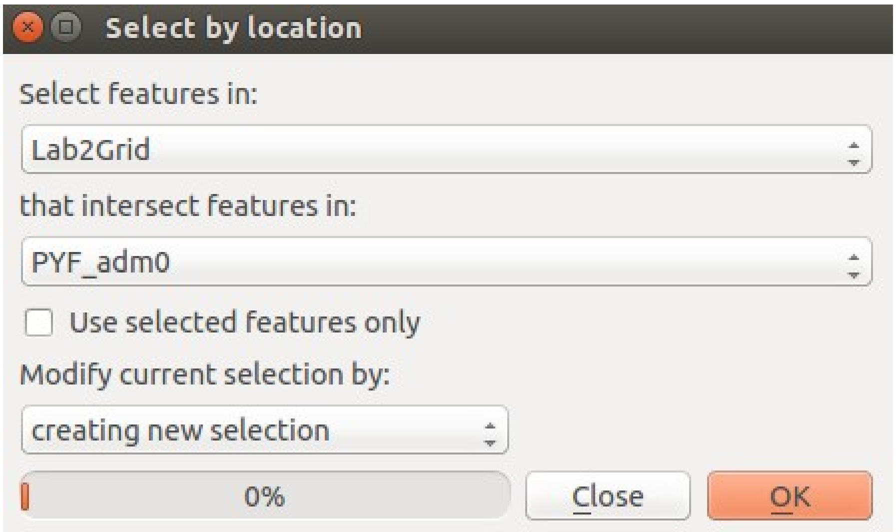
2. Now, create the grid using your French Polynesia boundary, as shown in the below image – make sure to save the "Output shapefile" on your H drive, and output the grid as polygons! Note you'll need to click "update extents from layer" before clicking OK. *Make sure you change the X/Y parameters to the values shown below – this is the spatial resolution of our analysis, and if you leave it at the default it will take your computer a VERY long time to process!* When prompted, add the new file to your table of contents.



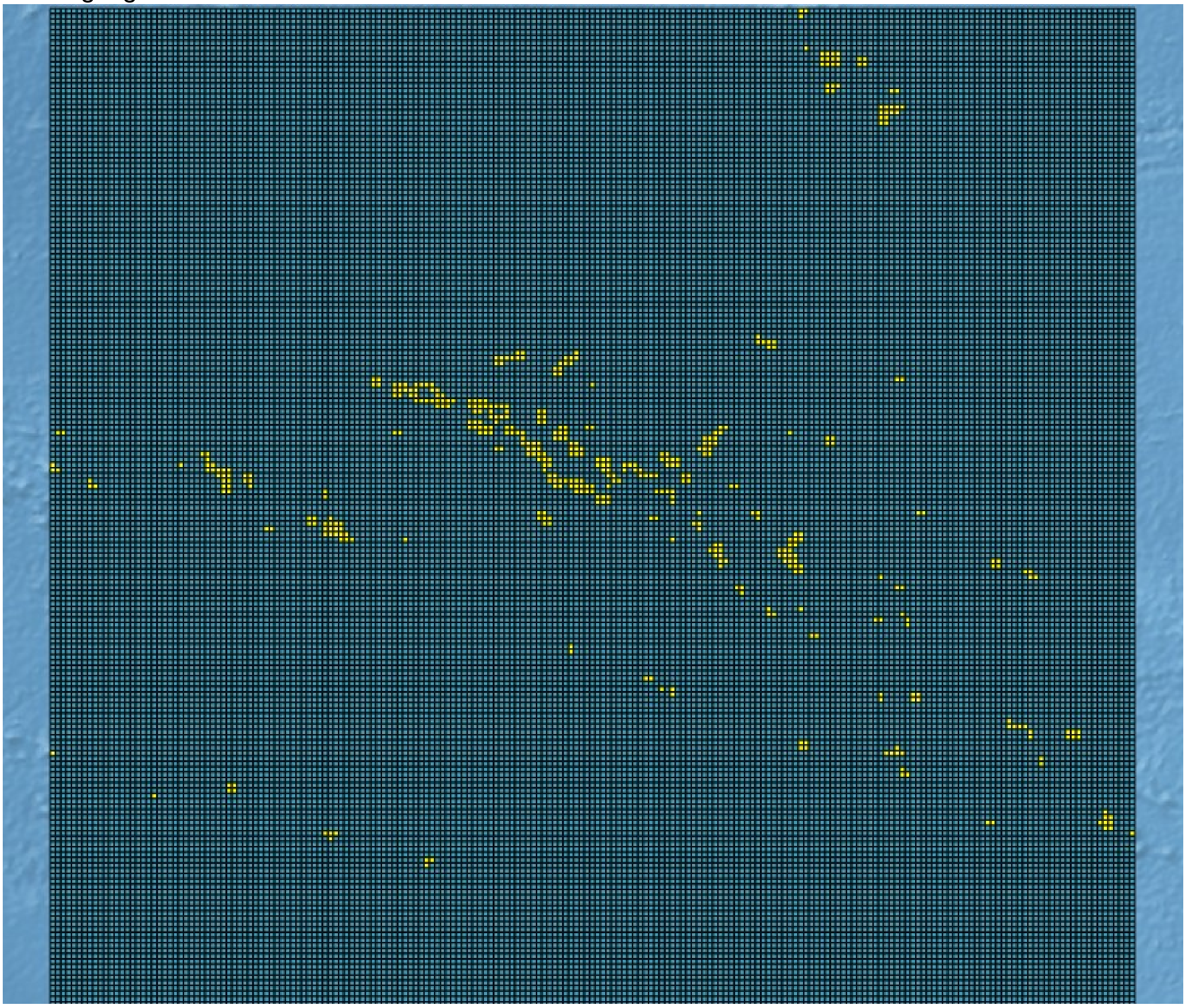
3. You're now staring at a gigantic grid! Most of these grid cells are not actually candidates for our villa, however, as they're in the middle of the water! Let's narrow down our candidates by only selecting areas that intersect with our islands of interest by selecting them based on location:



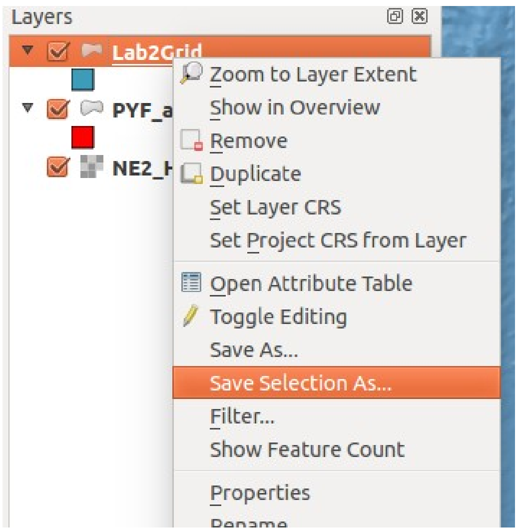
4. Now, we want to choose all grids in our grid file that intersect with the PYR_adm0 – note this may take a minute!

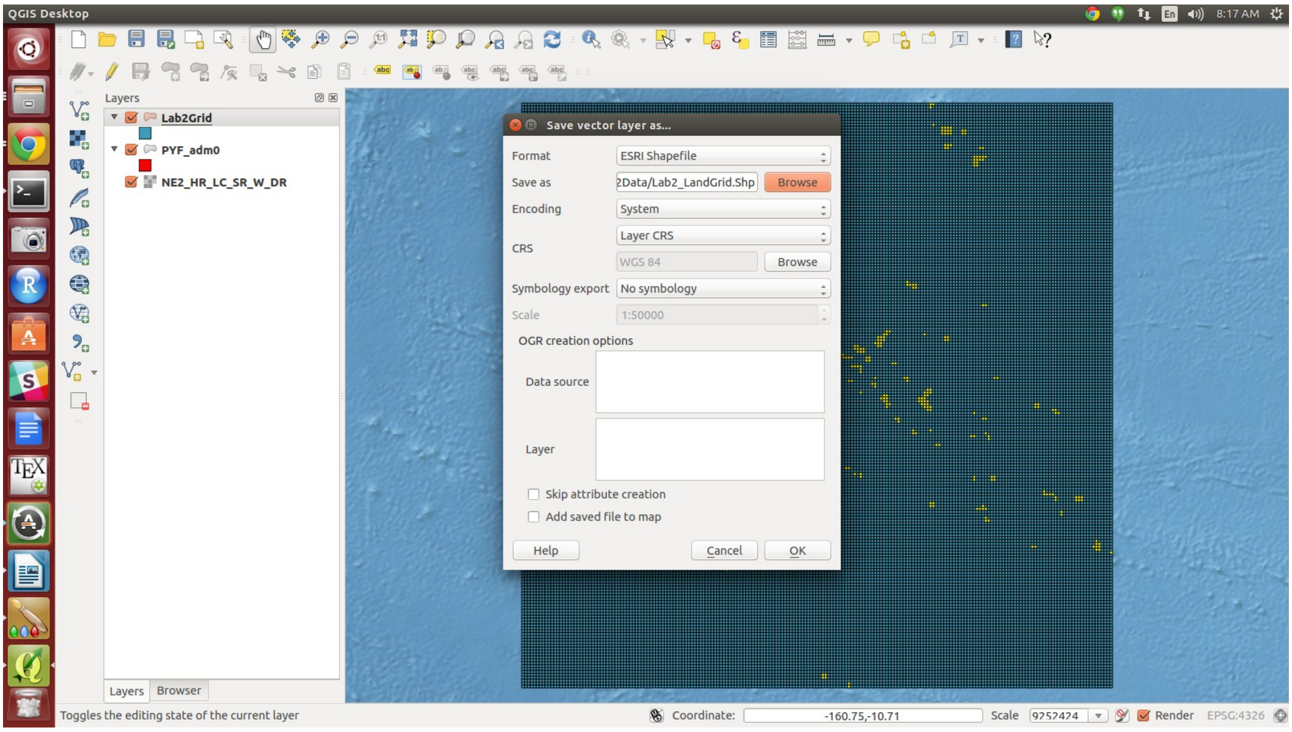


5. Your map will now look similar to this – note that the cells that are nearby atolls or islands are now highlighted:

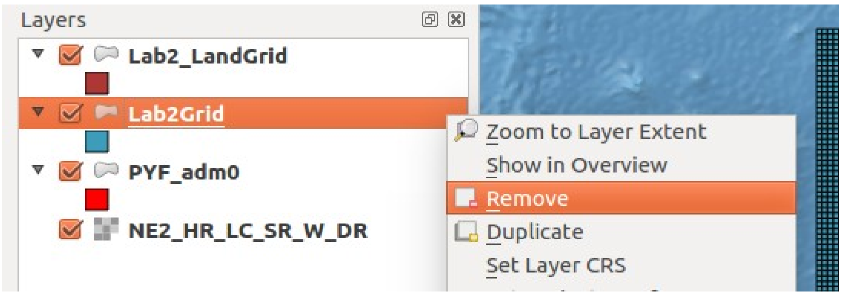


6. Now, let's save that selection as a new boundary file. To do that, go to your table of contents and right click on the grid you created back in step (2). Then click "save selection as...", and save this file on to your H drive somewhere you'll remember it. See the image on the next page to see how to save your file – all you should have to do is click "Browse" and then "ok"! (NOTE: If you're on a Mac, you need to click "Save As", then check the box that says "Save Only Selected Features.")

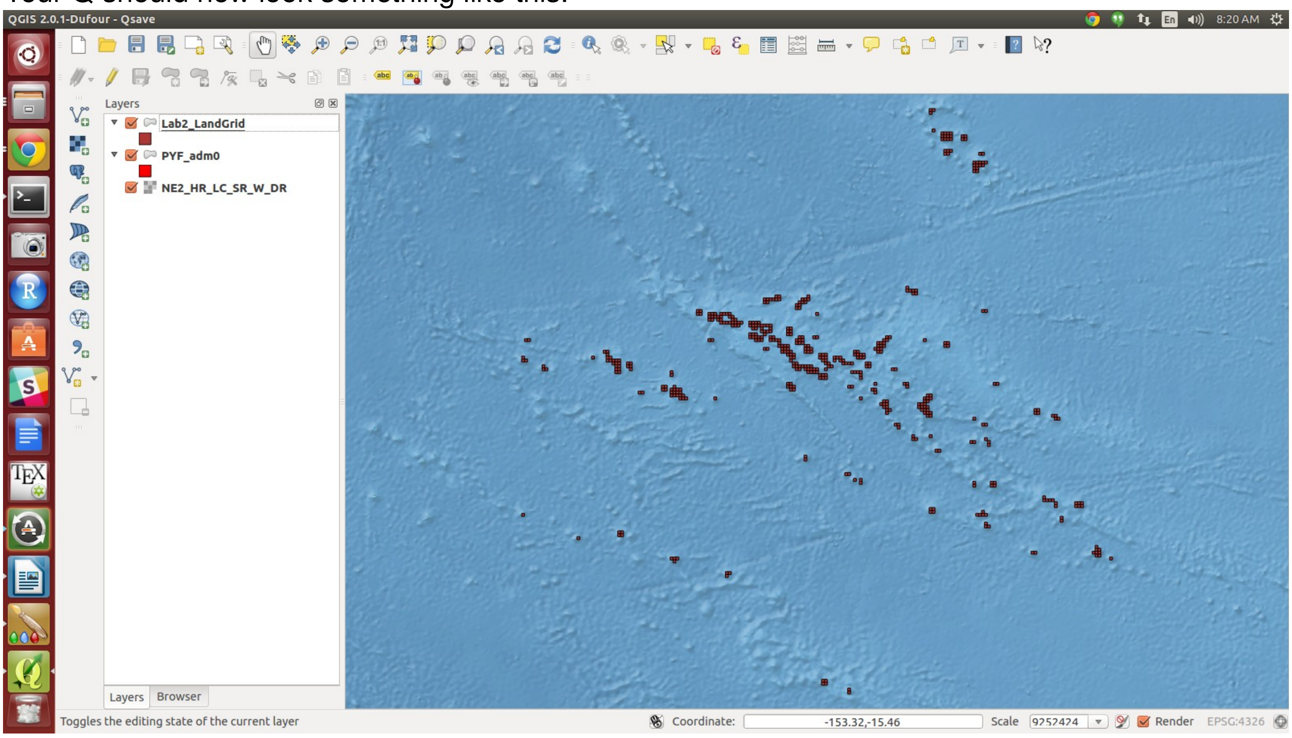




7. Now, just like you did in Part 1, step 4, you're going to add the vector layer you just created back into Q. Once you've added the new grid, remove the old one by right clicking on it and selecting “remove”.



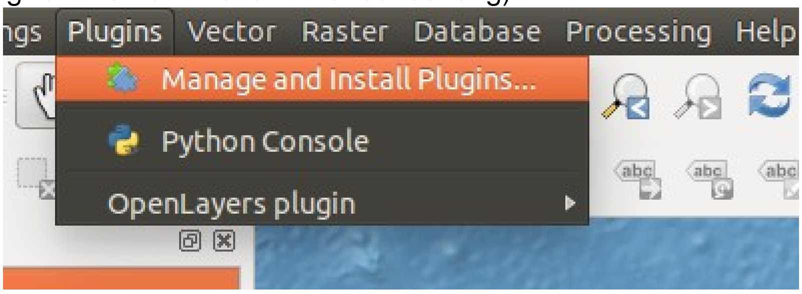
Your Q should now look something like this:



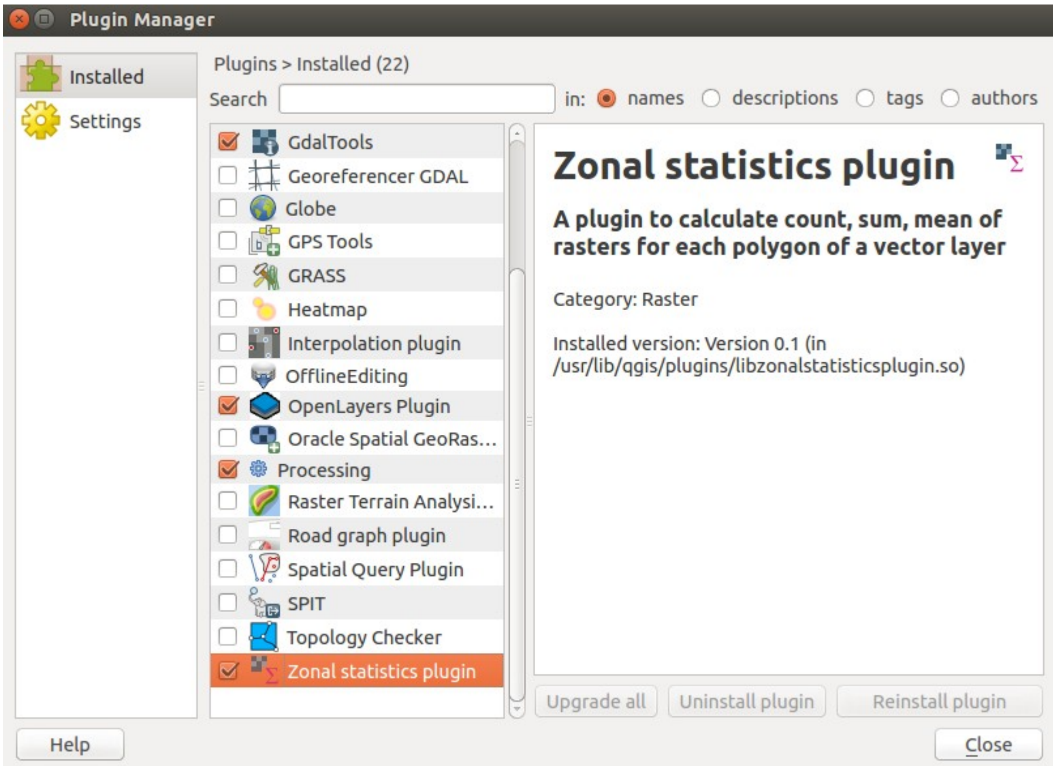
8. Spend a moment zooming around your newly created grid cells – if you uncheck then re-check the box in the table of contents next to your grid, you'll see grids now exist everywhere there is land according to your polygon defining French Polynesia (PYF_adm0).

9. Now, we're going to start assigning values to those cells that are meaningful for understanding human habitability! The first of these is the most straightforward – elevation. Higher elevation represents better resilience to sea level rise and flooding. This dataset was recorded using an instrument on NASA space shuttles as a part of the “Shuttle Radar Topography Mission”, or SRTM. Click on “Add Raster”, and choose the “elevation.tif” file.

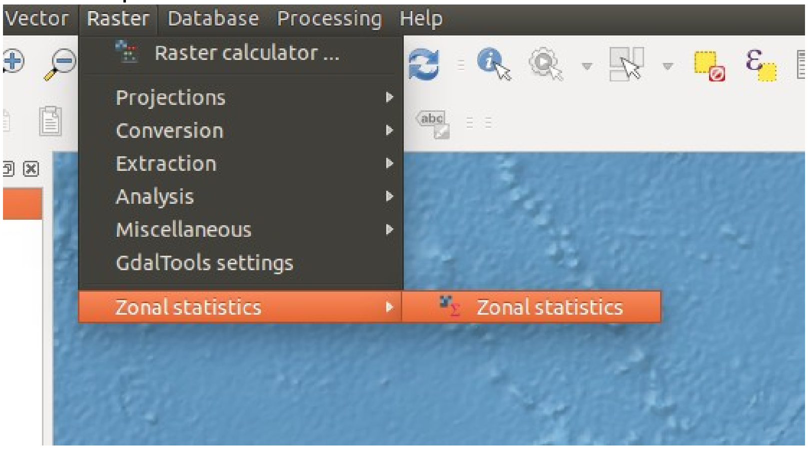
10. Zoom in to one of your islands and toggle the grid on and off – you should be able to see the elevation of each area now. For elevation we're going to calculate the average for each grid cell so that we can choose the “best” and “worst” grid cell options. To do this, we're going to use a plugin for Q called “Zonal Stats”. First, we need to enable the plugin. Go to the plugin manager (you can click “Abort Fetching” on the next menu if it takes too long):



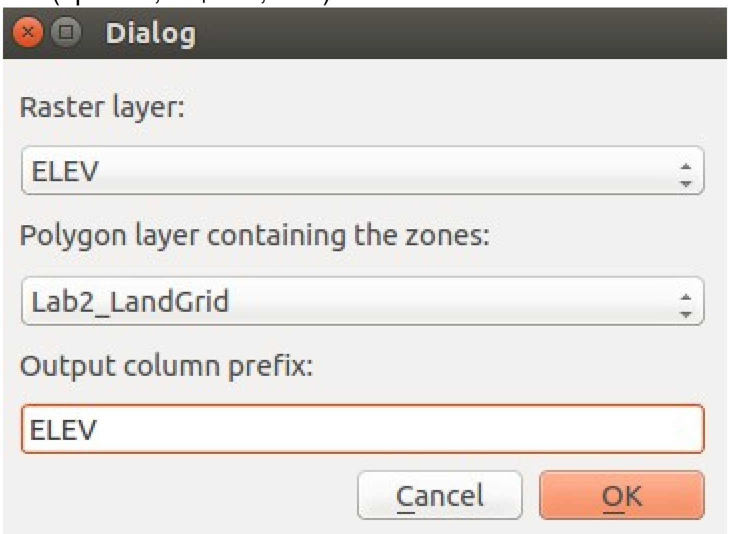
11. Now, scroll down the list of Installed Plugins and put a check beside “Zonal Statistics Plugin”, and click close:



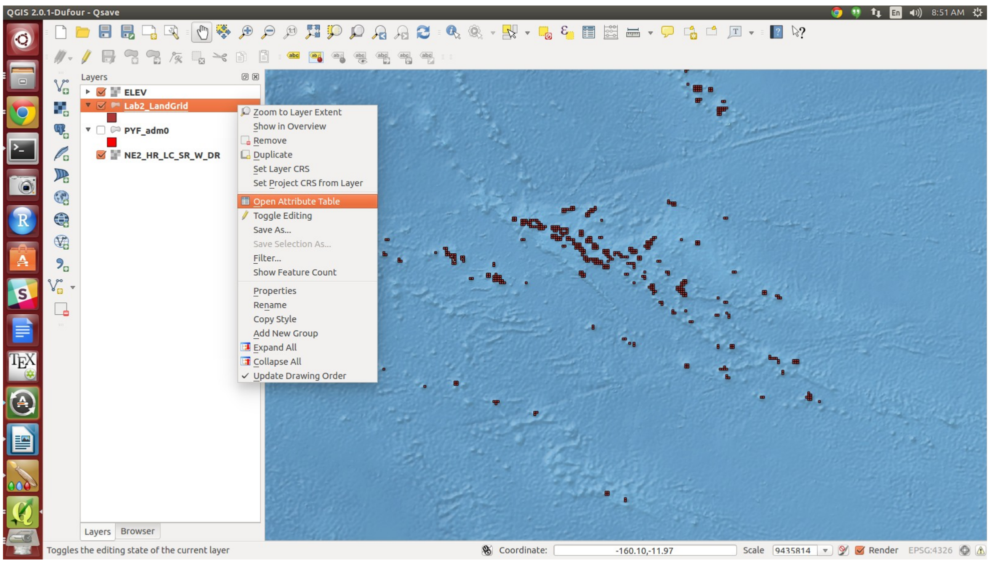
12. You'll now have a new option – Zonal Statistics – under the raster menu. Click on it:



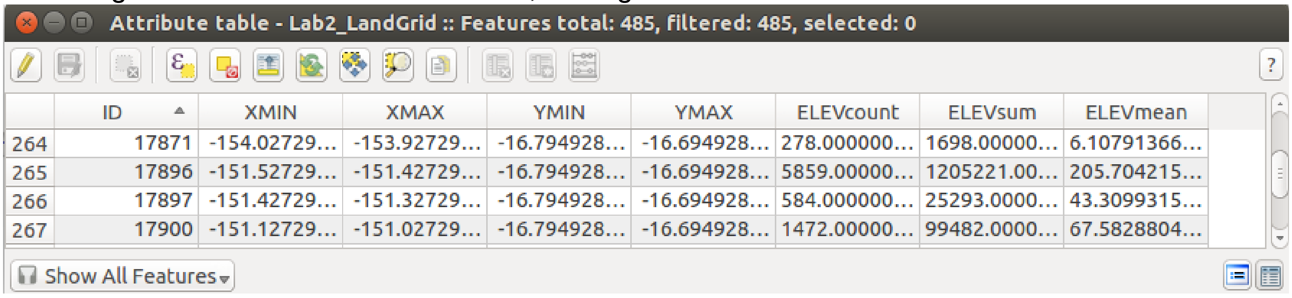
13. Set your statistics options so it looks like the below – what you're doing is telling Q to tell you, for every grid cell in your Grid, what the average elevation value is. In the output column, you're telling it to name that new data “ELEV”. You can call this anything you want, but keep it short and don't use odd characters (spaces, %$#&*, etc.)



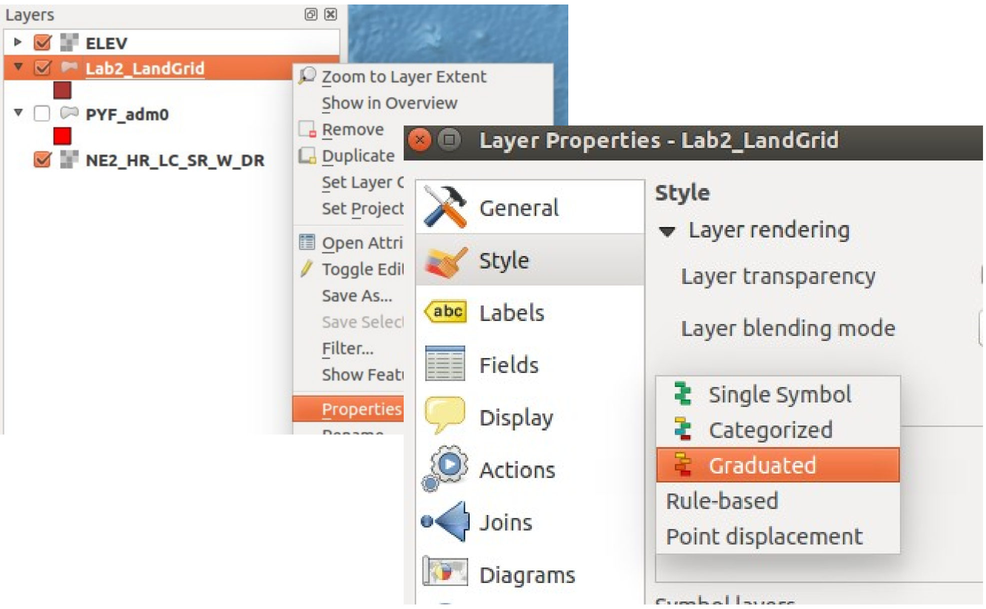
14. Let's make sure it worked – we're going to look at something called the “Attribute Table” of your grids, which is where all of the data is stored. Think about it like an excel sheet, where every row represents one grid on your map:



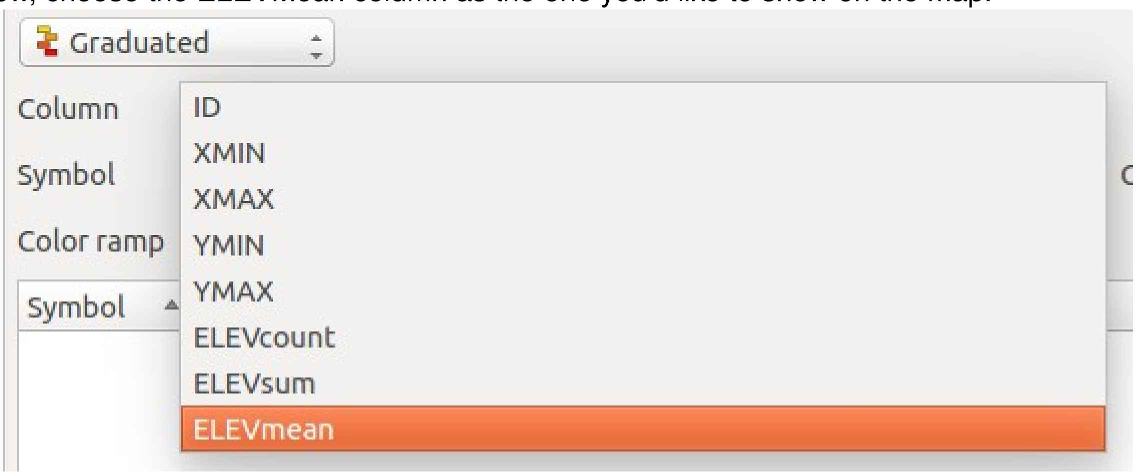
15. Your attribute table should have three new columns at the end now - ELEVcount, ELEVsum, and ELEVmean. The most interesting for us is ELEVmean, which represents the mean elevation within a given cell – as mentioned earlier, the higher this is the better!



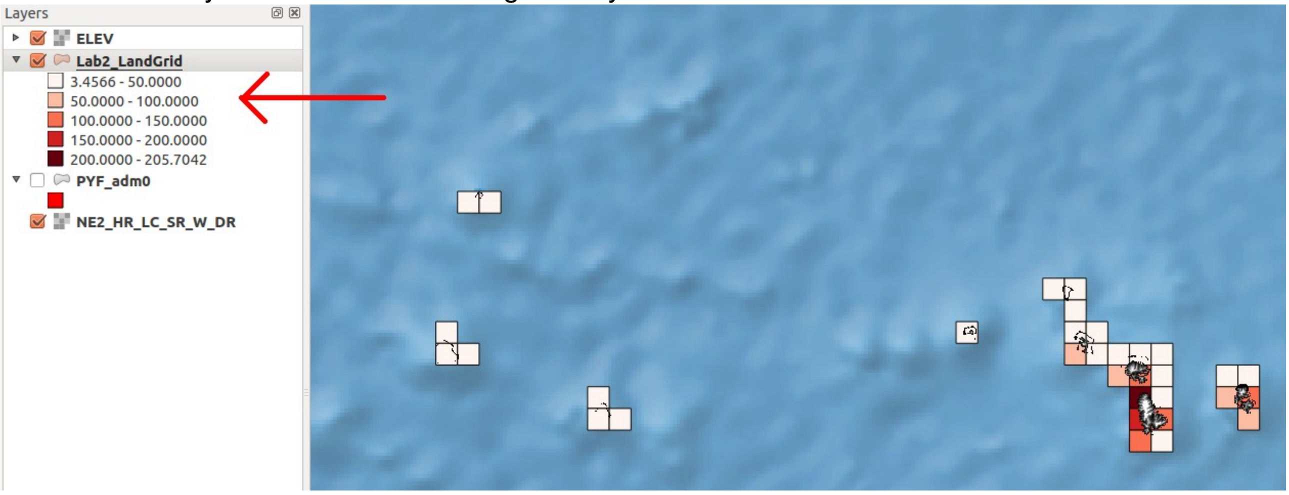
16. Now, let's use our map to symbolize the Elevation Mean so we can get a visual representation of our data (as the spreadsheet view make it hard to understand!). To do this, right click on your grid in the table of contents and choose “properties” again. Then, on the next menu you'll choose “Style” and “Graduated” (see below).



17. Now, choose the ELEVMean column as the one you'd like to show on the map:



18. You can click “OK” now to see what the map looks like with this symbolization, or you can play around with the color ramp, number of categories, or mode categories are made with (i.e., try “Pretty Breaks!”). Clicking “Apply” lets you preview changes. When you make your final map, you will want to symbolize things in ways that help to illustrate the arguments you're trying to make. Below is an example of what you'll see, zoomed in to one area using the red color ramp and pretty breaks. Notice you can also see the legend in your table of contents.



19. Repeat this process for the remaining data we'll be using – vegetation, temperature, and precipitation. This will take a few minutes!

Part 3. Analysis

In this part, we're going to build an index to identify the grid cells we believe are the best and worst candidates for human habitation. This is a three-step process. First, we have to scale all of our values so they're comparable – i.e., it wouldn't make sense to add temperature to elevation without doing something to make it possible! Second, we have to add the scaled values together. Finally, we map our output and identify strong and weak candidates.

1. Now that we have all of our data extracted, we're going to prepare to make an index that shows where human habitability would be best and where it would be worst based on the data we have. The first step of this is to make sure all of our values are comparable – i.e., we're going to translate values like "150 meters elevation" and "20 degrees Celsius" to values between 0 and 1 so we can add them. **For example, this makes no sense at all:**

Grid Cell	Elevation	Temperature (C)	Sum
A	150m	5C	155
B	50m	23C	73
C	25m	25C	50

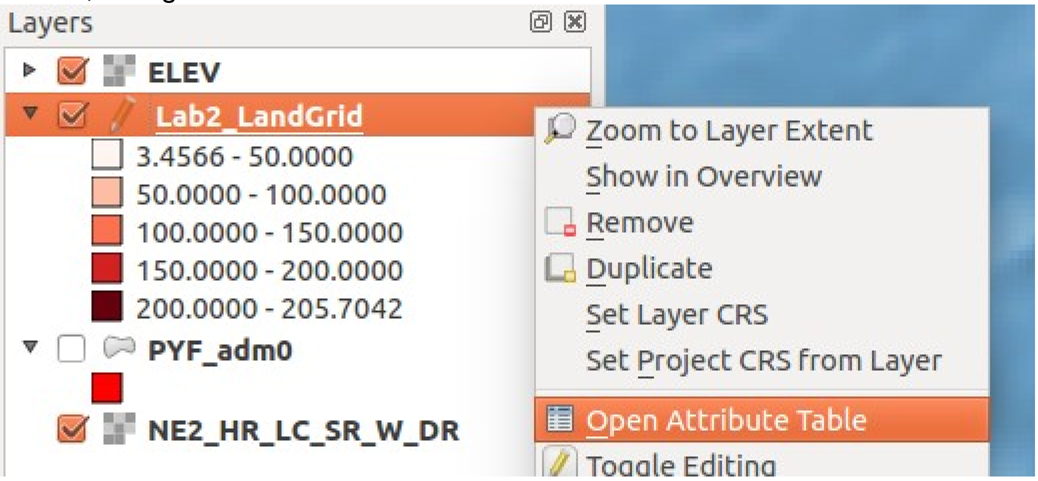
If we used this approach, we would argue that grid cell A is the "best" because it has the highest score, but we haven't standardized – elevation and temperature can't just be added together. Instead, we do something like this:

Grid Cell	Elevation	Temperature (C)	Standard Elevation	Standard Temp	Standard Sum
A	150m	5C	150/150 = **1.0**	5/25 = **0.2**	1.0 + 0.2 = **1.2**
B	50m	23C	50 / 150 = **.33**	23/25= **.92**	.33 + .92 = **1.25**
C	25m	25C	25 / 150 = **.166**	25/25= **1.0**	.166 + 1.0 **1.166**

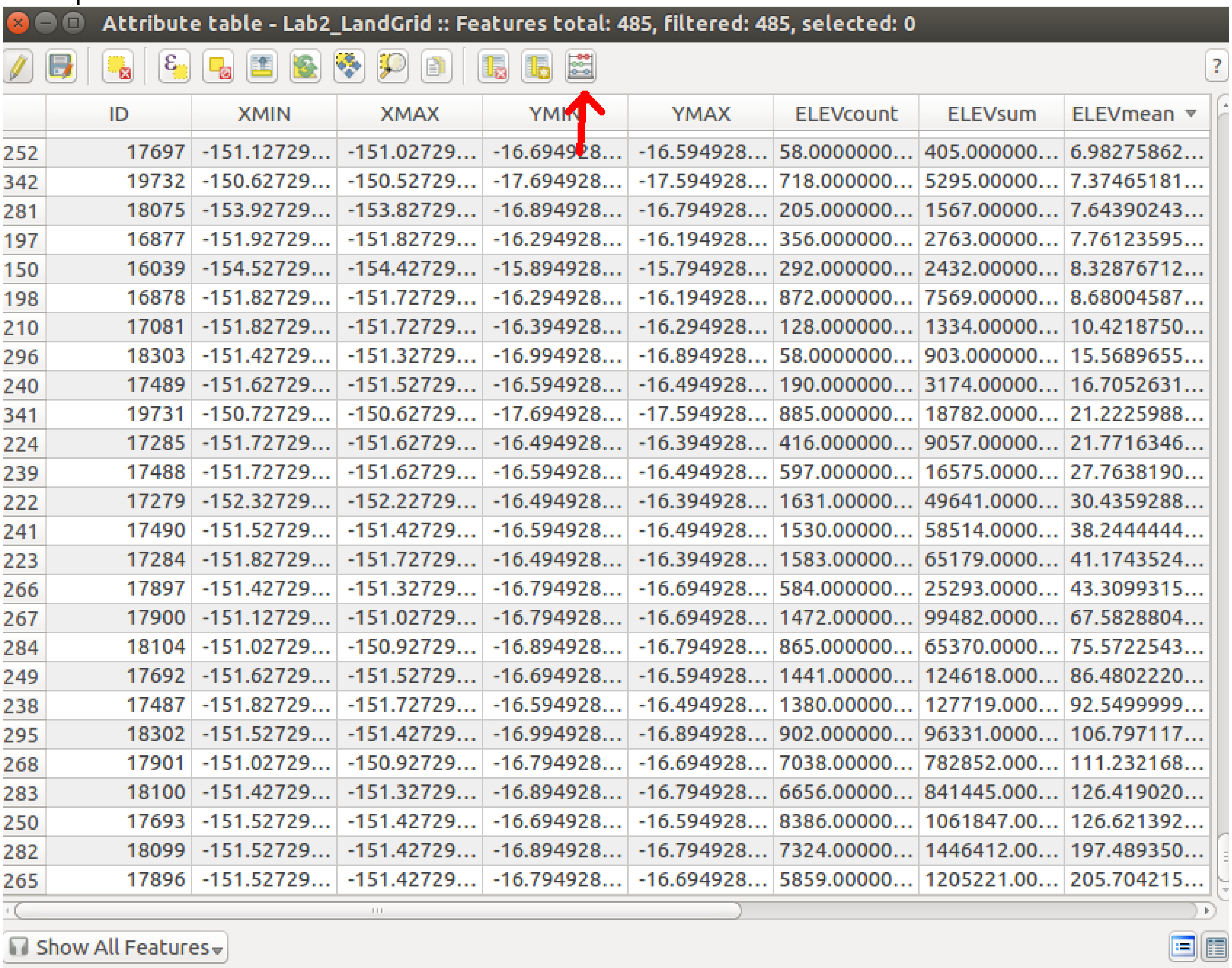
This represents the simplest way to do data "standardization" - there are many more complex methods, but we will not go into detail on those in this course. As you can see, after standardization you would reach another more reasonable conclusion – that grid cell B is the "best".

2. We'll be calculating standard values for every one of the datasets that you generated: elevation, vegetation, temperature, and precipitation. We'll make a very simple assumption that higher values in every "indicator" are always better – i.e., we always want more elevation, vegetation, temperature and precipitation. Of course, this will not always be true (you don't want hotter temperatures in a desert!), but serves as a simple starting point.

3. To get started, let's go back to our attribute table:

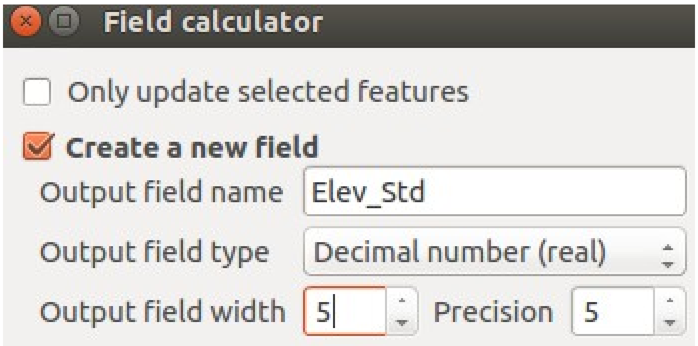


4. Now, click on the “toggle edit” button, and then “Field Calculator” - this is where we'll type in the equations to standardize our data:

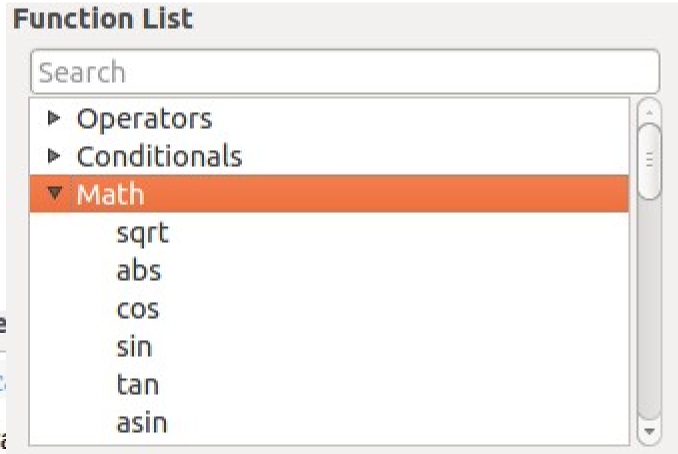


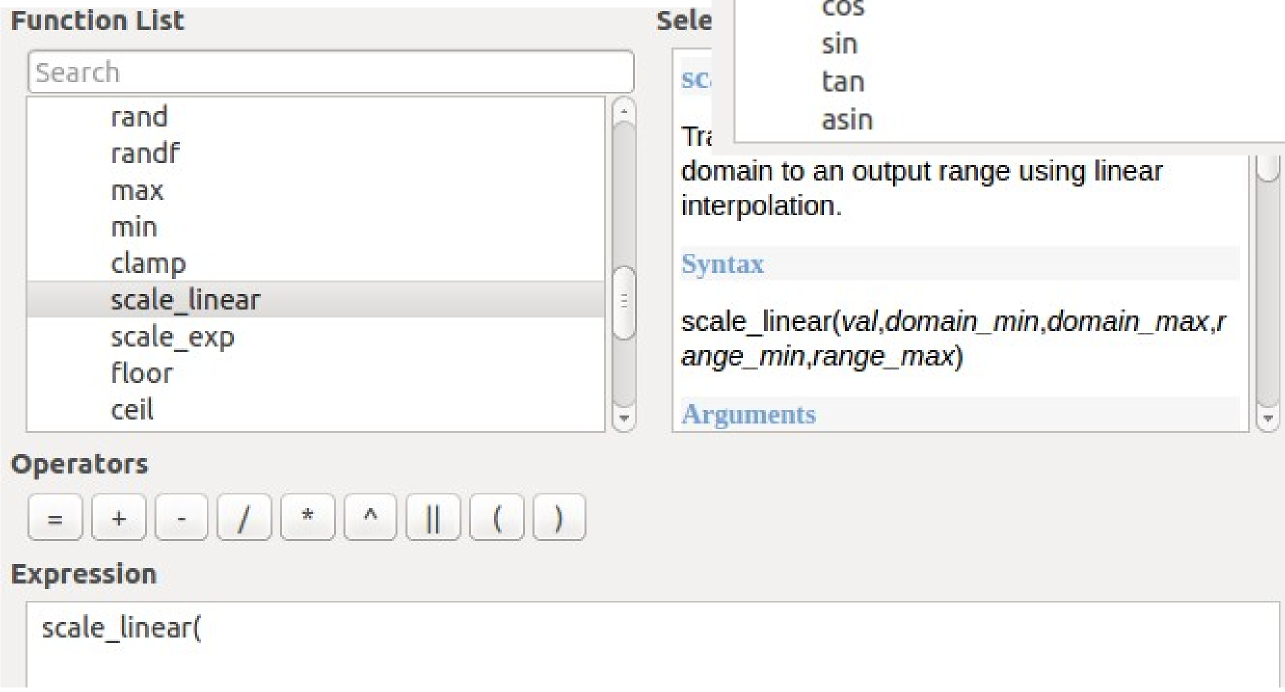
	ID	XMIN	XMAX	YMIN	YMAX	ELEVcount	ELEVsum	ELEVmean
252	17697	-151.12729...	-151.02729...	-16.694928...	-16.594928...	58.0000000...	405.000000...	6.98275862...
342	19732	-150.62729...	-150.52729...	-17.694928...	-17.594928...	718.000000...	5295.00000...	7.37465181...
281	18075	-153.92729...	-153.82729...	-16.894928...	-16.794928...	205.000000...	1567.00000...	7.64390243...
197	16877	-151.92729...	-151.82729...	-16.294928...	-16.194928...	356.000000...	2763.00000...	7.76123595...
150	16039	-154.52729...	-154.42729...	-15.894928...	-15.794928...	292.000000...	2432.00000...	8.32876712...
198	16878	-151.82729...	-151.72729...	-16.294928...	-16.194928...	872.000000...	7569.00000...	8.68004587...
210	17081	-151.82729...	-151.72729...	-16.394928...	-16.294928...	128.000000...	1334.00000...	10.4218750...
296	18303	-151.42729...	-151.32729...	-16.994928...	-16.894928...	58.0000000...	903.000000...	15.5689655...
240	17489	-151.62729...	-151.52729...	-16.594928...	-16.494928...	190.000000...	3174.00000...	16.7052631...
341	19731	-150.72729...	-150.62729...	-17.694928...	-17.594928...	885.000000...	18782.0000...	21.2225988...
224	17285	-151.72729...	-151.62729...	-16.494928...	-16.394928...	416.000000...	9057.00000...	21.7716346...
239	17488	-151.72729...	-151.62729...	-16.594928...	-16.494928...	597.000000...	16575.0000...	27.7638190...
222	17279	-152.32729...	-152.22729...	-16.494928...	-16.394928...	1631.00000...	49641.0000...	30.4359288...
241	17490	-151.52729...	-151.42729...	-16.594928...	-16.494928...	1530.00000...	58514.0000...	38.2444444...
223	17284	-151.82729...	-151.72729...	-16.494928...	-16.394928...	1583.00000...	65179.0000...	41.1743524...
266	17897	-151.42729...	-151.32729...	-16.794928...	-16.694928...	584.000000...	25293.0000...	43.3099315...
267	17900	-151.12729...	-151.02729...	-16.794928...	-16.694928...	1472.00000...	99482.0000...	67.5828804...
284	18104	-151.02729...	-150.92729...	-16.894928...	-16.794928...	865.000000...	65370.0000...	75.5722543...
249	17692	-151.62729...	-151.52729...	-16.694928...	-16.594928...	1441.00000...	124618.000...	86.4802220...
238	17487	-151.82729...	-151.72729...	-16.594928...	-16.494928...	1380.00000...	127719.000...	92.5499999...
295	18302	-151.52729...	-151.42729...	-16.994928...	-16.894928...	902.000000...	96331.0000...	106.797117...
268	17901	-151.02729...	-150.92729...	-16.794928...	-16.694928...	7038.00000...	782852.000...	111.232168...
283	18100	-151.42729...	-151.32729...	-16.894928...	-16.794928...	6656.00000...	841445.000...	126.419020...
250	17693	-151.52729...	-151.42729...	-16.694928...	-16.594928...	8386.00000...	1061847.00...	126.621392...
282	18099	-151.52729...	-151.42729...	-16.894928...	-16.794928...	7324.00000...	1446412.00...	197.489350...
265	17896	-151.52729...	-151.42729...	-16.794928...	-16.694928...	5859.00000...	1205221.00...	205.704215...

5. In the field calculator, the first thing to do is to check the “Create a new field” option, and name it something you'll remember – in this example, I'm calling it “Elev_Std”, which stands for Standardizes Elevation. You'll also want to set the output field type to a decimal number with a precision and width of at least 5:



6. Now we're going to do a *tiny* bit of coding. Under the “Function List” open the “Math” section, then scroll down to “scale_linear” and double click on it. You should see the word “scale_linear(“ pop up in the expression box at the bottom:



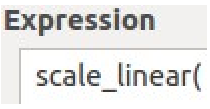


7. This is a “function” - just like the functions you were using in R. We're going to need a little bit of data to finish filling this function in: the minimum and maximum value for elevation in our dataset. For now, close the field calculator and go back to the attribute table. Find the column “ELEVmean” and double click on the top. Every time you double click it will change from “descending” to “ascending” order – i.e., the largest or smallest number will be at the very top. Write down what the largest and smallest elevation values are (**ignore any negative values smaller than -9000, as these indicate a lack of data**).

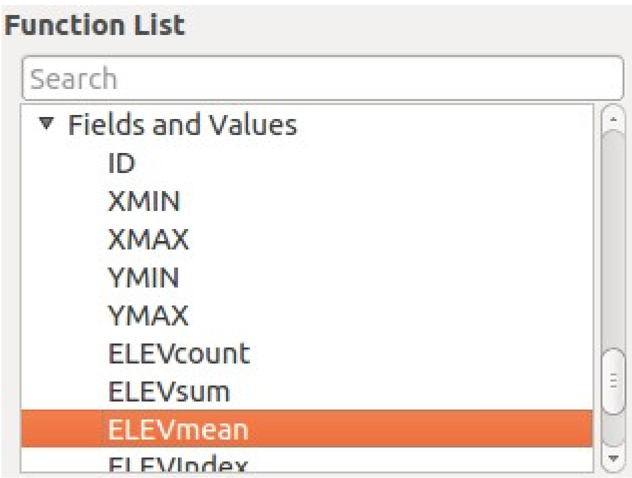
Attribute table - Lab2_LandGrid :: Features total: 485, filtered: 485, selected: 0

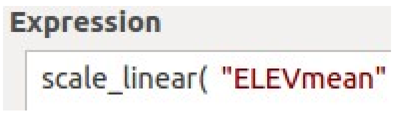
	ID	XMIN	XMAX	YMIN	YMAX	ELEVcount	ELEVsum	ELEVmean
265	17896	-151.52729...	-151.42729...	-16.794928...	-16.694928...	5859.00000...	1205221.00...	205.704215...
282	18099	-151.52729...	-151.42729...	-16.894928...	-16.794928...	7324.00000...	1446412.00...	197.489350...
250	17693	-151.52729...	-151.42729...	-16.694928...	-16.594928...	8386.00000...	1061847.00...	126.621392...
283	18100	-151.42729...	-151.32729...	-16.894928...	-16.794928...	6656.00000...	841445.000...	126.419020...
268	17901	-151.02729...	-150.92729...	-16.794928...	-16.694928...	7038.00000...	782852.000...	111.232168...
295	18302	-151.52729...	-151.42729...	-16.994928...	-16.894928...	902.000000...	96331.0000...	106.797117...
238	17487	-151.82729...	-151.72729...	-16.594928...	-16.494928...	1380.00000...	127719.000...	92.5499999...
249	17692	-151.62729...	-151.52729...	-16.694928...	-16.594928...	1441.00000...	124618.000...	86.4802220...
284	18104	-151.02729...	-150.92729...	-16.894928...	-16.794928...	865.000000...	65370.0000...	75.5722543...
267	17900	-151.12729...	-151.02729...	-16.794928...	-16.694928...	1472.00000...	99482.0000...	67.5828804...
266	17897	-151.42729...	-151.32729...	-16.794928...	-16.694928...	584.000000...	25293.0000...	43.3099315...
223	17284	-151.82729...	-151.72729...	-16.494928...	-16.394928...	1583.00000...	65179.0000...	41.1743524...
241	17490	-151.52729...	-151.42729...	-16.594928...	-16.494928...	1530.00000...	58514.0000...	38.2444444...
222	17279	-152.32729...	-152.22729...	-16.494928...	-16.394928...	1631.00000...	49641.0000...	30.4359288...
239	17488	-151.72729...	-151.62729...	-16.594928...	-16.494928...	597.000000...	16575.0000...	27.7638190...
224	17285	-151.72729...	-151.62729...	-16.494928...	-16.394928...	416.000000...	9057.00000...	21.7716346...
341	19731	-150.72729...	-150.62729...	-17.694928...	-17.594928...	885.000000...	18782.0000...	21.2225988...
240	17489	-151.62729...	-151.52729...	-16.594928...	-16.494928...	190.000000...	3174.00000...	16.7052631...
296	18303	-151.42729...	-151.32729...	-16.994928...	-16.894928...	58.0000000...	903.000000...	15.5689655...
210	17081	-151.82729...	-151.72729...	-16.394928...	-16.294928...	128.000000...	1334.00000...	10.4218750...

8. Now, go back into the field calculator and repeat steps 5 and 6 above, so that your formula looks just like this:

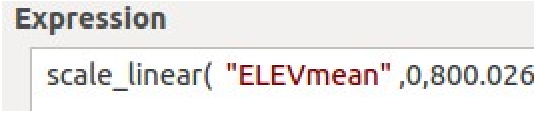


Now, we're going to choose the variable we want to scale – you'll go back to the function list, expand “Fields and Values”, and double click on “ELEVmean”. Your expression box should look like the second image below after this step.





9. Now, we want to put the minimum and maximum values that we wrote down from step 7 – this tells the program what numbers to standardize between:

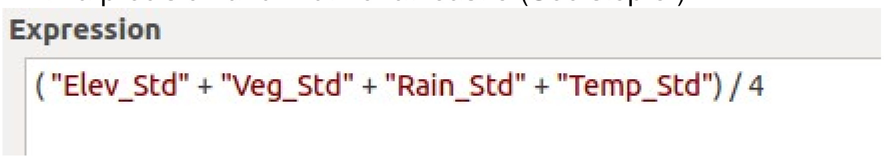


10. Finally, we put the numbers we want to “rescale” to, in our case 0 to 1. Close the parentheses and click OK:

11. Check your attribute table – you should now have a new attribute called “Elev_Std”. All of the values should be between 0 and 1, with higher elevations having higher numbers.

12. Repeat this for each of the remaining variables - vegetation, temperature, and precipitation.

13. Finally, we're going to use the map calculator to add these fields together. This is very straightforward, but note your variable names may be slightly different depending on how you named things in step 12. Call this new variable whatever you want, but make sure it is also a decimal number with a precision and width of at least 5 (See step 5!):



14. Take a look at your attribute table – you should now have a final variable that is a summary of all four of the variables. Using the steps in part 2 (starting with step 16), make a map symbolizing this final index. Using the attribute table and sorting by ascending/descending order, click on each row to identify where the best and worst conditions are for human habitation.

Lab Deliverables

The final version of this lab should be turned in on a single word document. It contain the following:

1. Answers to the following questions:
    1. In the index with four components (Elevation, vegetation, rainfall, temperature), what is the final index value for the "best" grid cell for human habitation? The worst?
    2. What is the highest elevation in the region, measured in a grid cell?
    3. What is the highest temperature in the region, measured in a grid cell?
    4. What is the lowest rainfall in the region, measured in a grid cell?

2. A single map showing your final index, which should include *at minimum*:
    1. A legend
    2. A scale bar
    3. An indicator of which grid cell is the lowest value in the final index.
    4. An indicator of which grid cell is the highest value in the final index.
    5. A base map

3. A two-paragraph description of your map, the data you included in the map, and why that data is effective in defining where you would choose to build your villa (or, alternatively, where human habitation is likely to be most successful).

Stretch Goals

1. Add importance weights to each layer – i.e., multiply each layer by a value so they are more or less important to the final index. Justify these weights with citations.
2. Add another raster layer to your analysis. For example, you might download the rasters at http://sedac.ciesin.columbia.edu/data/collection/gpw-v3/sets/browse .
3. Make a finer resolution grid (i.e., below 0.1) and conduct the same analysis – how do your results change?