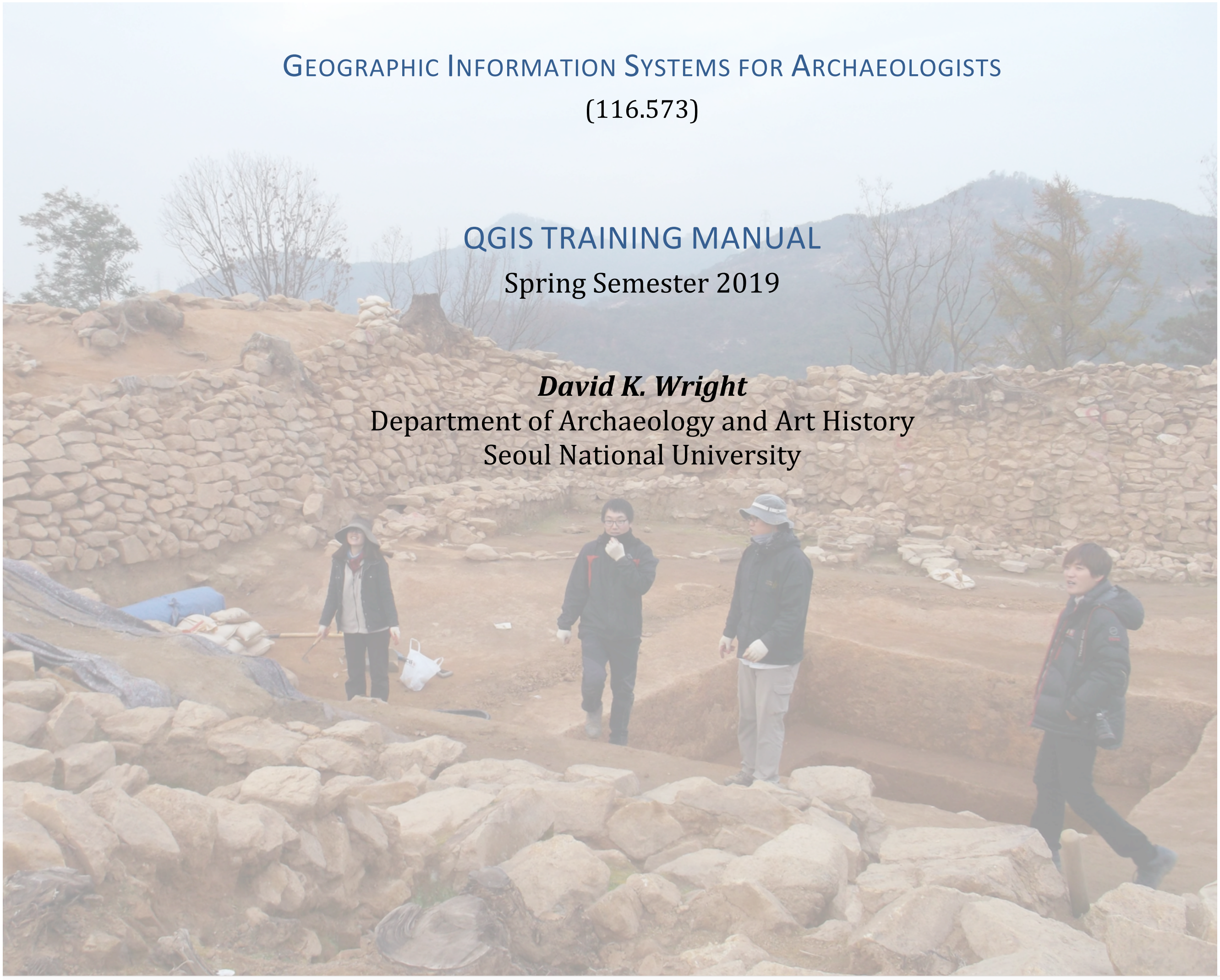
Geographic Information Systems for Archaeologists

(116.573)

QGIS TRAINING MANUAL

Spring Semester 2019

David K. Wright
Department of Archaeology and Art History
Seoul National University



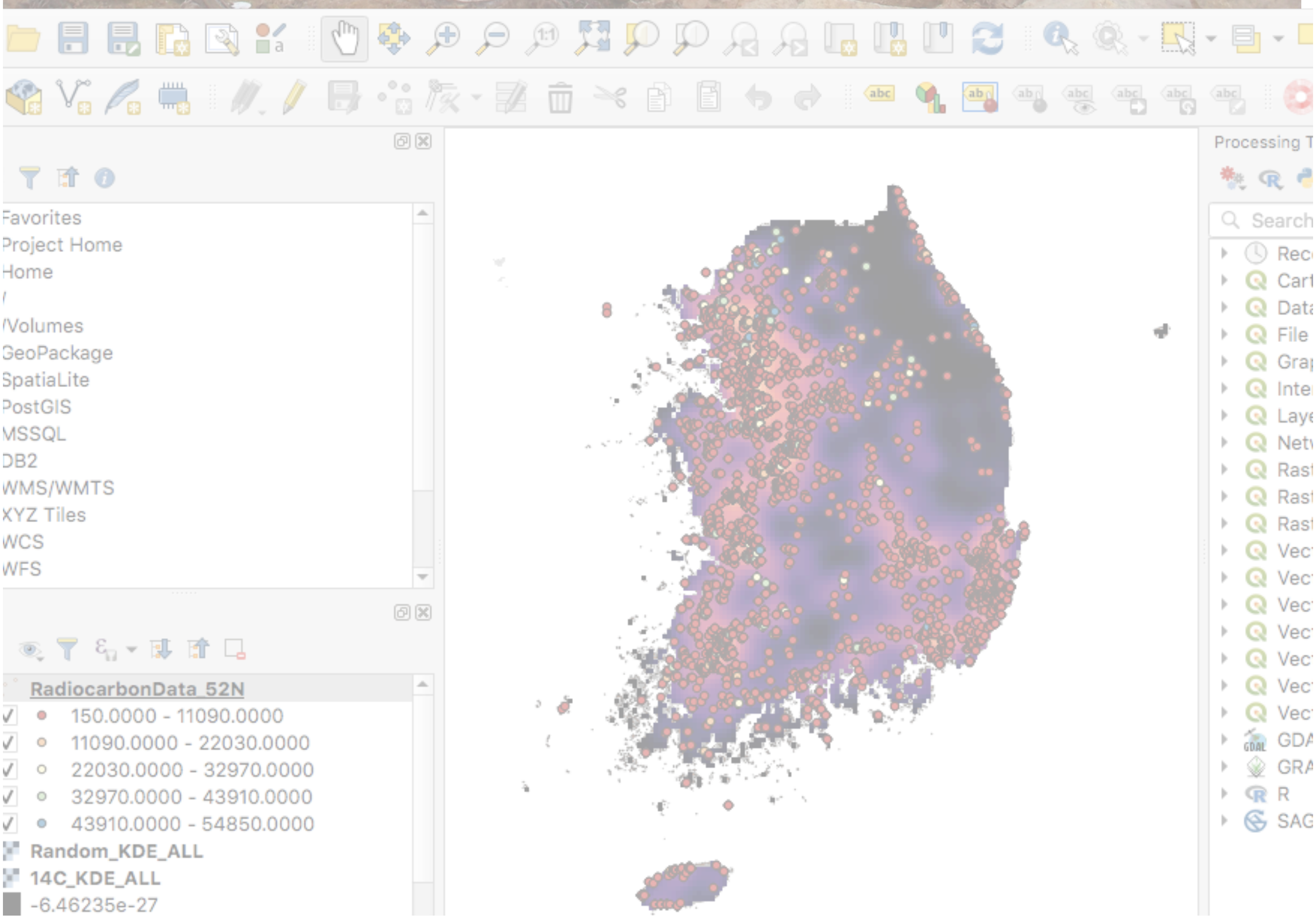


Table of Contents

NOTE ON THE PUBLIC VERSION OF THE GIS TRAINING MANUAL

This document was created for a course at Seoul National University in the Spring 2019 semester. In this document, there are references to the "eTL" which hosts files needed for the training exercises. The eTL is the university's online classroom environment, which is not available to non-SNU users.

Therefore, I have created space in my Google Drive, where members of the general public can access the files needed for the training. If you cannot access the information for some reason, please email me at daudi.wright@gmail.com and I will help you access the information you seek.

Many of the assignments refer to data that is not publically available (for example, the real 14C data from Korea, which the research group headed by Dr. Jangsuk Kim is working with). Unfortunately, these data are under study and cannot be released without prior consent. If you have a compelling reason to access the data, queries should be directed to Dr. Kim.

This manual is released in Word format to the general public and you are free to reuse and modify the content as you see fit, but this material cannot be copyrighted by any entity. I have created the content in the spirit of open dialogue and collaboration.

David K. Wright
Seoul National University
Department of Archaeology and Art History
14 May 2019

Chapter 1: QGIS Interface and Coordinate Systems

1.1. Background of the QGIS project

QGIS is an open-source Geographical Information System (GIS) platform written in C++ computer language. It is currently associated with the Open Source Geospatial Foundation (http://www.osgeo.org), which is a community of developers committed to providing a highly functional GIS environment with a Graphical User Interface (GUI) at no cost to users. There are a whole range of webinars and resources available for users on the website.

Compared to ArcGIS (https://www.arcgis.com), which is the most commonly used proprietary GIS platform, QGIS provides most of the functionality and ease of use as the industry flagship. In addition, QGIS has a robust user community who develop plugins (https://plugins.qgis.org), usually written in the Python computer language, for users to download. The plugins provide additional, targeted functions that specific users will find useful, but are not of general interest to most users. For example, pyArchInit (https://plugins.qgis.org/plugins/pyarchinit/) is a plugin that structures common archaeological database construction including stratigraphic units, sites, chronologies, taphonomy and multimedia interfaces. Since most of the users of QGIS are not archaeologists, this plugin is not of general appeal to the community, but is of great value to users like us.

Because it is an unpaid developer-based platform, QGIS has evolved as a hybrid program that blends together several different open-source initiatives. For example, GRASS GIS (https://grass.osgeo.org) is a powerful, non-GUI GIS platform that is designed to interface with other open-source software such as R (https://www.r-project.org) and operate in 3D-raster (voxel) environments. Users with familiarity in Ubuntu or Linux command-lines will feel comfortable in GRASS. The most recent version of the software also has a GUI. Since GRASS and QGIS are both part of the Open Source Geospatial Foundation, all of the GRASS tools can be installed into QGIS, which provides the user with significant statistical operability.

1.2. Installing the Software

Go to www.qgis.org and download the most recent version of the software.

Fig 1.1. QGIS download interface.

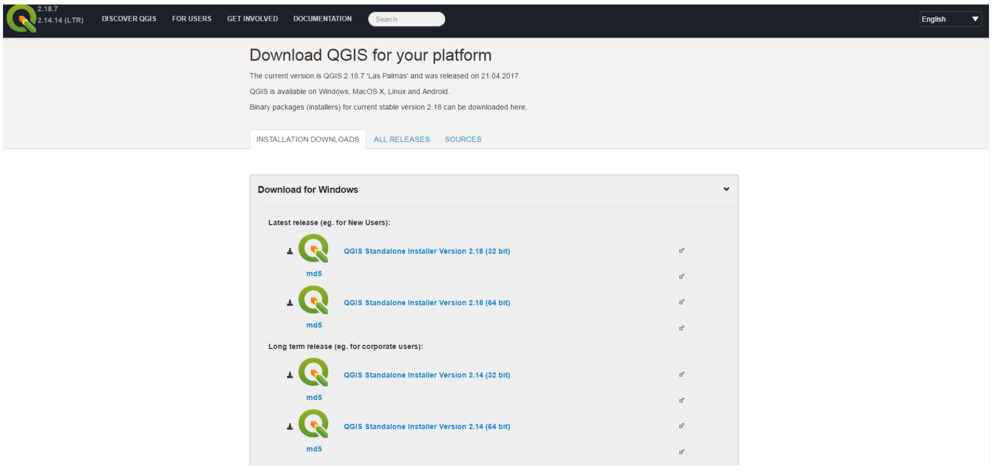


Fig. 1.2. Install your software.

Install the program into your C:/ drive (Programs folder).

Click on the Q icon ( ) at the bottom of your desktop. The program will open (Figure 1.3) and you will create a new project by clicking the " " icon in the upper lefthand corner. You now have a blank template in which to begin working.

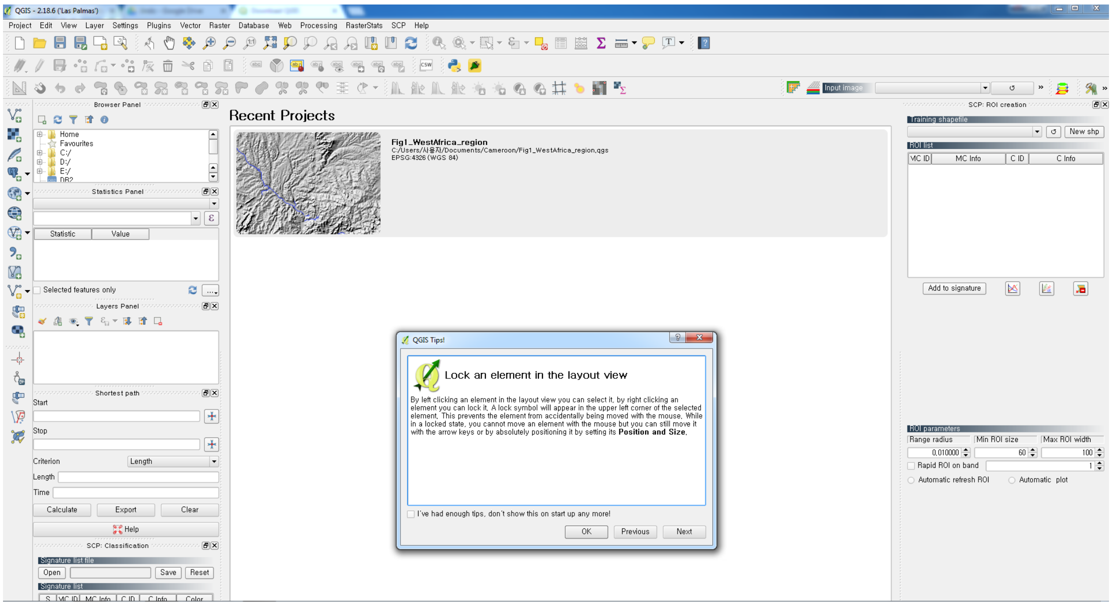


Fig. 1.3. Open the program.

1.3. Plugins

One of the best features of QGIS is the robust set of free developer plugins that can be installed to help you with particular aspects of your work. Go to Plugins > Manage and Install Plugins… to see the full list. It is usually easier to type a keyword for something you want to do.

For example, later in the course, you will be checking the **topology** of polygons you will be creating. Topology is the spatial concordance of information within your database. There are a number of plugins that can assist you with this task. In your Manage and Install Plugins window, type "topology" to see a list. Download the Topology Checker plugin and after it downloads, click the icon to see a tool button appear in your tool menu (left hand side of your screen).

1.4. The QGIS Interface (General Overview)

QGIS is comprised of a GUI (graphical user interface) that can be generally subdivided into three components (Figure 1.4): (a) the **toolbars** provide shortcuts to the basic functional tools that are used in a GIS environment. The top horizontal bar includes saving, viewing and calculation tools. The second row includes **line** and **point** tools for vector-based operations. Toward the right of these, labeling functions are hosted and the far right includes a tool that allows you to access the Plugins and undertake a Principle Components Analysis quickly. The final row includes tools for measuring, creating and manipulating **polygon** vectors. The remaining tools to the right are for **raster** manipulation and the generation of raster statistics.

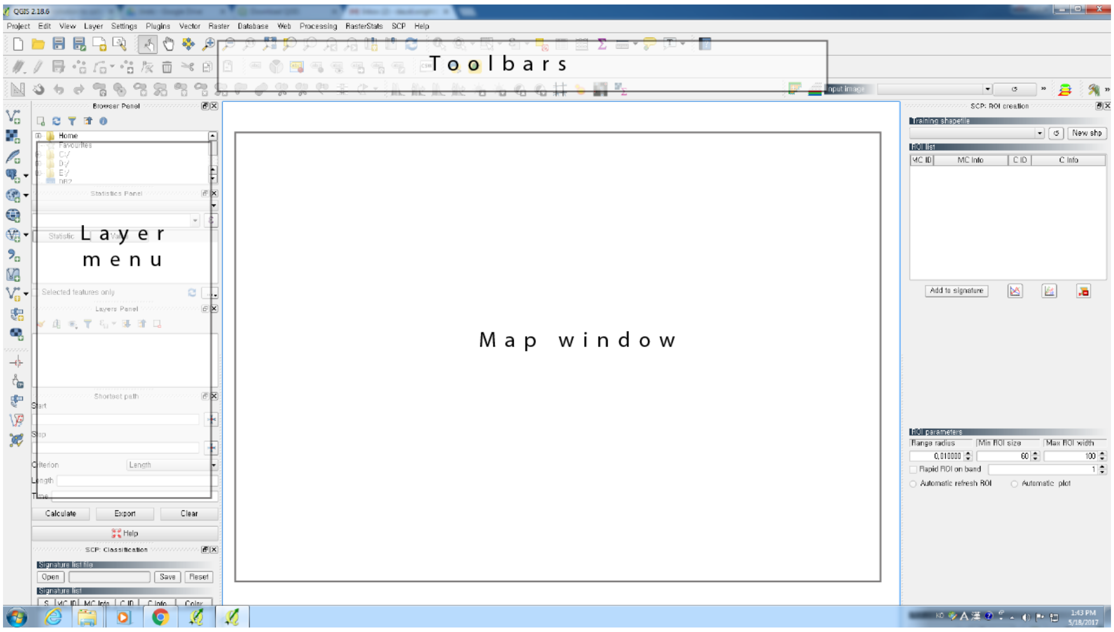


Fig. 1.4. GUI interface components to QGIS.

On the left vertical side of the screen are tools for adding vector, raster and various types of database files. You can also import GPS data and perform various calculations between multiple types of vector files from this toolbar.

The **layer menu** allows you to quickly access your files from the computer and/or network. The panels give you a quick view of the files you can work with. Clicking and unclicking the files only affect the viewing environment of QGIS. A quick view of statistical outputs can also be viewed and copied here.

Finally, the **map window** shows the user (you) the layers that are active and operational in QGIS. This view is not printable or exportable in this environment because GIS is not explicitly about making maps—the operational environment is designed to allow users to visualize and calculate relationships between datasets. Exportable maps can be created, but we will not worry about that initially.

Now that you know the basics, let's get started! In the project drop-down menu at the top left-hand corner of the QGIS window, select *Project Properties*. There are a number of menus here, which range from simple to complex in application. The most important thing for you initially is to "Enable 'on the fly' CRS transformation," (Figure 1.5) because it will give you the most flexibility when dealing with different coordinate systems simultaneously.

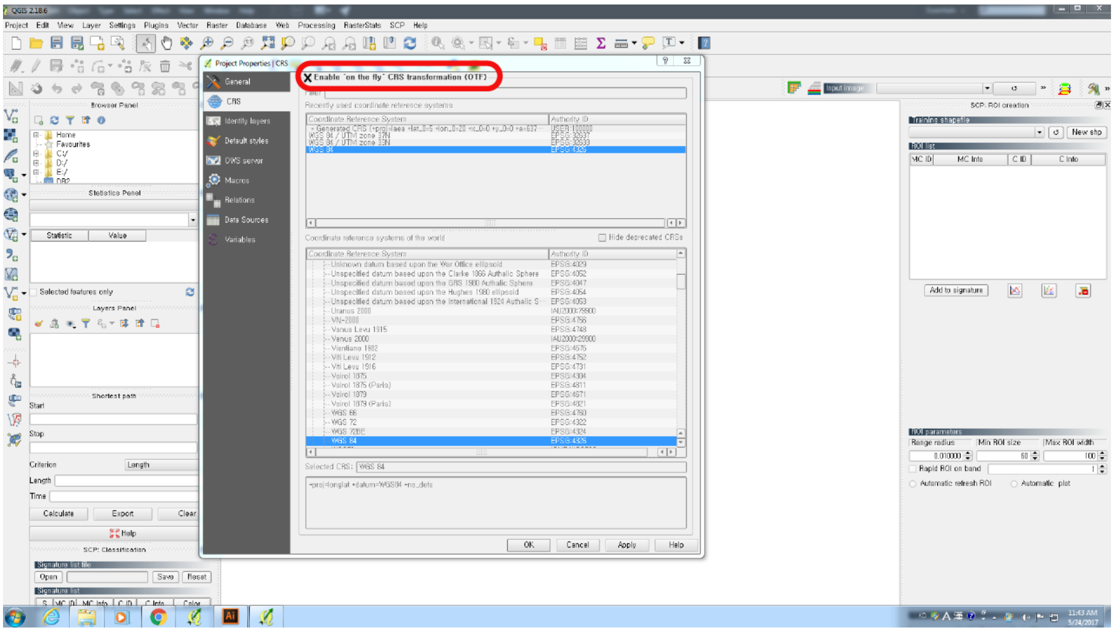


Fig. 1.5. Project properties window.

The most important thing with any GIS is to maintain good file organization. GIS works by linking files together in a single operating environment and it is very difficult to move files around a system without creating a mess in your GIS operating environment. The project files you create (*.qgs) will not have any self-contained data—it only serves as a map of paths to where your data is hosted.

Therefore, you will want to create a simple, but expandable architecture in your hard drive or network that can be modified through time. An example of such a basic structure is provided in Figure 1.6.

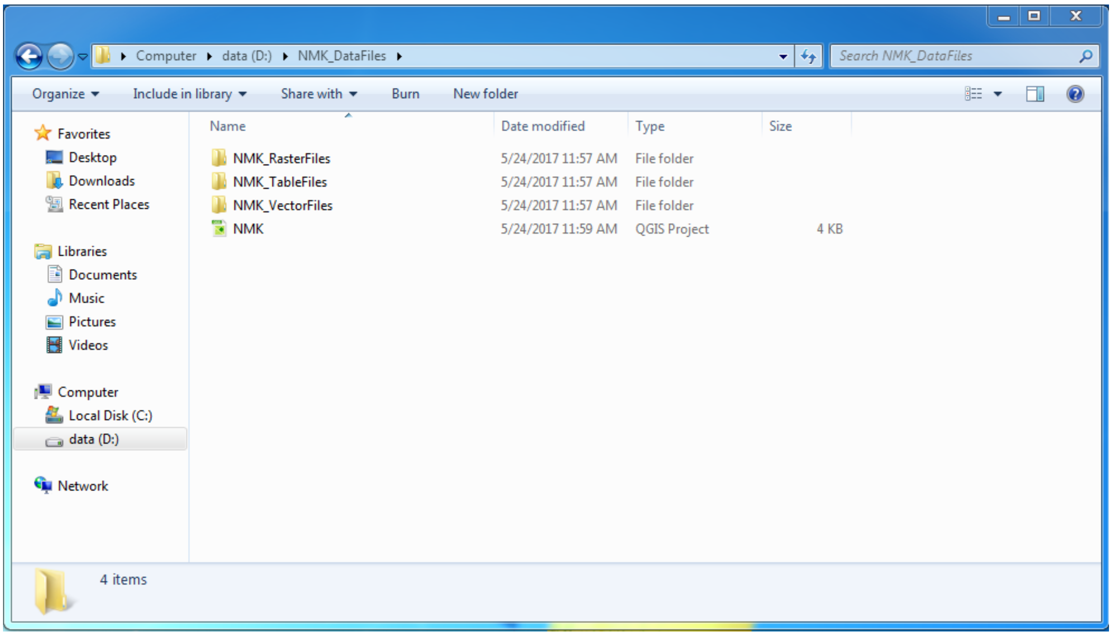


Fig. 1.6. Example of flexible data structure architecture in setting up your GIS.

1.5. Dealing with Coordinate Systems

There are many ways to measure space and one of the challenges of representing space in two dimensions is that the earth is a three-dimensional object! Translating three-dimensional space into two dimensions involves "projecting" the space. There is a good summary of projections here: https://www.gislounge.com/map-projection/.

Reprojected space is measured in **coordinates**. The coordinate system you are probably the most familiar with is **latitude** relative to the equator and **longitude** relative to the prime meridian in Greenwich, UK, but you probably aren't aware that since 3D space is projected in different ways, there are different interpretations of which coordinates apply to the space you are sitting on. This is because with different topographic and gravity models of earth's surface, a radian degree (the units that latitude/longitude are measured in) is not always the same distance on the ground because all radians are measured up to 360°. Until recently, there was no precise way to measure the entire circumference of the earth and there are many legacy coordinate systems that reflect old imprecisions. Here is a link nice summary of coordinate systems, projections and the issues involved with translating spatial dimensions.

In the lower right hand corner of your Map Window, you will notice that there is a little gray circle after which it says, "EPSG:4326." This is your coordinate system and datum (lat/long using the WGS84 datum....Google it if you don't understand what this means). This is the default coordinate reference system (CRS) used in QGIS and most GIS programs.

But, there are many instances where you will not want to have your data in EPSG:4326 (cf., this website). This is because archaeologists typically work in meters, but EPSG4326 is in degrees radian, which uses a geographic coordinate system and is difficult to translate into meters. For the most part in this class, we will be working with a different coordinate system (EPSG3093): UTM Zone 52N in which the measured units are in meters. It is always best practice to reproject all of the imagery you collect into the CRS you are using.

In this first exercise, we will get a digital elevation model (DEM) from a public repository and reproject it. Let's go to https://earthexplorer.usgs.gov. You are going to need to create a user name and password in order to download data. Once you have done that, navigate to the Korea region and create a polygon box around Seoul (Figure 1.7).

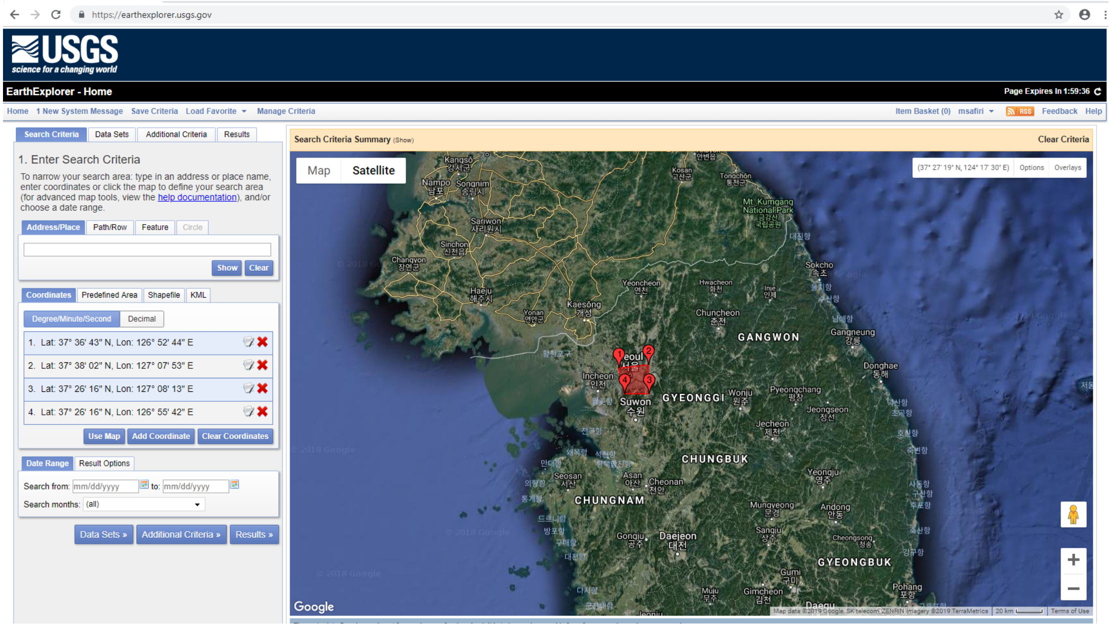


Fig. 1.7. The Navigation pane in the USGS's Earth Explorer (https://earthexplorer.usgs.gov).

Just to make things easy, let's all perform exactly the same parameter search on Earth Explorer. Hit the tab that says "decimal" under "Enter Search Criteria" and you will see that your latitudes span from about 37.60°N to 37.45°N and longitudes are between 126.87°E and 127.14E°. Now click "Clear Coordinates" and then click "Add Coordinate." A box will pop up that says "Add New Coordinate." Enter all of the coordinates above to make a manually bounded box (Figure 1.8).

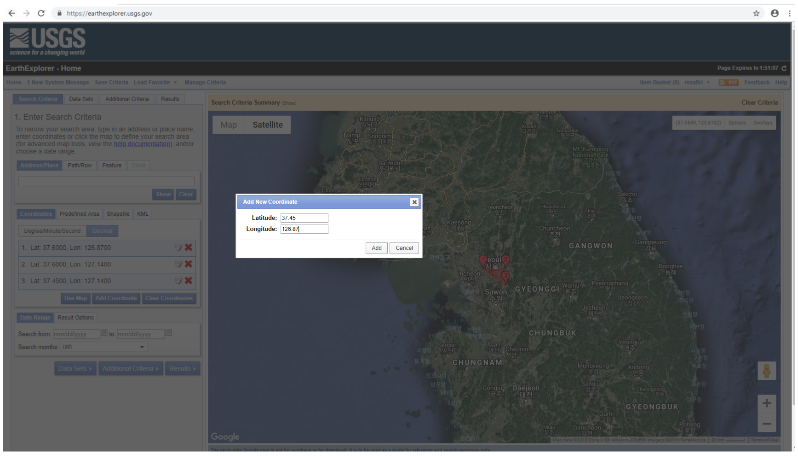


Fig. 1.8. Manually add coordinates in the USGS's Earth Explorer Navigation pane.

Now go to the "Data Sets" tab and scroll to "Digital Elevation" and select "ASTER GLOBAL DEM." (Figure 1.9). Click the "Results>>" tab at the bottom of the data selection menu.

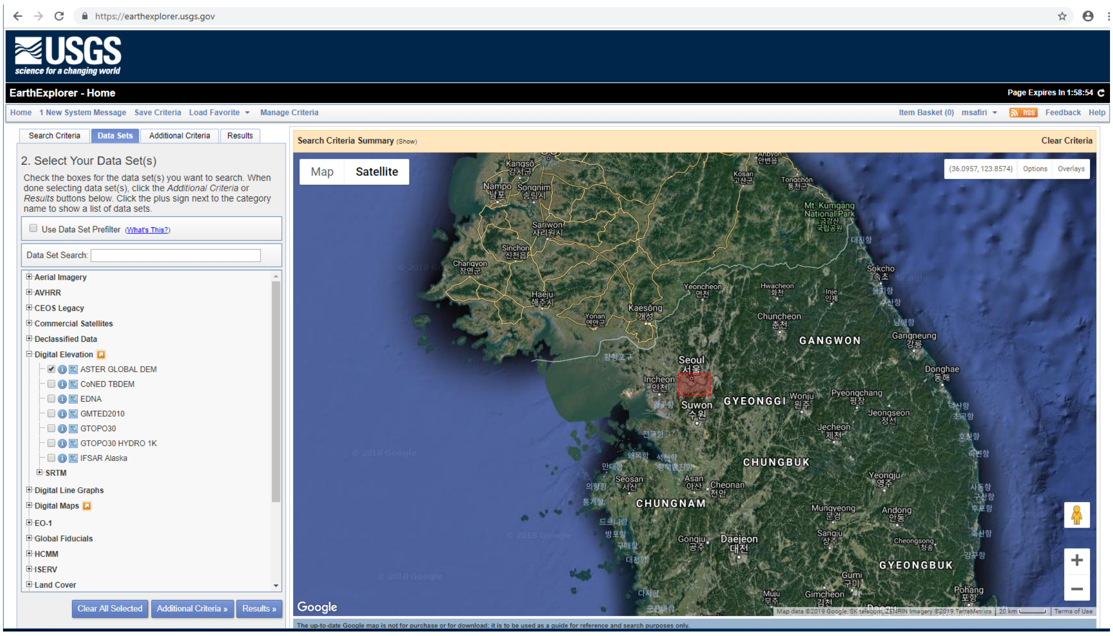


Fig. 1.9. Select the data platform you want to download in the USGS's Earth Explorer.

You will see two DEMs appear. Click the little feet to see the footprint of where the DEMs cover (Figure 1.10). You can zoom in on the map using the mouse wheel or the + sign in the lower right hand of the map display window. Click on the ASTGDEMV2_0N37E126 entity. It will download as a zip file into your downloads folder on your desktop.

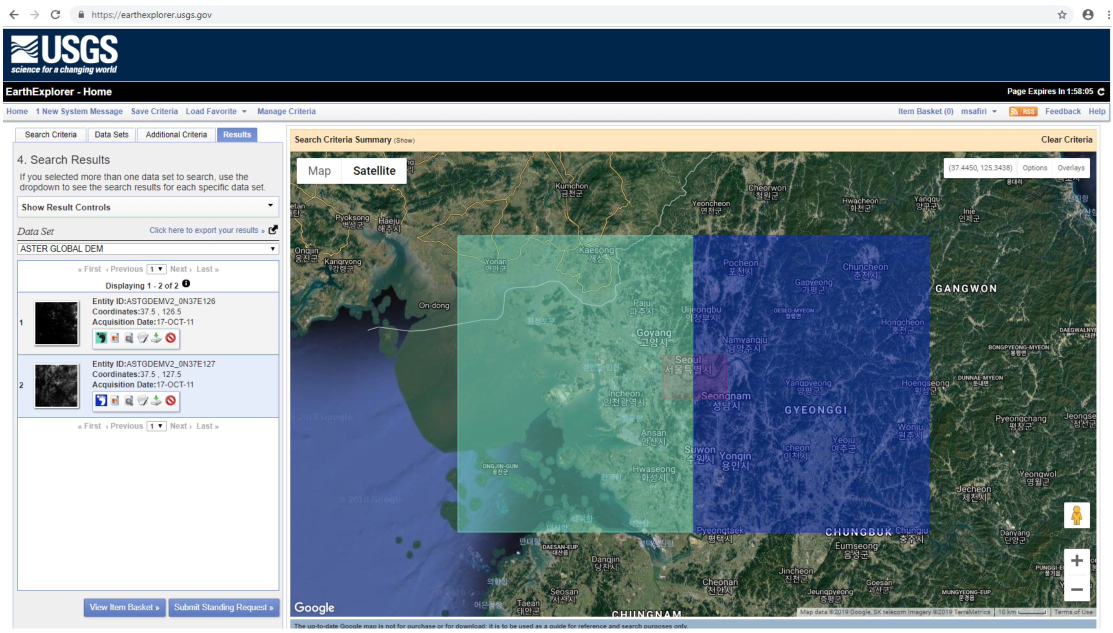


Fig. 1.10. View the data coverage of your selection in the USGS's Earth Explorer Navigation pane.

You will want to save the zip file in your D: drive so you can work with it. In Figure 1.11, you can see how I saved mine. I recommend that you have three master folders as shown in Figure 1.6 and since I work in many places in the world, I will keep mine separated in a folder named "Korea." I recommend that you just make a Korea country folder in your D: drive even if you plan to work your entire career in Korea. (We are going to make a Malawi folder later in the semester anyway. Later, you will also create a “Global” folder for bigger, global data sets.) This will allow you to keep things conceptually separated. DEMs are raster files, so save the zip file in your Rasterfiles folder. Unzip the file in the same location and delete the zip file. You will see three files: two tiffs and a PDF. If you click on the PDF, there will be a description of what you have downloaded.

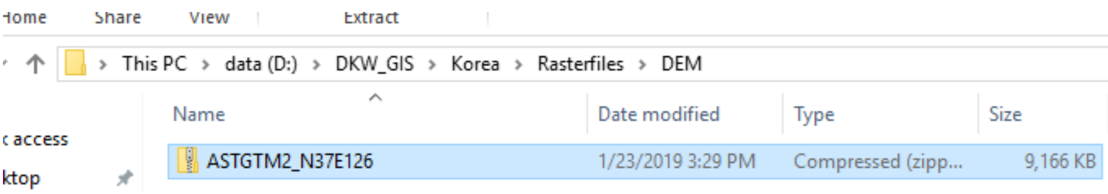


Fig. 1.11. Save raster data in your D:/ drive.

We want to work with the DEM in QGIS. The simplest way to open it up is to drag the DEM (not the QA file) into the Layers Panel of QGIS (read the PDF if you don’t understand this). You can also add the layer manually by clicking Layer>Add Raster Layer… and navigating to the location on your drive (Figure 1.12). Note that the EPSG is 4326 and the Coordinate window at the bottom registers the coordinates in latitude and longitude.

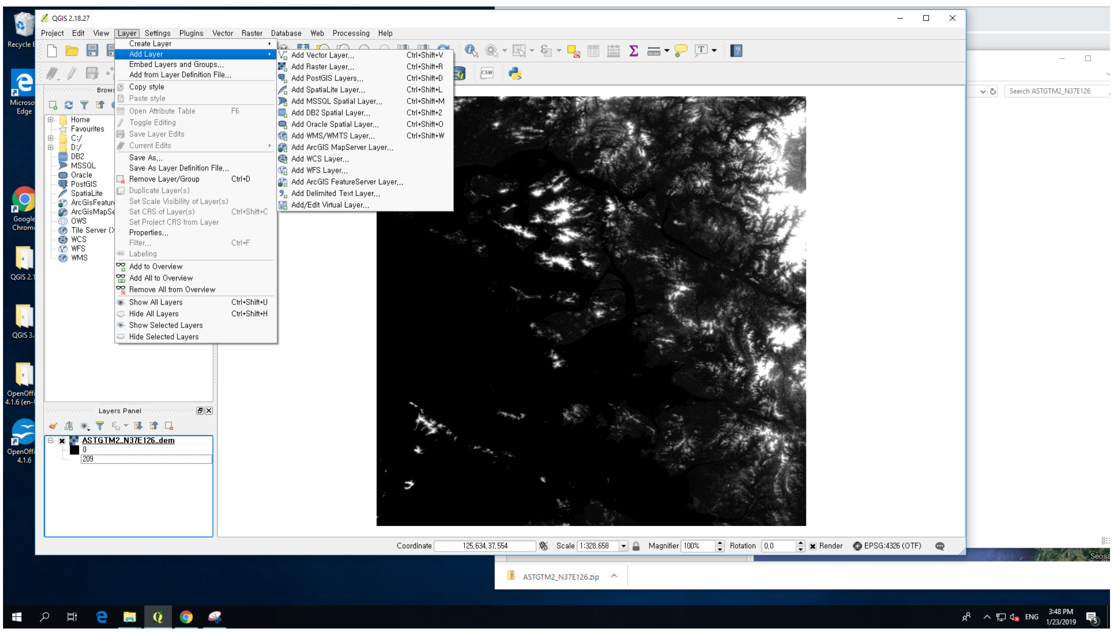


Fig. 1.12. Add raster data to the Layer panel in QGIS.

We want to reproject the image into a new coordinate system, but first let’s change the coordinate system in the project properties (Project>Project Properties…). In Filter, type in 3093 and you will see the Tokyo / UTM 52N option appear. Click that and then click Apply. When you close the window, you will notice that the DEM has changed appearance slightly. Also notice that the EPSG is different and the Coordinate window has a different set of numbers. This is all good. It means that “On The Fly” reprojection you selected earlier in the exercise is working.

However, when working in a GIS, there are many instances where you will need to have the file reprojected into the coordinate system and saved as a different file for analytical purposes. In QGIS, this is done quite simply in an open source application called "GDAL." Click Raster>Projections>Warp (Reproject)… to open a dialogue box (Figure 1.13).

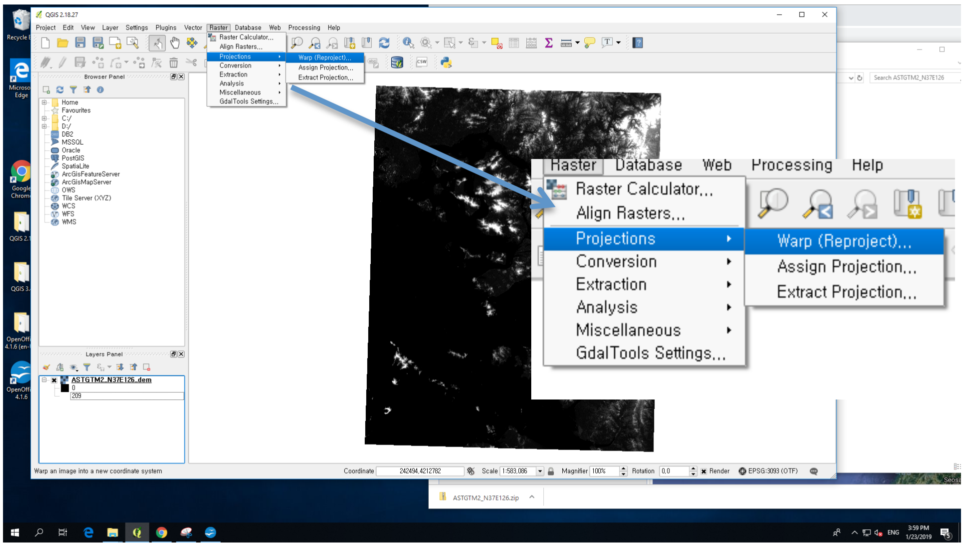


Fig. 1.13. Reproject the coordinate system using GDAL in QGIS.

In the dialogue box, you see some options. In this instance, the Input file is selected automatically for you. You will want to create a new output file by navigating to the file location and renaming your file with "UTM" or "UTM52N" at the end of it. The source SRS is automatically selected for you since QGIS can detect what the native coordinate system of the file is. You need to change the Target SRS to EPSG:3093. If you click Select next to this option, it will probably be one of two options available to you. Resampling is important. For DEMs and raster images, I generally prefer to use Cubic or Cubic spline. If you use the default Nearest neighbor or Bilinear methods, your image will look really choppy at close resolution. If you clicked on the PDF after you unzipped the DEM, you will read on Page 5 that no data values are -9999 (0 is water, and that is data you will most certainly want to have included). So, change your "No data values" parameter to -9999. Figure 1.14 shows you how everything should look before you click OK. Once you have clicked OK, you will see two files in your Layers Panel. Remove the EPSG:4326 file by right clicking it and left click "Remove."

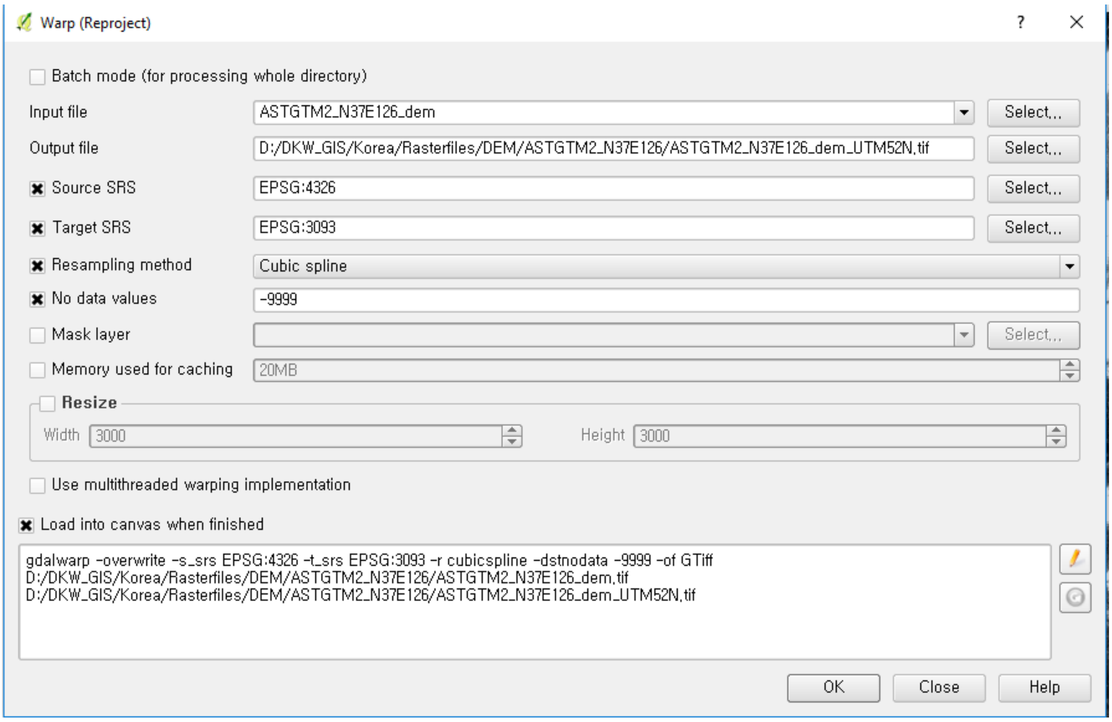


Fig. 1.14. Reprojection dialogue using GDAL in QGIS.

It's not a very pretty image just yet, so let's make it nicer. Right click the DEM file in the Layers Panel and go to Properties. In the Properties, click Style in the left panel. In Render type, change the type to "Singleband pseudocolor." If you click the Color palate, you'll see that there are many options to choose from. I like the default (Spectral) one, but I inverted the color so that red represents the high spots and blue is the low spots. The problem is that anything below 1m is the ocean, which is typically depicted as blue. The easy way to change the display so that only water is blue is to click on the 51.5 value in the classification window, then hit the green + sign and add the value of 1. It will give you a default color of pink or something horrid, so double left click on the color tab and adjust the color so that it is a lighter color of green than the 51.5 class (Figure 1.15).

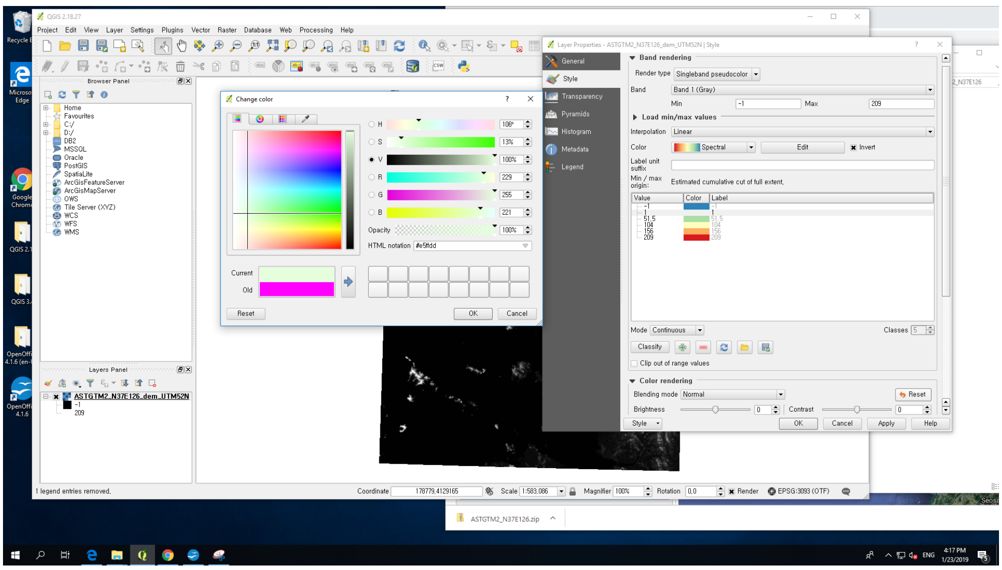


Fig. 1.15. Adjust the color display.

Click all of the OKs and you should have a nice image similar to what is shown in Figure 1.16. Be sure to save your project somewhere. I like to have a separate "qgs" folder in my root directory where I save the qgs folders.

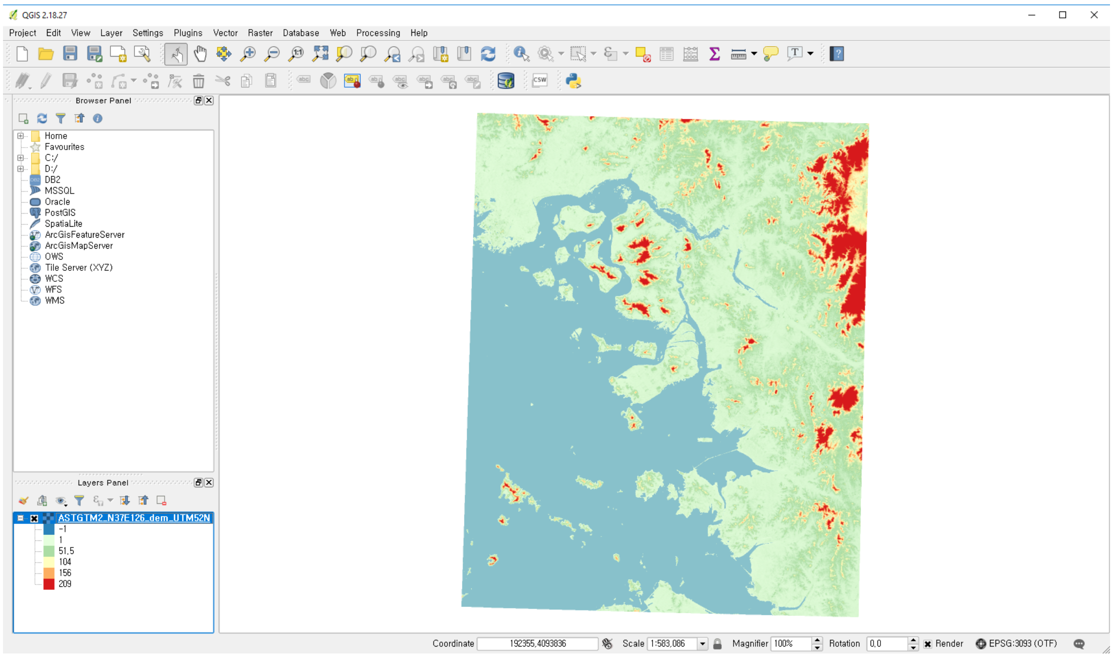


Fig. 1.16. The final display of your DEM.

1.6. EXERCISE

Download the adjacent DEM (ASTGTM2_N37E127) and add it to the project you have started. Rescale the image so that it matches the coordinate system of the E126 file. Adjust the views of the images so that they are the same (Figure 1.17 should be your final image). You will need to adjust the Mode from Continuous to Equal Intervals and then adjust the values accordingly. You will need to add some values for the maps to make sense. The important thing is that I shouldn't see a seam in the image when it is fully zoomed out.

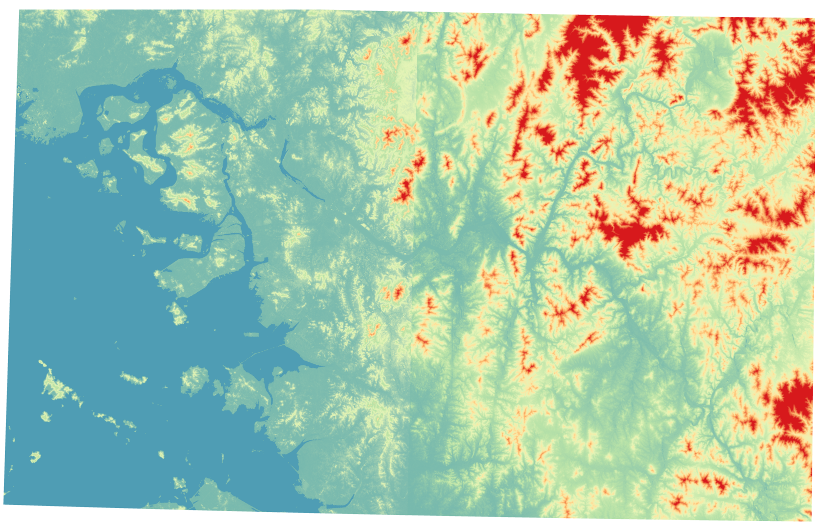


Fig. 1.17. Matching DEMs of Seoul/Gyeonggi-do seamlessly in QGIS.

Export the final view as a .png file. Click Project>New Print Composer and give it a title. Click the "Add new map" icon on the left side of the panel then drag the cursor across the screen while holding the left mouse button. To adjust the view, click Item properties in the right panel and go down to Extents. I have provided an example in Figure 1.18 on how to adjust the extents to something appropriate. Then click the "Export as Image" icon (also found under the Composer menu). Save the file as a .png (*YOURNAME*_Ch1.png) and **upload it to the eTL by March 11 at 9:00 am** (5 points for completing the assignment correctly and on time). **Your image should look pretty similar to the example I have provided here,** so you will need to do the proper adjusting to get it to work. If you get stuck, there are lots of online resources to help you figure it out (see Appendix A), and, of course, you can ask me. Ideally, you should get used to solving your own problems, though, so just try to figure it out. In addition to the image, answer the following questions (saved as *YOURNAME*_Ch1.docx):

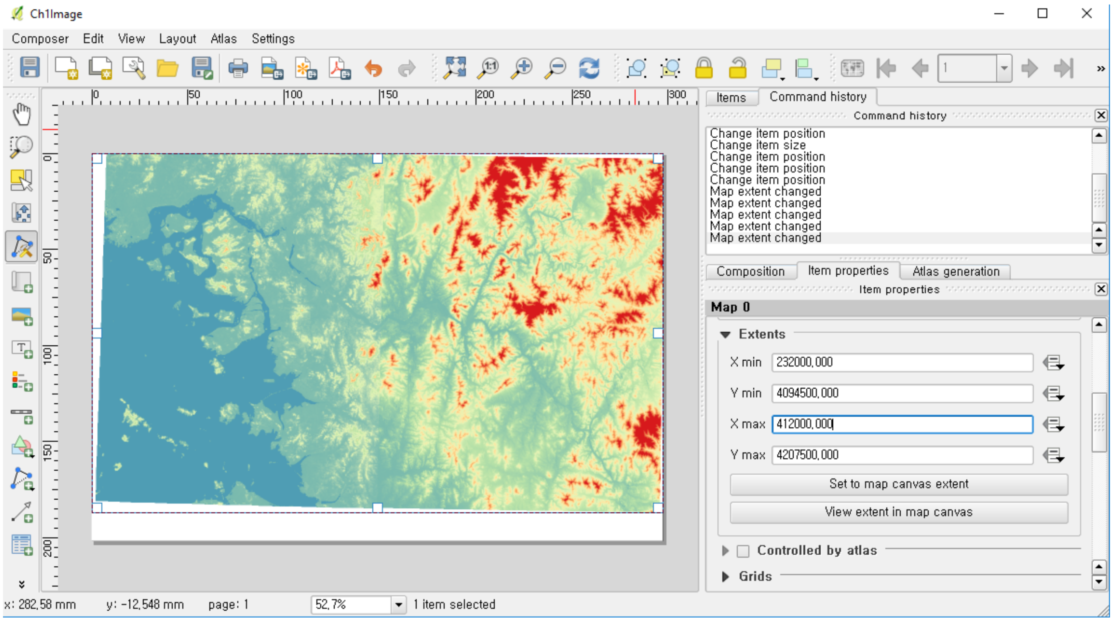


Fig. 1.18. The Print Composer window in QGIS.

STUDY QUESTION (1): What does a GIS provide a user that a map does not (refer to Conolly 2016 and Chapter 1 of Conolly and Lake 2006)? (2 points)

STUDY QUESTION (2): Provide one example drawn from Conolly (2016) that illustrates how you could use your DEM to discuss the landscape archaeology of the Seoul/Gyeonggi region. In other words, how could you apply your DEM to develop a landscape archaeology project? (3 points)

Chapter 2: Data Layering and Making a Map

2.1. Cartographic Principles

The art of map making is called **cartography**. It is important to emphasize that making a map is one output of GIS, but it is not normally the most important output. As you will learn in this class, GIS provides a power set of analytical tools that are used to better understand spatial relationships within a temporal setting.

The important thing about making a map is to have an idea of what you want to create before you start the creative process. Part of this is to know what information you want your map to convey. So, if you want your audience to know the locations of archaeological sites as they relate to modern infrastructure, you will need to create a map that has all of those elements in a clearly **legible and scalable** manner. You need to pay attention to avoid visual clutter, which can obfuscate your map. Present imagery in clearly contrasting colors that stand out from one another. Be sure to include a scale bar, north arrow and legend to communicate the basic information you want to convey. There is a nice overview of basic cartographic principles in this online PDF.

2.2. Working with Multispectral Imagery

Landsat data is tricky to deal with because it is **multispectral**. Multispectral data capture multiple wavelengths of reflected light from earth's surface ranging from spectra you can see with the naked eye to those you cannot. We will work with multispectral images a lot in this class. The first thing you are going to want to do is download Landsat 8 files from https://earthexplorer.usgs.gov. Let's use the same area that we selected for Module 1.5. This time we want to select the Landsat Archive Collection 1 Level-1 data, Landsat 8 OLI/TIRS C1 Level-1 (Figure 2.1).

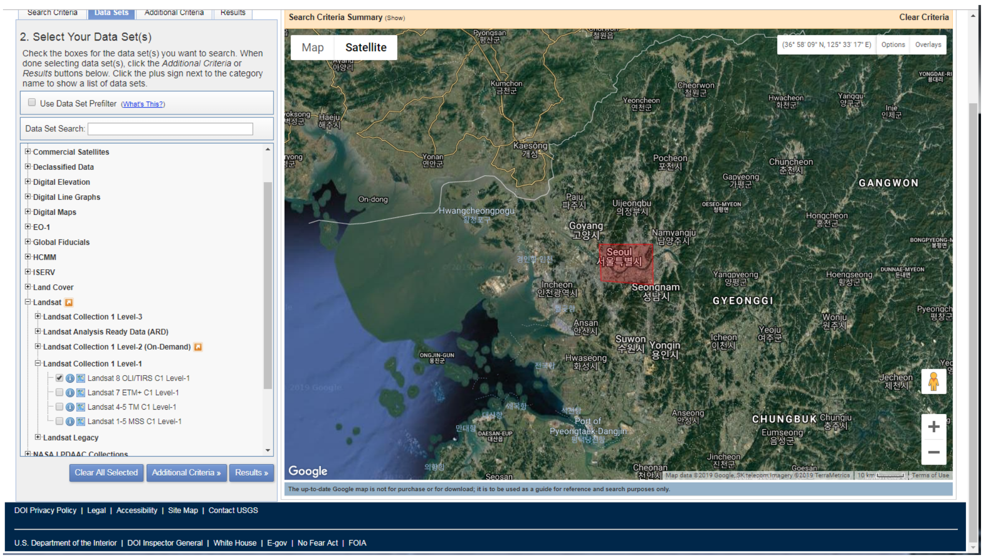


Fig. 2.1. Select Landsat archive data.

You will want to select “Additional Criteria” and be sure to select Land and Scene Cloud Cover as <10% (Figure 2.2). Then click the Results tab. Select a scene that looks good to you and has a footprint that best conforms to the Seoul/Gyeonggi-do region you are analyzing. Click the little black foot under the scene frame to see where the footprint falls. Click the Download button and select the Level-1 GeoTIFF Data Product for the image dated 20-Jan-19. Be sure to sign in if you aren’t already signed in. If you are analyzing more than one scene of Landsat data, be sure that the scenes were collected on the same day so that you get consistency in the light reflectance from the earth’s surface.

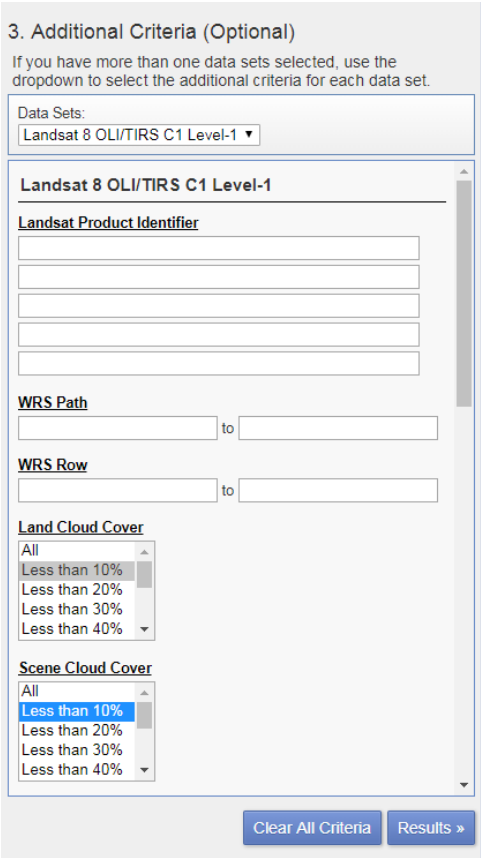


Fig. 2.2. Refine your search for Landsat data.

You are going to download a scene that will end in “.tar.gz”. These are zipped files that take about 30 minutes to download. You need to extract in your raster folder after the download is complete. I recommend making a separate folder called “Landsat8” in which you can place your zipped files. Then extract the files into new folders within the Landsat8 master folder. You will have a folder with 14 separate files as shown in Figure 2.3. They appear as black and white images in the thumbnails, but they are multispectral and the bands of those spectra are designated at the end of the file name. For a comprehensive explanation of the band types of Landsat satellites, visit this website. (I highly recommend spending a few minutes to familiarize yourself with the bands in Landsat. This subject is going to come up a lot in the class and you will not be able to work independently without some understanding of how they work.)

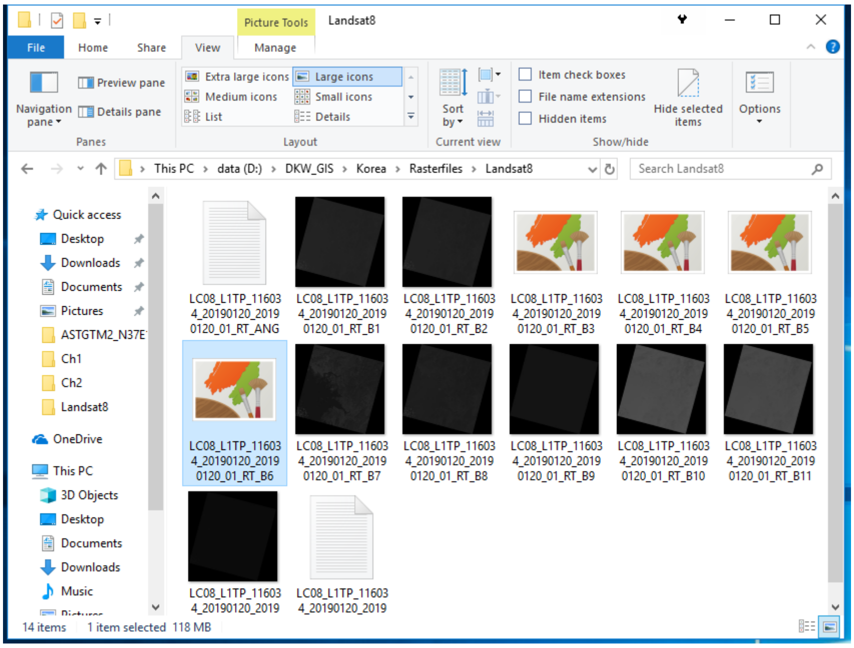


Fig. 2.3. Multispectral Landsat data saved in the raster subdirectories.

In order to make these files usable, we need to add them in as separate files that can be viewed simultaneously. There is a very simple way of doing this in QGIS. Go to Raster > Miscellaneous > Build Virtual Raster (Catalogue)… (Figure 2.4).

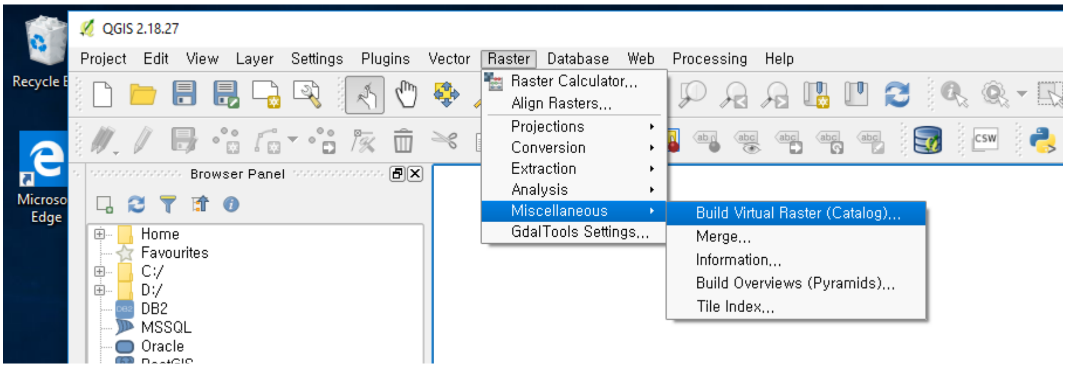


Fig. 2.4. Make a Virtual Raster Catalogue.

Here is a nice series of suggestions for how to layer up Landsat 8 images for different purposes: https://blogs.esri.com/esri/arcgis/2013/07/24/band-combinations-for-landsat-8/. So, following that suggestion we are going to make a Natural Color map out of Bands 4, 3 and 2. So when you build your virtual select the appropriate bands from your Landsat8 subfolder (Figure 2.5) and create a .vrt file that is appropriately named (example: Gyeonggi_432). Double check that you have selected Separate and Allow projection difference from the menu options and then click OK.

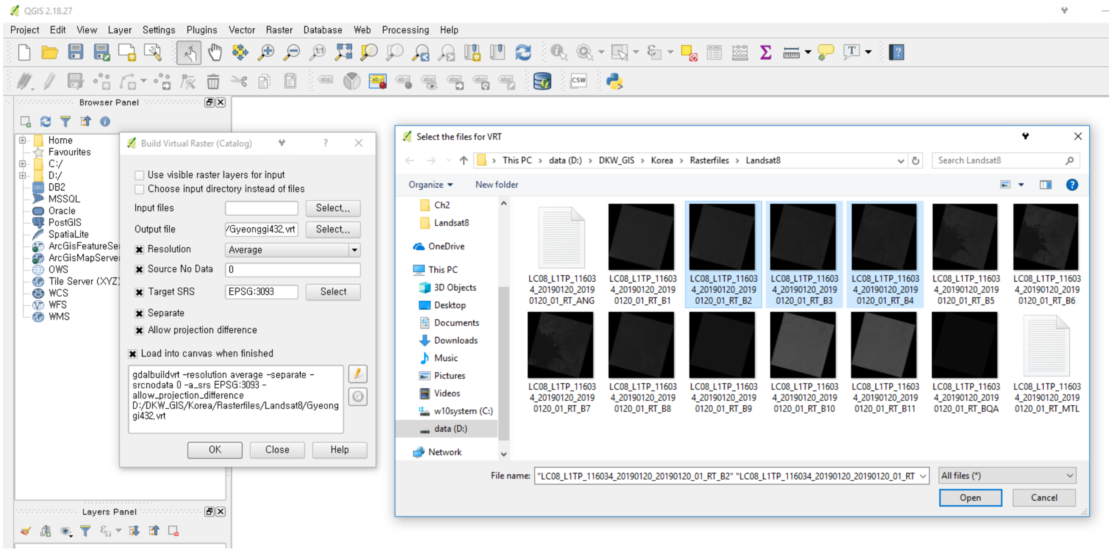


Fig. 2.5. Select the appropriate bands.

Once you have done this, you will get a strange looking image such as Figure 2.6.

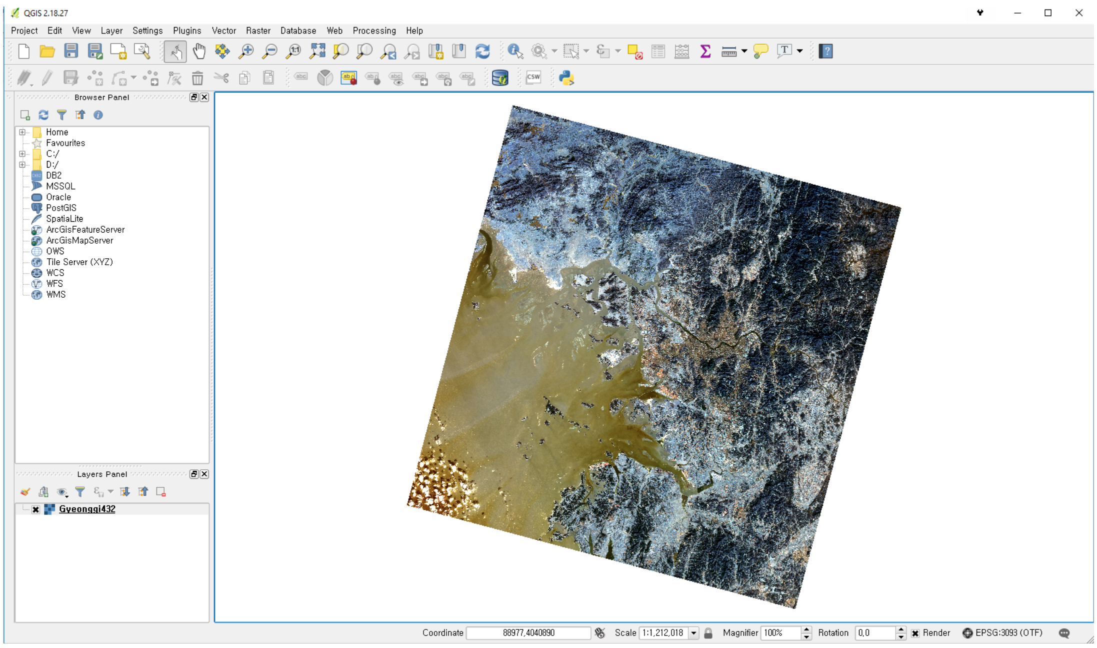


Fig. 2.6. Bands 4, 3, 2 initial output in QGIS Map Window.

You'll need to adjust the bands so that they read correctly. This has to be done manually. Open up the layer properties and you will see the band colors are automatically selected (Figure 2.7a). You will need to cut and paste the values in the R,G,B bands manually and adjust the bands as in Figure 2.7b.

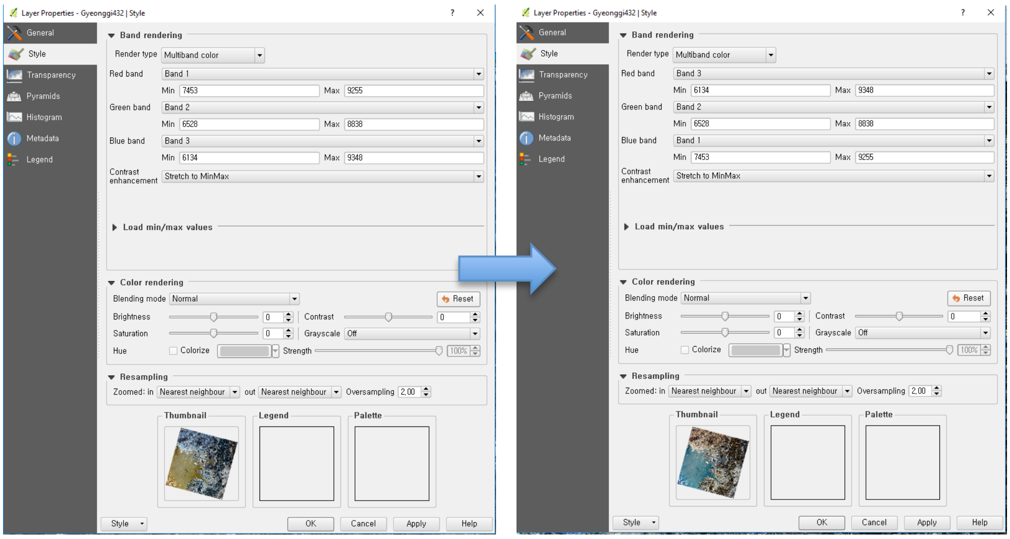


Fig. 2.7. (a) Left: Automatically selected layer properties. (b) Right: Manually adjust the bands in the Layer Properties dialogue box. Note the inversion of values between Red and Blue bands, which changes the color of the preview image in the lower left corner of the dialogue box.

Just for fun, add the DEMs you created in Chapter 1 in order to add in some hillshading to your image. This time, add the two DEMs as a VRT to your Layers panel (HINT: Add the DEMs separately to your Layers Panel, then build the VRT from the visible layers, then remove them). Create a hillshade model from the bands (Figure 2.8).

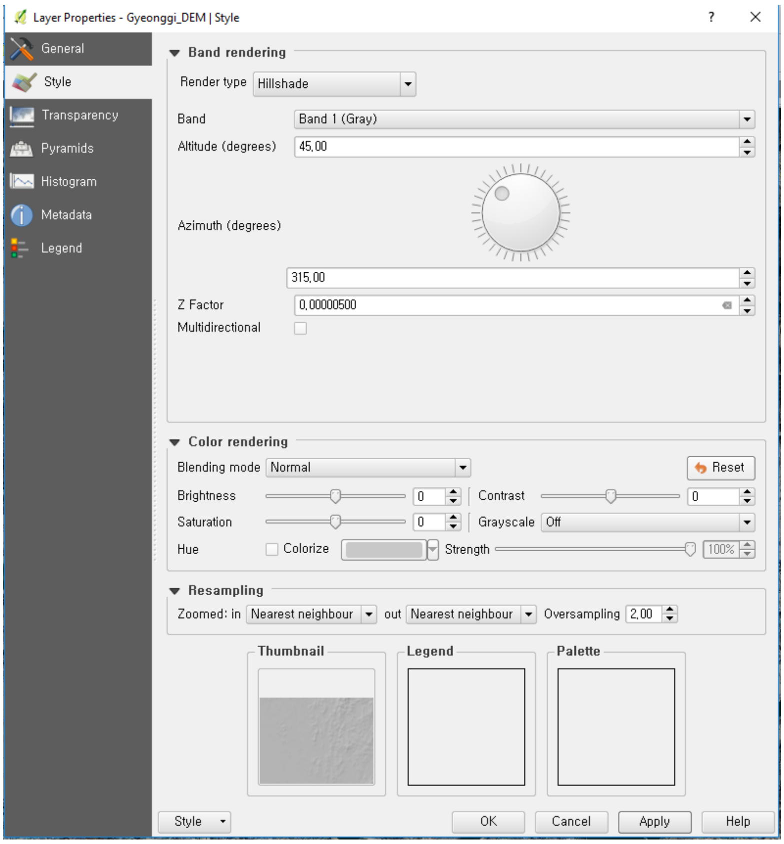


Fig. 2.8. Create a hillshade from your DEM.

Open the Layer Properties again, and for Color rendering, choose Multiply and this time adjust the Brightness to 70 and make the Contrast 20. You can play around with these numbers a little bit until you get something you like. Be sure that your hillshade map created from the ASTER DEM is the raster image directly below your Landsat 8 multispectral image in the Layers Panel (Figure 2.9).

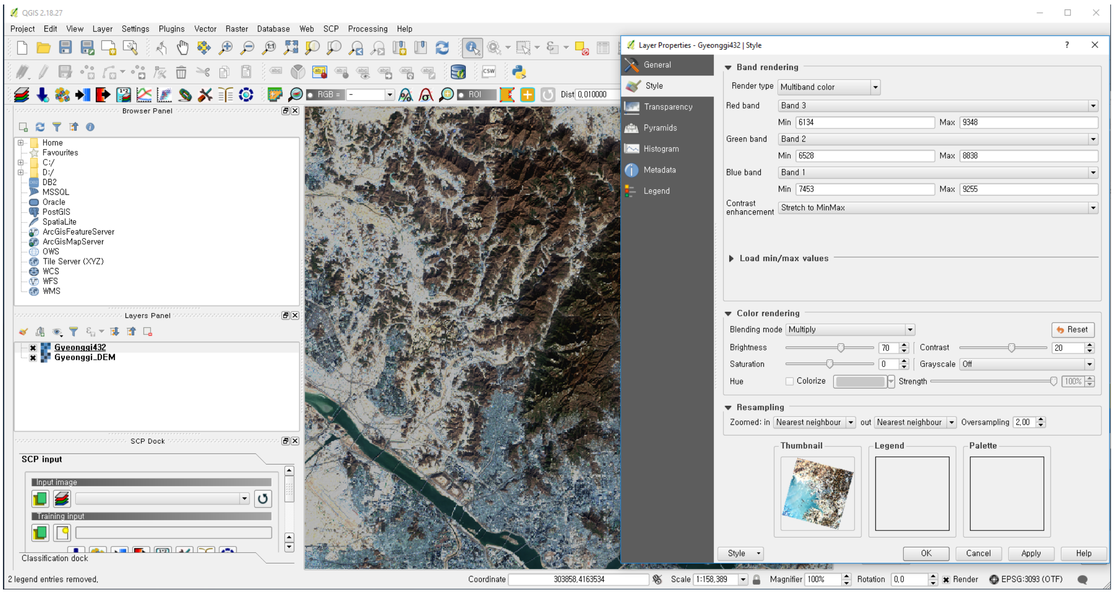


Fig. 2.9. Fine tune and layer the image for a higher level of analysis.

There is an alternative way of assembling multispectral imagery, which is summarized well in this video: https://www.youtube.com/watch?v=dxu2URXPr68. Their procedure is to Merge the directory of the raster files into one image and then you can select the bands you want to use more easily. The downside of this method is that it consumes more storage space than a .vrt file because it creates a new .tif file. But, in cases where you are uncertain which combinations of bands are going to provide you with the view you need, this procedure could be the better option.[1]

One final step is to pansharpen your image. This resamples the resolution of your image from 30-m to 15-m resolution. It makes for a much nicer visual image, but is generally of low mathematical analytical value. Unfortunately, this is not a straightforward operation in the PC version QGIS, so I recommend doing it in SAGA, which is packaged with your OSGeo4W download.[2] To get started, load in bands 2, 3, 4 and 8 into SAGA (Figure 2.10).

---

[1] *NOTE:* I was unable to get this procedure to work in QGIS 2.18, but it worked in QGIS 3.4.4.

[2] Pansharpening is quite straightforward to execute in the Mac version of QGIS. You can use the automatically loaded GRASS tools, which you can locate in your Toolbox dialogue ("i.pansharpen"). SAGA also works better in the QGIS platform on a Mac and you can execute the steps in this forum: https://gis.stackexchange.com/questions/63367/how-to-pansharpen-landsat-8-in-grass. Because QGIS is open source and relies on developers' free time to put together, there are often things that work on a PC that don't work on a Mac and vice versa. The point is that if you are having a hard time getting a simple thing to work, it is likely to be a problem related to the version of software you have. Try older or newer versions or a different computer operating system if you are frustrated. Also, check for an answer on Stack Exchange (https://gis.stackexchange.com). More often than not, the answer you seek can be found there.

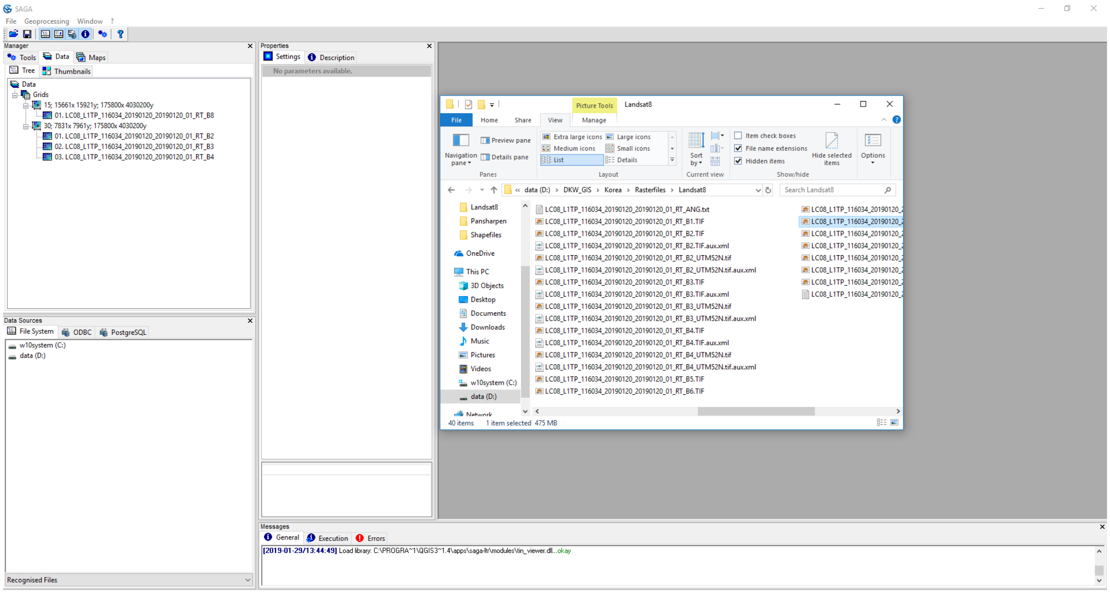


Fig. 2.10. Pansharpen your image in SAGA, step 1.

Go to Geoprocessing> Imagery>Image Sharpening and pick one of the options (you can experiment with what looks best to you), hit Okay (Figure 2.11). The tool may take a minute or two to calculate the image. *NOTE:* If you create the image as I have done without reprojecting the files (recommended), the coordinate system is EPSG:32652 and you will need to reproject it later.

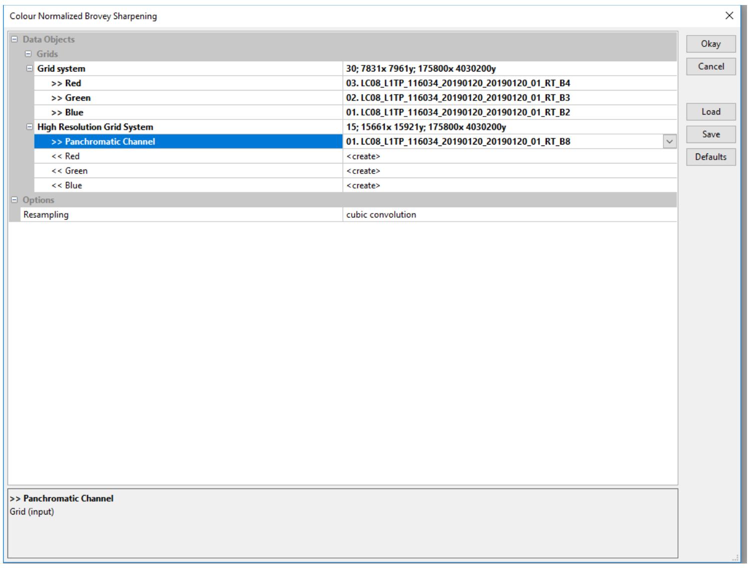


Fig. 2.11. Pansharpen your image in SAGA, step 2.

Now, create an RGB composite from Visualization>Grid>RGB Composite (Figure 2.12). Then double click the image and you will see a very nice image. Right click the "Composite" Grid and Save as Image…, which will be a tif file. Make sure you select to Save Georeference. Open up a new QGIS project and reproject it into EPSG:3093. You can also do this in SAGA (Geoprocessing>Projection), but this is a bit complicated. Layer in the DEM under it and color multiply the pansharpened image (set the brightness to 0 and increase the contrast to 50 or so).

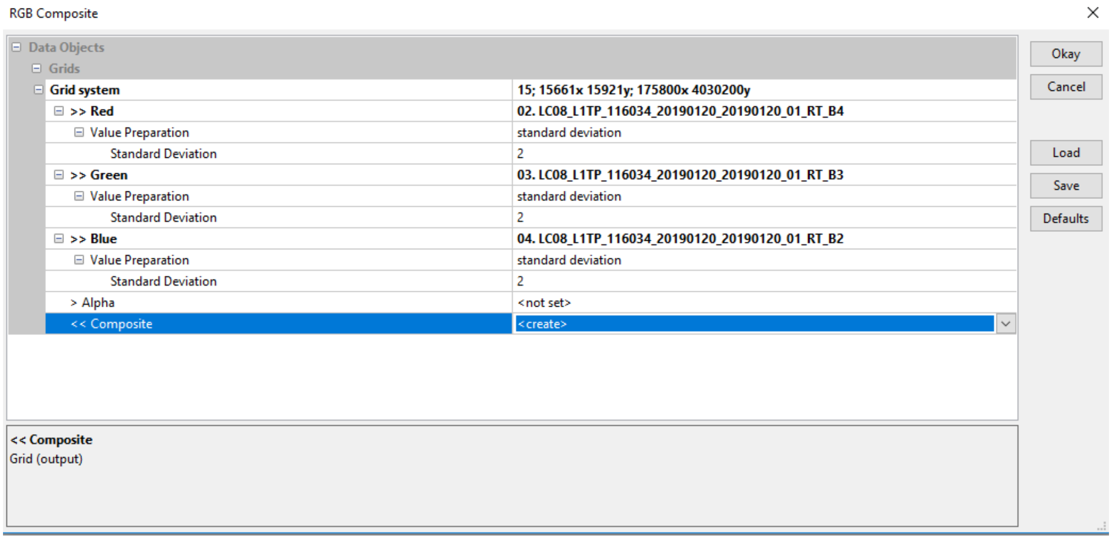
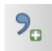


Fig. 2.12. Pansharpen your image in SAGA, step 3.

2.3 Adding a Vector Layer

Let's add in some point data. I have given you a csv file called 14C_TrainingData (in the eTL). In this file is a randomized dataset based on the known quantity and distribution of recorded archaeological sites with radiocarbon dates in Korea. The radiocarbon dates were created using the rexp() function in R (see code in Appendix B). The site locations are randomized within a 1-km buffer of all of the surveyed properties within the Ministry of Cultural Heritage database. There are 16,101 points we are going to use, which accords with the real data you are going to use in your assignments. *NOTE: If you just got an .xlsx file, you can open it in Excel or Open Office and save it as a .csv file (Save as…).*

I found that this portion of the exercise did not work in QGIS 2.18 (Las Palmas), but works fine in QGIS 3.4.4 (Madeira). So, if you have not done so already, but sure that program is installed on your PC.

Once that is done, go to QGIS and click the  (add delimited text layer) icon in the left toolbar or in Layer>Add Layer> Add Delimited Text Layer….. Navigate to your 14C_TrainingData.csv file and add it with the parameters clicked as shown in Figure 2.13. Note that the Geometry definition is defined as x and y in the csv file. If you put the table together as latitude and longitude as the column fields, your X field would be latitude and your Y field would be longitude. You might need to manually select those depending on how the fields were set up.

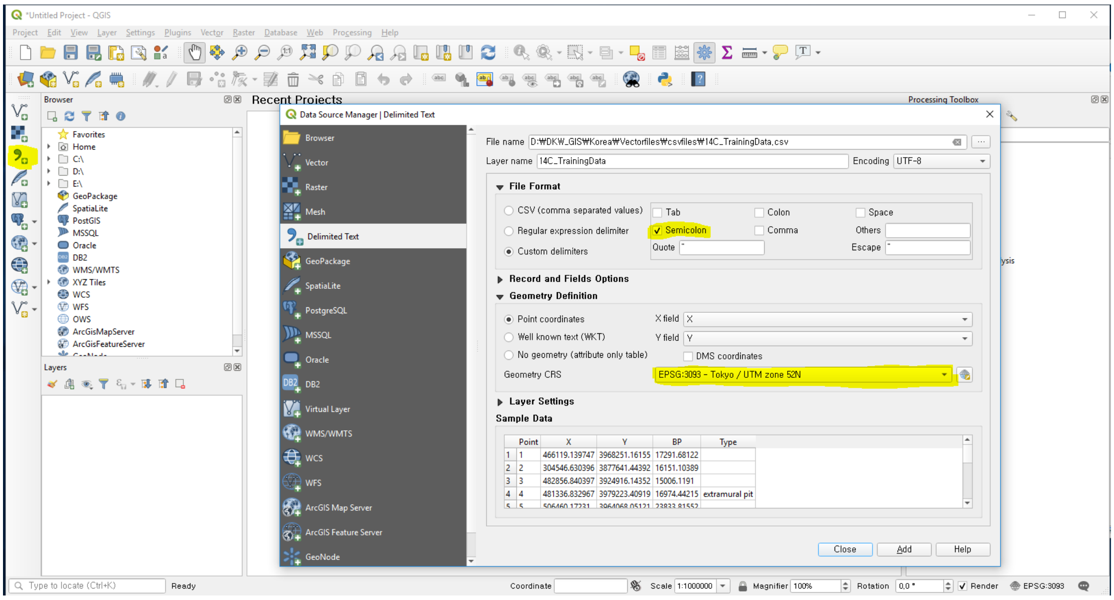


Fig. 2.13. Import a CSV in QGIS 3.4 (Madeira).

You will need to save the csv file into a format that can be spatially manipulated, such as a shapefile. In order to do this, right click on the file, and save it as a shapefile from Export>Save Feature As… (Figure 2.14). Be sure to change the CRS to EPSG:3093. You can close out Madeira and go back to QGIS 2.18 (Las Palmas).

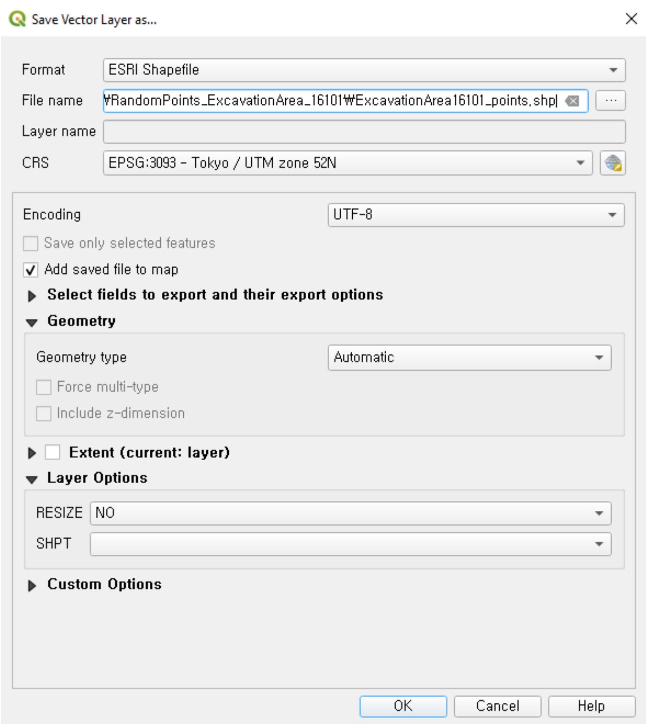


Fig. 2.14. Export CSV file as a shapefile.

Just for fun, zoom in on an area that has a lot of points and use the Measuring tool to measure the distance of sites to roads or to each other (Figure 2.15). Now you see the potential power of GIS!

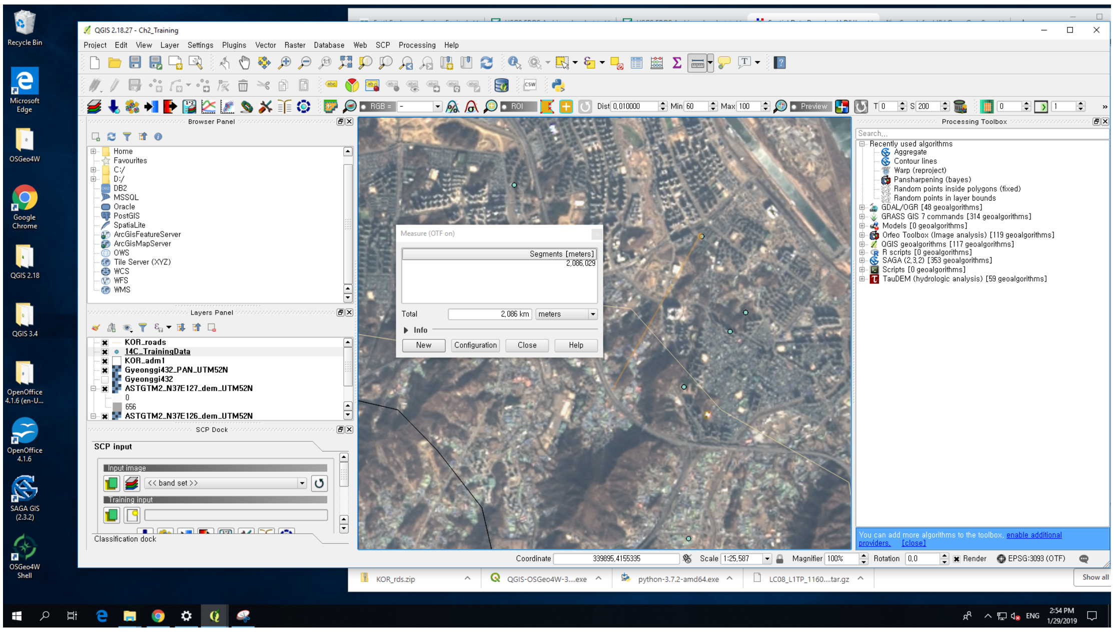
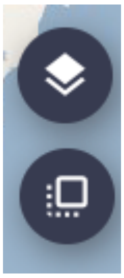


Fig. 2.15. Measure the distances between sites and roads in QGIS.

2.3 EXERCISE

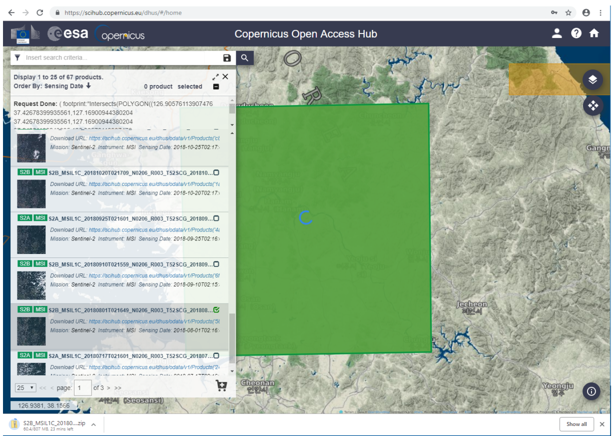
You are going to create a map on a multispectral image. This time, you will use the European cluster of satellites to obtain the images you will work from (https://scihub.copernicus.eu). You will need to create a user name and a password, then click the button that says "Open Hub." You will be taken to a map that begins in Europe, but you can click and drag your way over to Korea using the left mouse button.

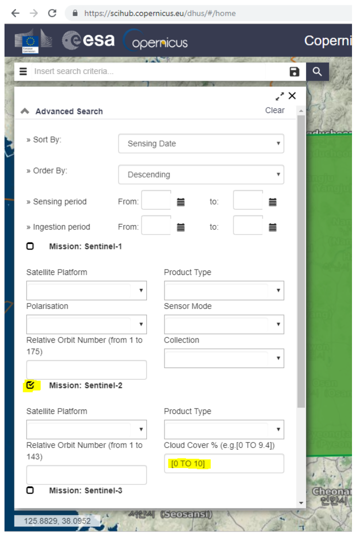


Just for fun, use these tool bars to change the views around. They are located on the right side of your screen. The upper one gives you terrain and panchromatic overlays.

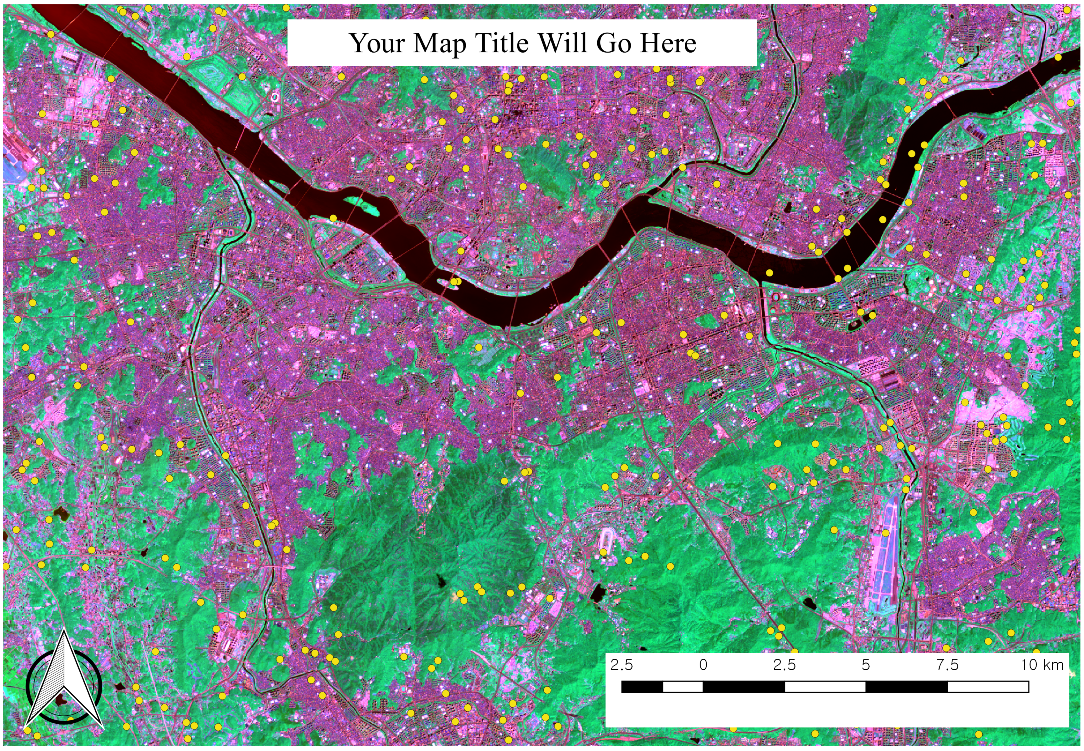
The lower button allows you to select an area you want to work in. Use the lower button to select an area around Seoul for downloading an image. You can zoom in on the area using the middle scroll wheel of your mouse.

You will want to download images from a Sentinel-2 satellite. Note the variables (left) I have highlighted here in yellow as guidance. I selected and downloaded the data from 2018.08.01 (below).





Here is a link to a forum that has some good information about Sentinel-2 bands: https://forum.step.esa.int/t/list-of-band-combinations-for-sentinel-2/1156/7. Create a Vegetation Analysis map (11-8-4) with your archaeological sites overlain onto them. Be sure to enhance the contrast of the points so that I can see your image (example below). If your image does not look like the example (e.g., yellow and orange), you probably added the data in the wrong order. Make a nice map as you did for the Week 1 exercise. This time, you are going to add a title, scale bar and north arrow to the map (see example).



The instructions for making a map with a legend can be found here: http://www.qgistutorials.com/en/docs/making_a_map.html. Be sure your map title tells me *what* the displayed elements are (e.g., say something about the types of points you have and something about the raster you are using). You must have a north arrow. You must have a scale bar in *kilometers* (not meters).

Save the file as a .png (*YOURNAME*_Ch2.png) and **upload it to the eTL by March 18 at 9:00 am** (5 points for completing the assignment correctly and on time). **Your image should look pretty similar to the example I have provided here,** so you will need to do the proper adjusting to get it to work. In addition to the image, answer the following questions (saved as *YOURNAME*_Ch2.docx):

STUDY QUESTION (1): What are the advantages and disadvantages of raster data in relation to this week's assignment (refer to Conolly 2016 and Conolly and Lake 2006)? Specifically, you need to refer to your experience working with the data in this exercise. What worked well, what didn't work so well or took a long time to perform and why? (2 points)

STUDY QUESTION (2): Based on your view of the data, what do you think the pink and green areas in the 11-8-4 VRT of the Sentinel 2 satellite represent? Technically speaking, what makes some areas pink and others green in terms of the output of the image? Where do the randomized archaeological sites tend to occur in relation to these values? Why do you think this is? (3 points)

Chapter 3: Creating a Landscape Model

3.1. Landscape Principles

In general, archaeologists work on landscapes of some sort in order to manage or interpret cultural resources. Landscapes are venues where humans and nature intersect and given the scale of change that humans have brought to earth's lithosphere, it is virtually impossible to separate natural and anthropogenic processes. In modeling archaeological settlement on landscapes, archaeologists who use GIS are able to obtain large amounts of data in a relatively short period of time by feeding spatial variables into the GIS and asking the program to produce the needed information. While this task may seem banal to you now, in the old days before GIS was widely available, obtaining such information was cumbersome and imprecise.

In last week's module, you were asked to import Sentinel and Landsat maps and provide me with a general interpretation of your impression of the data based on your visual inspection of the maps. In this week's module, we will add to our site database and determine the sites' environmental (i.e., landscape-derived) variables with more precision. You will perform basic statistical analyses on the points in your database and begin to tell me some more nuanced information about the ecological variables that characterize the sites in South Korea.

3.2 Adding Vector Layers

Let's start putting in some real vector data into your GIS and explore how to input and change data within it.

Go to www.diva.gis.org/gdata to access the county boundary administrative units in Korea. Download the file and add it to your vector folder and unzip it (Figure 3.1). *It is very important that you never change the names of these files or move them around individually*. They are all linked to one another and if you remove one of the files, they will not work.

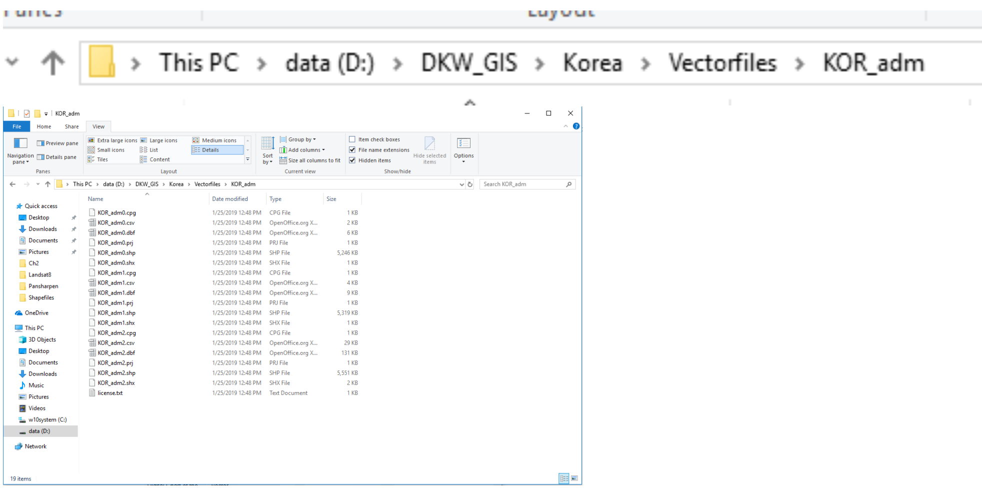


Fig. 3.1. Vector folder hosting administrative shapefiles in your data drive.

Return to QGIS and add a vector layer (Layer > Add Layer > Add Vector Layer… or click on the icon in the upper left corner of the Layer Menu). Navigate to the KOR_adm folder and add the file entitled KOR_adm1.shp that has 5,319KB. If you add the wrong file, it will not work.

Right click on KOR_adm1 in your Layers Panel and click Zoom to Layer. Now you have a map of South Korea with the county administrative boundaries in polygon form. Right click the file again and go to Layer Properties. Click on Style in the Tools Panel. Change the symbolizer to Categorized, change the Column to NAME_1 and then click Classify (Figure 3.2).

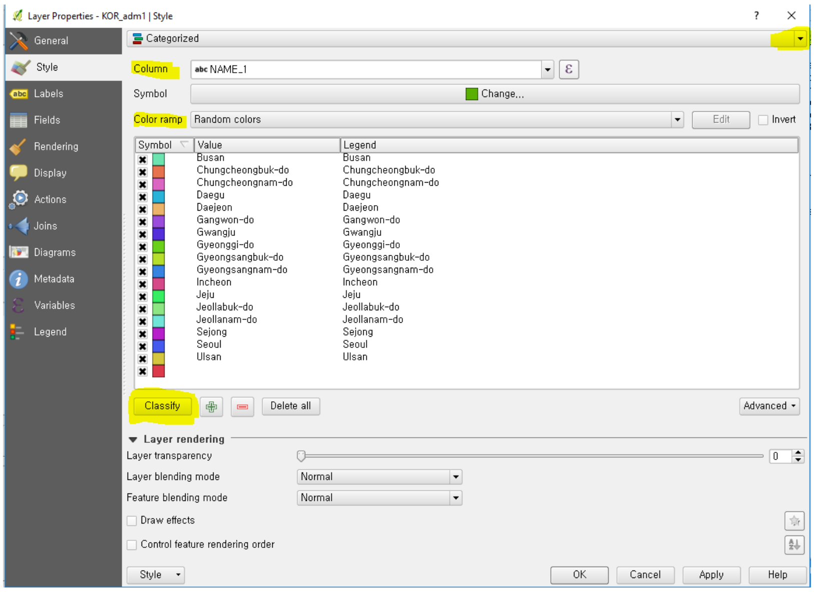


Fig. 3.2. Adjust the symbology on the administrative shapefile of Korea.

Click Apply, then Okay and your end result should look something similar to Figure 3.3.

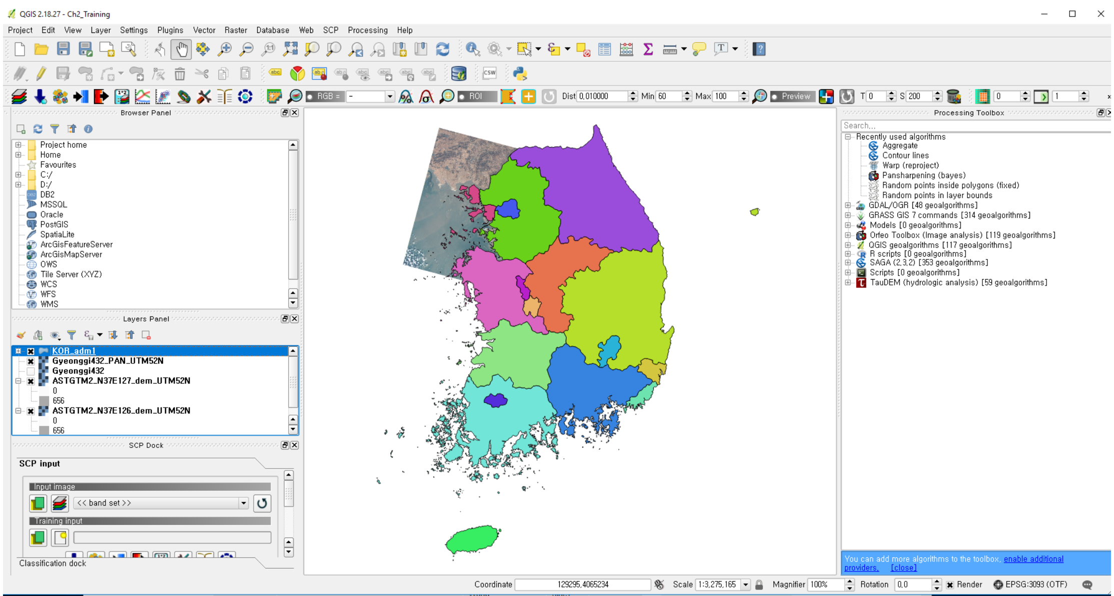


Fig. 3.3. Adjusted county view of South Korea.

Right click on the file again in the Layers Panel and click on Open Attribute Table. Here you can see a number of values in the table that include names and types of regions. Let's add some roads in. Go to back to diva-gis and download the roads zip file. Once the file has downloaded, add it to your vector data folder. Now adjust your adm1 file symbology so that the fill is a Single Symbol, then double click the Simple fill category and make the fill transparent. When you are done you will have something similar to what you see in Figure 3.4.

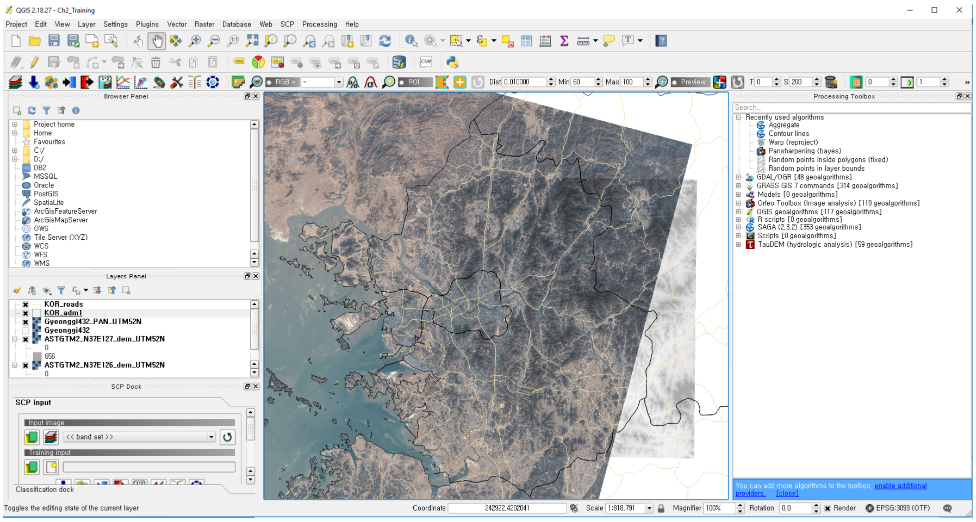


Fig. 3.4. Combined Roads and transparent-fill Admin shapefiles view in Map Window.

3.2 Querying Vector Data

3.2.1. Boolean searches

The power of GIS is ultimately in its ability to communicate information to the consumers. Therefore it is important for the technicians to be able to create and distill information for the consumers. As a professional archaeologist, you will want to be able to search for specific types of information through the millions of data points that are going to be in the GIS.

In order to do this, QGIS uses Boolean operators, and most particularly the terms "AND," "OR" and "NOT" in order to query data.

Let's make a simple Boolean search from our training dataset you worked with in Chapter 2 in order to get an idea of how to effectively use the GIS. Add then, right click your 14C Training Data shapefile and open the Attribute Table. There are 16,101 sites in this shapefile and the attribute information tells us the calibrated mean radiocarbon age of the point. Let's say that you have a researcher visit the museum who is interested in sites with radiocarbon dates from structural timber and dolmens in Korea. So, you want to find these features recorded in the database in order to help the researcher find the collections she is interested in.

In order to do this, you will want to "Select features using an expression" by clicking the  icon on the attribute table. A menu will appear. In order to look for structural timber or dolmen, lets execute the proper Boolean search. You need to use the Expression generator in order to build the proper query. By clicking the > arrow next to the "Fields and Values" field, you will see all of the attribute columns in our attribute table (Figure 3.5).

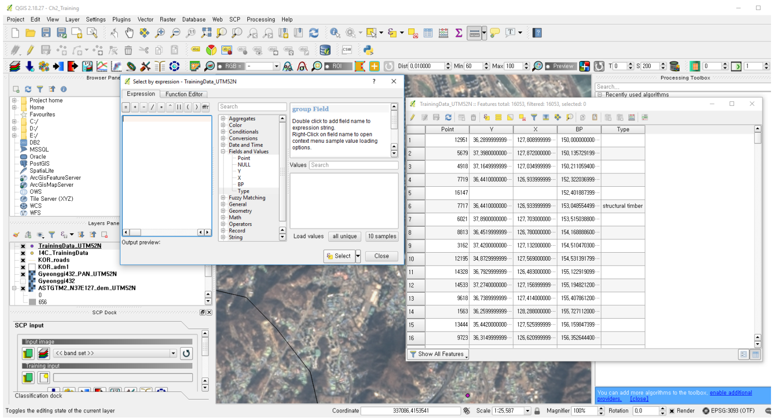


Fig. 3.5. Select by expression in the attribute table.

In order to look for structural timber and dolmens you will need to perform the following Boolean operation:

"Type" = 'structural timber' OR "Type" = 'dolmen'

From Figure 3.6:

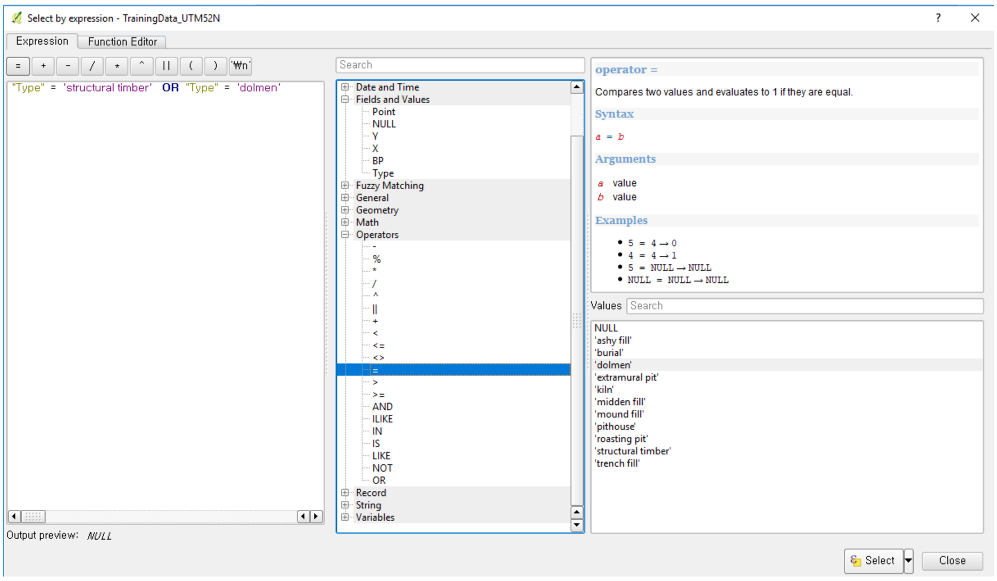
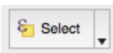


Fig. 3.6. Query site attributes based on Boolean operators.

"Type" was selected from "Fields and Values"; "=" was selected from the operators menu; and "structural timber" and "dolmen" was selected after hitting "all unique" after Load values. Boolean terms are found under >Operators. Once you have what you want, hit  and then close in the lower-right corner of the screen. What you see in the attribute table is every site in the database that includes the terms structural timber or dolmen in the Type attribute field.

The logic of this operation is that you want to query for the site types you are interested in and using "OR" is a way to include as much information as possible regarding these features. If you wanted to limit your search, you would use AND or NOT. For example, if you searched for sites that had the term structural timber AND dolmen, you wouldn't get any sites because there isn't a site that has *both* of these terms in the attribute field.

3.2.2. Average distances

Now, let's say we want to try to find the distances of sites to roadways. This is a multistep process. Your roads shapefile is in lat/long, WGS84 (EPSG:4326), so you need to reproject it into the 3093 coordinate system by right clicking the shapefile > Save As…. Give the new shapefile a new name that ends in "UTM52N" or "3093"…whatever you will remember. I recommend saving it in the same folder as your KOR_roads shapefiles. When you select the CRS, select EPSG3093.

The next step is to create a column in your point (shape)file (Figure 3.7). Enable the editor by clicking the little yellow pencil (). Right click the file in the Layers Panel, click Properties > go to Fields in the left panel, then click Add field (leftmost option under ∨ Fields). Call your new attribute something like "DistRoad" and make sure the length of the integer field can accommodate the maximum number of meters your sites are from a road. Click Okay and turn off the editor by clicking the pencil again.

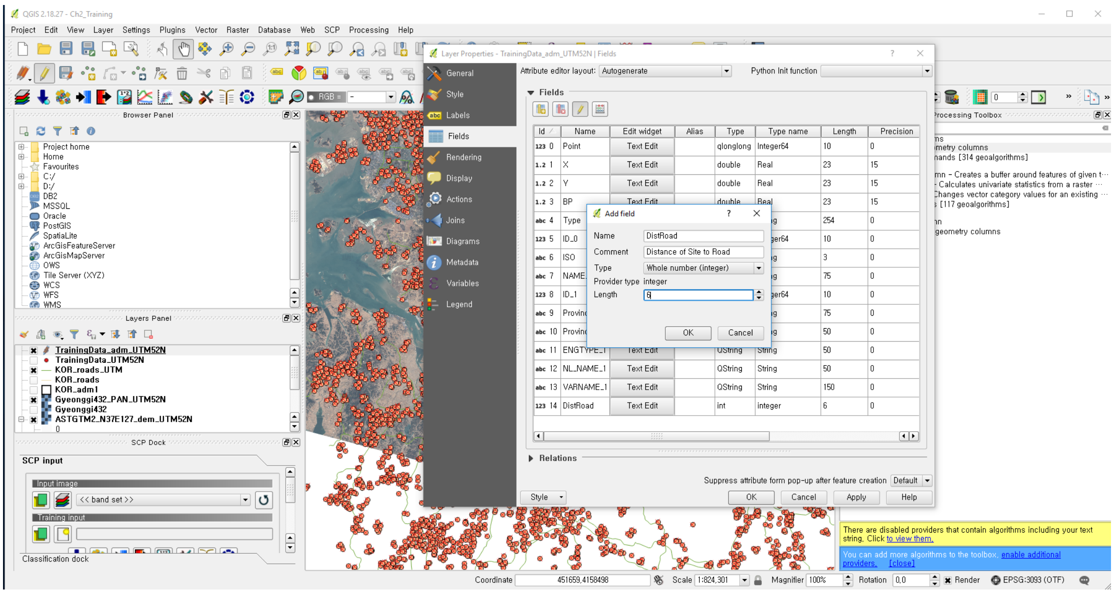


Fig. 3.7. Create a column in your attribute table.

Next, in the Processing Toolbox, call up the v.distance tool in the GRASS commands (Figure 3.8a). In the 'from' field, add your point file, in the 'to' field, add your roads line file (make sure they are both in UTM 52N). Be sure your "Column names(s) where values specified by 'upload' option will be uploaded" has DistRoad or whatever your attribute was that you created in the last step. In the advanced parameters, create new shapefiles. Click Run. If you right click the new shapefile, open up the attribute table and scroll to the right, you will see a new field (in my case, DistRoad) that has the shortest distance of the site from the nearest road. Now, to calculate the average distance of the sites to roads, click Vector>Analysis Tools>Basic statistics for numeric fields. Enter the fields as shown in Figure 3.8b, and navigate to the folder where you saved your html file and you will see that the average distance of your sites to a road is 2604 m and other related statistics.

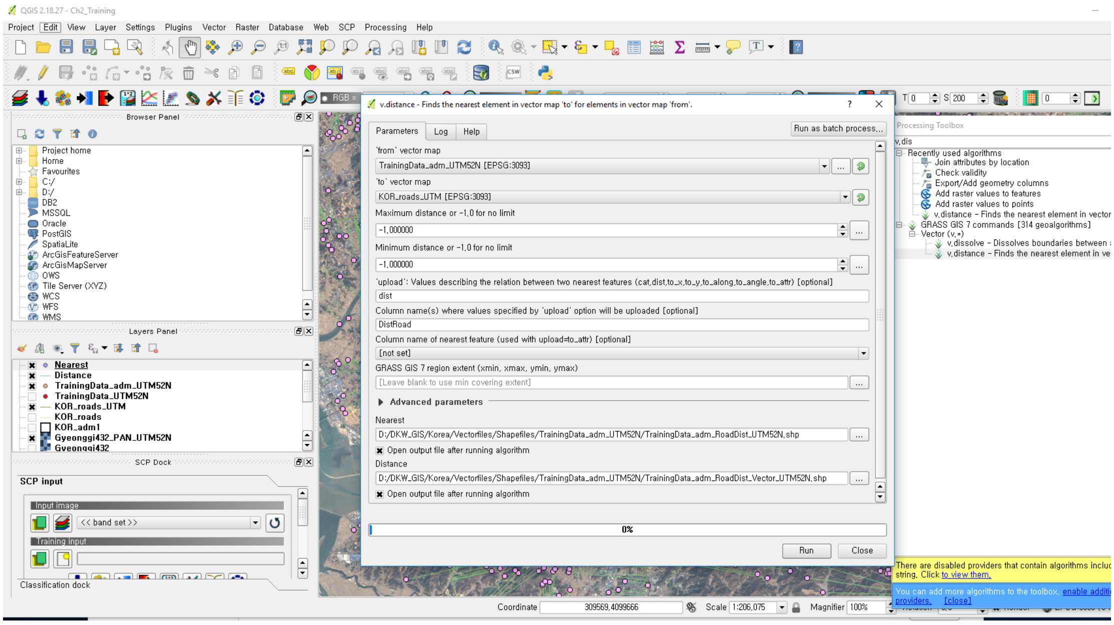


Fig. 3.8a. Locate the v.distance tool in the QGIS toolbox.

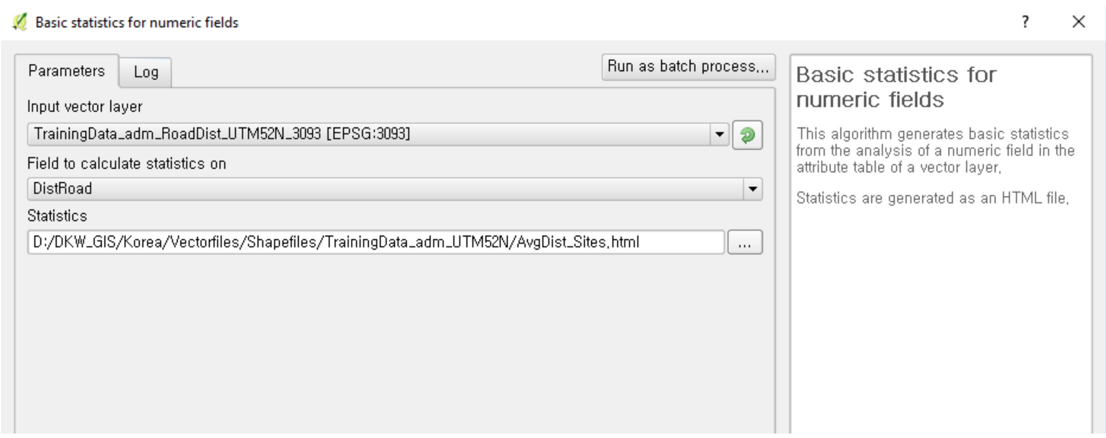


Fig. 3.8b. Field input in the v.distance dialogue.

3.2.3. Spatial joins and queries

Let's try another, more complicated exercise. Let's try to figure out which sites are found in Gyeonggi-do above N4150000. This takes a few steps performing an operation called a **spatial join**. Your administrative districts shapefile is in lat/long, WGS84 (EPSG:4326), so you need to reproject it into EPSG:3093.[3]

Once the file is in EPSG:4326 or 3093, go to Vector>Data Management Tools>Join attributes by location[4]. Choose TrainingData_4326.shp (or 3093) as your target vector layer, the Join vector layer is KOR_adm1 (Figure 3.9). Then select “within” as the Geometric predicate (you are telling to computer to merge the site points that fall within counties. “Take attributes of the first located feature” as your Attribute summary and “Keep all records (including non-matching target records)” in the joined table. You want an output file to be created after the algorithm is generated and then Run the operation. A new file will appear in your Layers Panel called “Joined layer.” Open the attribute table and you will see that all of the County data that were spatially associated with each of those points is now included in the attribute table.[5] If you have had to work in EPSG:4326, reproject the file into EPSG:3093 using the Save as… function explained above.

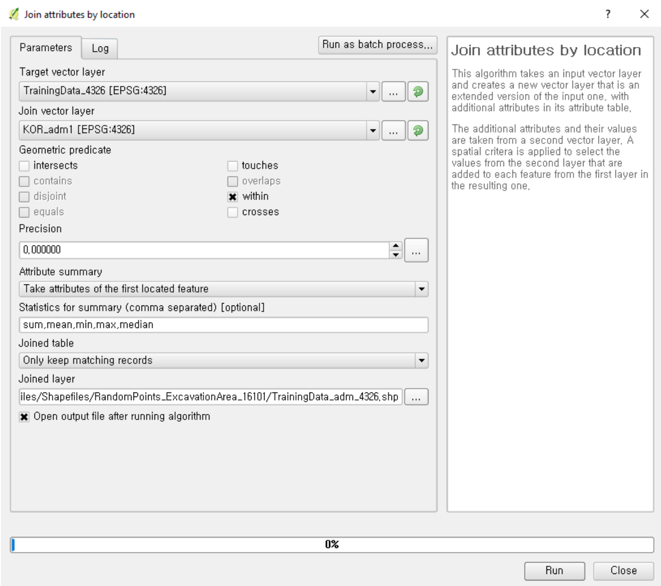


Fig. 3.9. Make a spatial join based on locations of features.

In order to do our search, we need to Select features by expression () and type in the following search:

```
"NAME_1" = 'Gyeonggi-do' AND "Y" > 4150000
```

When you go back to your map view, you see all of the sites above N4150000 that are located in Gyeonggi-do highlighted (Figure 3.10). If you wanted sites below N4150000, you would make the expression “y” < 4150000. You can save this selection by right clicking the shapefile, Export > Save Vector Layer as… and click

[3] *NOTE*: I was unable to get the polygon KOR_adm shapefile to reproject successfully into EPSG3093 on one of my tests. If this happens to you, my advice is to open up a new project in QGIS, open the TrainingData.shp file up and reproject it to EPSG4326 and save it (right click the shapefile, Save As… and select the 4326 CRS renaming the file appropriately). Be sure to work with files that are all in the same CRS!
[4] *NOTE*: When assembling this module, I encountered a problem with QGIS 2.18 that causes the computer to crash whenever a spatial join is performed on a .csv file. To get around this, close QGIS and go to the .qgis2 folder in C:\Users\computer name\.qgis2. Delete this folder and reopen the program. Your spatial join will now work.
[5] If you get an error that says “QgsFeature.setGeometry…” you will need to fix the geometry of your source files. Click Vector>Geometry Tools>Check Validity and fix the errors that appear.

the little box under Encoding that says "Save only selected features." Be sure to give the new file a name that indicates that these are sites north of N4150000 within Gyeonggi-do.

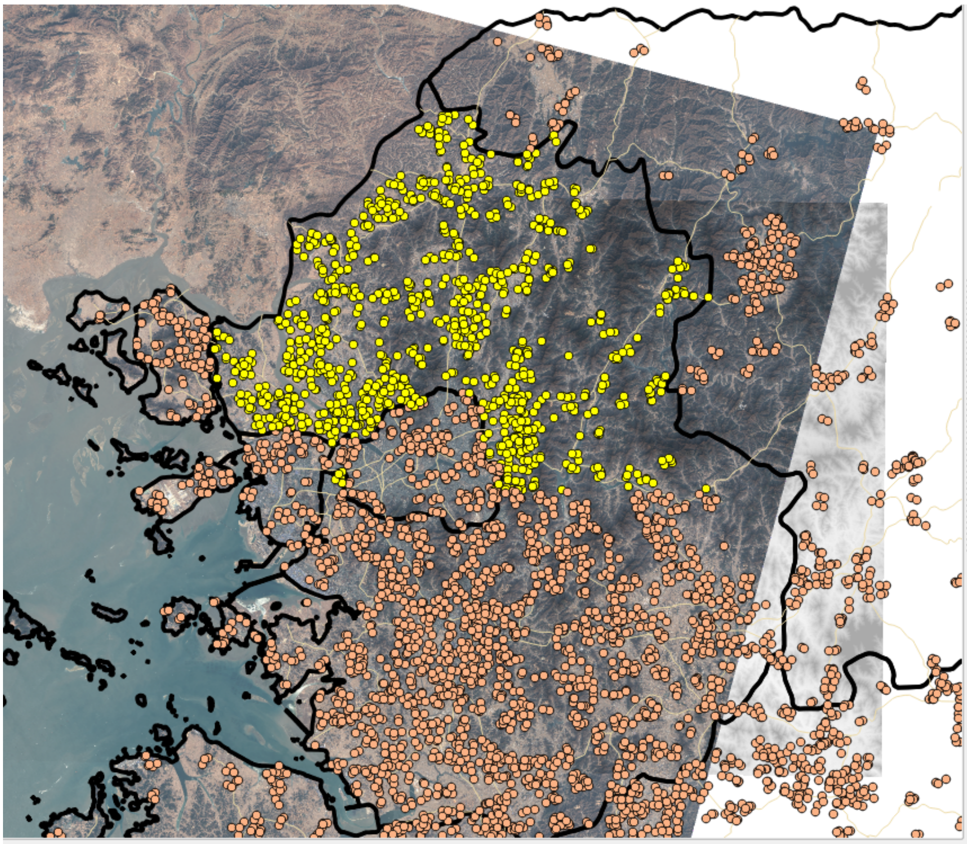


Fig. 3.10. Randomly placed sites within 1km of known site survey boundaries in Gyeonggi-do north of 4150000N.

If you are unhappy with the names of the fields, you can enable the editor and change them in the project properties. To do this, click the little yellow pencil again when the TrainingData_adm_UTM52N file is selected in the Layers Panel. Then, right click the TrainingData_adm_UTM52 file, open the Layer Properties Fields (Figure 3.11b). You can then rename the fields to whatever you like, but since we are working with shapefiles, the limit of the text fields is normally 12 characters, so use them wisely.

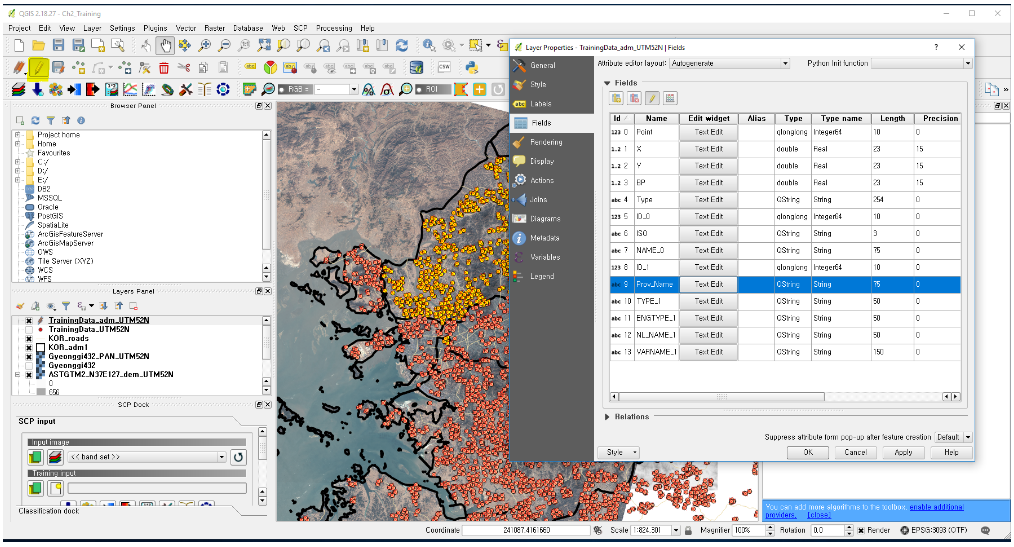


Fig. 3.11. Rename the fields in the Properties dialogue.

3.3 EXERCISE

You are provided another an Excel file with real archaeological point information. These data are in EPSG:4326 (latitude/longitude, WGS84). You need to get a GLCC map (under the Land Cover category of the Subject) from the diva-gis server (www.diva-gis.org). I want to see two maps from you: one with all of the locations of pithouses in south Korea with an inset of Gangwon-do before 2000 BP and one with all of the pithouses in South Korea with an inset of Gangwon-do after 2000 BP against the GLCC background. Make a nice map as you did for the Weeks 1 and 2 exercises. This time, you are going to add a title, a double box scale bar and north arrow to the map.

Include the province names in the map similar to how the guy who made the tutorial on map making put in city names. When you create the inset be sure to follow the directions in Steps 17 and 18 of the previously linked website. Also, be sure your map title tells me *when*, *where* and *what* the displayed elements are (e.g., Korean Pithouses before 2000 BP). You must have a north arrow. You must have a double box scale bar in *kilometers* (not meters).

The legend codes for the GLCC2000 dataset can be found in the Word document here. I have also placed a copy in the eTL.

You will get 4 points for the maps if they are done correctly (saved as a png file, title, scale, north arrow, inset, correct data shown, visually clear display). I also want a zip file of the shapefile that includes all of your point data reprojected into EPSG:3093 (1 point).

Save the file as a .png (*YOURNAME*_Ch3.png) and zip the shapefiles (*YOURNAME*_Ch3.zip) and **upload them to the eTL by March 25 at 9:00 am** (5 points for completing this portion of the assignment correctly and on time). In addition to the image and zipped shapefile, answer the following questions (saved as *YOURNAME*_Ch3.docx):

STUDY QUESTION (1): What are the advantages and disadvantages of raster and vector data in relation to this week's assignment (refer to Conolly 2016 and Conolly and Lake 2006)? Specifically, you need to refer to your experience working with the data in this exercise, but you can draw a comparison to the way it worked last week when you answered this question. What worked well, what didn't work so well or took a long time to perform and why? (2 points)

STUDY QUESTION (2): You have created two maps for this week's assignment. What are the differences and similarities between the locations in relation to modern land cover of pithouses before and after 2000 years ago? In general, on what kinds of land cover are the earlier sites located? In general, on what kinds of land cover are the later sites located? What do the changing temporal distributions of pithouses on the different land cover types imply about human subsistence behavior? (3 points)

HINT: If you want to better understand the distributions, you can open up the files in Madeira (QGIS 3.4.4), go to the SAGA application in the Processing Toolbox, click Vector <-> Raster, then Add raster values to points…. Select the shapefile in which you are working with for your Points file, select KOR_msk_cov for your Grid and make a shapefile in a different folder. You will then have an attribute field in this new shapefile called "KORmaskcov" that has the field codes of the GLCC map. You can then manipulate the views of the data using the Select Features by Expression tool.

Chapter 4: Working with Databases

4.1. Database Principles

For the most part, practitioners of GIS work with spatial databases. You can think of a database as a container that holds a lot of bits of information and it is your job to extract that information and make sense of it. In a normal workplace setting, a common GIS database would be hosted on a remote server and we would use PostGIS or some other kind of client to retrieve the data from the server. However, in their infinite stupidity, the SNU IT Department denied your lowly professor access to university server space for the purpose of your education, so we will have to improvise. We will work collaboratively next week, but this week, we will learn to perform some basic tasks that you can later apply to collaborative efforts.

In this exercise, you will learn how to take some typical archaeological data and join it to your shapefiles using relational fields. You already learned some of these tricks in last week's exercise, but now we are going to use some different techniques to visualize typical archaeological data.

4.2. Table Joins

Open a new project and add the TrainingData_UTM52N.shp file. Just to keep yourself oriented, also add the KOR_adm1_UTM52N file. Adjust the symbology so that you can keep track of where you are located.

In the eTL, I have placed two tables: Ceramic_Random.ods and RimDiameter_Random.ods. Place these files into a "Tablefiles" folder in your GIS folder. Once they are stored in your D: drive, open the files up using Open Office (you must download and install this free program on your computer if you don't already have it). You will notice that Ceramic_Random.ods has four categories of attributes: "Sherd_ID", "Sherd_class", "Sherd_Type", and "Site_Id". Sherd_ID is the type of attribute field that would be created at a museum when curating ceramics. I made random designations of Sherd_Type ranging from A to D, but you could call these Chulmun, Mumun or whatever you want. The Site_Id field correlates to the Point field in your TrainingData csv (this is the **relational field**). RimDiameter_Random.ods has two categories: "Sherd_Id" and "Diameter (mm)". These categories are roughly similar to Conolly and Lake (2006:54). Save the files as csv (just accept all of the default options for Character set, Field delimiter and Text delimiter) (Figure 4.1).

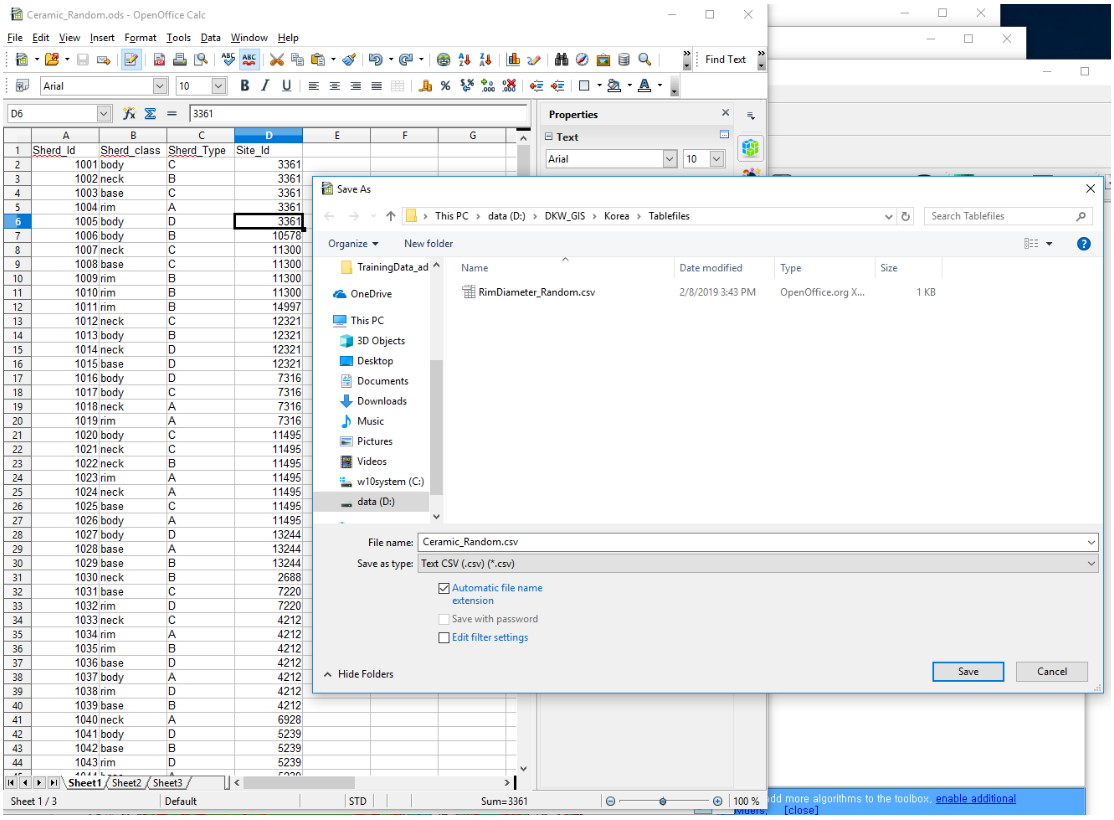


Fig. 4.1. Save the CSV for later import into QGIS.

Now, you are going to add these csv files to your workspace as you did in Week 2, but be sure you do not add a geometry field (attribute only table)(Figure 4.2).

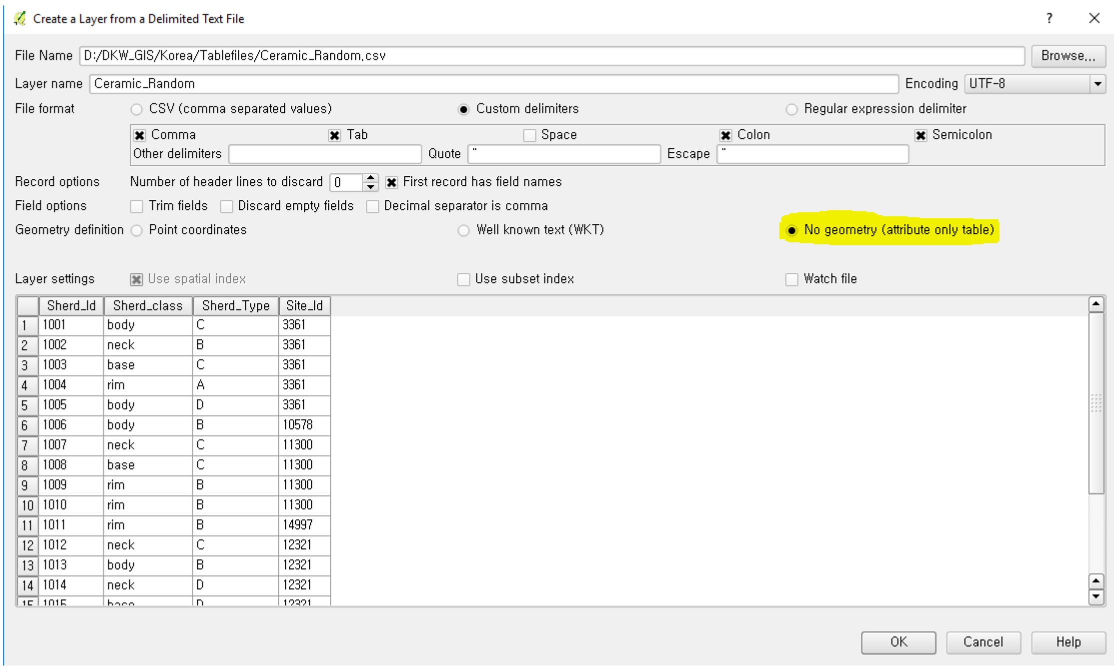


Fig. 4.2. Add the CSV as a layer in QGIS.

Because you have more than one sherd per site, you need to create a new shapefile based on the data. Download the plugin "MMQGIS" from the Plugins manager (Figure 4.3).

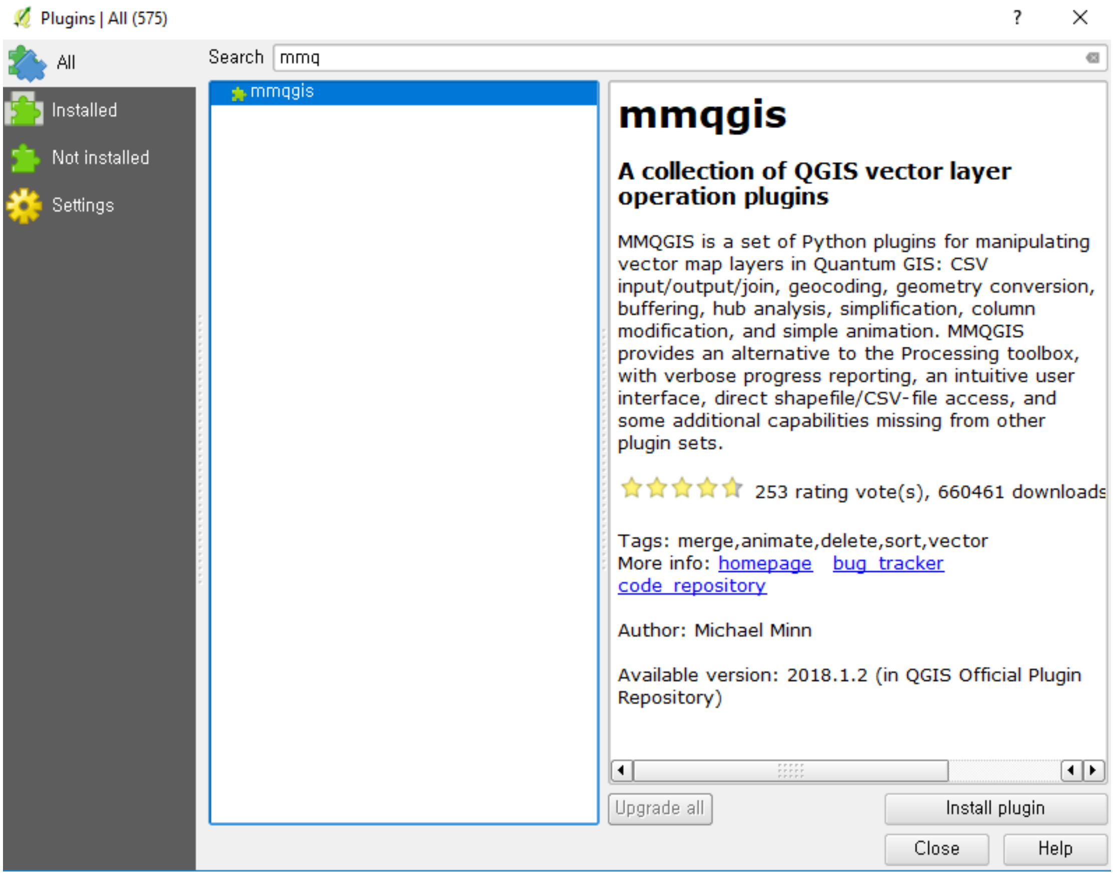


Fig. 4.3. Install the MMQGIS *plugin into QGIS.*

Then in the File menu go to MMQGIS>Combine>Attributes Join from CSV file. Select the Ceramic_Random.csv file to join by attribute and join the layer attributes as shown in Figure 4.4.[6]

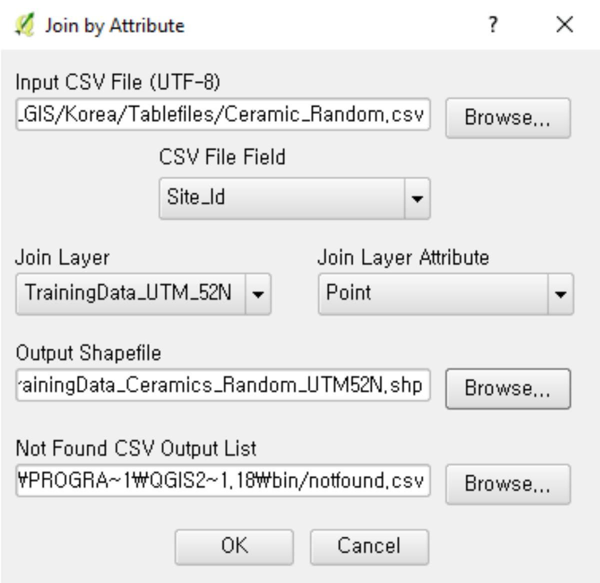


Fig. 4.4. Make an attribute join in MMQGIS.

[6] Ideally, you would like to have your relational fields coded the same so that you don't make a mistake when performing spatial joins. In this example, I have deliberately made the relational fields different so you can see how it works in a non-ideal setting. However, I encourage you to use standard nomenclature when setting up your database so that you can tie together relational fields easier.

Add the created shapefile to your map and you will see that only those sites that have ceramic sherds are part of your new shapefile (Figure 4.5).

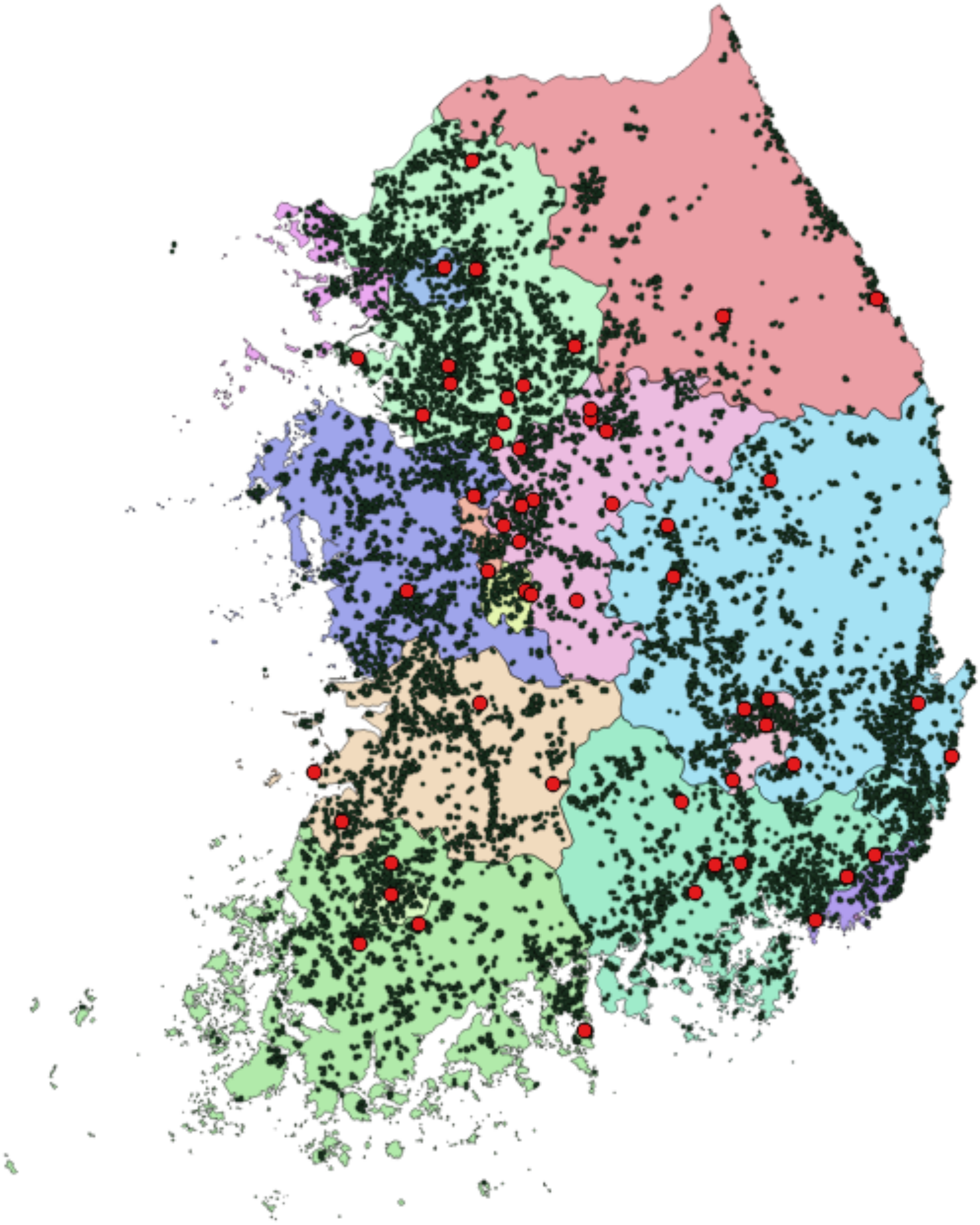


Fig. 4.5. Sites from your randomly generated database that have ceramic sherds are red dots.

Now you need to join the RimDiameter_Random and Ceramic_Random tables together using the Sherd_ID field. Since have already joined Ceramic_Random to your TrainingData file, you can perform the next join simply through the Properties. Right click your TrainingData_Ceramics_Random shapefile and select Properties. In the Layer Properties dialog, select Joins and hit the little green + sign at the bottom (Figure 4.6). Join the RimDiameter_Random table to the TrainingData_Ceramics shapefile. The Join field is Sherd_Id and your Target field is Sherd_Id. If you open the attribute table of your TrainingData_Ceramics file, you will now see a rim diameter attribute column, which is the aperture of the rim in mm. You can rename the columns to make more sense. For example, I deleted the field "Ceramic_Ra" because it is the same as "Sherd_Id". In my case "Ceramic__3" is the averaged value of "RimDiameter" for all attributes that include rims. So, it isn't useful and I deleted it.

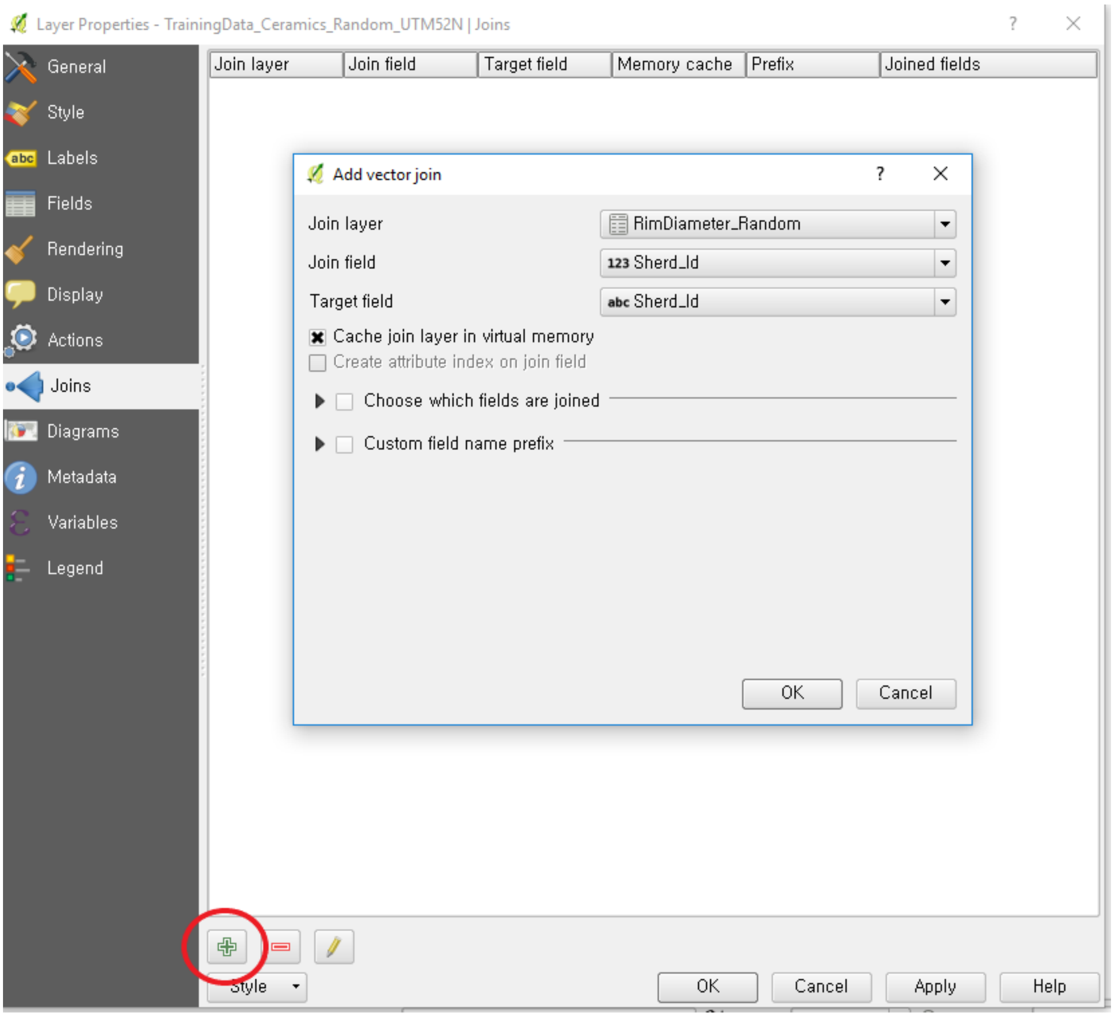
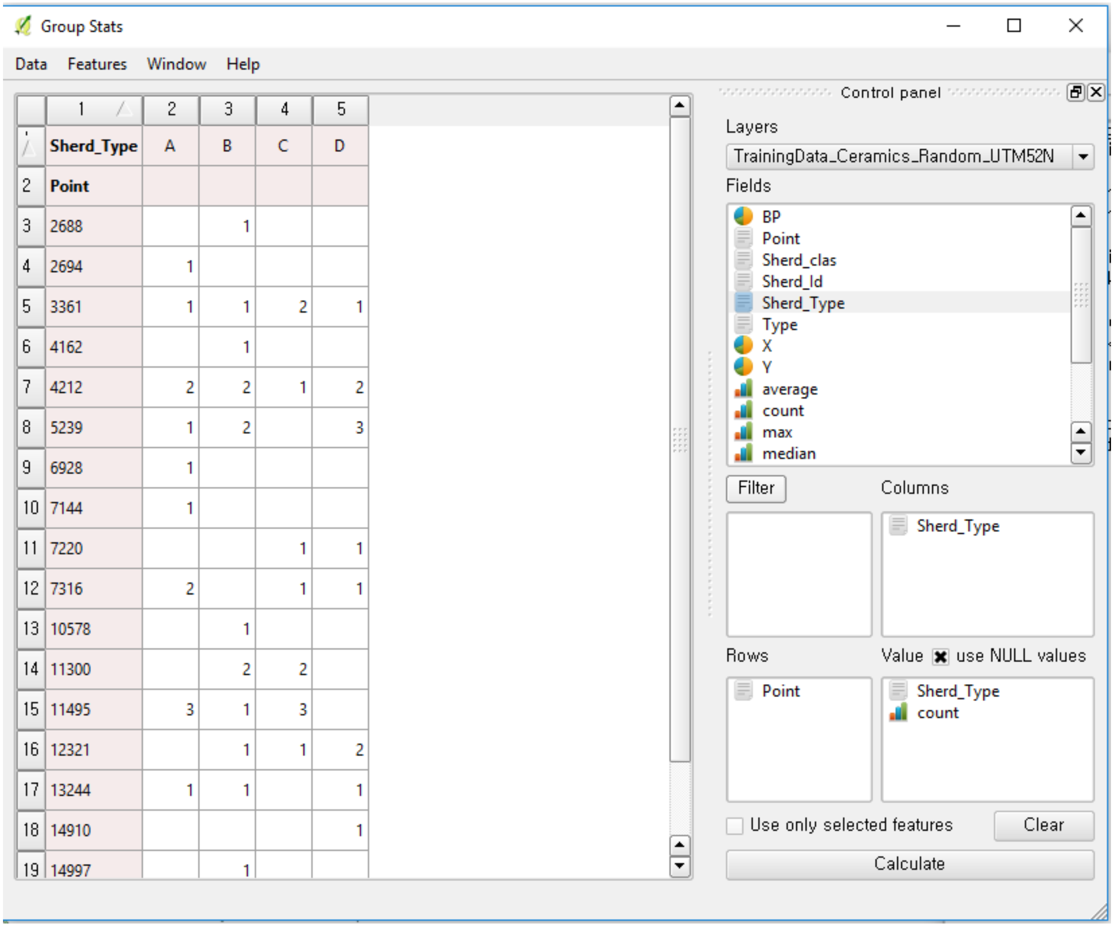


Fig. 4.6. Make an attribute-based join in the Properties dialogue.

4.3. Making Sense of Tabular Data in Graphical Form

Your computer screen now has a display of all of the sites in Korea that have (randomly assigned) pottery from features that span 2377-2732 cal BP. But, what you will notice when looking at the attribute table is that most of the sites have more than one sherd (Points = sites). Let's say you are working for a museum and you want to better understand the distribution of pottery types over space. This would be relatively straightforward if each site had only one type of pottery, but this isn't the case. So you need to map the data proportionally in a graphical format.

This is a multistep process that begins when you download the "Group Stats" plugin in the Plugin Manager.[7] Once that is activated, open the plugin (Vector>Group Stats). Drag Point to the Rows field, Sherd_Type goes to the Columns category and in the Value category, drag Sherd_Type and count (Figure 4.7). Click calculate. Then click Data>Save all to CSV file and name your file TrainingData_Ceramics_Counts.csv. I saved mine in my Tablefiles folder in my D:/DKW_GIS/Korea/Tablefiles subdirectory.



1	2	3	4	5
Sherd_Type	A	B	C	D
Point				
2688		1		
2694	1			
3361	1	1	2	1
4162		1		
4212	2	2	1	2
5239	1	2		3
6928	1			
7144	1			
7220			1	1
7316	2		1	1
10578		1		
11300		2	2	
11495	3	1	3	
12321		1	1	2
13244	1	1		1
14910				1
14997		1		

Fig. 4.7. Use the Group Stats plugin in QGIS to organize your aggregated data.

Just to make things easier for yourself, open up the newly created csv file with Open Office and relabel the Point column (Figure 4.8).

[7] Note that in order to activate a plugin, you need to select the plugin using the X field. In Korean, X means "no" or something is not active. But, in QGIS, X means yes—the application is active.

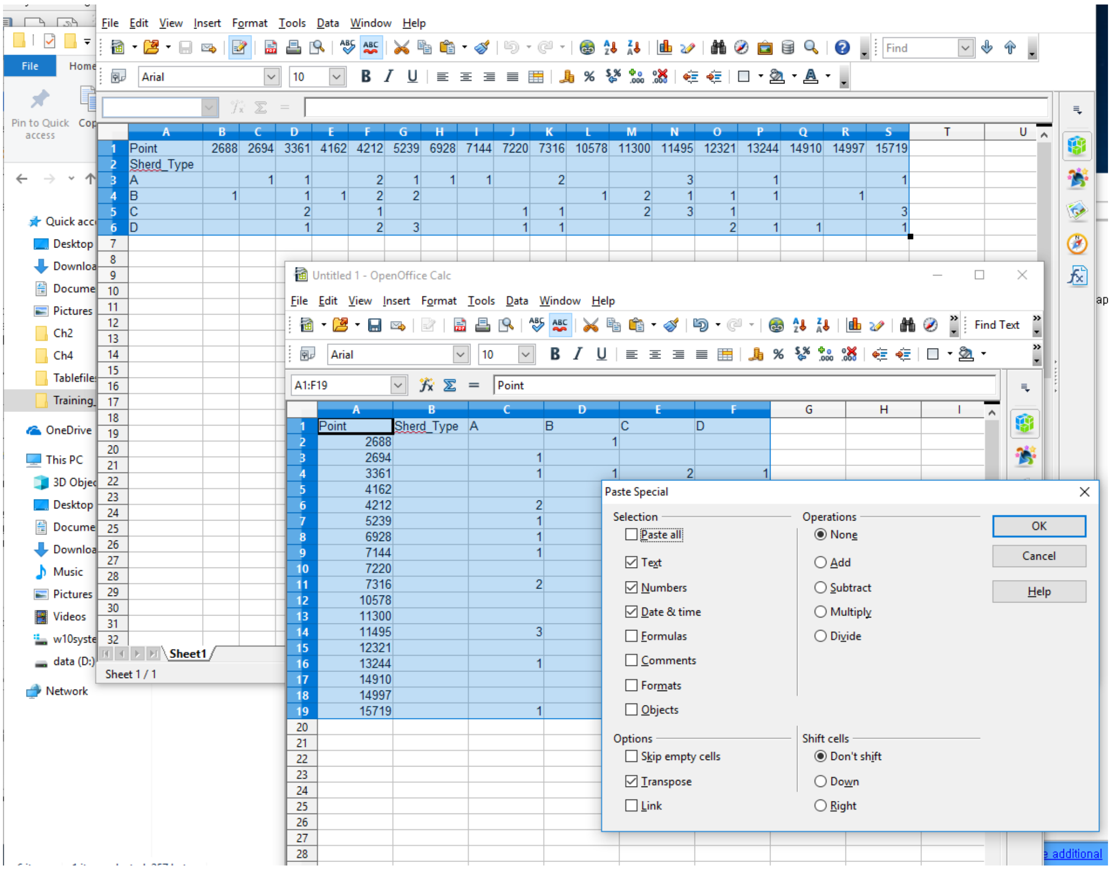


Fig. 4.8. Relabel a column using Open Office.

Then, add the csv as a Text Delimited Layer (Chapter 2.3). Be sure "No Geometry" is selected when you import it (Figure 4.9). Now Join the table to your TrainingData_Ceramics_Random_UTM52N shapefile. Your Join and Target fields are going to be Point. The X and Y fields need to be selected under "Choose which fields are joined." Check the attributes now and you should see a bunch of new fields added to the table.

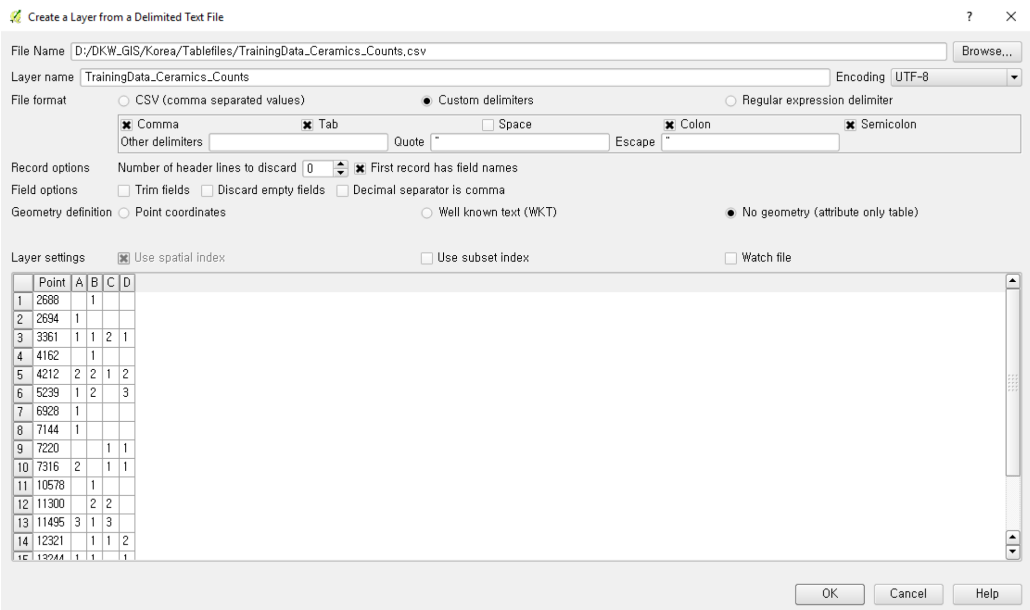


Fig. 4.9. Import the CSV with "No geometry."

In order to map your pottery types proportionally, click Properties again and go to Diagrams. Select Pie chart from the upper left drop down menu and individually add the attributes using the green + sign (Figure 4.10). Adjust the colors so that they are distinct from each other. Go to placement and increase the Distance to 3.00 and make the Priority high. You will see proportional pie charts appear in your Map Window after you hit Okay.

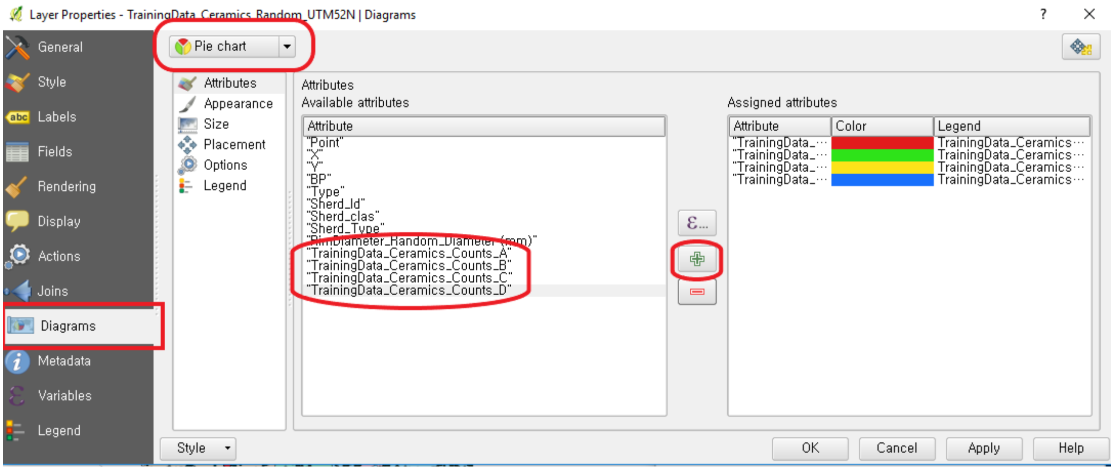


Fig. 4.10. Make a pie chart diagram in the Properties dialogue.

There is a problem with this map, though. If you zoom in on the sites that have multiple sherd types, you will see that there are as many pie charts as there are ceramic types. This is because you are making the pie chart from the TrainingData_Ceramics_Random_UTM52N shapefile, which has as many points as there are sherd types. So, the most efficient thing at this point is to export the table as a CSV file (Figure 4.11) with a CRS.

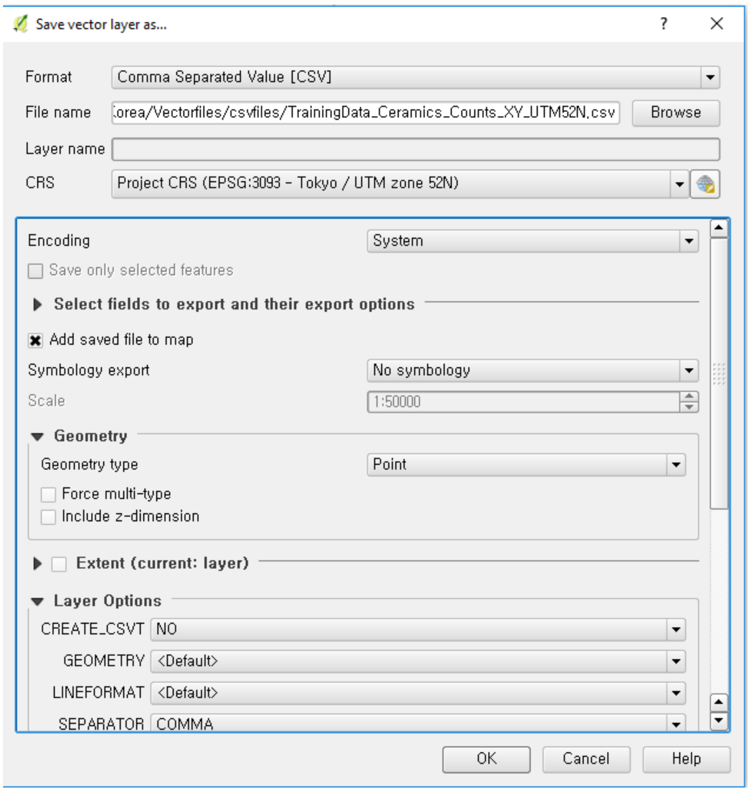


Fig. 4.11. Export your CSV with a CRS.

Once this file is exported, you will import it as a CSV with geometry (Figure 4.12). I would advise that you then save (export) this imported CSV file as a shapefile and import that for the next steps.

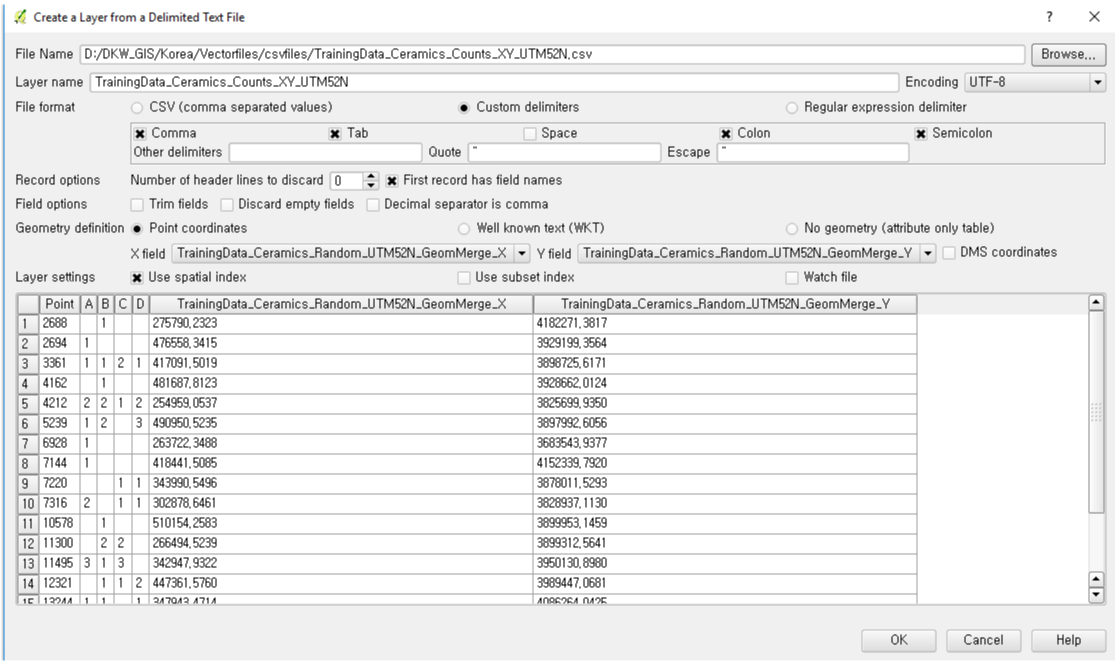


Fig. 4.12. Import your CSV with geometry.

Once you have made a shapefile (or not), you can make a new pie chart in the Diagrams dialogue (Figure 4.13), and when you zoom in on the sites, you will only see one pie chart.

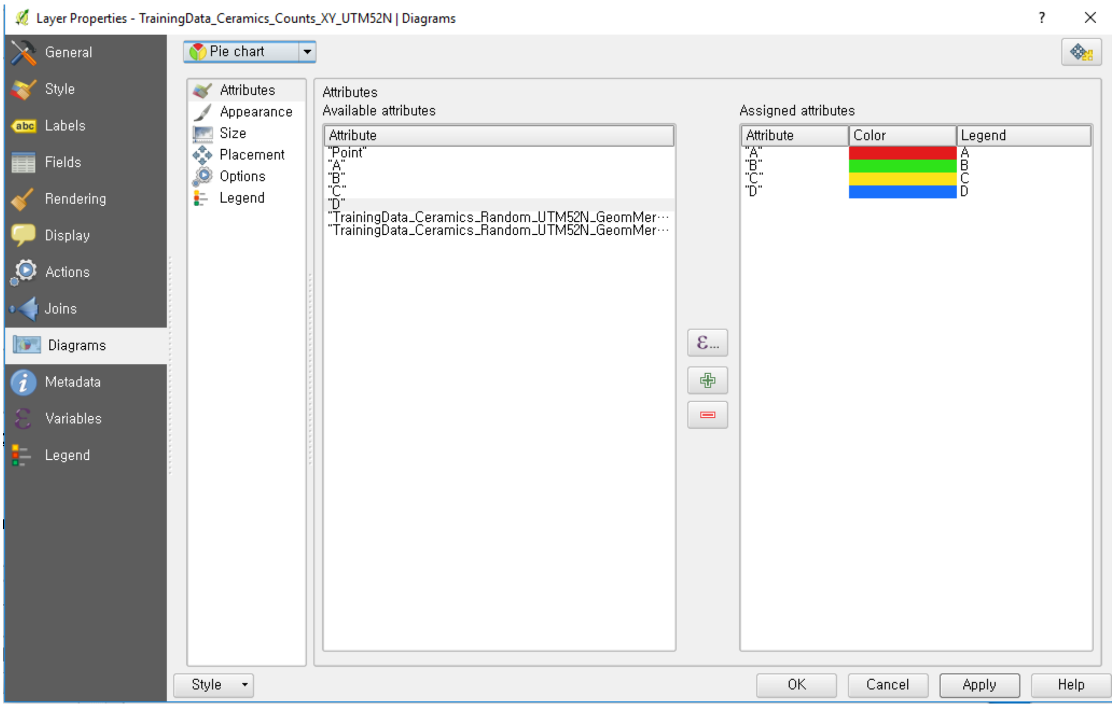


Fig. 4.13. Make a new Pie Chart Diagram.

When you look at the Map Window in full extent, you will notice that the charts are pretty close to each other and the map is not easy to interpret. To make it easier to understand which chart belongs to which site, we need to increase the spacing of the charts from the points. You can do this in the placement options (Figure 4.14).

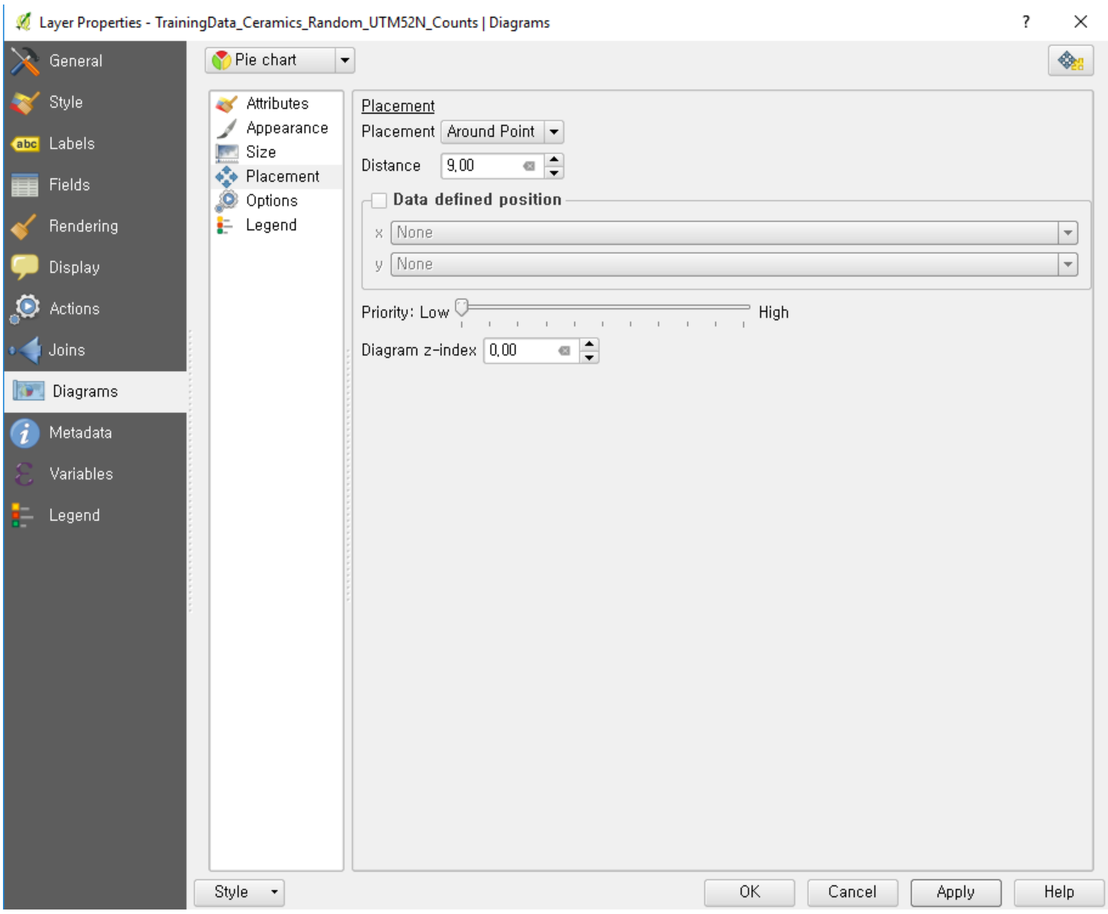


Fig. 4.14. Adjust the placement options for the charts in the Diagram dialogue.

Unfortunately, QGIS does not inherently recognize your pie charts as having geometry, so you need to create it. There are several techniques for doing this, and none of them are easy.

On my Mac, I was able to get this technique to work:
https://gis.stackexchange.com/questions/243058/moving-pie-chart-in-qgis-and-creating-leader-line
It is important (but not clearly stated) to create duplicate shapefiles to work with. One shapefile hosts the chart geometry and pie charts, the other shapefile has the points you want to draw the lines to. It is also important to move every single pie chart at least once to get the x_coor and y_coor fields to populate in the attribute table. You can copy this markdown code into the Expression Builder window and it should work:

```
make_line(   make_point(  "x_coor" , "y_coor" ),    centroid(
$geometry ) )
```

However, I was unable to get this technique to work on a PC because my Move Label and Diagram tool would not activate in the Editing functions. This may be a bug in the PC version of QGIS2.18. So, you can use the following work around, although it is much more time consuming.[8]

Just as in the Stack Exchange example, add X and Y fields in the Properties that will host your pie charts' central locations. In this example, I have called it Chart_X and Chart_Y, but you could make it x_coor, y_coor or whatever you like (Figure 4.15).

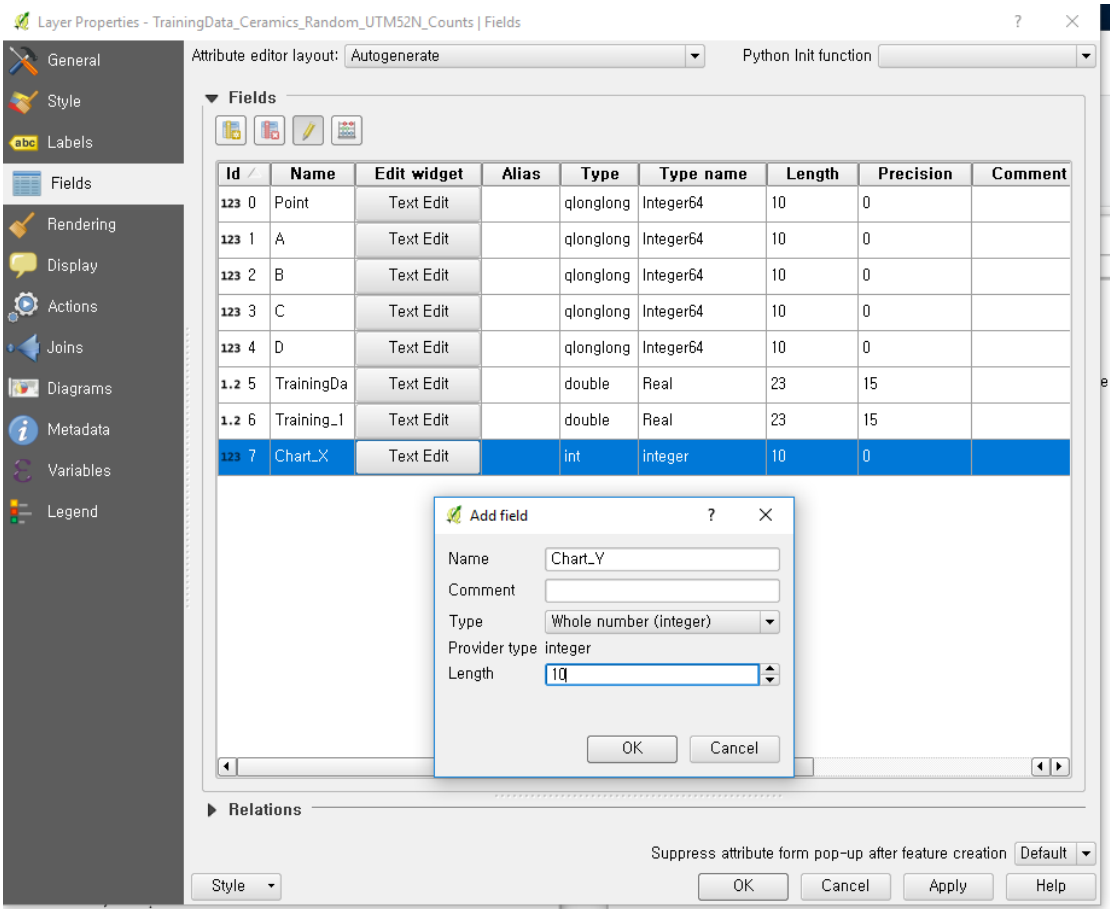


Fig. 4.15. Add a new attribute column in the Properties dialogue.

For the next phase, you need (1) be sure your map is set to the view you want to produce for your final product—if you move it, the lines will not connect to the points, and (2) to be sure your Coordinate Capture plugin is activated in your Plugins menu. It should appear in the lower left corner of your screen below the Layers Panel. Now click in the center of a pie chart you want to connect a line to. In the Coordinate Capture window you will see two sets of UTMs display (Figure 4.16). Click the "Copy to Clipboard" button. Then go to the attribute table and paste the values one-by-one into your sites[9], deleting the extra numbers that do

---

[8] If you toggle editing and can move the pie charts around with the move tool, take the first option. This next option is quite time consuming.

[9] This task is much easier if you have added labels to the sites. I used the Points field to navigate. Just click the Labels tab, then ""Show labels for this layer", Label with Point.

not belong in the given field. Note in the figure that you need to paste the values into the Expression Builder dialogue box above the table. You need to have the appropriate row highlighted and the field as Chart_X or Chart_Y and hit the "Update Selected" button. If you hit Update All, the fields will all populate with the value you pasted.

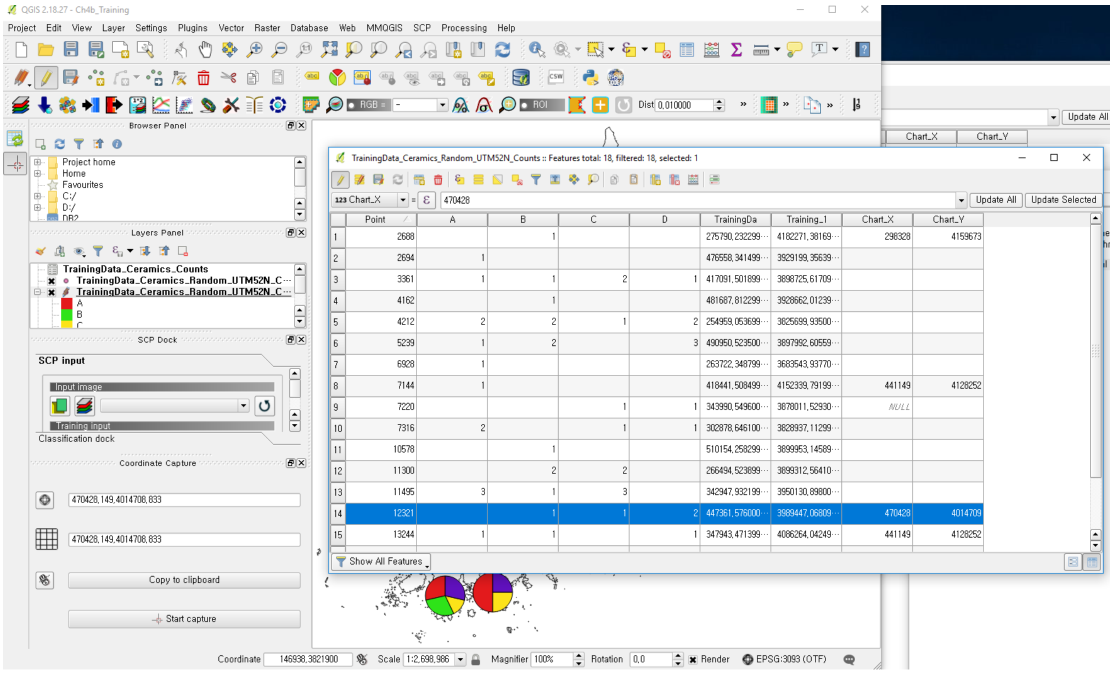


Fig. 4.16. Copy and paste coordinates from the Capture Coordinates plugin.

Now, you can create lines to your points by adding a Geometry generator field (lines) using the code (Figure 4.17):

```
make_line(   make_point( "TrainingDa" , "Training_1" ),   make
point("Chart_X" , "Chart_Y" ) )
```

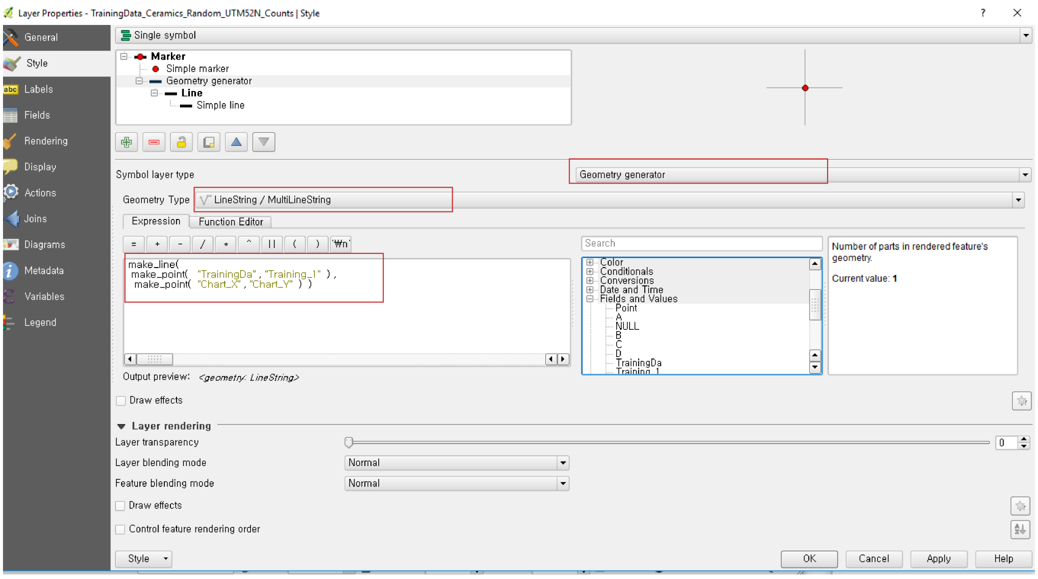


Fig. 4.17. Set geometry points of your charts. Note that if you change the zoom level of your view in the Map Window, these geometric points will not automatically adjust.

In the PC version of this exercise, I was able to generate a Map Window view as shown in Figure 4.18.

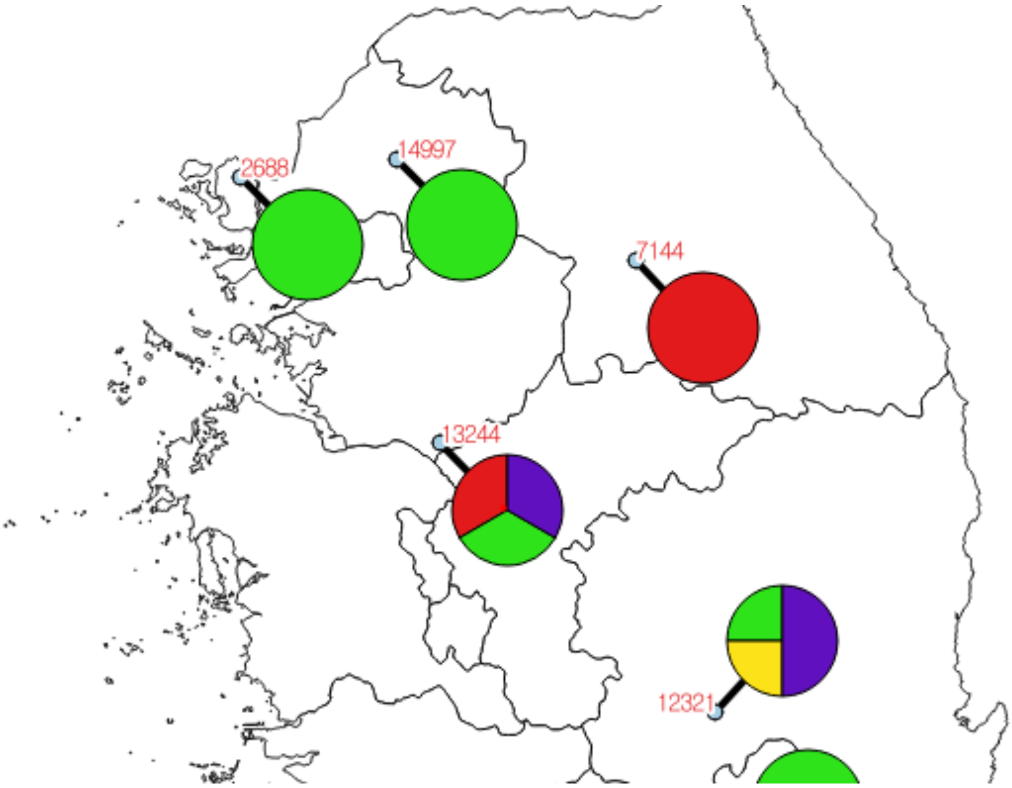


Fig. 4.18. Example of charts created with leader lines connecting to the sites.

You can create a legend in the Print Composer. You should have sufficient skills to make a legend that has the pottery types properly displayed (as well as scale bar, north arrow and other important elements to include in a map.

4.4 EXERCISE

You have been given an Excel file with a real set of point data that has information on Chulmun pithouse dwellings. Many of the points have multiple structures with different attributes. Create bins of the "min floor area(m^2)" category: 0-10m^2, 10-20m^2, 20-30m^2, 30-40m^2, 40-50m^2, >50m^2. Display these bins using bar charts from the Diagram dialogue. Create one map for before 5200 BP and one for after 5200 BP. Scale the maps for the best visibility of all of the data. Be sure to have a title, north arrow, double box scale bar (in kilometers) and modern-day administrative boundaries (use the KOR_adm1 shapefile). Have a DEM background that is visually pleasing and easy to interpret. There must be a legend for your diagram colors and DEM.

**NOTE*: Be sure to select Histogram > Size > Maximum value so that your bar charts are readable (Figure 4.19).

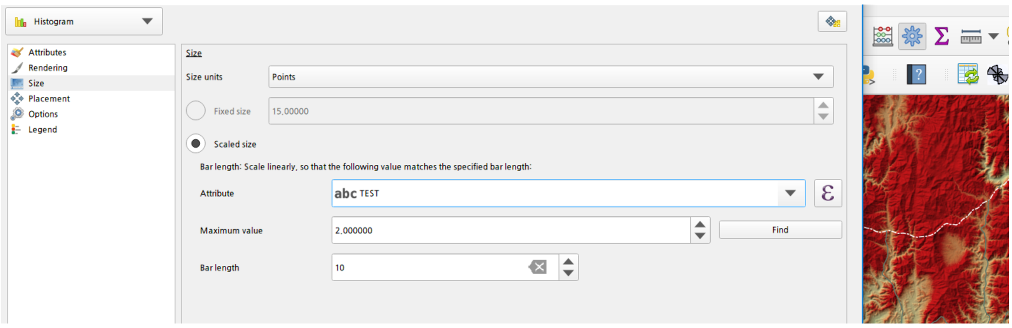


Fig. 4.19. Select the maximum size option for the histogram so that the bar charts are legible.

Save the file as a .png (*YOURNAME*_Ch4.png) and **upload it to the eTL by April 1 at 9:00 am** (5 points for completing this portion of the assignment correctly and on time). In addition to the image, answer the following questions (saved as *YOURNAME*_Ch4.docx):

STUDY QUESTION (1): Conolly and Lake (2006) discuss four functions (p. 51) and stages (pp. 56-57) of setting up a database. As you prepare for your master's thesis (or final project in this class), provide me with a brief conceptual outline of how you will prepare your data for analysis. Be specific. (2 points)

STUDY QUESTION (2): In the map you created for this week's assignment, what do the different sizes of pithouse floors tell you about Chulmun settlement overall in this portion of the country? In a paragraph or two, explain the trends you see in pithouse floor sizes over space and time. (3 points)

Chapter 5: Digitizing Maps

5.1. Principles of Digitization

Until the 1990s, most spatial information was hosted on paper maps created by professional **cartographers** who would assemble relevant bits of information and transfer it into two-dimensional (2D) formats. With the introduction and proliferation of personal computers, this situation has now changed and maps are only one spatial product used for detailed analysis. As we have learned, a GIS includes layers of data that can be manipulated and analyzed to provide the consumer with specific types of information for specific needs. For archaeologists, we can look at how sites fit into ecological or geographic systems, or, as in the last exercise, you were able to look at how the distributions of pottery rim sizes changed over space and time, which tells you something about social conditions.

Because the use of GIS is relatively new in the scope of the total history of cartography, there are many old maps that need to be digitized in order to be useful in a GIS. These **legacy data** are often invaluable records of research and need to be included into a GIS. Conolly and Lake (2006:78) list six principles that must be considered when digitizing an old map. Before you start digitizing anything, review these principles and keep them in mind throughout the entire process. Failure to account for these will result in a significant waste of your time. Always bear in mind that digitizing old data is inherently fraught with imprecision. It is your job to minimize uncertainty by keeping careful records of where the potential flaws in your data lie.

5.2. Georectifying to Digitize

Digitizing is a task that can be completed many ways in QGIS. To get you started, we are going to do the first digitization exercise in class together.

If you forget principles of digitization we learn in class, I recommend this website as a good basic tutorial.

First, we need to acquire the base images that we are going to use to digitize. Download and install the plugin called "TMS for Korea." In the eTL, I have placed a zipped shapefile (Namhansanseong.zip) that has the location of the Namhansanseong UNESCO World Heritage site. Open this file and navigate to the location. When you open up the TMS for Korea plugin Naver Street map, you can see the location of the mapped walls of the site, which are identical to the view provided in the Cultural Heritage GIS Service database. Note that when you open up the Naver Street map view, the site reprojects into EPSG:5179 (Figure 5.1). Let's just keep the coordinate system set to this (Korea 2000/Unified CS) for digitization purposes. We will reproject the data later.

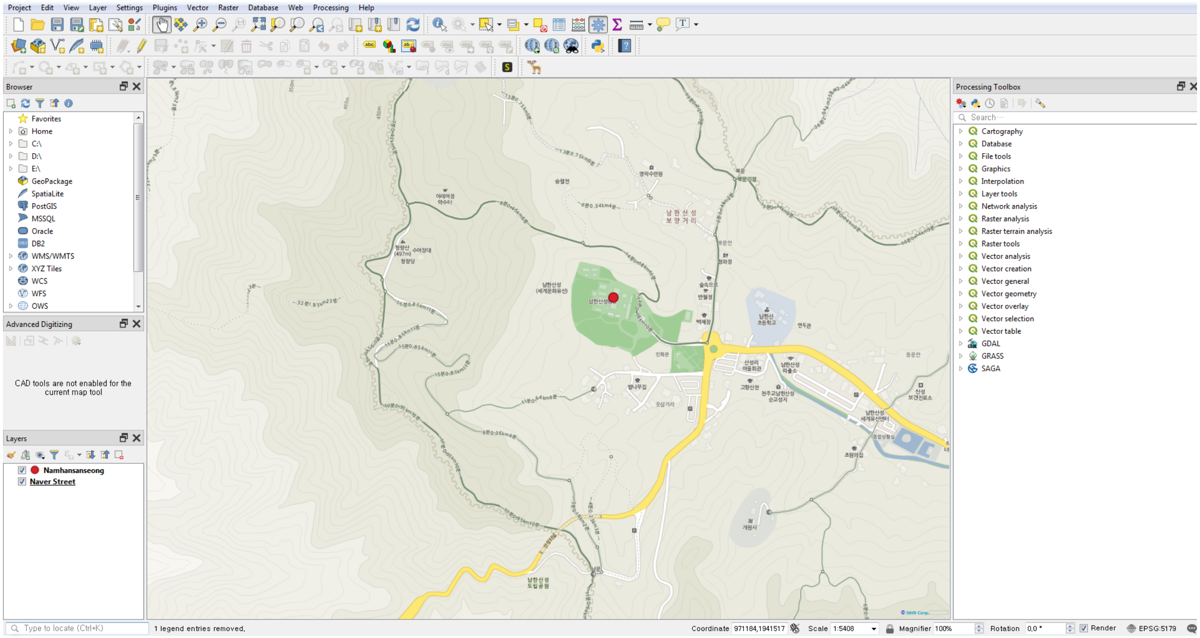


Fig. 5.1. Namhansanseong in with a Naver Map view.

Next, download an old map of Namhansanseong (http://lib.snu.ac.kr/find/collections/old/chosun). This map is said to have been made between 1911 and 1924. This is the file that we are ultimately going to digitize. I have Photoshopped the image and placed it in the eTL for your convenience (Figure 5.2). Note that the image is upsidedown in that south is at the top of the screen and north is at the bottom. It will be a good idea for you to change the aspect of the image by rotating it 180° before you open it in QGIS.

Fig. 5.2. Old map of Namhansanseong

We need to georectify this map based on the outline of our modern (georectified) Naver Map. For the purposes of this exercise, we will call Figure 5.2 the "old map" and the Naver Map is the "new map." Overall, this is not an easy process because portions of the old wall have been remodeled over the years. We will need to do our best to identify old portions of the wall still standing that we can connect to the old map.

The best way to georectify a map is to find "sharp features"—those areas where there are distinct identifiers that separate a portion of the map from the noise that surrounds it. If you notice in the southeast portion of the wall, there is what appears to be a parapet feature that juts out from the rest of the wall. Click Raster>Georeferencer… (if working from QGIS 2.18, you may need to download the Georeferencer plugin). A separate window will appear and you will click the little raster button in the upper left corner. Navigate to your old Namhansanseong image and select it and it will appear in the window (Figure 5.3).

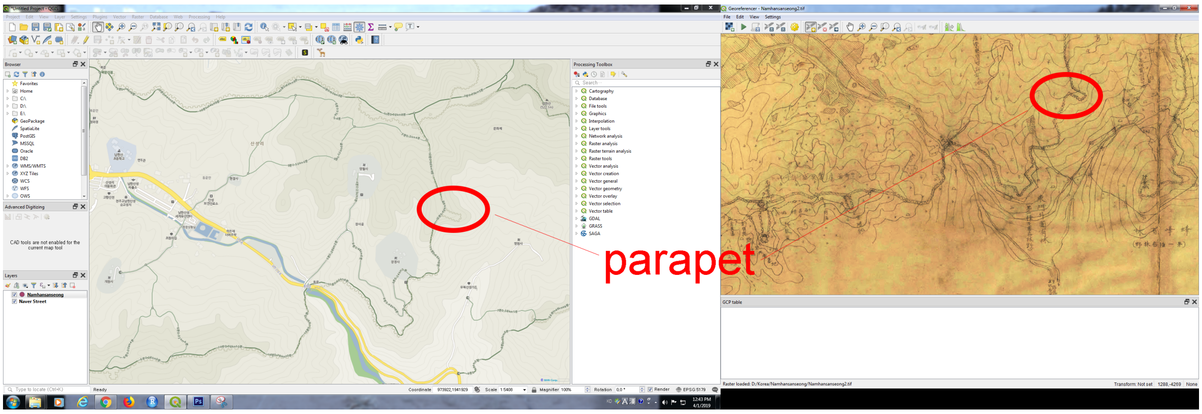


Fig. 5.3. Old Namhansanseongmap open in Georectifier (Madeira 3.4.5). The parapet is identified in both the Naver Map and Georectified view.

When you have your maps lined up, click the "Add Point" button (also available in Edit>Add Point). Select a portion of the two maps that you can match up pretty well. Click the spot on the old map first, select the option "From map canvas" when the Enter Map Coordinates dialogue appears, the Georeferencer window will disappear, then click the same feature in the new map. The dialogue will reappear populated with an X and Y coordinate. If you think this is right, click OK and a new Ground Control Point (GCP) will be added.

Once this process is completed (I made a total of 18 georeference points), click the green Start Georeferencing button to the right of the Add Raster button. A dialogue will appear that says "Please set the transformation type," and you click OK. Be sure the target SRS is EPSG:5179 and your transformation type is Linear and click OK (Figure 5.4). A bunch of red lines will appear. **Hit the Start Georeferencing button again** (don't skip this step or you will lose your work). The image will appear in your Map Window now.

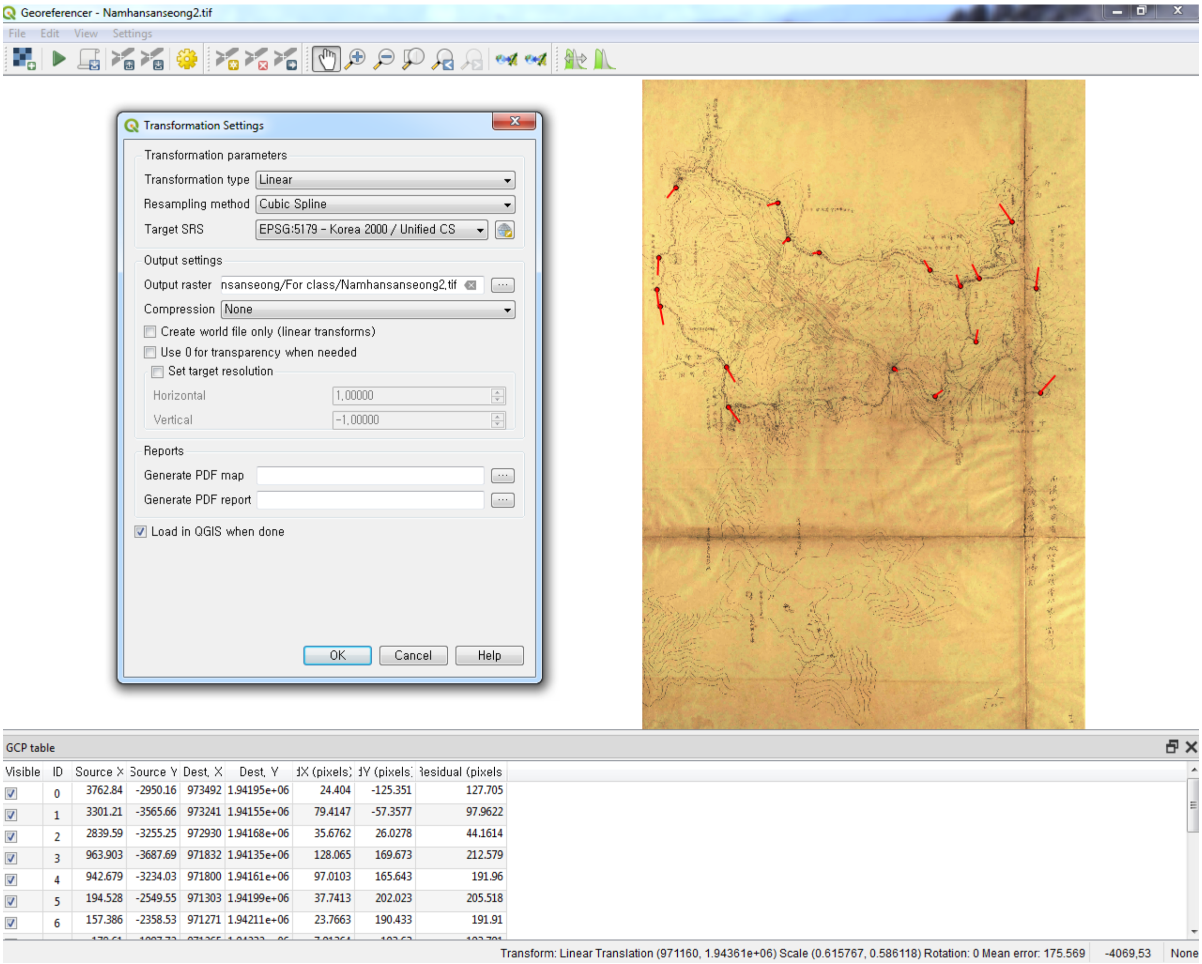


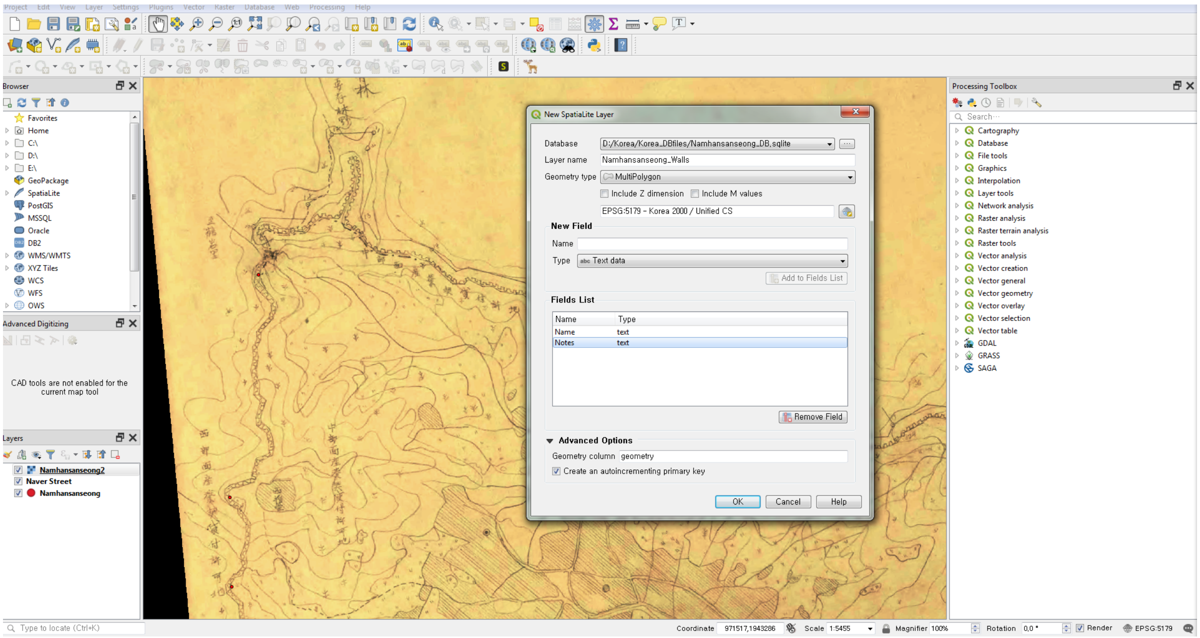
Fig. 5.4. Adjust the Georeferencer transformation settings.

If you think you did something wrong, I have provided the GCPs in my Google Folder. You can load those into the Georeferencer (File>Load GCP points...) and compare yours to the ones I made. It is easy to make a mistake. In order to double check your work, double click the old map you have georeferenced and in the Layer Properties dialogue, you can adjust the opacity to 50% in the Transparency tab and see how it lines up with your Naver map.

It is important to note that old maps are not as accurate as new maps, and so you will need to manipulate the map as best as you can. I also recommend trying another transformation type such as Helmert or one of the Polynomials to see what works the best. Georectifying old maps is one of the most important things you can do as an archaeologist, but also one of the most challenging. The more your old map lines up with the new map, the more accurate your digitization project will be.

5.3 Digitization of Namhansanseong

With the old Namhansanseong map in front of your Map Window, you are going to create a new Spatialite database from which you will digitize your points, lines and polygons. Go to Layer>Create Layer>New Spatialite Layer.... You will need to create a new database folder in your D:/ drive (I called mine Korea_DB files). You will then give the Spatialite layer a name (I called mine Namhansanseong_DB.sqlite) (Figure 5.5). This is the location where you will store all of your vector files with this digitization project. Before you continue, you will create a Multipolygon layer “Namhansanseong_Walls” (select MultiPolygon in Geometry type). Be sure to adjust the CRS to EPSG:5179! Add a couple of fields (one called Name and one called Notes). These are text data fields. Unlike shapefiles, Spatialite layers have unlimited text characters available, so you don't need to adjust this parameter. Now click OK.



In the Layers Panel, you will seeNamhansanseong_Walls with a default fill. It is terribly difficult to digitize a polygon when there is a fill selected, so I recommend you adjust the fill settings using a Layer Properties dialogue (Figure 5.6).

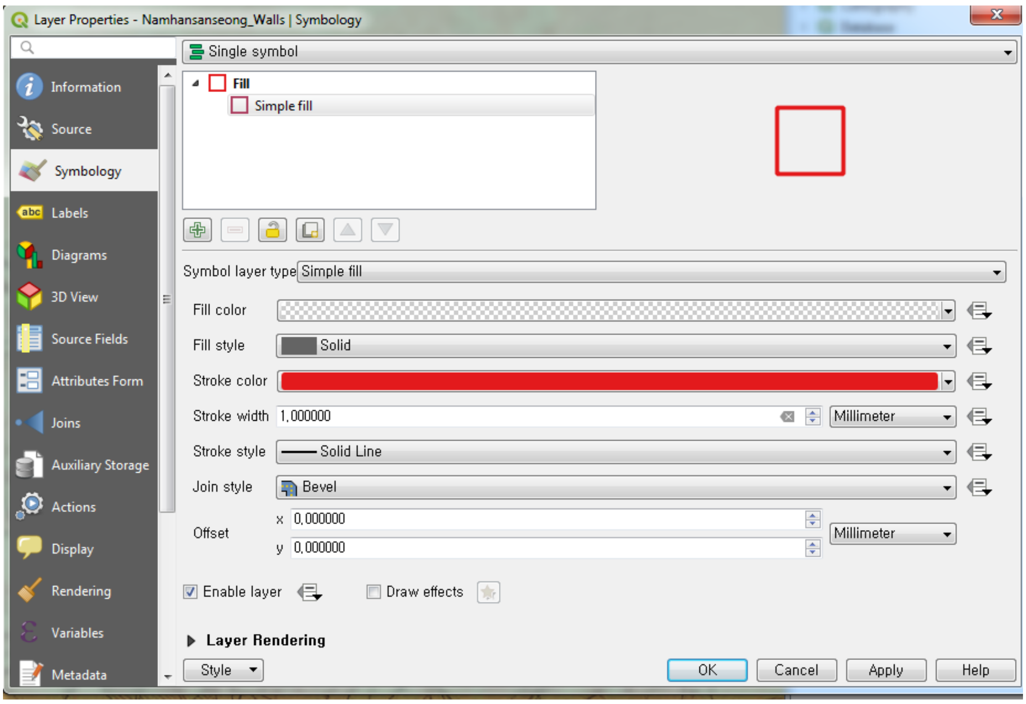


Fig. 5.6. Adjust the Symbology of Namhansanseong_Walls. Select a transparent fill in Fill Color.

Now, you are ready to digitize! Click the little yellow pencil to toggle the editing options. Once you click it, your digitizing tools will light up (I have optional ones selected in my screen menu…these can be activated in View>Toolbars>DigitizingTools). The second button to the right of the Toggle Editor button is the Add Polygon Feature button. Click this. A little target will appear in your screen. Line the target up with the walls of Namhansanseong and begin left clicking around the edge of the wall (Figure 5.7). This tool takes a little time to get used to. Unfortunately, Ctrl+Z/Cmd+Z will not do anything here. If you screw up, you have to fix it later. We will learn how to do this later.

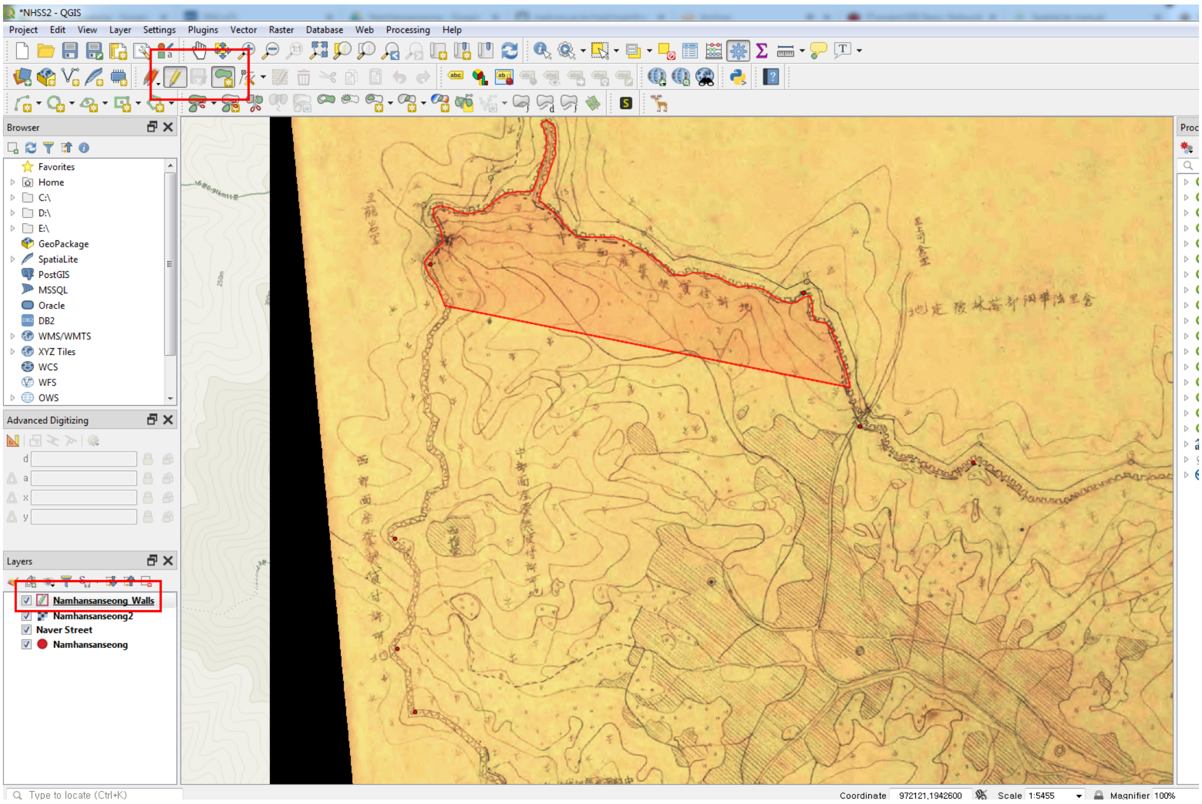


Fig. 5.7. Begin digitizing the walls of Namhansanseong as a polygon.

When you have made a coherent perimeter around the site, you will right click at the last spot and a dialogue will appear (Figure 5.8). You can give the wall a name (I called mine "Main") and add some notes to clarify what, specifically, you have digitized.

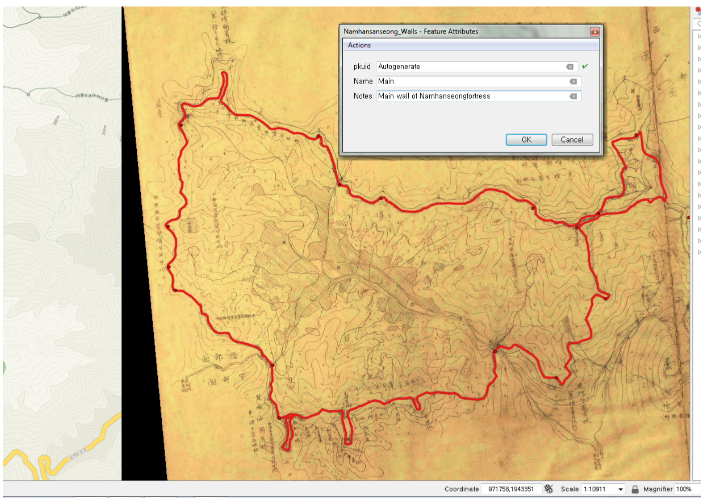


Fig. 5.8. End digitizing the first wall.

After you have created the polygon, you should go around the perimeter and identify mistaken locations of nodes. First, select the Vertex Tool (to the right of the Add Polygon Feature button) and select it. If you notice that a node is not located on the wall, hover over it and when it lights up with a big dot, click it. You can move your cursor to the location where you want to place the node and left click it (Figure 5.9). Another unfortunate feature of QGIS is that moving nodes is not very flexible, particularly if you have to just move them short distances. However, you can always zoom in on it using the middle wheel of your mouse and it gives you better resolution for adjusting the node positioning. The node can also be slid down the segment, which is convenient for small adjustments.

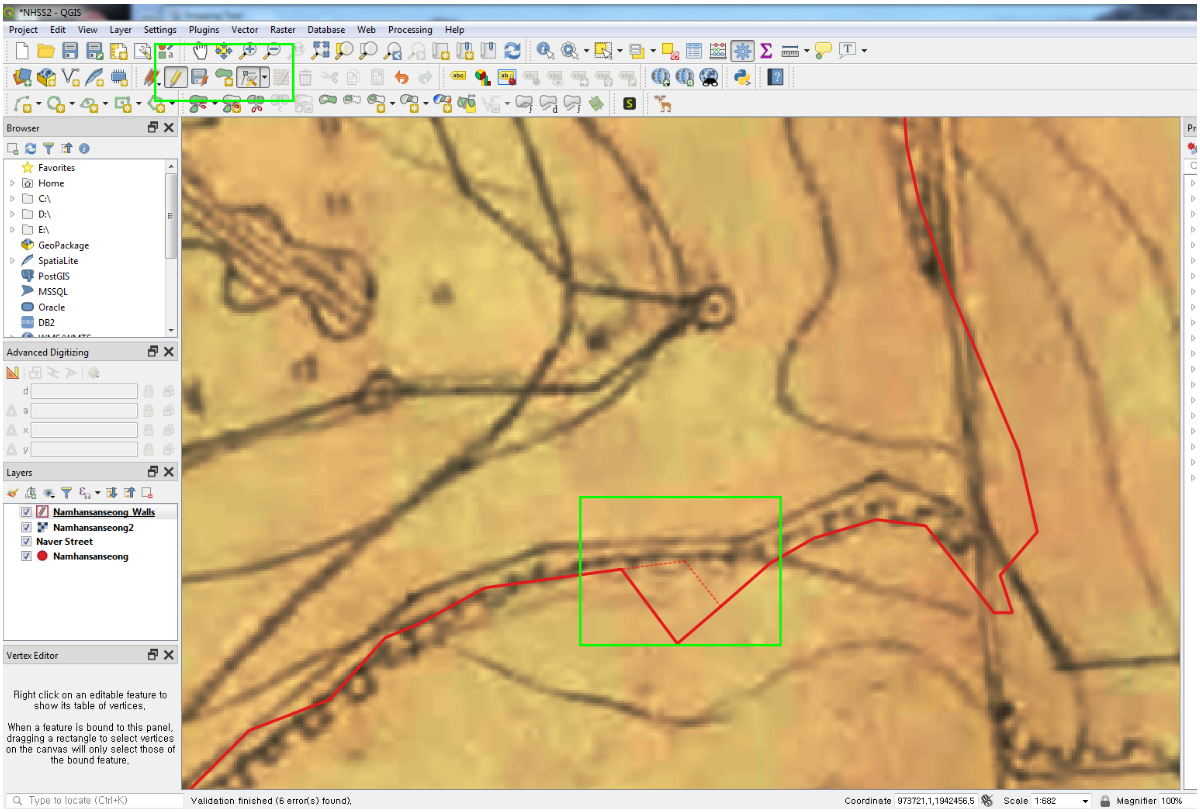


Fig. 5.9. Edit the vertexes of the Namhansanseong digitized wall.

Some portions of the wall have compound walls and these must be digitized with care. Here, you will need to refresh your memory about **topology**, which was covered in previous readings. The idea in this situation is that walls that are shared between features should be commonly shared boundaries and not separate. Navigate to the southwest corner of the wall and you will see one such feature. Click Add Polygon Feature and digitize the extra wall (Figure 5.10). You will notice that as you approach the wall you already digitized, the digitizer tool is attracted to the vertices. If this is too sensitive, you can adjust the setting in the Processing Toolbox, all the way to the right hand side there is a wrench (Options). Along the left panel of the Options is a tab that says Digitizing. You can disable "snapping" by unselecting "Enable snapping by default." More easily, go to Project>Snapping Options... and a dialogue will appear in the lower left side of your screen below the Layers panel. If you click the magnet, it will deactivate snapping. If you need snapping back on, click the magnet again. You will have to experiment to see what you need...it depends on the circumstance. The important thing is that the newly digitized polygon should line up to all of the vertices of the common topological feature they share.

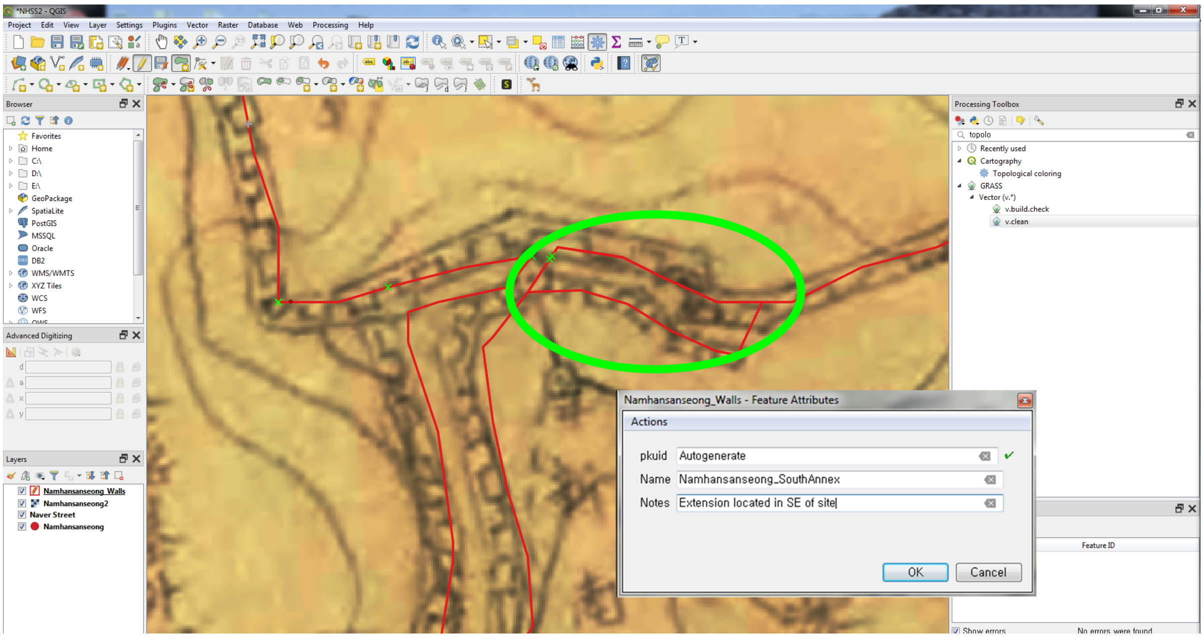


Fig. 5.10. Make common topology for digitized polygons.

You will need to digitize all of the Namhansanseong walls. In addition, for this week's assignment, you need to digitize lines and points. I have digitized the road that approaches the site from the north by adding a new Spatialite layer (Layer>Create Layer>New Spatialite Layer.... This time the Geometry type is Line. Be sure the coordinate system is set properly because it will automatically default to EPSG:4326). Once you have added the new layer, toggle the editor and start digitizing the line. You can adjust the colors in the Symbology editor (Figure 5.11).

Fig. 5.11. Create a road.

Moving forward, you can digitize buildings from the Naver Map (Polygon or Multipolygon), old Namhansanseong settlement area (Multipolygon) and streams (Lines). There are points in some of the TMS for Korea maps (e.g., 사지 in the NGII Street Map). The cadastral maps also have a lot of information that can be digitized. It is too much to digitize the cadastral boundaries of the entire Namhansanseong Fortress unless this is your final project for the class, but it may help you find the temple boundaries.

5.4. EXERCISE

Either continue the Namhansanseong digitization exercise or begin a new one that you will use for your final paper. If you decide to continue with Namhansanseong, you will need to get some more maps or data to digitize from additional sources. You need to digitize at least 100 features in total and there must be all three classes of data (points, lines, polygons). You must also have some bordering polygons, which have been checked for their topology. Some advice on how to do that can be found here. I recommend double-checking your map using the Topology Checker plugin (because I will when I grade you). You can digitize from any medium you like, but you have to do the work yourself. The data must be projected into EPSG:3093 (or another UTM system if you are working in a different country beside Korea). Create an attractive map with a nice panchromatic background (the TMS for Korea plugin has some great satellite images you can work from). Your map needs to have a title, north arrow, scale (probably in meters), and a zebra border with UTMs shown (see this website for step-by-step instructions on how to do that). Save the maps as *YOURNAME*_Ch5.png, also zip the Spatialite file along with your answers to the study questions (*YOURNAME*_Ch5.zip) and upload them to the eTL by **April 15 at 9:00 am**. You must turn in only one Spatialite file. You will receive 5 points for creating an accurate, legible, visually attractive map that has at least 100 features as well as the Spatialite file in the proper format. If you decide to digitize from another map beside Namhanseong, please also provide a geotiff (tiff file with a Worldfile [.wld]) that I can use to crossreference your vector files. In addition to this assignment, answer the following questions and save your answers to a single document (*YOURNAME*_Ch5.docx):

Study Question (1): Explain the process you used to digitize the map you created. How were the original data collected (see Conolly and Lake 2006:61-76 for examples of what I mean)? What are some of the limitations to the digitization you undertook? In other words, what other kinds of data would you like to have accessed to improve the accuracy of your map? (3 points)

Study Question (2): Explain the concept of **topology** and how you accommodated this issue in creating your digital map. (2 points)

Chapter 6: Surface Models

6.1. Principles of Creating a Surface Model

Surface models are interpretations of how points connect in space. A surface model with which you are likely familiar is a digital elevation model (DEM), which is created on the basis of algorithms that connect together absolutely measured points in space. It is important to realize that if you are standing on a ground surface, looking at Google Earth (from which the elevations are created based on a DEM) on your phone, it is highly likely that there is a significant (±3 m) difference between the interpolated value and the actual value. However, the more absolutely measured points you have to work with, the more precise the interpolation values are.

Similarly, georectified images are types of surface models as the pixels from which they are constructed must be stretched over 2D space. Panchromatic images, such as those that can be accessed from Bing and Google Earth, provide "naked eye" views of earth from space. The colors that comprise the pixels are interpretations of chromatic values (think R,G,B or C,M,Y,K) of the ground surface after they pass through earth's atmosphere. On irregular surfaces like mountains or steep river valleys, significant interpolation on a ground surface model is needed to make the view seem realistic to the human eye.

In this module, we will learn some basic principles of generating surface models and stretch a georectified image over one such model. You will learn how to blend topographic and panchromatic images together to create an interpreted view of a ground surface.

6.2. Importing the Data and Getting an Overview

For this week's in class exercise, we will go to the country of Malawi in south-central Africa and build a surface model from some topographic points followed by creating a georectified image underlay. In the eTL, you will see a folder entitled "BruceShpFiles"[10]. You will need to create a new country folder in your GIS files (Malawi) and then make three subfolders as you did in Korea: Rasterfiles, Tablefiles and Vectorfiles (Figure 6.1). Import the BruceShpFiles folder into your Vectorfiles directory.

[10] These data were collected by the Malawi Early Middle Stone Age Project (MEMSAP, https://memsap.org/) under the direction of Dr. Jessica C. Thompson in cooperation with the Malawi Ministry of Civic Education, Culture and Community Development. An outstanding team of local crew worked on all excavations along with their cohorts at the Catholic University of Malawi. Fieldwork and analysis were funded by National Geographic-Waitt Foundation grant W115-785 10 (JCT), Australian Research Council Discovery Project DP110101305 (JCT), Wenner-Gren Foundation grant 8539 (JCT), the University of Queensland Archaeological Field School (JCT), the Korean Research Foundation Global Research Network Grant 2012032907 (DKW) and Emory University (JCT).

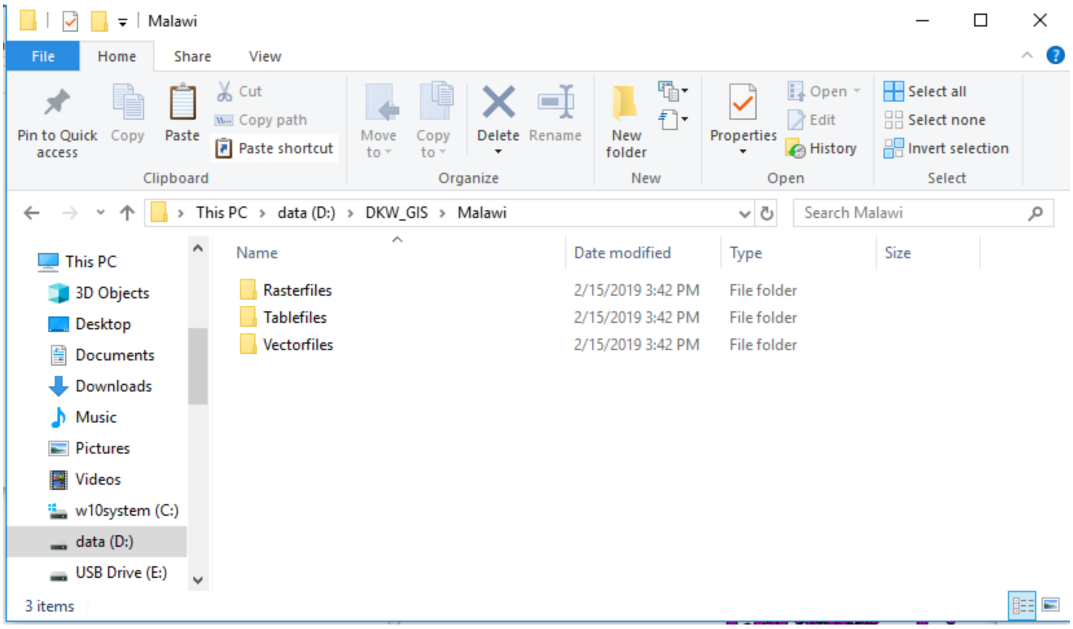


Fig. 6.1. Create a Malawi folder with appropriate subfolders.

Open up QGIS2.18 and start a new project. Adjust the CRS to WGS84/UTM Zone 36S (refer to Ch. 1.5). Be sure "on the fly" CRS transformation is enabled. Add the shapefile entitled "Bruce_Topography" as a layer. This can be done by dragging the .shp file to the Layers Panel, or if that doesn't work, you can go to Layer>Add Layer>Add Vector Layer… (refer to Ch. 3.2). In your Map Window, you will have a series of points as in Figure 6.2. Note that there is a dense concentration of points in the center of the screen.

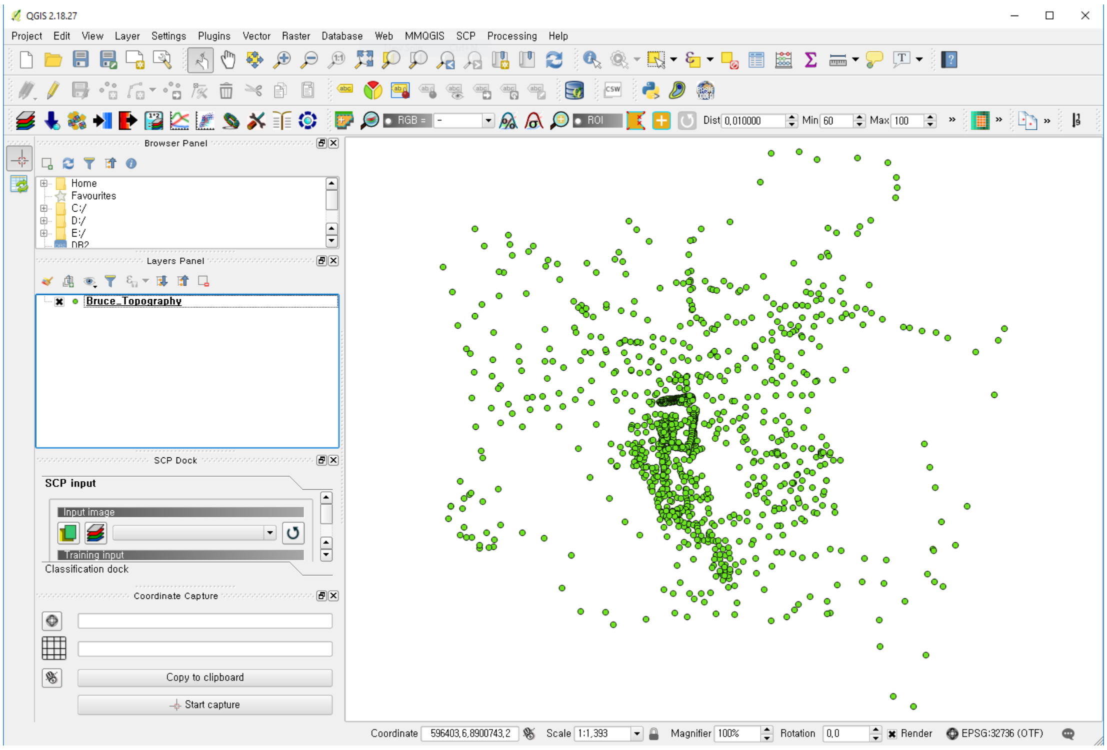


Fig. 6.2. Add the point shapefile of the Bruce site to your Layers Panel.

If you right click the file and open the attribute table, you will see 1116 points with a Northing, Easting and Elevation recorded to the scale of a millimeter between 25 July and 9 August 2012 using a few different instruments. If you click on the Elevation field you will see that the numbers span between 510.169 and 526.783. These are recorded in meters above mean sea level. You should already know that the UTM coordinates are in meters relative to a transverse mercator grid that spans the planet. The project area is about 300 m east to west by about 275 m north to south.

There are a couple of ways to interpolate a raster from elevation points in a vector. The simplest tool is the v.surf.idw GRASS tool. Be sure your Processing Toolbox is enabled (View>Panels>Toolbox or hit Ctrl+Alt+T). Type in "interpolation" and you will see the v.surf.idw tool. Double click this and the dialogue box will appear.

Your Bruce_Topography file should already appear in the Input vector layer field. The default number of interpolation points is 12, which is fine. We can leave the power parameter at 2, which squares the value of the weight given to surrounding points in constructing the surface model. The field "Attribute table column with values to interpolate" needs to be changed to "Elevation." Change the GRASS GIS 7 region cellsize to 1.0, which is telling the computer to make a 1-m scale grid. You will need to make a new folder in your Rasterfiles folder called "Bruce" and then give the file a name that tells you what kind of surface model you have created (Figure 6.3).

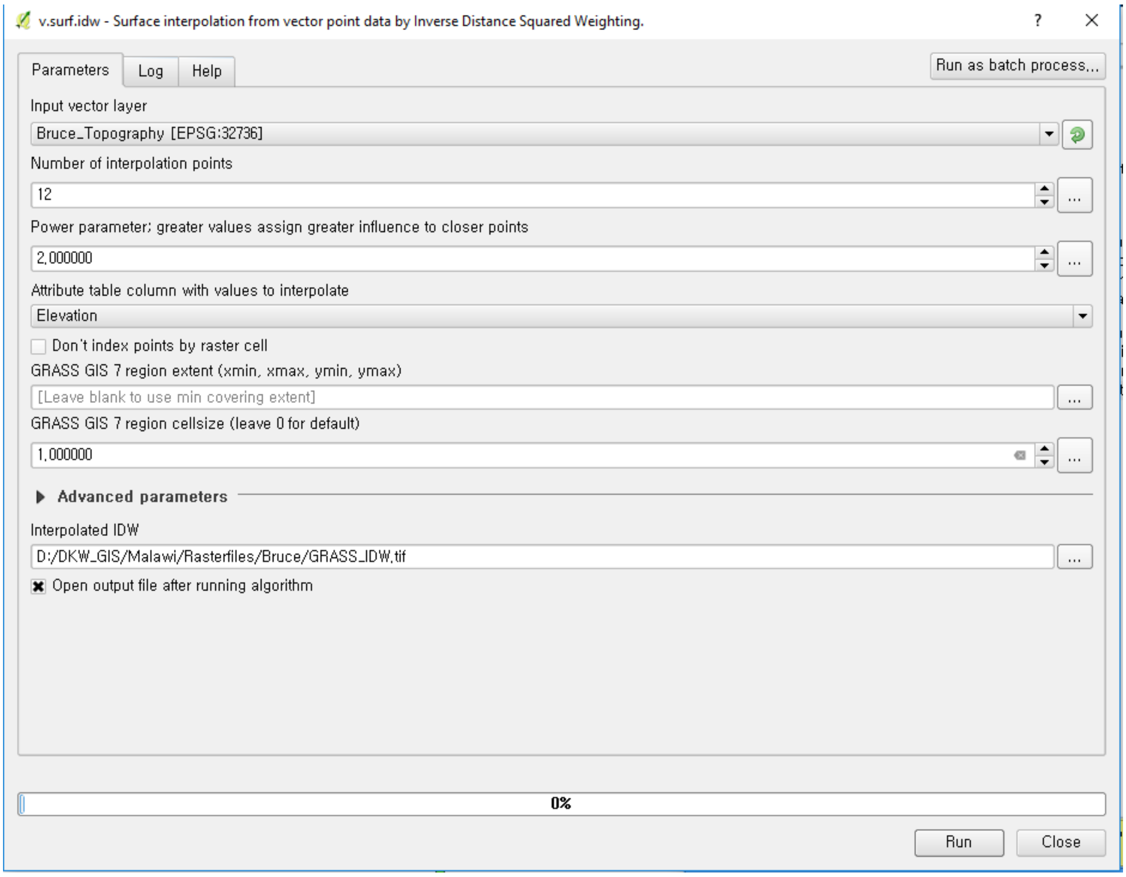


Fig. 6.3. Create an IDW surface model of the Bruce site.

You now have created a surface model. If you move your shapefile above your tiff in the Layers Panel, you can get a sense of how it fits together (Figure 6.4).

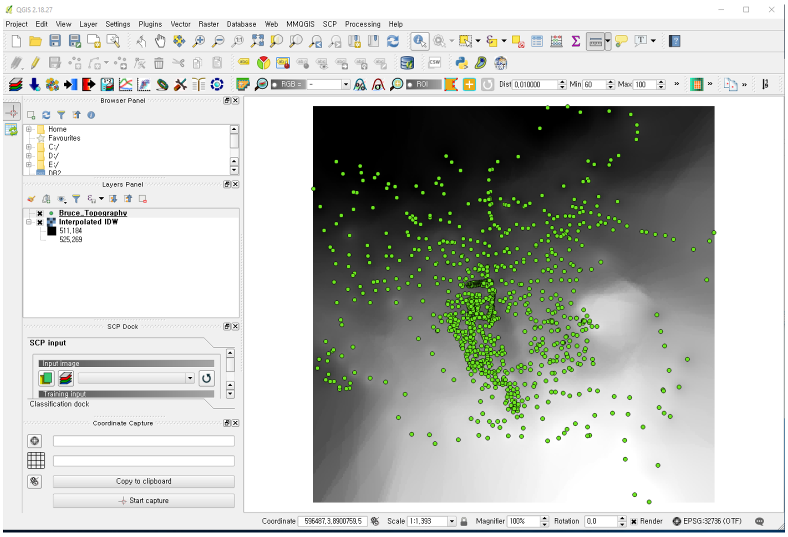


Fig. 6.4. IDW model of the Bruce site with point data.

Of course, this view isn't very easy to interpret, so we will play with the color a bit. Open the Layer Properties and select Singleband psuedocolor. You can pretty much use whatever color values you like, but you need to be sure to click the "Classify" button after you adjust the color scheme, and, as always, for rasters, I recommend changing the Resampling to Cubic when zoomed in and Average when zoomed out (Figure 6.5). This will give you a nice smooth appearance rather than blocky. Click okay and now things should be a bit clearer.

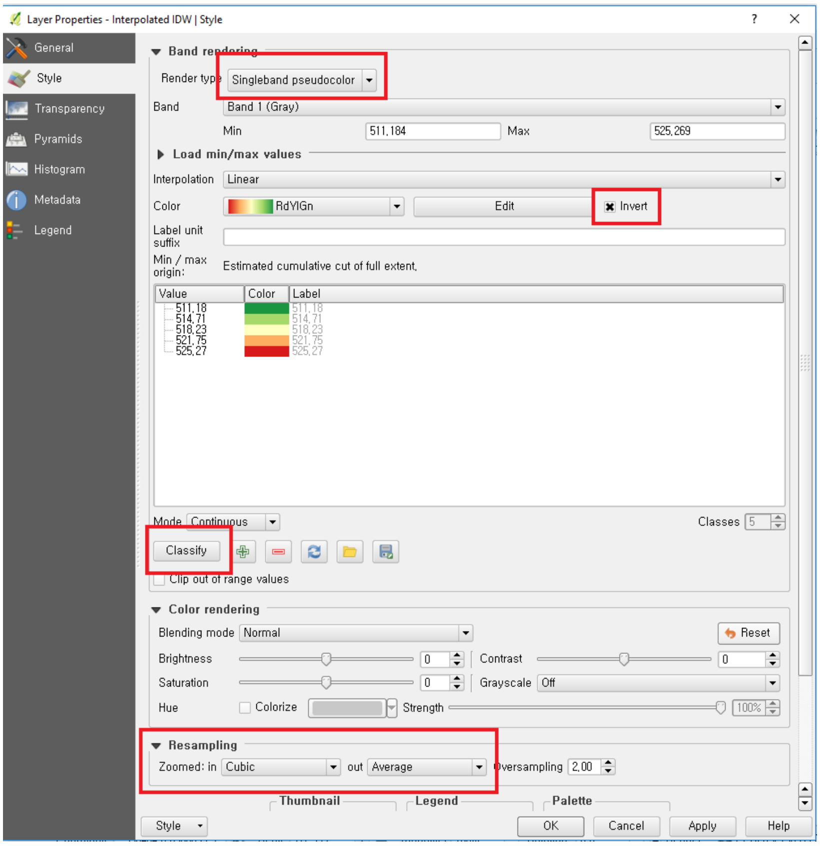


Fig. 6.5. Adjust the pseudocolor scheme of the surface model.

With this raster map, we can now create a contour map. Type "contour" into the Processing Toolbox and you will see the GDAL tool Contour (you can do the same thing with the SAGA toolbox if you have a Mac or). Make a 2-m elevation contour map by setting up the parameters as shown in Figure 6.6.

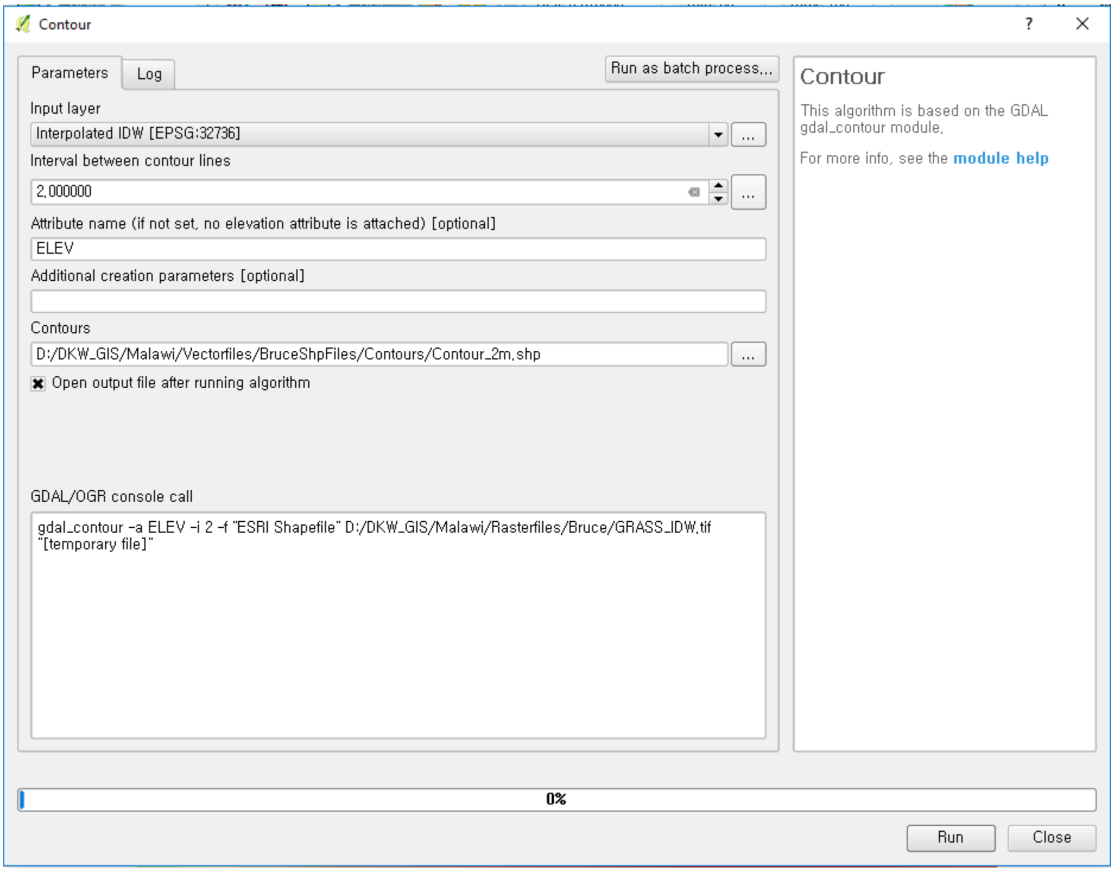


Fig. 6.6. Create 2-m contour vectors of the Bruce site.

Create a contour folder in your Vectorfiles/BruceShpFiles folder. After clicking Run you will have an image similar to Figure 6.7.

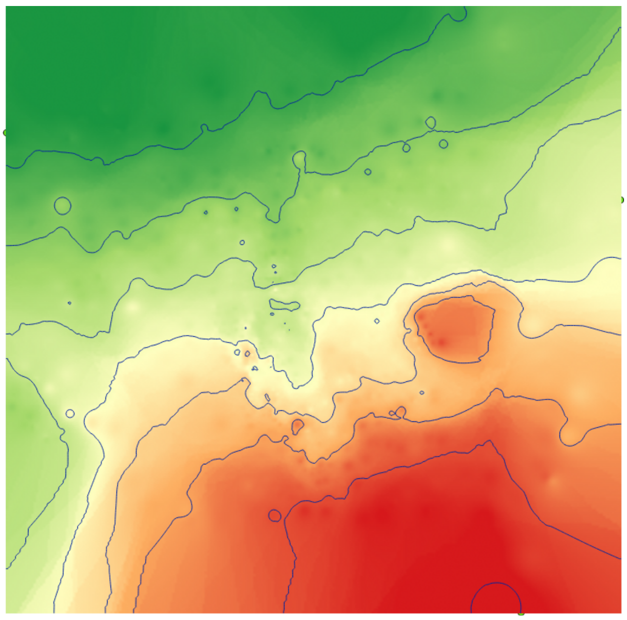


Fig. 6.7. IDW surface model and 2-m contours of the Bruce site derived from total station point data.

For comparative purposes double click the SAGA “Contour lines” tool and set up the analysis as show in Figure 6.8. Be sure that your Equidistance is set to 1 (which is 1-m) and click Run.

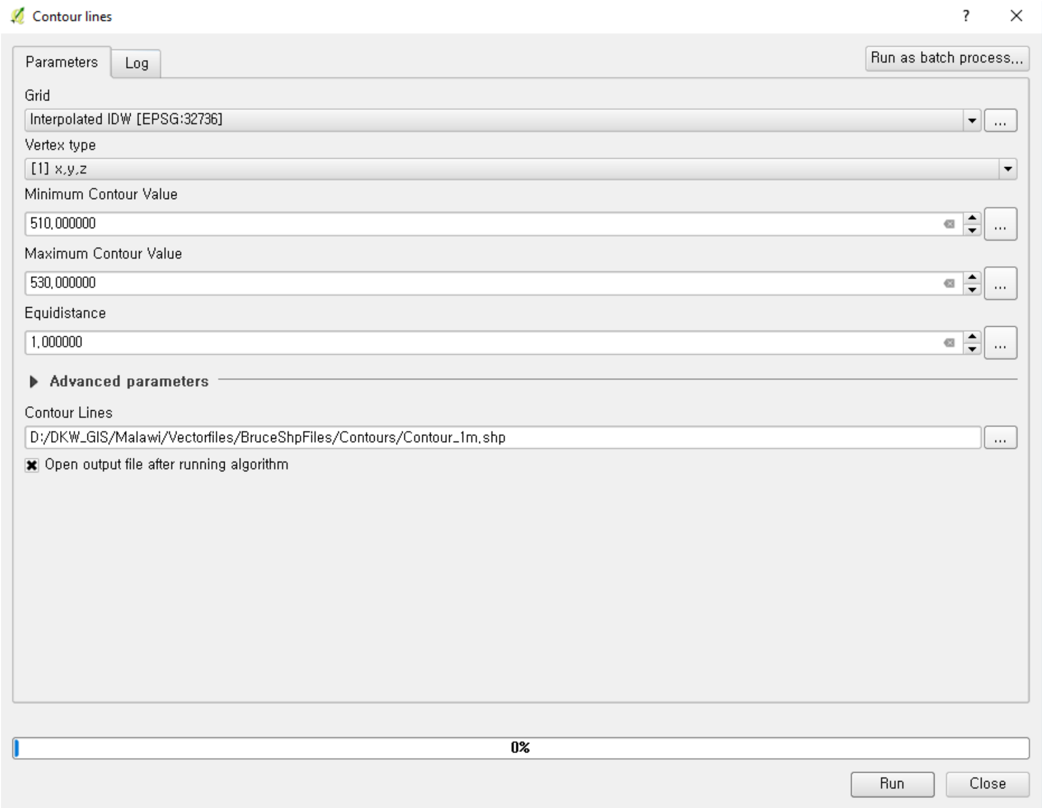


Fig. 6.8. IDW model of the Bruce site created in SAGA.

When the contours load, they have significant overlap (Figure 6.9), but if you zoom in closely on the 2-m lines, you can see some small differences in the way the algorithms worked.

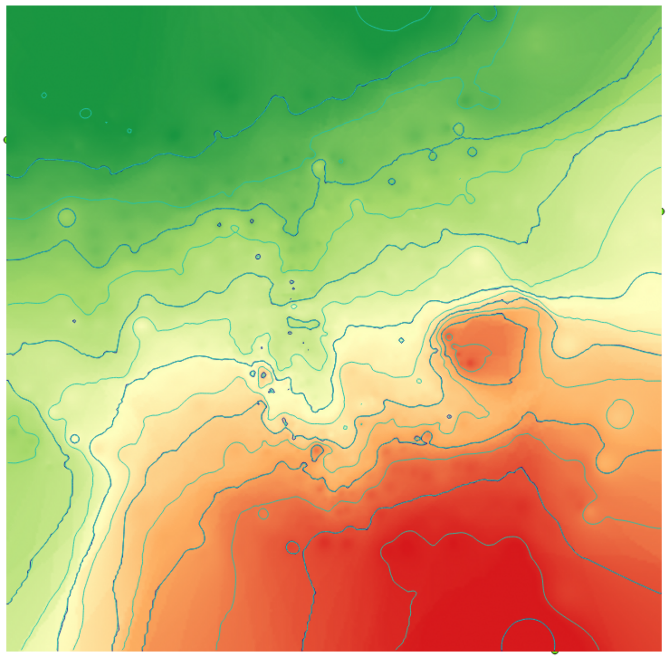


Fig. 6.9. IDW surface model created in SAGA and 1-m contours of the Bruce site derived from total station point data. Zoom in on the contour lines to notice subtle differences in the surface models' outputs.

Following a similar logic, but getting back to raster interpolation, type "interpolat" into the Processing Toolbox search bar again and open up the Interpolate (cubic spline) tool. Enter the values as shown in Figure 6.10 being sure to adjust the Cellsize to 1. Click Run.

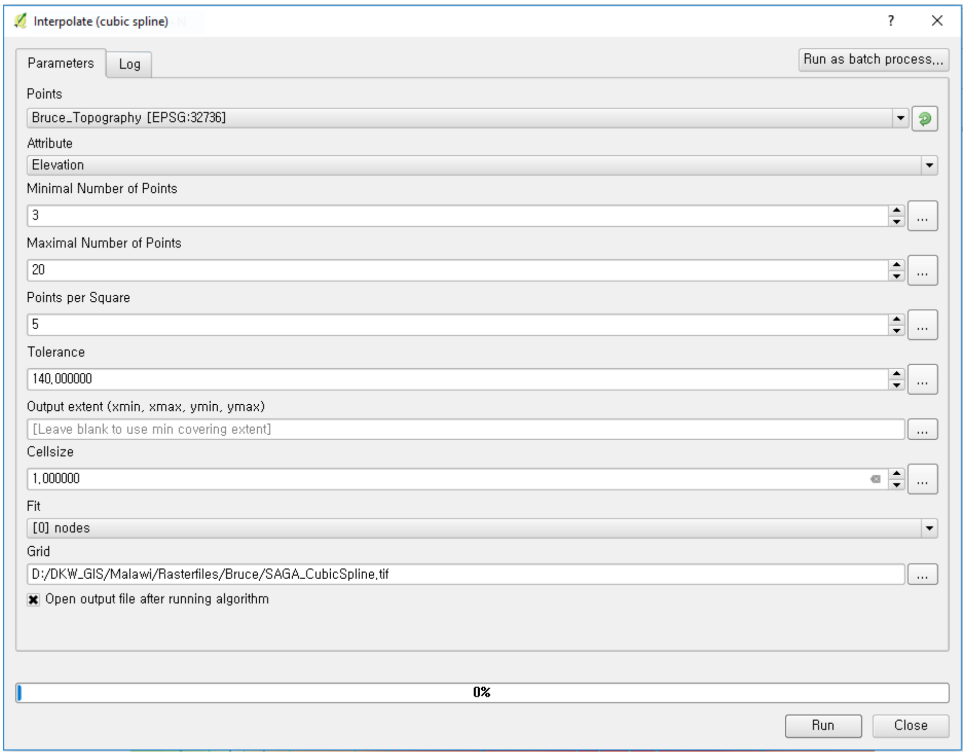


Fig. 6.10. Set up a cubic spline interpolation of the Bruce site.

When the image is created, adjust the Style properties to the same parameters as what you had in your IDW model (Figure 6.11). If you click back and forth between the views of the different models, you can see different surface representations inherent in those models. Remember, surface models are <u>representations</u> of the space between topographic points.

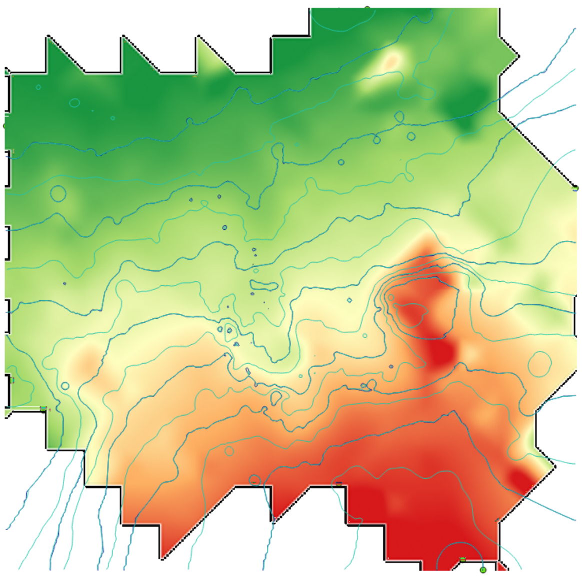


Fig. 6.11. Cubic spline surface model and 2-m contours of the Bruce site derived from total station point data.

6.3. Integrating a Google Earth Image

6.3.1. What is Google Earth?

In recent years, there has been probably no easier tool developed for land surface reconnaissance than Google Earth. Google Earth uses a satellite cluster called Quickbird to create panchromatic images of earth's surface. Users can now download a free version of Google Earth Pro, which allows you to create high-resolution imagery from screenviews of particular areas. Google Earth is an excellent first resource to access when you are thinking about working in a new project area and the resolution of the images are typically <2m.

For all of its advantages, there are two significant limitations to Google Earth: (1) the data is panchromatic only, meaning there is no multispectral dataset available, and (2) unlike files you download from the USGS or European Union, the images are not georectified, meaning you have to do this yourself. As you learned in last week's exercise, this introduces some measure of imprecision and uncertainty in the accuracy of your images.

6.3.2. Downloading Google Earth Pro imagery

With full realizations of its limitations, Google Earth images are some of the highest resolution images you can find, so it is sometimes worth the effort to integrate it into your interpolation model.

The first step is to download and install Google Earth Pro onto your computer if it is not already present (https://www.google.com/earth/versions/). Update your version if it is already installed on your computer. You will not want to use the Google Earth for Chrome option. You will need a Google account (e.g., Gmail) to download and install the program.[11]

The second step of the process is to define your project area scope and download the right size image to accommodate your needs. You are probably unfamiliar with the location in Malawi where you have created an initial interpolation model. I can tell you that it is near the city of Karonga, and you can type that into the Google Earth Pro search bar and it will get you close to the area, but you will notice that there aren't a lot of landmarks to navigate you into an exact location.

The easiest way to find your project area is to export your Bruce_Topography shapefile as a .kml file, which is a spatial format readable by Google Earth. Right click the Bruce_Topography shapefile and under the Format option, select "Keyhole Markup Language [KML]", assign the file a new name like Bruce_Topography_32736 and be sure to select the CRS as EPSG:32736. Once that is done, drag the file from the folder where you have saved the file over to the Places panel in Google Earth. The program will zoom in on the project area. If you hover your pointer on the navigation tool on the right side of your screen, you will notice a North indicator and a small eyeball in the center. If you hold the left button on your mouse down and pull downward you will notice that the view

---

[11] *NOTE:* The Quick Map Services (QMS) plugin gives you georectified Google Earth images, but we are performing an exercise here that you will need to know how to do. Follow the instructions provided for how to georectify "flattened" images because there will be many times you will need to know how to do something like this.

switches to plan view from a semi-profile view (Figure 6.12). If you double click the “N”, your view will be reoriented so that north is at the top of the screen.

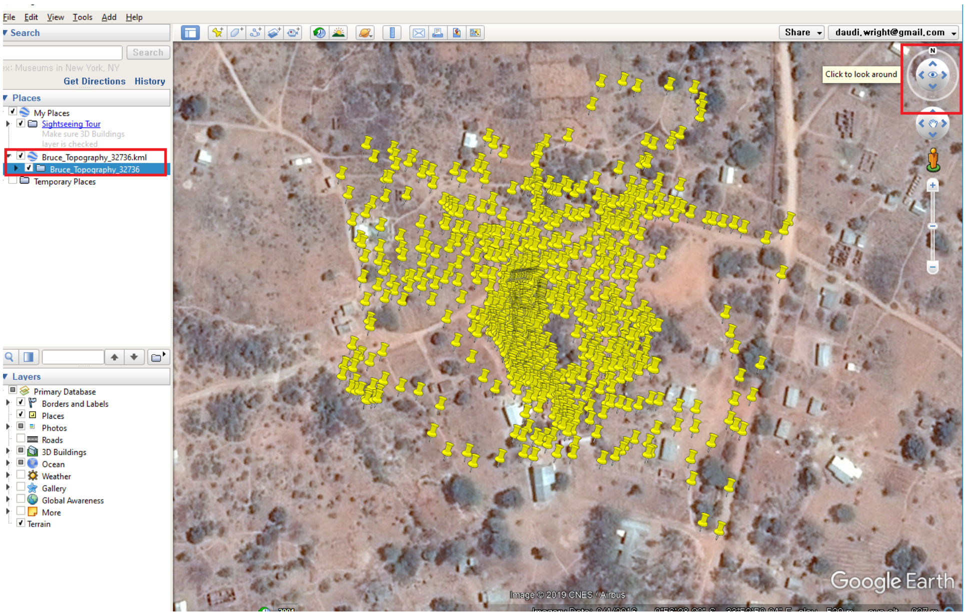


Fig. 6.12. Google Earth view of the Bruce site with total station points imported as a .kml file.

Now you have located your project area and if you unclick your kml file, you will see that you have a nice image of the area. You can see in the navigation pane at the bottom of your screen that the coordinate system is in degrees, minutes, seconds (WGS 84), which is the default CRS of Google Earth. Click Tools>Options… and under Show Lat/Long, select Universal Transverse Mercator (Figure 6.13). Be sure that Meters/Kilometers are selected as your Units of Measurement.

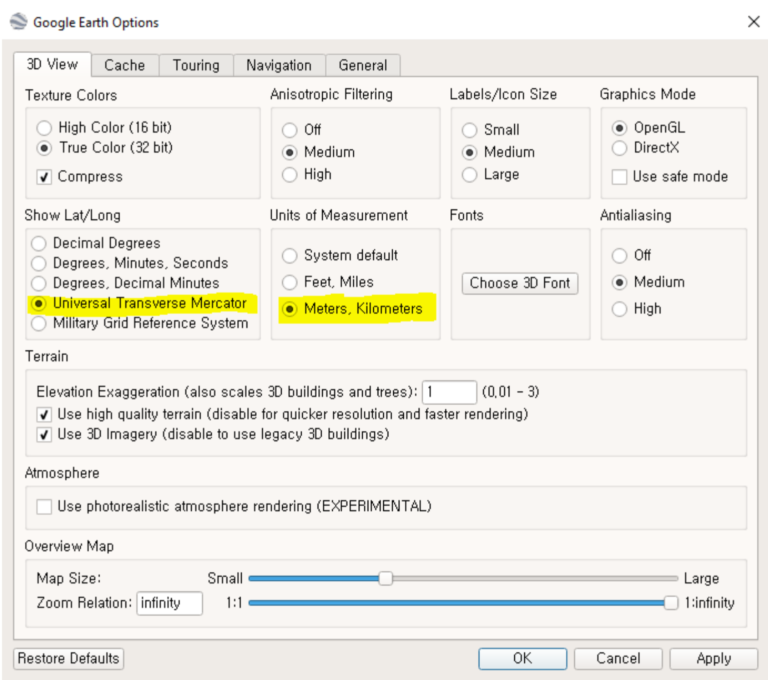


Fig. 6.13. Adjust the options in Google Earth to match the outputs we are going to use to georectify the image.

6.3.3. Preparing to georectify a Google Earth image

You are going to need to georectify this image, and so you need to provide some orientation points for yourself to make the process easier. If you left click on > arrow next to Bruce_Topography_32736.kml, a folder with the same name will appear. Left click on the > arrow to the left of the folder and you'll see a bunch (1116, to be precise) of yellow pins open below the folder. Right click the first one and select Properties. An edit dialogue will open up. Select the Style, Color tab and click the yellow pin to the right of Name (Figure 6.14). Make it simple and give it a new pin like 1 and rename the Placemark as 1 as well.

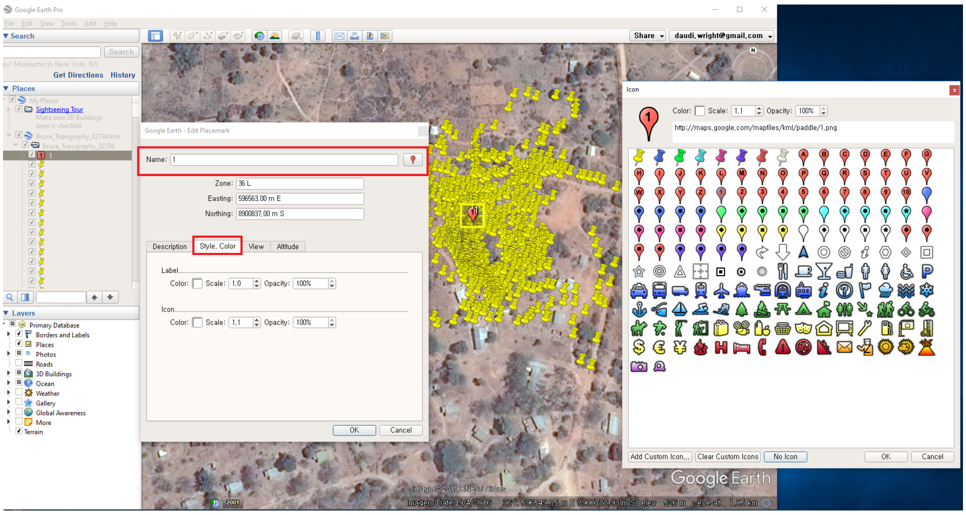


Fig. 6.14. Set precise points in Google Earth for georectification.

Navigate your selector hand around the area and left click pins around the edges and center of the project area that will give you good georectifying points. Perform the operation as above, dropping different pins in each spot. After you have 10-12 points, you can right click the Bruce_Topography_32736 folder again and select Sort A-Z. Your labeled pins will appear at the bottom of the folder. You can select those and drag them up to the top of the folder (Figure 6.15).

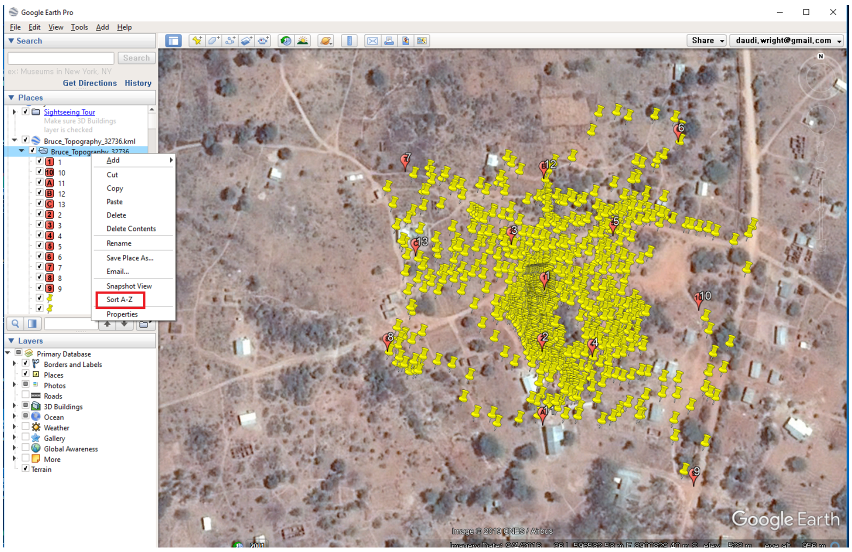


Fig. 6.15. Sort your pins for easy georectification.

Now you have orientation points established for georectification. Be sure your image is centered and unselect all of the options in the View menu. Under the View menu>Show Navigation, select Never. You can temporarily remove the Toolbar as well, but keep the Sidebar (Figure 6.16). This will give you the best overall image of the project area, but not mess up your orientation when you remove the pins in the later steps. Unselect all of your pins and then select only your orientation points in the view.

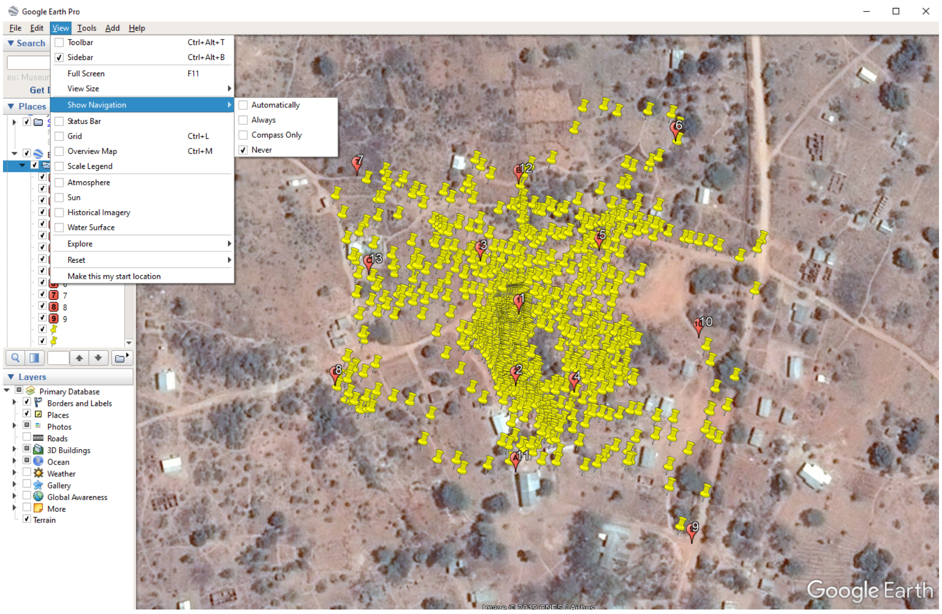


Fig. 6.16. Adjust the viewscape of the Google Earth image.

Now you are going to create the images you are going to georectify. The first image is for orientation purposes. Left click File>Save>Save Image… and under the Map Options that appear, remove all of the Elements like Title and Description, Legend, Scale, Compass and HTML Area. Change the Resolution to the Maximum (4800x3349). Left click Save Image… and save it in your Rasterfiles/Bruce folder. I made a new subfolder called GoogleEarth in which to save this image and labeled my jpeg Bruce_GE_NavPoints. Without moving any aspect of the screen or changing the orientation, unselect your Bruce_Topography_32736 kml file and save the file with a different name like Bruce_GE_NoNavPoints (but at the same resolution). After you have your images, be sure to go back to the View menu and reselect all of the options you unselected just before creating your image. If you fail to do that, the next time you open Google Earth, you will have difficulty navigating to the spots to which you want to go.

6.3.4. Georectifying a Google Earth image

Minimize, but don't close Google Earth Pro. In my opinion, the georectifying tools in QGIS 3 (Madeira) are a lot easier to use than in 2.18. Open Madeira and set the CRS to EPSG:32736. Click Add Raster Layer and navigate to your Bruce_GE_NavPoints jpeg. Because the image doesn't inherently have a coordinate system associated with it, it may reset the coordinate system to EPSG4326. If it does this, reset the CRS of the project and the layer again to 32736 and it should work. If you look at the Coordinate Window at the bottom of the Map Window, you will notice that the upper left corner of the image is at 0,0 and the lower right corner is 4800, -3349 which corresponds to your image resolution. However, you need to fix the coordinates so that the image relates to the UTM grid.

Be sure the Georeferencer GDAL plugin is active (if you don't see it, go to Plugins, look for the tool and activate it). Now go to Raster>Georectify and a new window will open. Click the little Open Raster button in the upper left corner and your image will load in. Click  to add a control point. Place your point at the point where the pin meets the ground surface and left click. A dialogue box will open that says "Enter Map Coordinates" (Figure 6.17). Go back to your Google Earth view and right click on the corresponding number, then left click Properties and a box entitled "Edit Placemark" will open. Here you can see the coordinates (Easting, Northing) and you can copy and paste these into the corresponding locations in the Enter Map Coordinates dialogue box in QGIS. Once you have entered both coordinates, click OK, and perform the operation again for all of your points. The more points you add to your georectification, the more accurate it will be. Be sure to only copy the numbers and not the "m E" or "m S." At the bottom of your digitizer, you will see the GCP table begin to fill with points. The Source X and Source Y are the current coordinates of your image (which correspond to the image resolution) and the Dest. X and Dest. Y correspond to the coordinates you are entering. When you start the Georeferencer tool, it will translate the Source points in the image to the Destination.

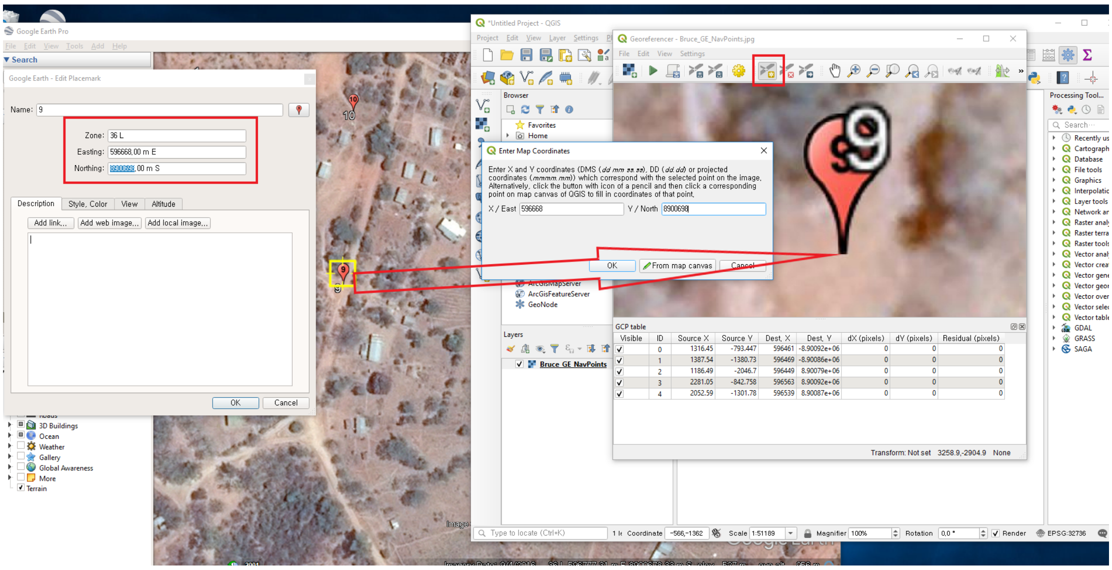


Fig. 6.17. Georectification of a Google Earth image of the Bruce site.

Once you have all of your points, click the little green Start Georeferencing () button and click OK to set the transformation type. There is no absolute "best" transformation type, but the safest transformation is Helmert. For resampling, Cubic will give you a nice, clean reprojected image. Be sure to set your target CRS (Figure 6.18). Click the Start Georeferencing button again and your image will reproject in the Map Window.

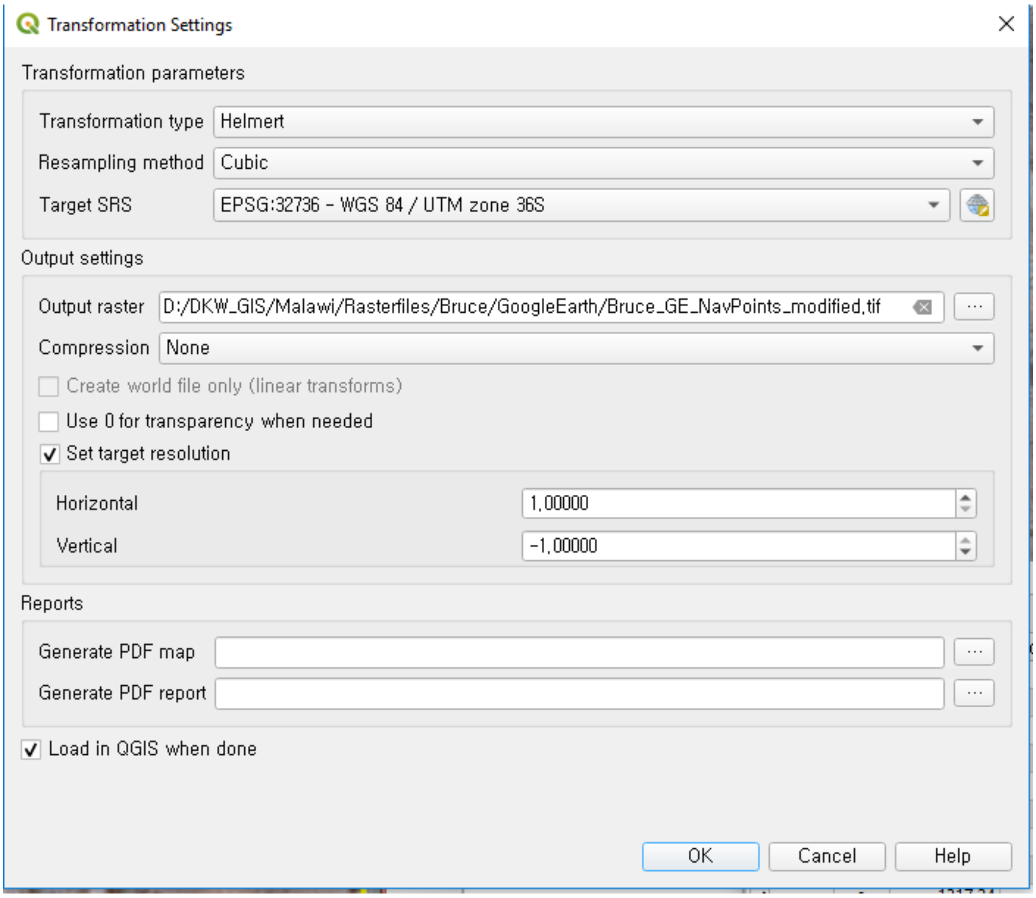


Fig. 6.18. Transform the Google Earth image into UTM coordinate space.

Before you close the Georeferencer tool, click the Save GCP Points as () button and save the .points file in the same folder as your raster image. Now you can close the Georeferencer tool.

After you close the Georeferencer tool, you will be back in the Map Window with two images loaded into your Layers panel. Remove the first layer you loaded in (Bruce_GE_NavPoints.jpg) and you will now see a blank screen. Right click the modified file and select Zoom to Layer and you will be transported to the image you just digitized and if you compare the coordinates that appear in your window to those displayed in Google Earth, they should match by location. If they don't or if your image is distorted, you did something wrong and probably need to start over. Sometimes you can edit the control points individually and fix the problem by loading them back into the Georeferencer, sometimes, it is just easier to start over.

In any case, open the Georeferencer again and you will see the image you just digitized. Click Open Raster again and open the image that does not have the control points visible. Now click the Load GCP Points button (to the left of the Save GCP Points as button) and the points you created in the last image should load precisely in the right spots of the new image. This is why it is so critical that you not move or resize your image between the time when you Save As… in Google Earth with the control points and when you Save As… without the control points. If you moved it, the GCP points will not correspond and you will digitize the image incorrectly.

Click the Start Georeferencing (green arrow) again and you will need to select a raster name (Figure 6.19). Unselect the Set target resolution button, make sure everything else is set up the same as shown in Figure 6.19 and click OK.

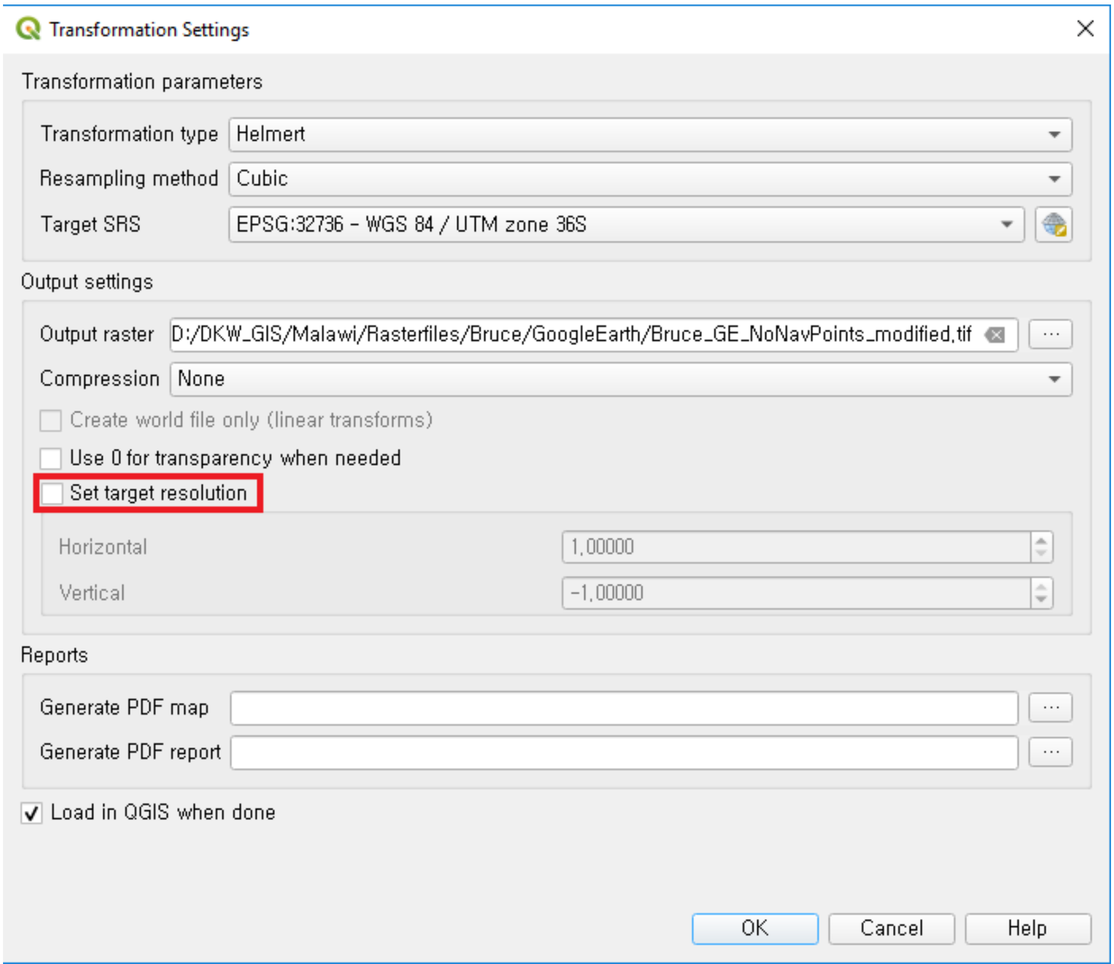


Fig. 6.19. Transform the "clean" panchromatic image into UTM coordinate space.

If you did things correctly, the images should precisely correspond to each other. To check this, click and unclick the top file and you can use the different resolutions of the images to check the locations (Figure 6.20).

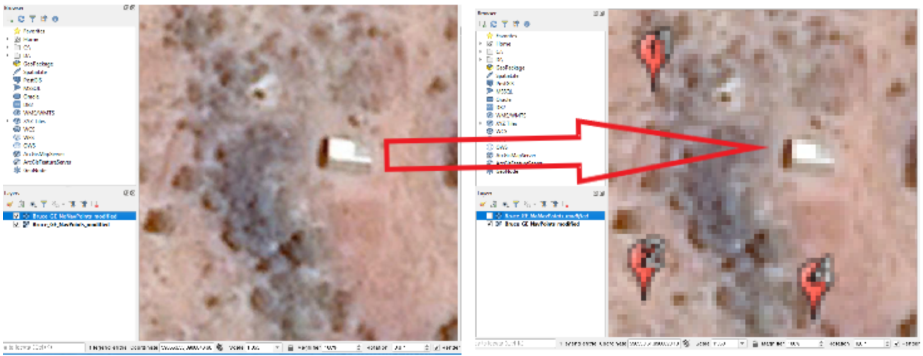


Fig. 6.20. Double check the transformations of the georectified images.

6.3.5. Color burn the Google Earth image

Go back to your interpolated images in QGIS 2.18 and add the modified tiff image to your Layers Panel. If you have the Google Earth image at the top of the Layers Panel, double click the layer and under color rendering, select Multiply and you will see the interpolated image underlay your panchromatic satellite image (Figure 6.21). This gives you a quick and easy way to separate the low from the high spots. If you adjust the underlying image to a Hillshade (Properties>Style, Render Type>Hillshade), you can make interesting underlays for panchromatic images.

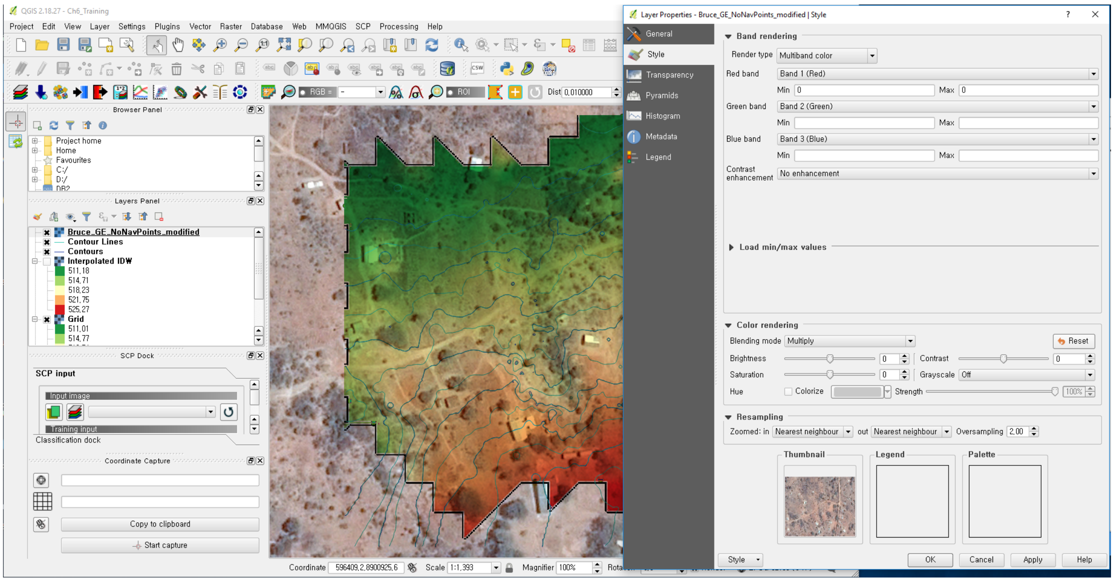


Fig. 6.21. Color burn the cubic spline image under the Google Earth image.

6.4. EXERCISE

You have been given two CSV files (GRID_1, GRID_2) in the eTL with the following attributes: ID numbers, Y (latitude), X (longitude) and Z (m above sea level). These are spot locations (WGS84) of the same area extracted in different ways. Create two sets of interpolation maps from each one of these files (as executed in 6.2). *HINT: X and Y in EPSG:4326 are in degrees radian, while Z is in meters, so you will need to reproject the files into a coordinate system that has a projected CRS like UTMs.*

Save four files as *YOURNAME*_Ch6.png and **upload it to the eTL by April 22 at 9:00 am.** Create one map (with a title, scale bar, north arrow and title) with a view similar to that in Figure 6.21. You can use Landsat images from Chapter 3 as your overlay. You will receive 5 points for completing this portion of the assignment correctly and on time. In addition to the images, answer the following questions:

STUDY QUESTION (1): Using Conolly and Lake (2006) and other resources you can find, briefly describe the differences between the two surface models (IDW and cubic spline) you created. What are the differences in the models created between GRID_1 and GRID_2? (2 points)

STUDY QUESTION (2): Import GRID_3 into your Map Window and perform one of the interpolations (either IDW or cubic spline). In your answer sheet, paste GRID_1, GRID_2 and GRID_3 side by side in a single row. Describe the differences between the GRID_1/GRID_2 and GRID_3. Why do you think there are differences in the ways these data are visually portrayed on an interpolated map even though the data are drawn from the same original DEM? (3 points)

Chapter 7: Data Analysis

7.1. Principles of Exploratory Data Analysis

Exploratory data analysis gives users opportunities to sift through massive amounts of data quickly. In a spatial environment, this gives GIS consumers means of making searches and queries on site parameters that would normally take hours, days or years to execute with a paper and pencil. Data can be statistically analyzed for correlations between site locations and extant parameters such as soil type, distance to streams or elevation.

Data analysis can be performed in a supervised way or an unsupervised way. The former method involves the user (you) manually inputting the parameters of the analysis and performing all of the analysis of the data. Unsupervised data analysis involves constructing models and allowing the computer to produce and analyze the data with no real direct input from the user. Agent based modeling is a new field of research in which the computer is programmed according to specific variables that occur probabilistically, and a simulation is run to understand the response of the outcomes on the functioning of the system.[12]

In this chapter, we will run basic statistics on the occurrences of variables within our training dataset. These are not explicitly spatial statistics, which we will focus on in the next chapter. Instead, these are basic statistics that can be generated from attribute data. You are expected to have a basic knowledge of elementary statistical principles (mean, media, standard deviation). If you need a review, this website provides an overview of the topics we will cover in this module.

7.2. Setting Up Your Model

Most of the work in calculating statistics from your data is in the setup. You have learned in previous chapters the complications that you can experience when coordinate systems conflict with each other. In this week's exercise, we will examine statistical relationships between our randomly derived dataset of archaeological sites in Korea in relation to three primary variables: elevation, slope and soil.

The first step is to obtain elevation data at a resolution sufficient for detailed spatial analysis. There are two ways to do this. The first way is familiar to you and that is to access and download the DEMs directly through the USGS Earth Explorer window. Once you have a polygon that covers the Korean peninsula, you will select the Shuttle Radar Topographic Mission (SRTM) 1-Arc Second dataset (Figure 7.1), which is the highest resolution data available for Korea. 1-arc second is a degree radian measurement (0°0'01") that roughly translates to 30-m in Korea (it is actually more like 25 m).

---

[12] See Crooks, A.T., Malleson, N., Manley, E. and Heppenstall, A.J. (2019), *Agent-based Modelling and Geographical Information Systems: A Practical Primer*, Sage, London, UK.

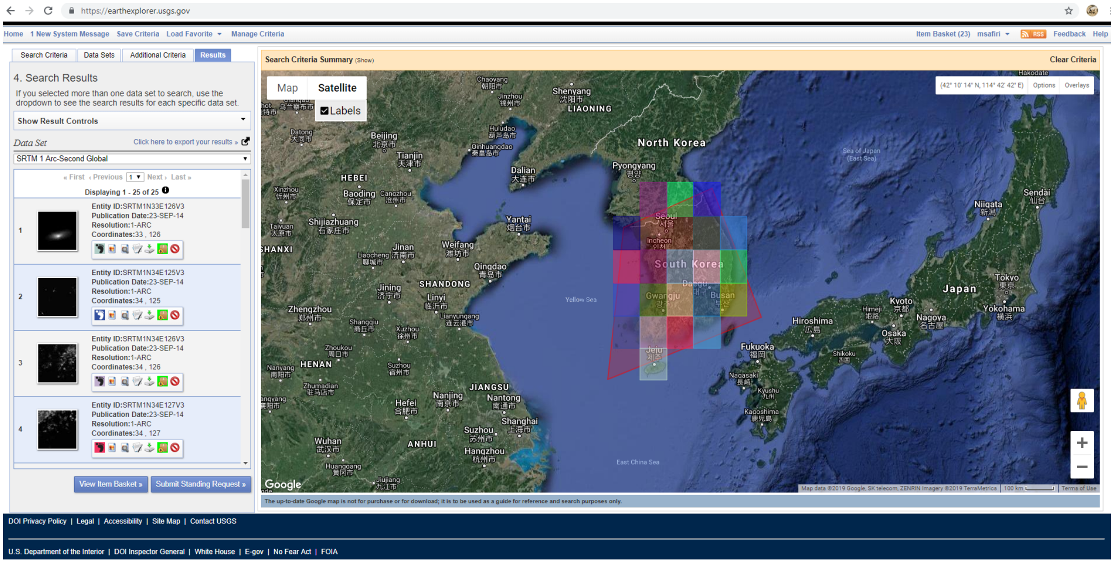


Fig. 7.1. Set the download region and datasets in USGS's Earth Explorer.

You can download these tiles individually, but the most efficient thing is to conduct a bulk download since these are large files. The USGS offers a Bulk Downloader Application (BDA) for free that is relatively easy to set up. Follow the directions and install the BDA in your program files. When you open it up, you will need to right click it and "Open as Administrator." The program will prompt you for a password (use your USGS login and password) and then direct the bulk download to your Rasterfiles folder. You will download the BIL files into separate folders (Figure 7.2).

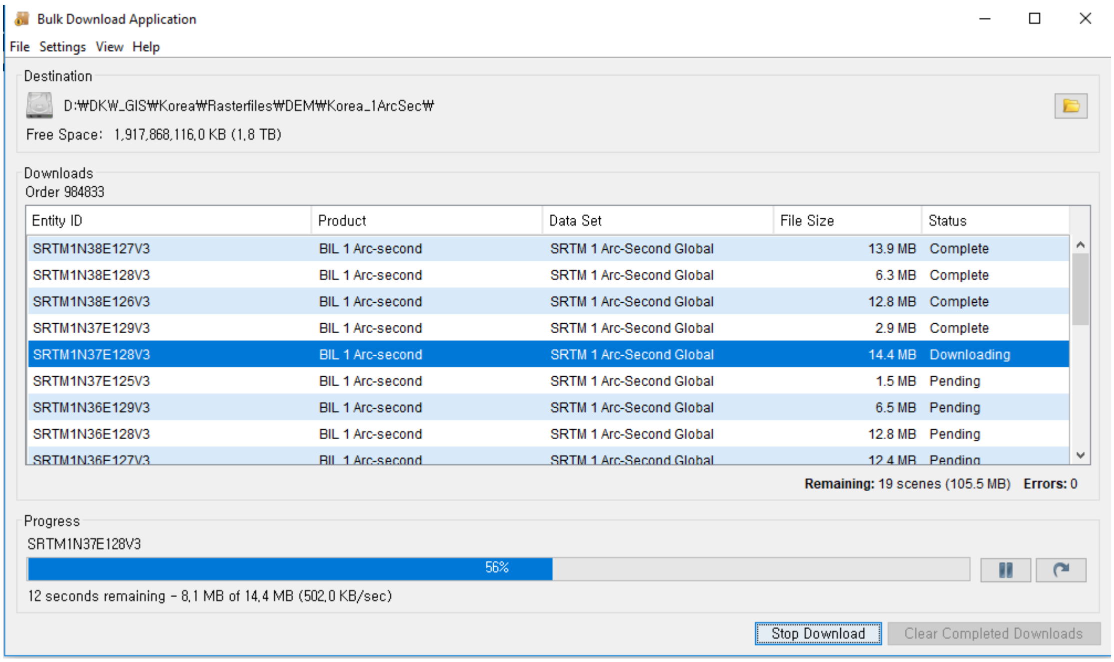


Fig. 7.2. USGS Bulk Downloader Application

There is an alternative method to access these data, which is to use the SRTM Downloader plugin. When I tried it, I did not have success using QGIS 2.18, but it worked well in 3.4 (Madeira), so the rest of this exercise will take place using Madeira, which will give you some familiarity with the newest member of the

QGIS family. You will need a NASA/Earthdata login (different than the USGS) before you attempt to use it. Go to: https://urs.earthdata.nasa.gov/users/new. Once you have a password, you can establish your search parameters from the KOR_adm file and the files will automatically download into the folder to which you direct them (Figure 7.3).

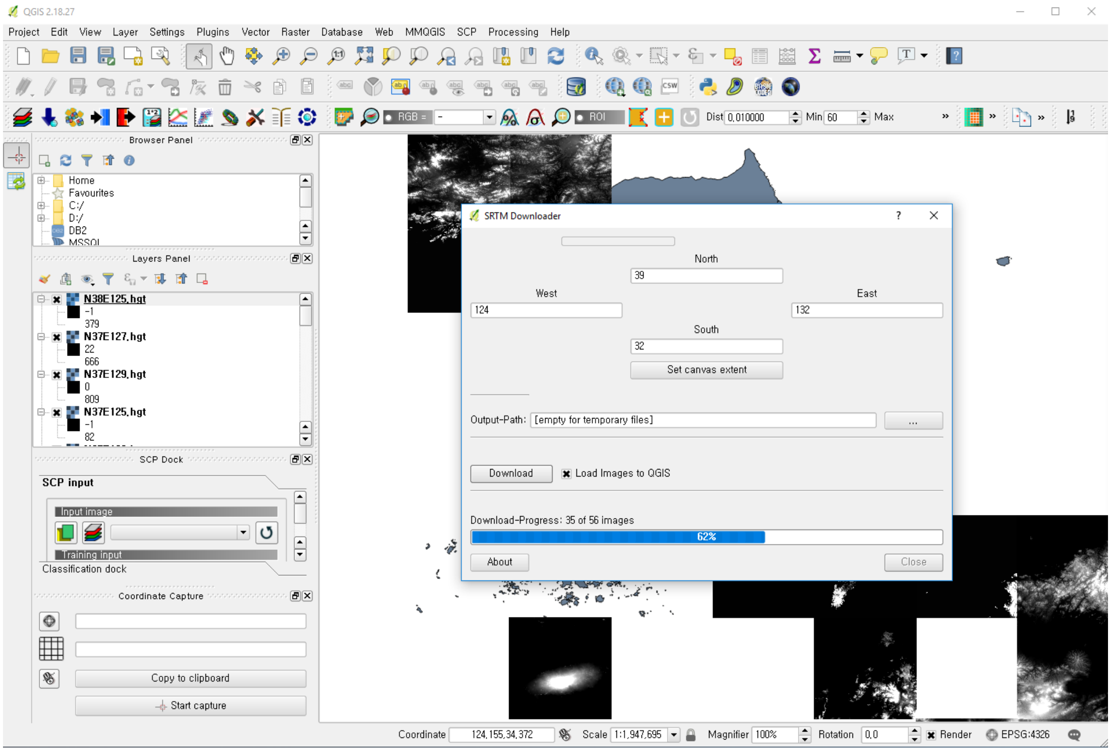


Fig. 7.3. SRTM Downloader in QGIS 3.4 (Madeira).

Once your data are downloaded, you can build a VRT from the files (Chapter 2) and you will have a seamless DEM of the Korean peninsula (Figure 7.4).

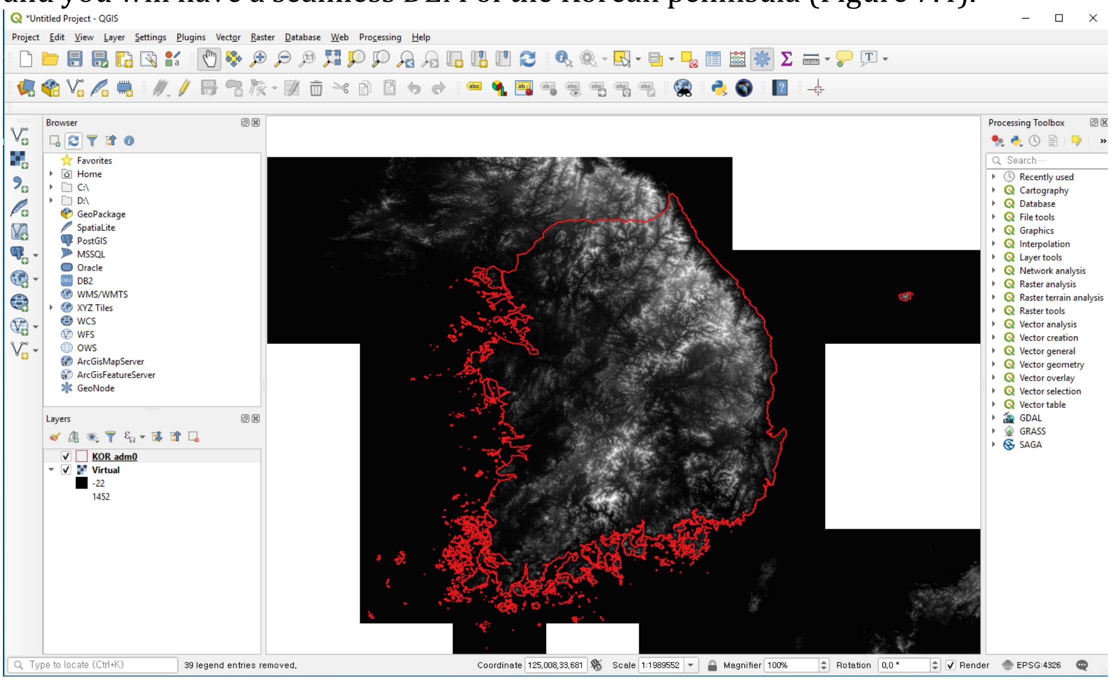


Fig. 7.4. Seamless VRT DEM of Korea.

Now, we want to access soils data from Korea. There is a very interesting project underway as part of an international consortium to better understand the distributions of soils across the globe called the International Soil Reference and Information Centre. They compile and host soil data at 250-m resolution from all over the world. The interface is not as easy to use as Earth Explorer, but with some persistence and practice, you can access what you need. Go to the Digital Soil Map of the world: https://soilgrids.org[13] and select Download data (Figure 7.5). On Soil Classification (TAXOUSDA), Select "Predicted most probable class". Download a 250m raster that covers South Korea (roughly 125°-130°E and 33°-39°N). You'll need to download two rasters because you will also need 34-39°N. The "TAXOUSDA_250m_ll.tif.csv" file located in the eTL has the key to what the raster tiles correspond to (downloaded from https://www.isric.org/explore/soilgrids).

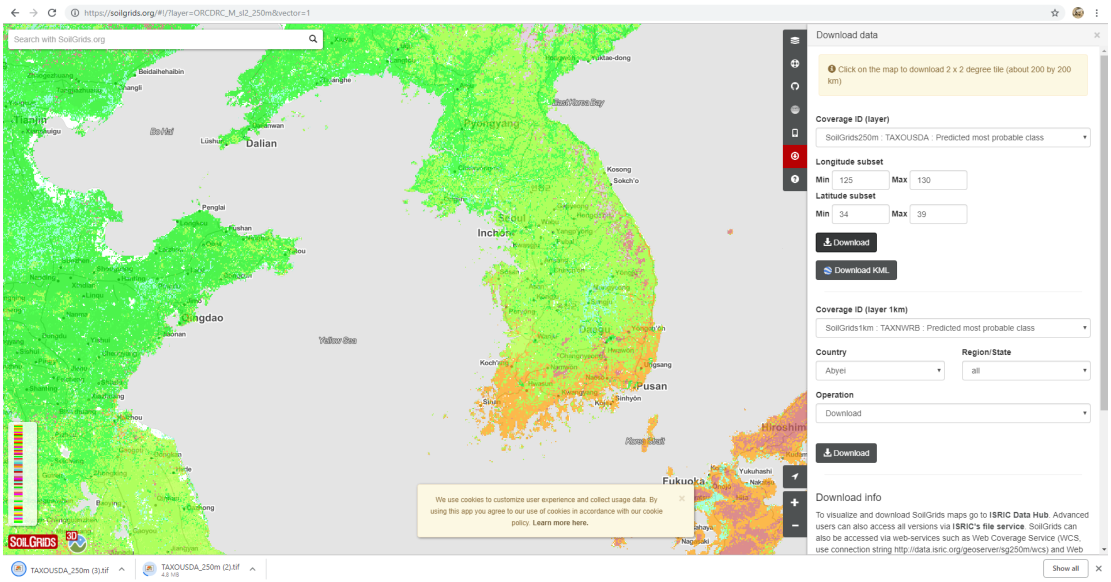


Fig. 7.5. Digital Soil Map of the world view of Korea (download selected).

A color classification should be created in Madeira (QGIS 3.4) because the color palettes in the Properties>Symbology menu allow you to create paletted/unique values, which is an improvement from QGIS 2.18. You can select random colors if you like, but a viridis palette makes sense because the soil types are aggregated by numeral proximity to one another (Figure 7.6). You will want to export your color palette by clicking Advanced Options (next to Delete All below the color palette window). Next, go to your 250-m resolution second soil map Symbology menu, click Advanced Options and import the color palette (.clr) into your Symbology editor.

---

[13] Hengl, T., Mendes de Jesus, J., Heuvelink, G. B.M., Ruiperez Gonzalez, M., Kilibarda, M. et al. (2017) [SoilGrids250m: global gridded soil information based on Machine Learning](). PLoS ONE 12(2): e0169748. doi:10.1371/journal.pone.0169748. A more local assessment of soil mapping can be found here: https://www.iuss.org/19th%20WCSS/Symposium/pdf/0736.pdf. This provides an overview of the soil categories present in Korea and the general coverage of those soils in relation to surface area.

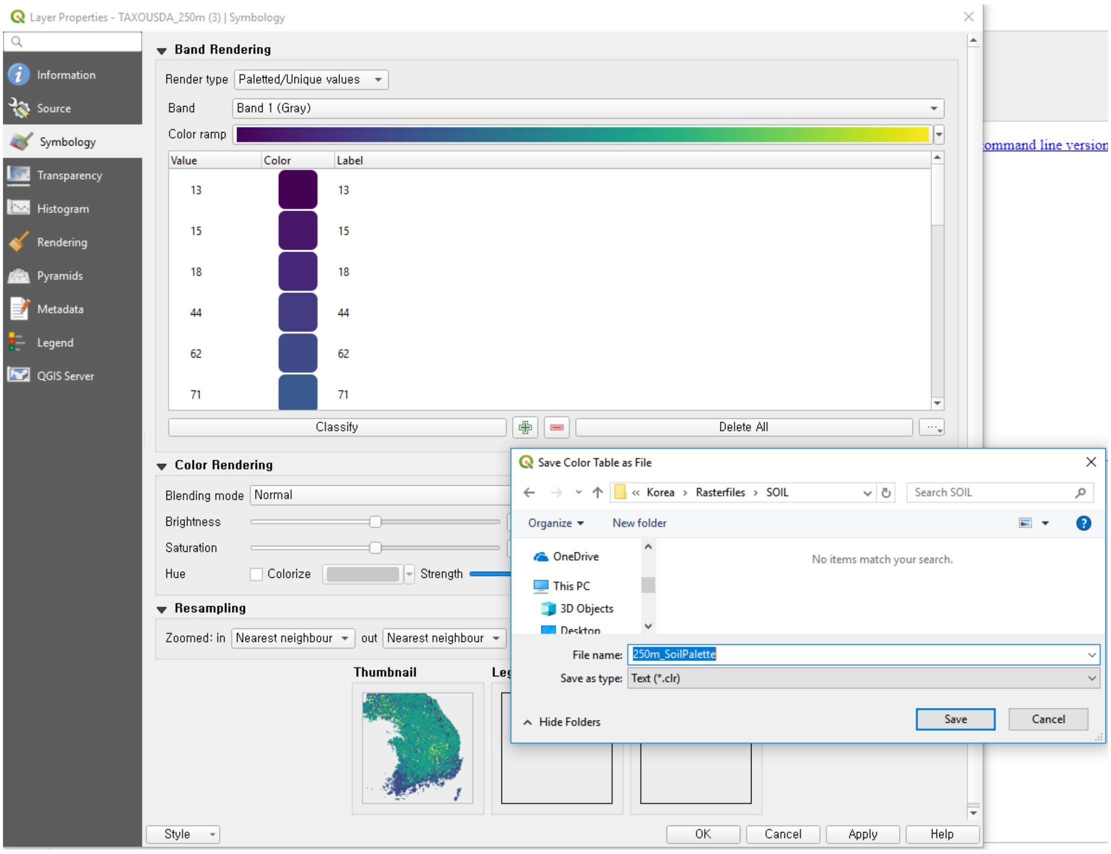


Fig. 7.6. Save color palette for extraction in adjacent rasters.

Now that we have the two rasters we need, we can extract raster values and fix them to our shapefile that contains the 16101 random points within 1 km of known survey regions in Korea (TrainingData shapefile). There are many ways to perform this task in QGIS. For the soil map, use the SAGA>Vector <-> Raster>Add raster values to features tool. Be sure to select Nearest Neighbor (Figure 7.7).

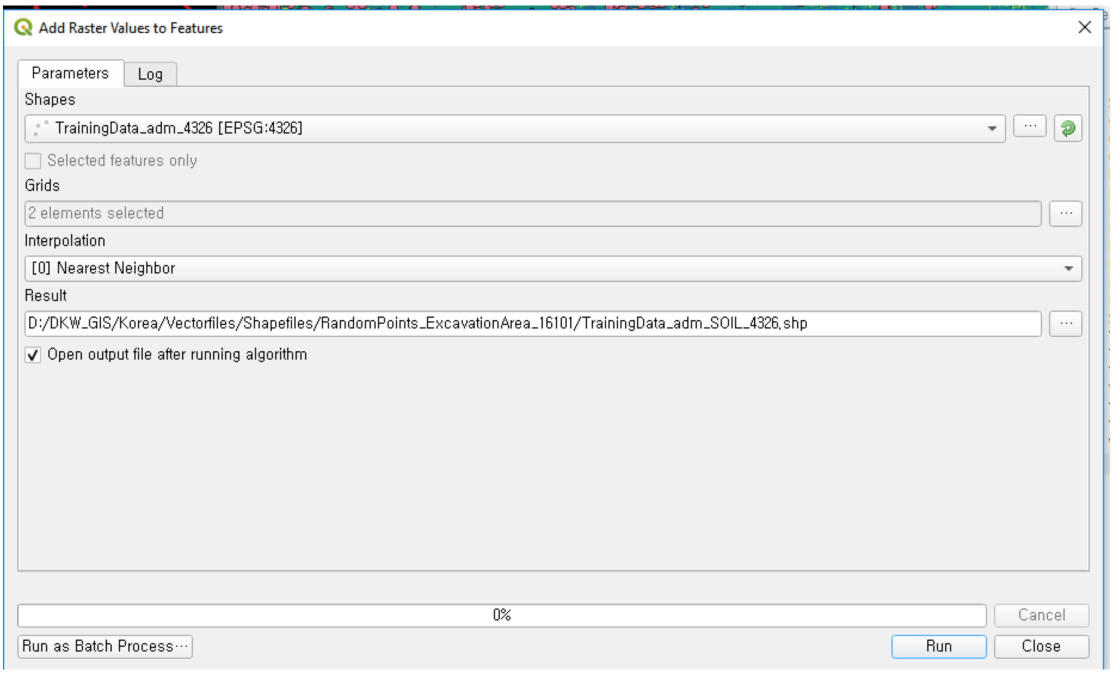


Fig. 7.7. SAGA Add Raster Values to Features tool in Madeira.

Another tool that can be used is the v.what.rast GRASS tool (type v.what into the search bar of the Processing Toolbox), but this can only sample one raster at a time.

Now use the Field Calculator to coalesce the attributes from the two different soil maps to create a new attribute field "Soil_250" (Figure 7.8). Note that I had 292 points that did not have soil values. This is fine. Those are from sites that are within 250 m of water and so they are projected to occur on *NULL* values. These just won't be calculated in the statistical model.

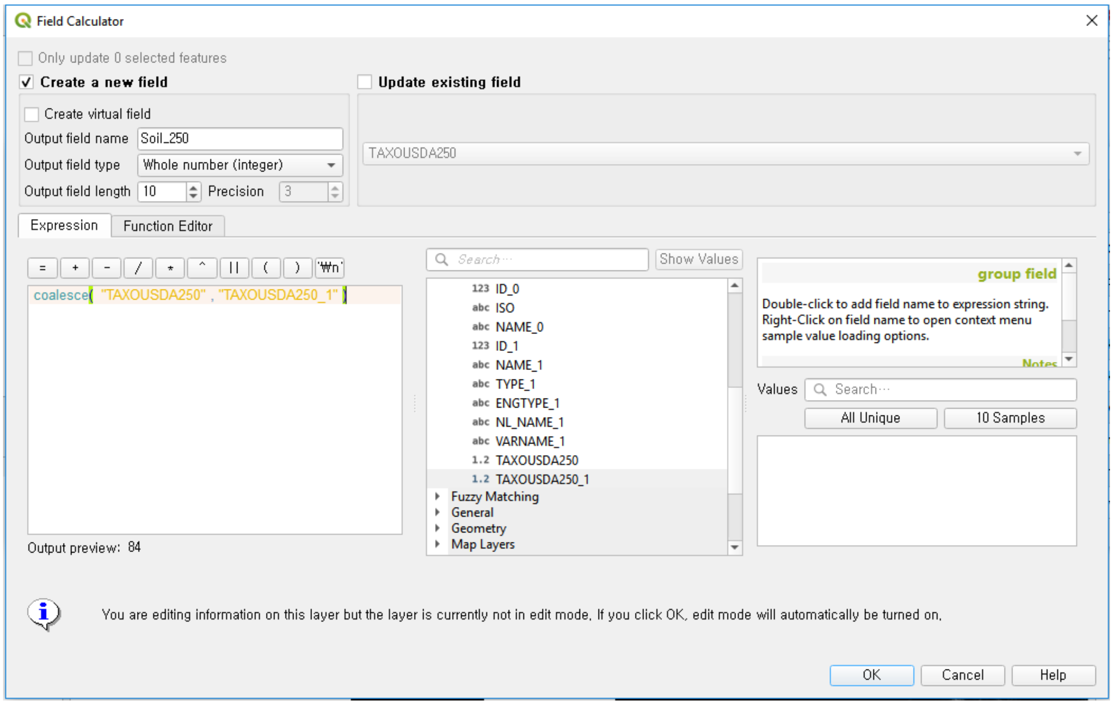


Fig. 7.8. Use the Field Calculator (accessible from the attribute table by hitting Ctrl+I or the icon). Note that if you perform this operation outside an editing session, you will not be able to undo the actions.

Now, extract the meters above sea level (m.a.s.l.) values from the DEM, but use a Bicubic spline interpolation model instead of nearest neighbor (Figure 7.9). Don't calculate the features that didn't have soil values. If you select the 292 points that had *NULL* soil values, click the invert selection ( ) button (or Ctrl + R) and only the sites with attribute values will be selected. Use only the selected values in the extraction model.

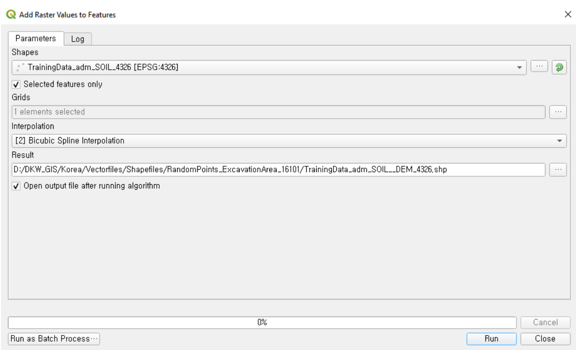


Fig. 7.9. Use a bicubic spline interpolation to determine elevations of sites. This is because sites are 2D in projected space, but we are using 1D point data. A bicubic interpolation gives average values of surrounding cells.

You may need to rename your DEM field (I called mine “masl” for meters above sea level) in the Properties>Source Fields dialogue. Be sure to toggle editing mode prior to changing the field name. Note that some of the elevations are below sea level! This is because of the bicubic interpolation model, which accommodates the surrounding values. This is fine. Save your reproject and save your DEM in UTM coordinates (EPSG:3093).

7.3. Running Statistical Models on Rasters

As we have learned in this class, rasters are quite elegant because of their relatively simple data structures. This makes them quite easy resources from which to generate basic statistical models. They can be a bit complicated to set up for archaeologists who are interested in specific spatial ranges of data, but once you set up your data, the possibilities are limitless.

The first consideration an archaeologist must have is that her/his 1D point data is not representative of the total range of environmental considerations prehistoric people dealt with. In the case of elevation and soil, while people may have been living in a rather restricted location (site), their fields and foraging ranges were far afield from the actual site locations. There is much debate on site catchment areas and how small or large they should be for different populations,[14] but for this exercise, we will make a simple 1000 m buffer around our sites to accommodate the variability in the soils surrounding a specific site. If you are going to make buffers for your final project, you must ground the construction of your buffer sizes in some kind of empirical basis. Arbitrary 1000-m buffers are not normally acceptable forms of categorizing big datasets such as these.

However, for the purpose of this training exercise, we are not going to worry about the reasons, we are just going to execute a task. To get started, you're going to add one of your TrainingData_adm shapefiles from Chapters 3 and 4 and create 1-km buffers (will need to reproject the shapefile into EPSG:3093/UTM 52N first if it is not already). This task is simply executed by going to Vector>Geoprocessing Tools>Buffer… and creating a new polygon file from the buffers derived. To get started, we are primarily interested in the geometry of the created file, so we will dissolve the boundaries between the buffers to make one continuous vector that we will use for the purpose of clipping. So, our attribute information will be a bit useless in this newly created file. To perform this task, go to Vector>Geoprocessing Tools>Dissolve (Figure 7.10). The created file is one that is not a series of 16101 individual polygons, but one large polygon and some smaller ones that occur on isolated islands (including 독도!).

---

[14] Jones, E. E. (2017). Significance and context in GIS-based spatial archaeology: A case study from Southeastern North America. *Journal of Archaeological Science*, *84*, 54-62.

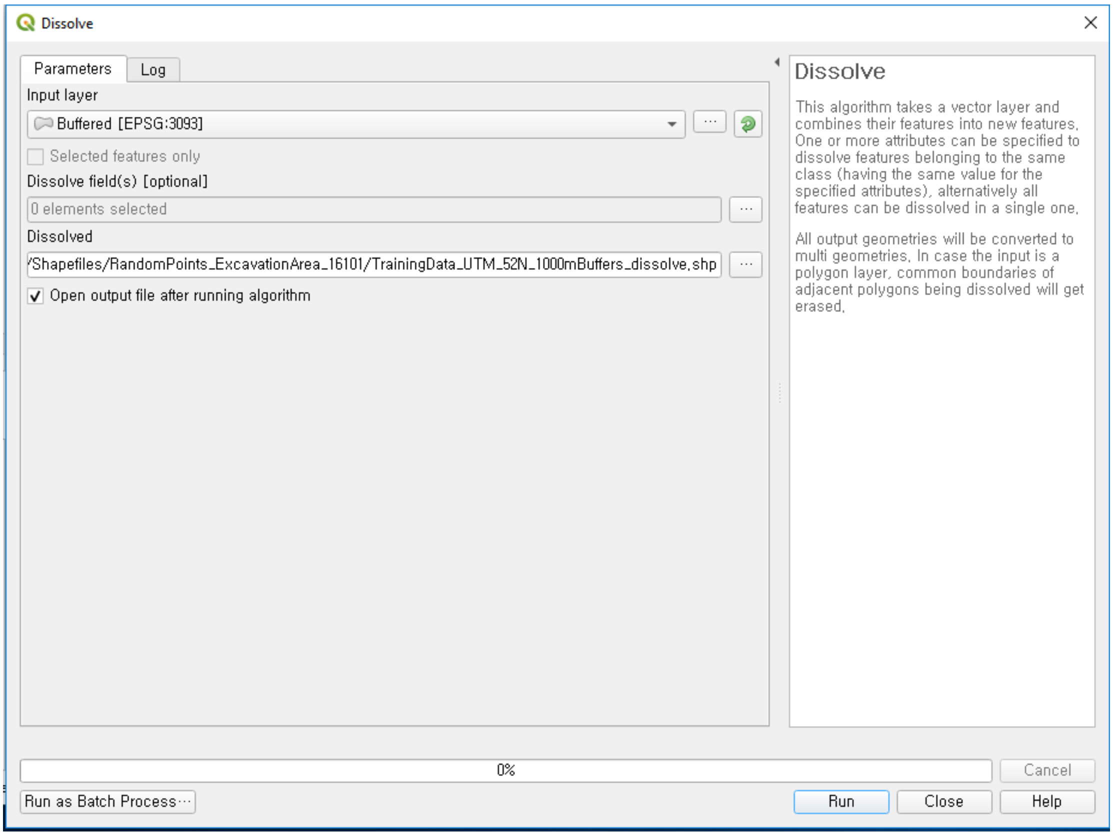


Fig. 7.10. Dissolve the buffers.

Next, your will want to make a new raster that is resampled from your buffer vector. Go to Raster>Clip Raster by Mask Layer and clip the Korea_Soil_250 by the dissolved 1km buffer (Figure 7.11). Save this as a new raster file. I called mine RandomPoints_1kmBuffer_4326.tif.

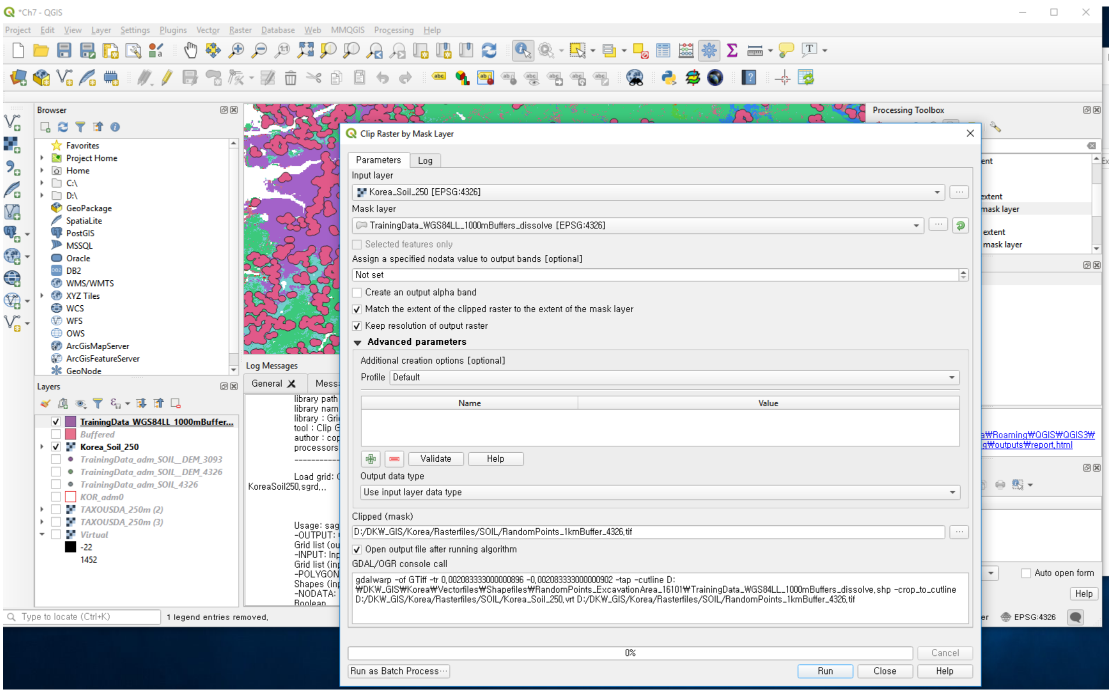


Fig. 7.11. Clip the soil raster by the dissolved buffer vector mask layer.

Now we are ready to get some basic statistics from the raster. You have created a new raster of soil taxa within 1 km of randomly created points that are within 1 km of known previously surveyed boundaries. These data are your controls for how a normalized set of archaeological variables should look if human beings made choices on a random basis (within the parameters of modern human development projects).

Let's start simply. Use the r.stats tool to get overview summary of statistics from soil (Figure 7.12). You can find this GRASS tool by searching the Processing Toolbox search bar. Your element selected will be your

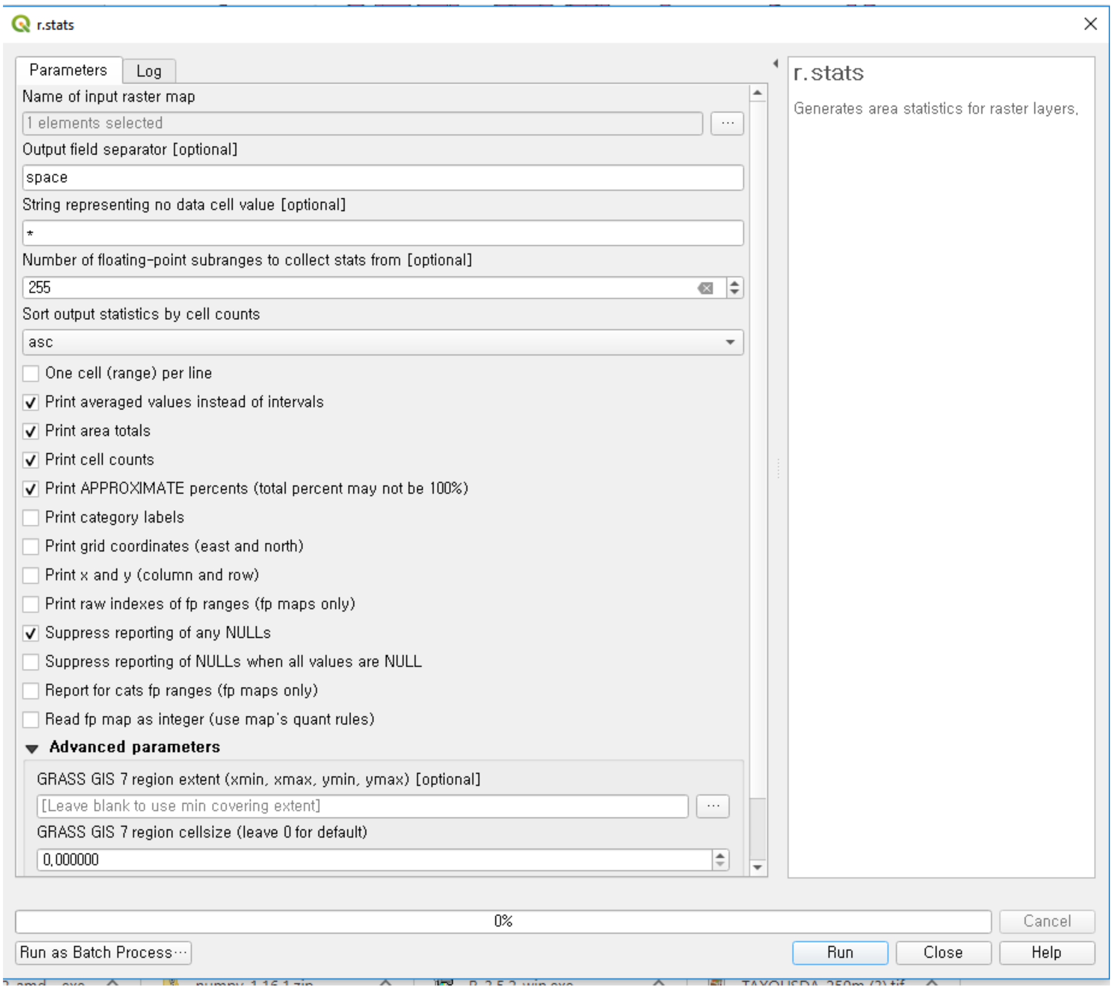


Fig. 7.12. Example of basic statistics available from v.rast.

The resulting HTML file can be pasted into OpenOffice (Right click, Paste Special… Unformatted Text… Be sure to click the Space option, Click OK (Figure 7.13). Add in the values from the TAXOUSDA_250m_ll.tif.csv. You can see that in the random distribution of points, most of the area where site areas are located are Udalfs (51.03%) and Udults (26.53%).[15]

---

[15] Soil taxa are a bit complicated, but quite informative of the nutrient capacity and formation processes (https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/edu/?cid=nrcs142p2_053588). I teach a geoarchaeology class in which we discuss these briefly, but if you are interested in soils, please see me and I will help you find resources that are targeted for your interests. A fairly simplified overview is available on Wikipedia, but there are likely a lot of terms you don't understand without some background knowledge.

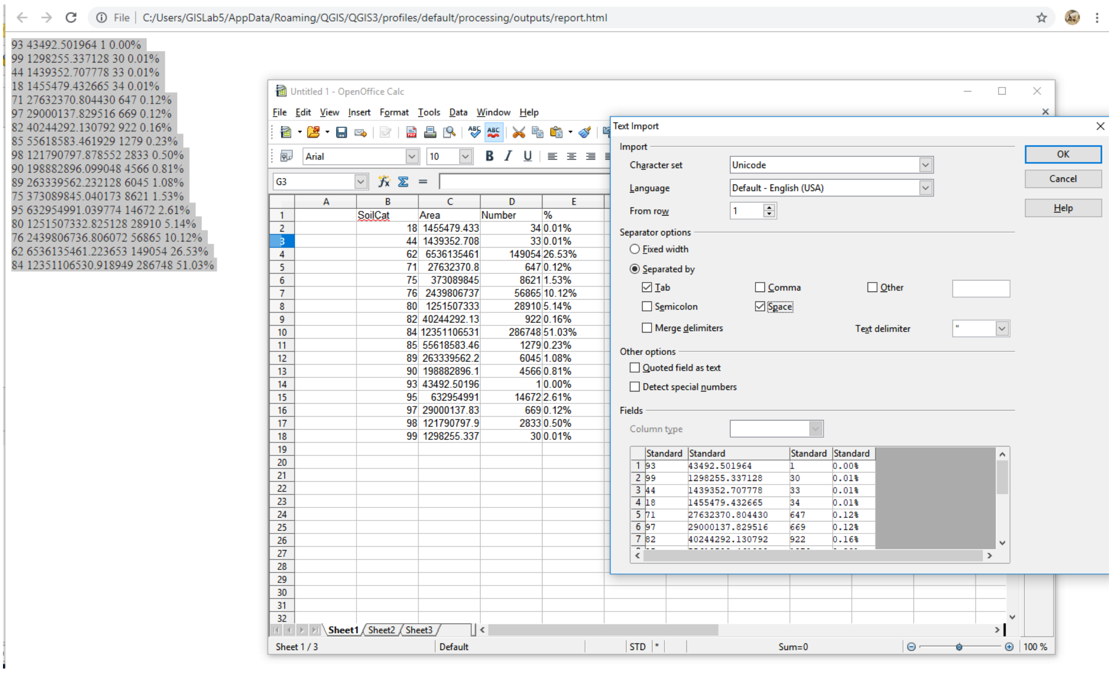


Fig. 7.15. Paste the HTML raster statistics into a spreadsheet.

Note that the total area estimations are not really accurate because of the coordinate system clash (it is effectively squared degrees radians). To calculate the total geometric (projected) area, open the Field Calculator in the Attribute Table in the UTM-derived shapefile, Create a new field called "area" and build the expression `$area/1000` (Figure 7.14). You see that the final result is 24,936,169 km². Alternatively, you can calculate the area from v.report (search for the GRASS tool in the Processing Toolbox) using the EPSG:4326 file. This produces an HTML file in which the last category will be the m² of the shapefile.

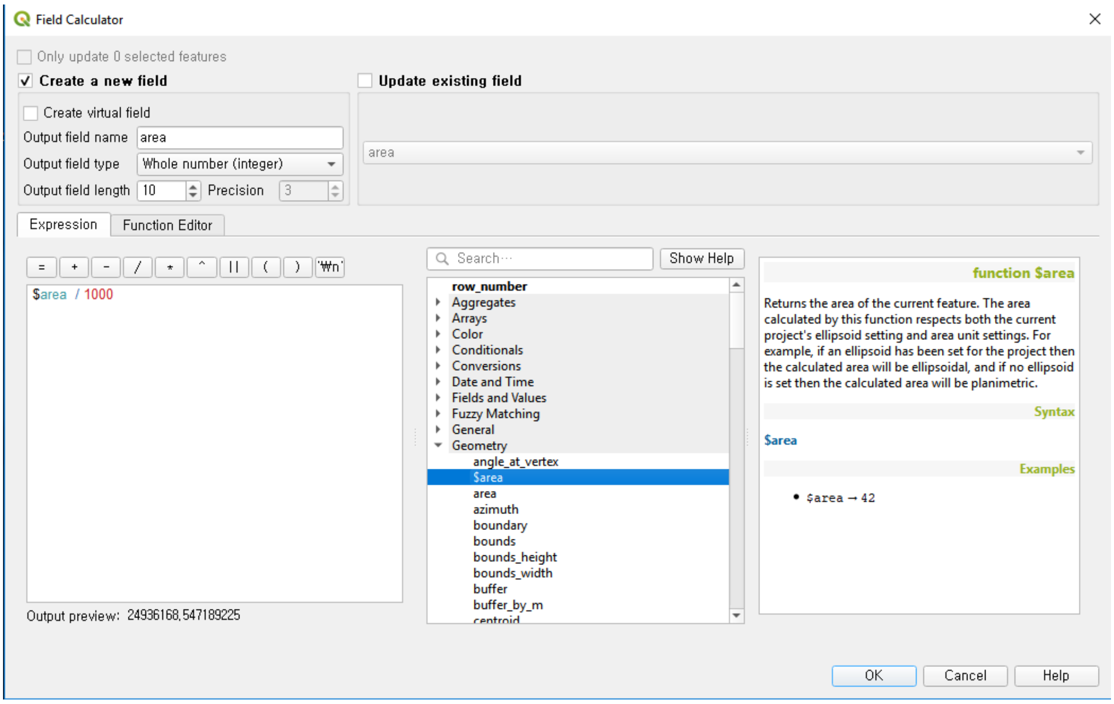


Fig. 7.14. Calculate the total area using the Field Calculator.

7.4. Running Statistical Models on Vectors

The soil taxa raster you worked with above is comprised of **nominal data**, which can be tricky to quantify. On the other hand, it is easier to create statistical models for **interval data**, which are numerically scaled data in which the differences between the values are known and the data do not have an absolute 0 and **ratio data,** which does have an absolute 0.

When examining site selection considerations, you can calculate the slope of the sites (which is an important consideration for site location, particularly for agricultural people). To do this, use the reprojected DEM (EPSG:3093). The GDAL Raster Analysis tool (type "slope" in the Processing Toolbox search bar) gives you the best range of options (Figure 7.15). This process can take several minutes to produce results, so be patient.

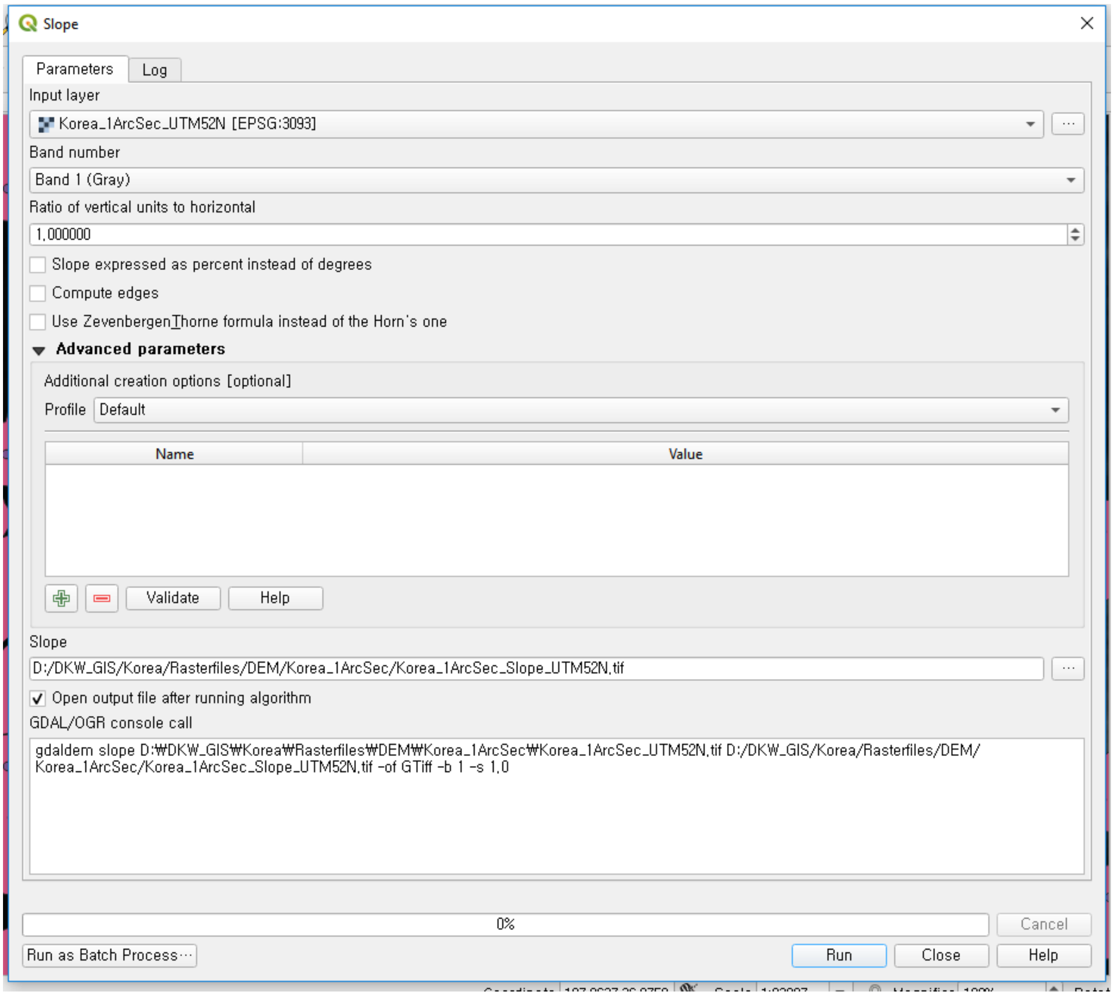


Fig. 7.15. Calculate the slope from your DEM.

Once you have the slope values calculated for the study region, add a "Slope" column to your TrainingData_adm_SOIL_DEM_3093.shp file (as you know by now, it can be done through the Properties dialogue or in the New Field (Ctrl+W) in the attribute table when the editor is toggled). Run the v.what.rast tool again with the overwrite option to update the Slope column of your attribute table (Figure 7.16). Interpolating values from the surrounding four cells is important. It gives you the general slope of the area surrounding your point.

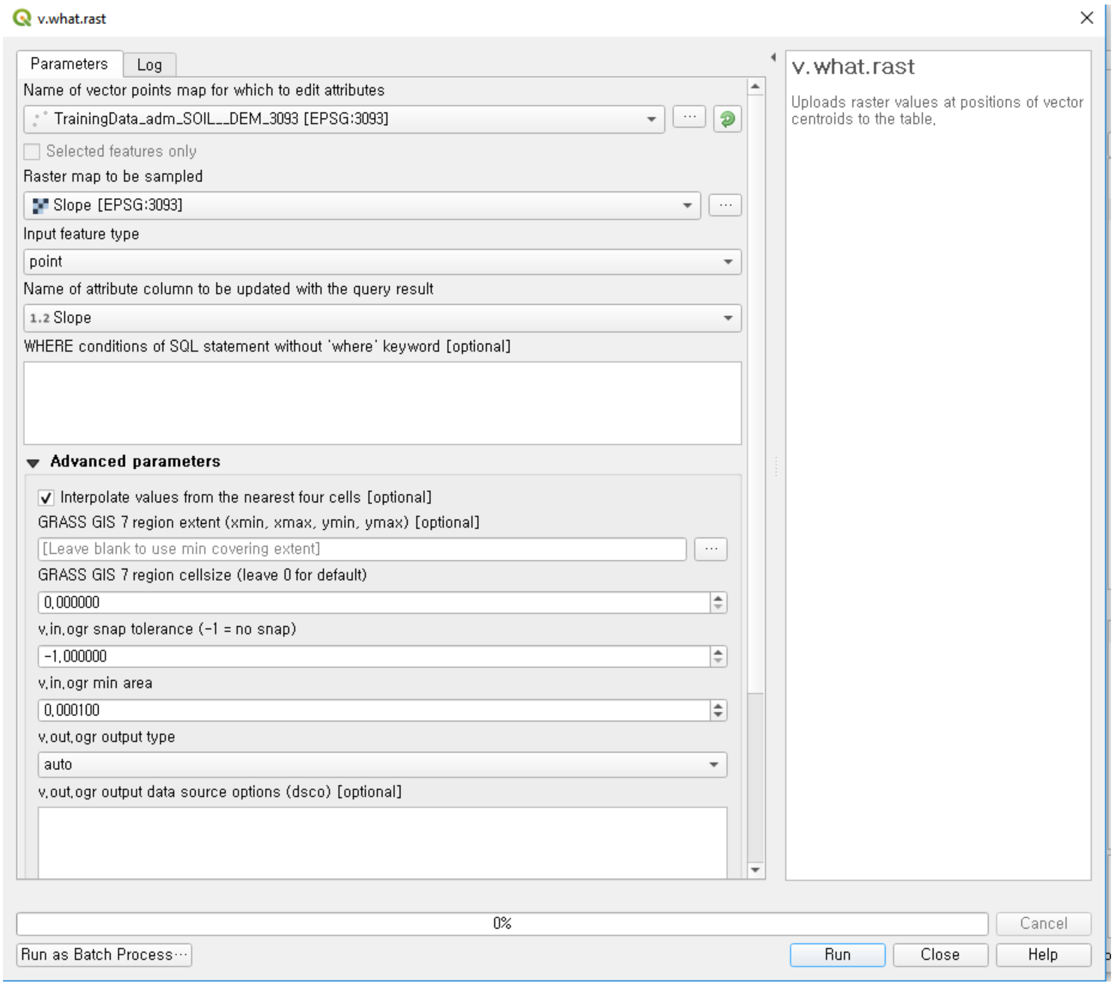


Fig. 7.16. Attach the slope value of the sites to the attribute table.

If that doesn't work, use the "Add raster values to points" SAGA tool (under Vector<->Raster). Extract the values from the DEM and add them to the attribute table.

To calculate the statistics for the randomly distributed sites, use the v.univar GRASS tool (Figure 7.17). An HTML file is produced in which the minimum (0°) and maximum (66.2893°) slope is identified, the mean slope is 9.68593° with a standard deviation of 8.28096°.

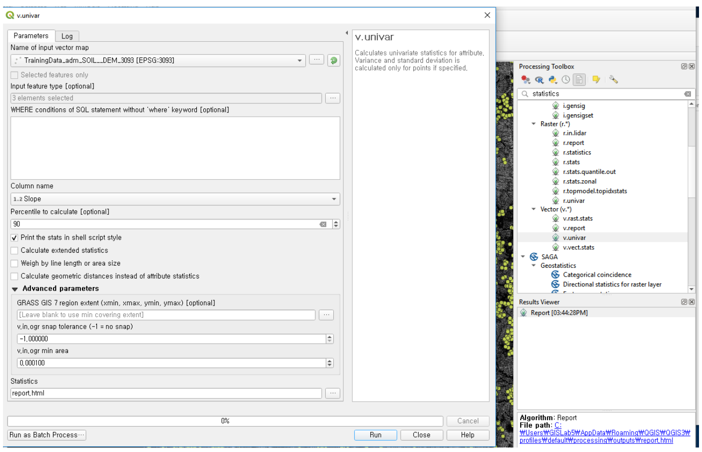
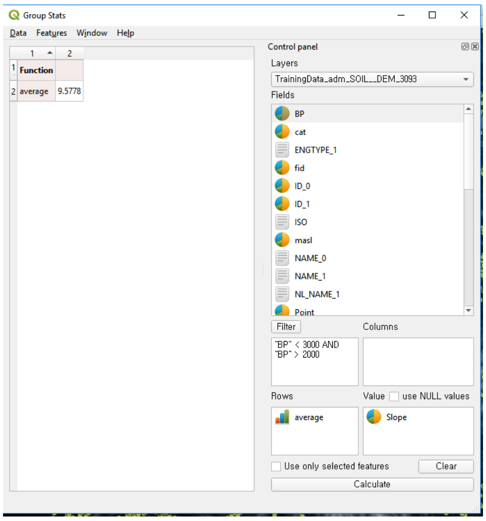
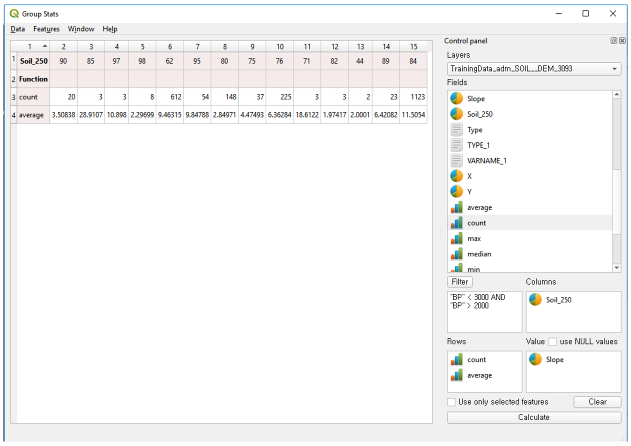


Fig. 7.17. Calculate the vector statistics of bicubically interpolated slope values of randomly situated archaeological sites in Korea.

Using the “Group Stats” plugin in either Las Palmas (QGIS 2.18) or Madeira (QGIS 3.4), you can perform finer sifts of the data in a “pivot table” workflow. For example, if you want to know the average slope for sites between 3000 and 2000 years BP, you can perform a quick calculation (Figure 7.18). Let’s say you are interested in how these correspond to soil type and you wanted a count of the number of occurrences, you can easily modify the data analysis (Figure 7.19). Of course, these are random variables, so you won’t expect any significant statistical differences





Left: Fig. 7.18. Use the Group Stats plugin to find average slope for randomly situated archaeological sites between 3000 and 2000 years BP. Right: Fig. 7.19. Count of randomly situated archaeological sites on different soil types.

There are dozens of tools and plugins for statistical analysis in QGIS. In Las Palmas (QGIS 2.18), the [LecoS tool] is a powerful ecological predictor, particularly useful in developing Bayesian models. Also in Las Palmas, Toolbox > QGIS Geoalgorithms > Vector table tools > Frequency Analysis can provide graphical frequency distributions of attributes such as radiocarbon dates, ceramic counts lithic types, etc.

7.5. EXERCISE

Return to soilgrids.org and this time download raster maps for “soil organic carbon content (fine earth fraction) in g/kg” that fit within the boundaries of the KOR_adm0 shapefile. Be sure to select it at 5 cm depth. Soil organic pools are important predictors of land coverage and agricultural productivity.[16] We can hypothesize that prehistoric people situated their settlements according to high soil organic content because those would have been the most productive parts of the ecosystem. Similarly, we can hypothesize that it was more important for farmers rather than hunter-gatherer-fishers to seek the most carbon-rich nutrient pools on which to settle.

---

[16] See Guimarães, D. V., Gonzaga, M. I. S., da Silva, T. O., da Silva, T. L., da Silva Dias, N., & Matias, M. I. S. (2013). Soil organic matter pools and carbon fractions in soil under different land uses. *Soil and Tillage Research*, *126*, 177-182 and Burke, I. C., Yonker, C. M., Parton, W. J., Cole, C. V., Schimel, D. S., & Flach, K. (1989). Texture, climate, and cultivation effects on soil organic matter content in US grassland soils. *Soil Science Society of America Journal*, *53*(3), 800-805.

Test these hypotheses using the actual radiocarbon dataset. Create 1-km buffers around all of the sites and give me the mean, median and standard deviation values of carbon content (g/kg) in the following time slices: 0-1000, 1000-2000, 2000-3000, 3000-4000 and >4000 years BP. Also calculate the bicubic interpolated slope values (mean, median, standard deviation) for the sites themselves in the same time increments (done on vectors points as in Section 7.4). Provide all of the results in a table format *YOURNAME*_Ch7.xlsx and **upload it to the eTL by April 29 at 9:00 am.** You will receive 5 points for completing this portion of the assignment correctly and on time. In addition to the spreadsheet, answer the following questions (and save the file as *YOURNAME*_Ch7.docx):

STUDY QUESTION (1): How do you interpret the evolution of site selection bias over time of archaeological sites based on the statistics you generated? How do the results generated on real data differ from the random points? (3 points)

STUDY QUESTION (2): Conolly and Lake (2006) provide quite a few examples of different types of statistical analyses (parametric and non-parametric) that can be applied to studying archaeological assemblages and sites. Provide me with two examples that could be applied in a study design to the radiocarbon dataset you have been working with. How would you need to set up the data to make these analyses work? (2 points)

Chapter 8: Spatial Analysis

8.1. Principles of Spatial Analysis

Spatial analysis is meant to test relationships in space. Archaeologists add the time dimension, so it can be conceptualized as a 3-dimensional view of how data co-vary. Discriminating events in time is not simple—Bayesian methods provide probabilities and frequentist measures test hypotheses.

As applied to archaeology, spatial analysis is primarily concerned with how points relate to each other in terms of clustering or dispersion, and often we weigh attributes as factors in determining these relationships. This week, you will begin the "real work" of GIS and begin learning geostatistical modeling of spatial data. This skill is one that takes a lifetime to master, so be patient with yourself.

8.2. Simple Spatial Analysis

The simplest forms of analysis are those that merely look at how points relate to each other in space. Typically, this is done in a Bayesian simulation environment in which sets of completely randomly distributed points are created within an identical or analogous space that your actual data exists. So, for example, a typical Markov-Chain Monte Carlo (MCMC) model of the distribution of sites in Korea would use the KOR_adm polygon as the bounding region and make 9 or 99 or 999 or 9999 permutations of random sites against which you can test the actual distribution of your data.

We will start our analysis with a [Nearest Neighbor Analysis]() of the distribution of points to determine clustering (a value near 0), dispersion (a value near 2.15) or completely random (a value near 1). This simple but powerful tool is normally a good first spatial test. Load in your TrainingData_UTM_52N.shp file and I have placed a DEM background and the KOR_adm0 polygon just to keep myself oriented. Call up the QGIS geoalgorithm (Figure 8.1) and you will want to start a new folder in your Tablefiles folder to keep track of statistics.

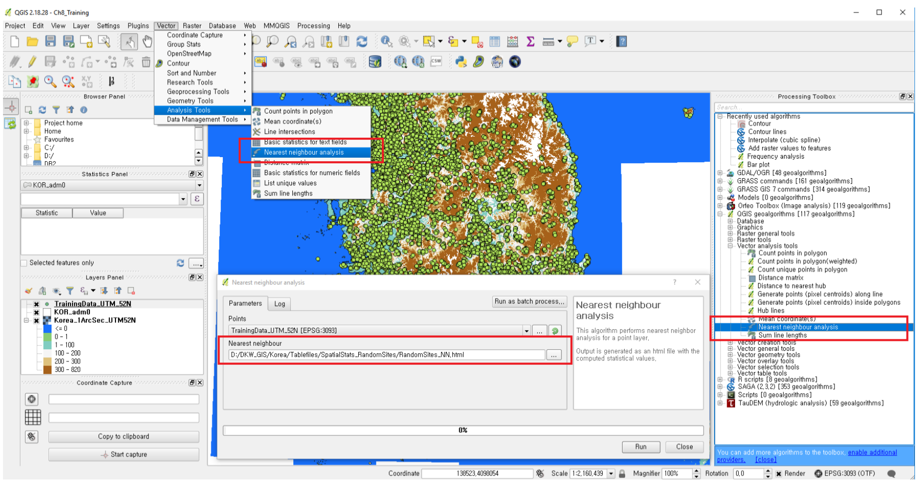


Fig. 8.1. Nearest neighbor analysis.

Click Processing>Results Viewer… (Ctrl+Alt+R) and you will get your NN value of 0.325 and a Z-score of -163.652, which is closer to random than clustered, but not by much. Non-randomness in the distribution is due to the fact that I restricted the creation of points to within 1-km of the boundaries of known archaeological surveys (which do not occur randomly). So, we already know there is a slight clustering bias in the distribution of our random data, which we can assume is due to how the data was generated in the first place.

Next, let's get the mean coordinates of our data. This is the geographic center of distribution of the population of data. This tool is most powerful when attributes are weighed. Starting simply, we will see what the overall mean coordinate is (Figure 8.2). This analysis produces a shapefile, so you will want to save that point in your Vectorfiles folder.

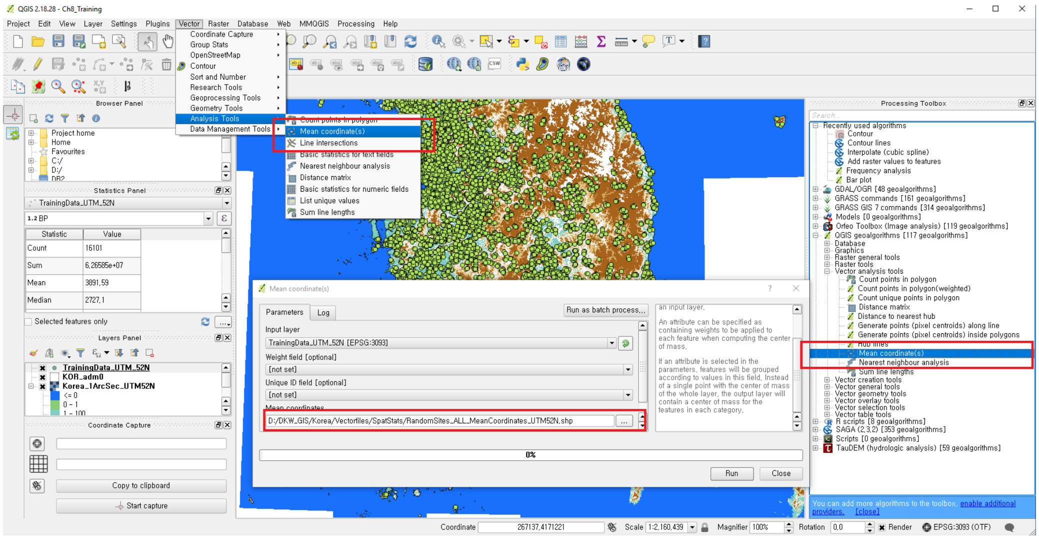


Fig. 8.2. Identify the mean center coordinate of your data.

After you have added your mean center coordinate, you will need to adjust the symbology of the point to make it visible. In Figure 8.3, I have added a red star from the symbology editor to emphasize the location.

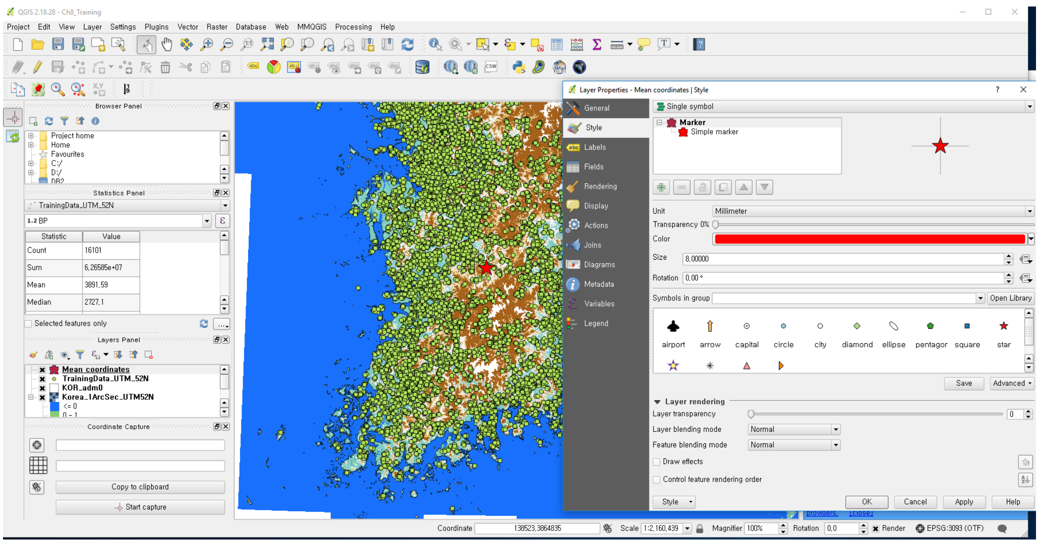


Fig. 8.3. Adjust the symbology of the mean coordinate to make it visible.

Adding a label to the screen will help us keep the information understandable. In these data, you need to add a new attribute column and create a label (I called mine "All data"). Set a buffer and place the label at a distance of 3 (pixels) from the center point.

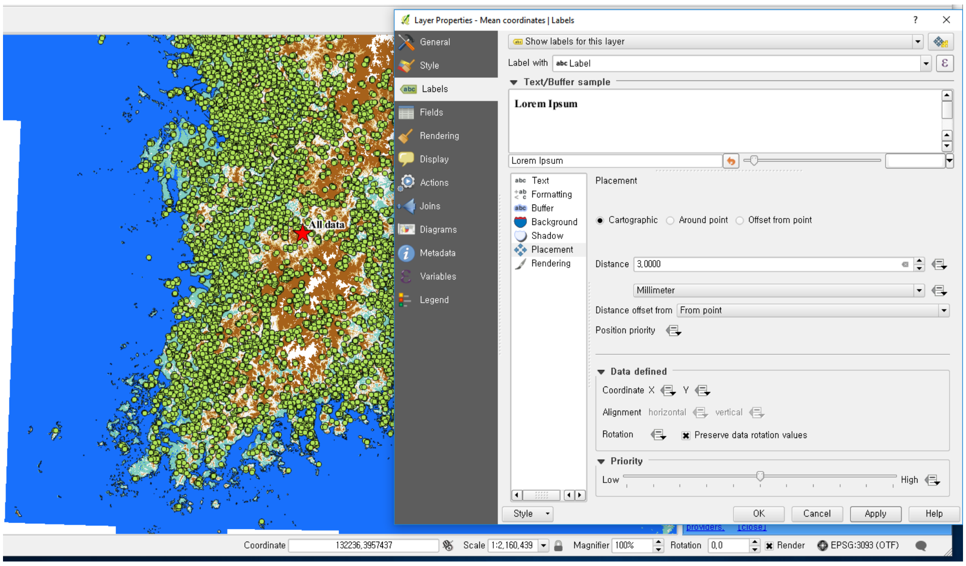


Fig. 8.4. Add a label to the mean center point.

The information we have is interesting, but further exploration of the data can reveal important trends. The QGIS Boolean tools are well equipped to enable detailed queries of the data. In Figure 8.5, I have created a search using the Query Builder (Select by expression) of sites that are younger than 1000 years BP by entering `"BP" <= 1000`. Hit Select and your sites that have the BP attribute <= 1000.000 will be highlighted.

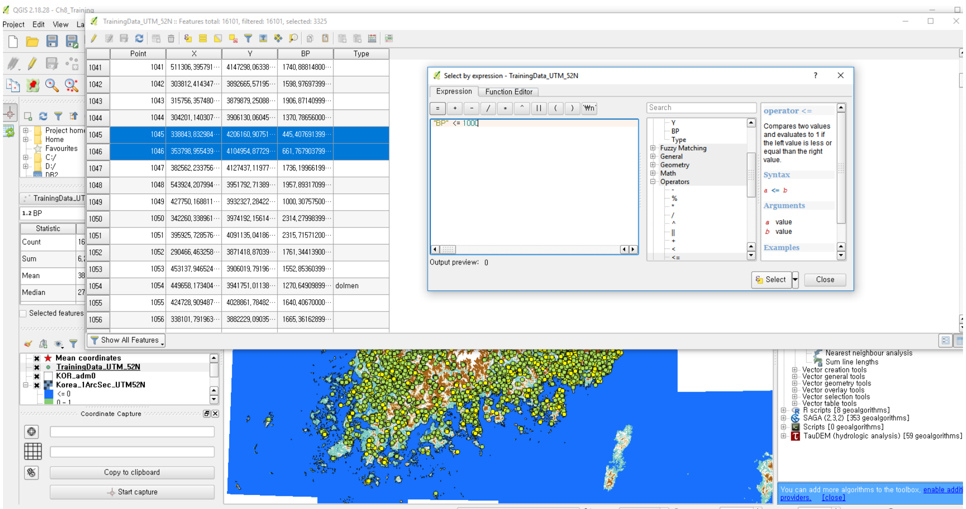


Fig. 8.5. Sift the data by time using Boolean operations

The next step is to make a new shapefile that has only the ≤1000 year BP sites (Figure 8.6). Right click the TrainingData_UTM_52N.shp file and Save as… and create a new folder for Under1000BP.

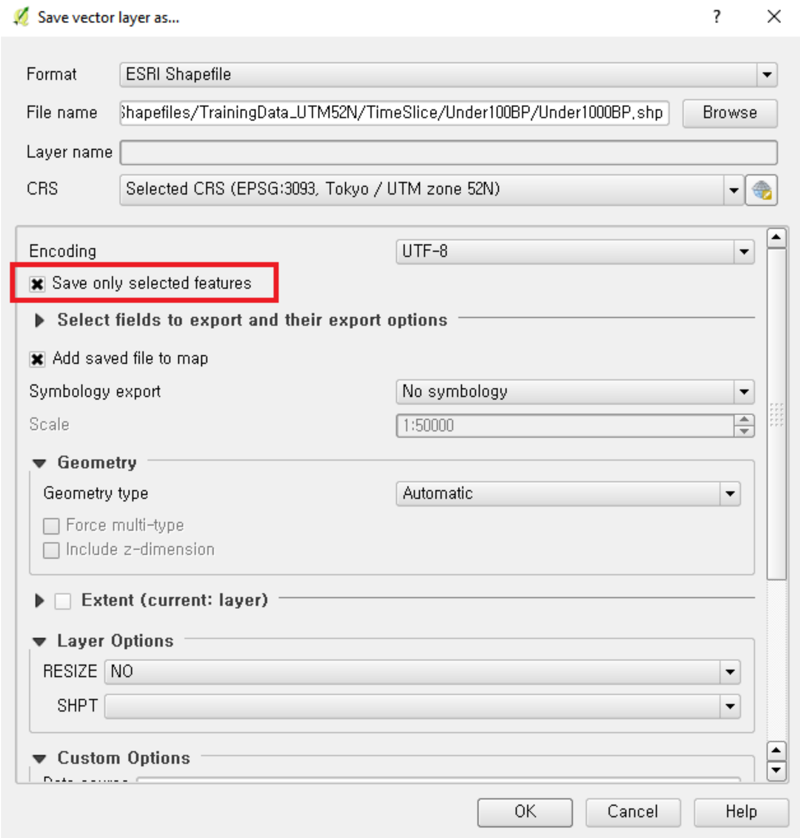


Fig. 8.6. Save attribute selection. Right click the TrainingData_UTM_52N shapefile and save selection.

From your new shapefile, calculate the mean coordinate. Label the point and drag it to the top so that it overlays your All data point. You can see that the point has not moved much (1.3 km, in fact, if you zoom in on it and measure it). Once again, this is because of how the data were created. When you start to use the real radiocarbon data, you will notice a significant shift in the location of the mean point over time, which will reflect real differences in the distribution of populations in southern Korea over time.

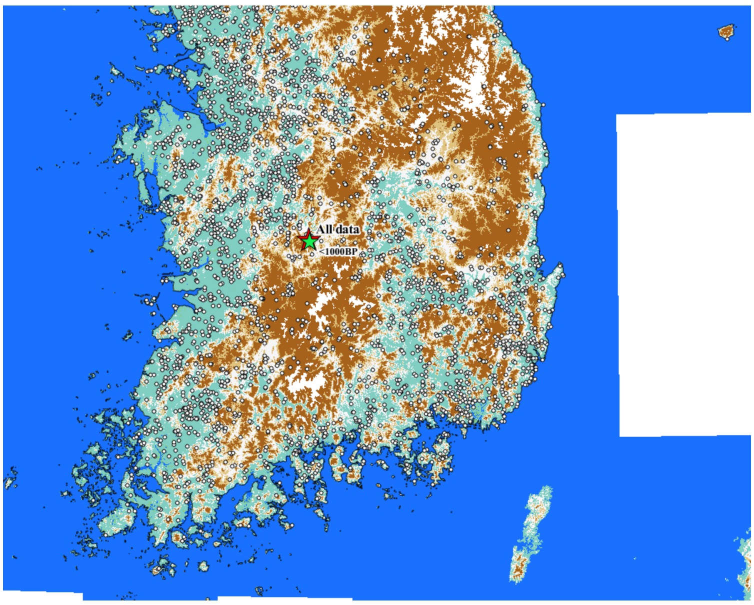


Fig. 8.7. Mean center of the ≤1000 year BP data.

Just to better understand the data, let's calculate the mean center of the KOR_adm polygon file (Figure 8.8). This gives us a sense of where the geographic center of the country lies. Note that it is almost 110 km southwest of the mean center of the data points. Can you think of the reason for this based on what you know of how the data was assembled?

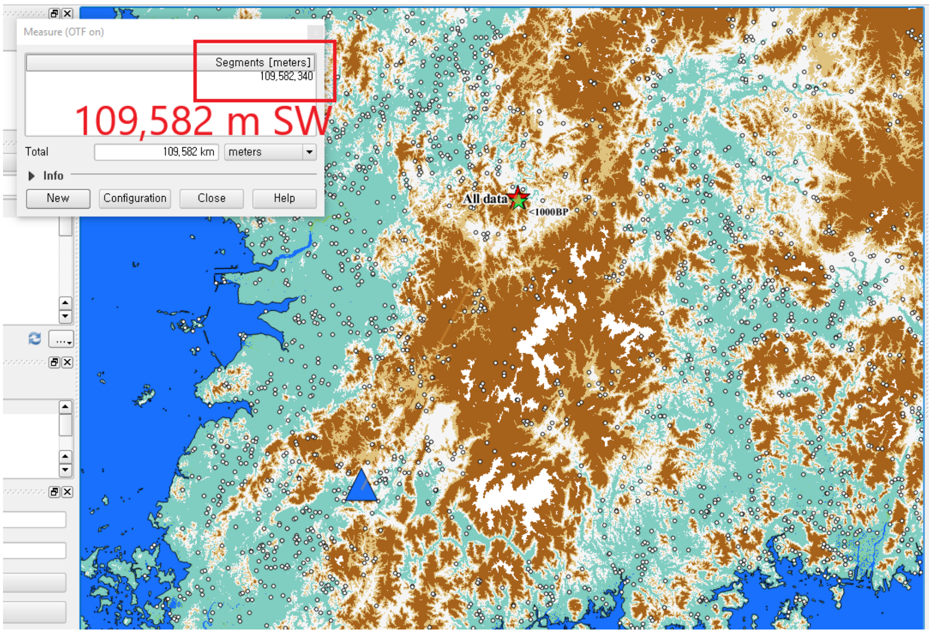


Fig. 8.8. Mean coordinate of the KOR_adm0 polygon file (blue triangle) is 109.6 km SW of the mean coordinate of the random sites.

8.3. Advanced Spatial Analysis

This class only offers you a small taste of advanced spatial statistics. While QGIS has some nice tools, much of the energy in developing advanced spatial statistics is committed to R. We do not have sufficient time in this class to undertake training in R, but if you are serious about pursuing advanced spatial statistical analyses of archaeological data, you will need to learn how to use R.

Fortunately, many R tools are available in the GUI environment of QGIS. So, we will perform some of these to give you an idea of what can be done. Setting up R in QGIS is not straightforward. Figure 8.9 shows you how to activate and configure up QGIS in QGIS 2.18 (Processing>Options… Providers + R scripts x Activate and note the R path and folders designated. You are required to have installed R on your computer already. If you have placed R into a different location on your drive than what is shown here, you will need to navigate to the correct path.

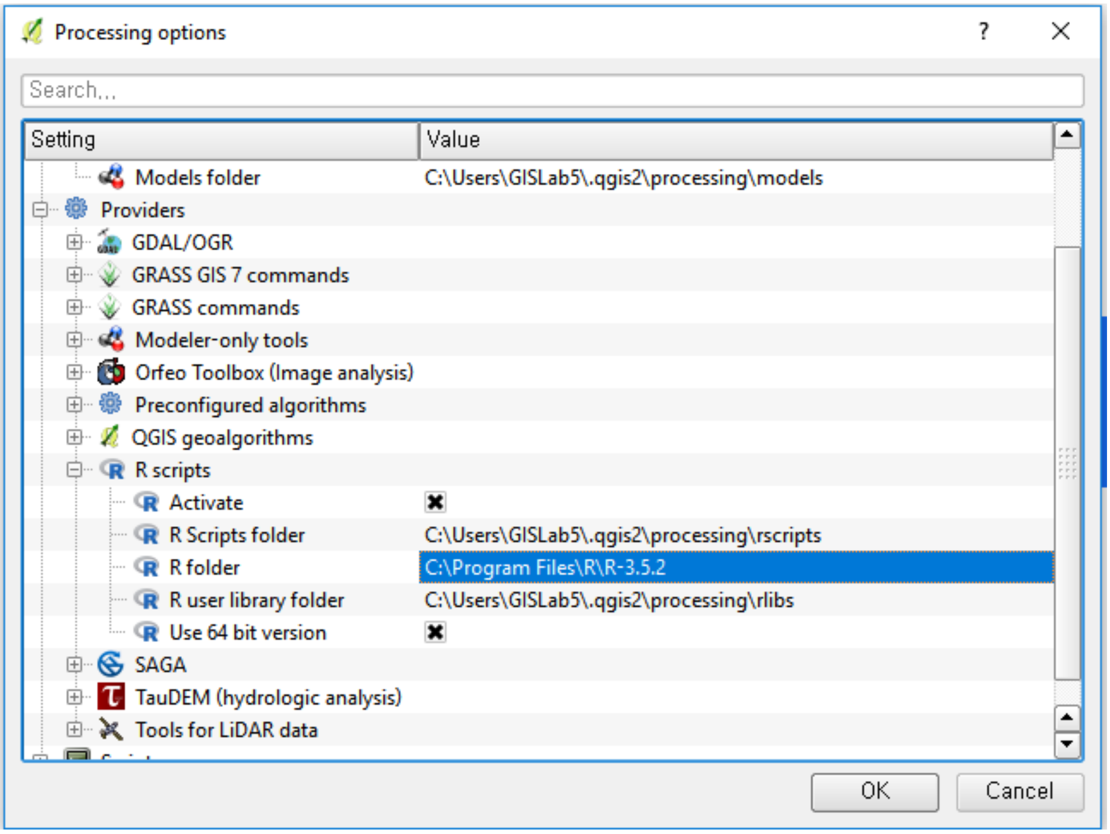


Fig. 8.9. Setting up R in QGIS 2.18.

Figure 8.10 shows how to set R up in QGIS 3.4. There is something wrong with the R protocols as they have been established in Madeira, so I recommend setting the script paths up to the 2.18 environment. If you like working in Madeira and want an R tool, you can simply open up Las Palmas, get the tool, close Las Palmas and the tool will be accessible by Madeira. I am certain that this issue will be worked out more elegantly in coming months, but using QGIS requires that you develop workarounds for development problems, and this is the workaround I have worked out for this particular problem.

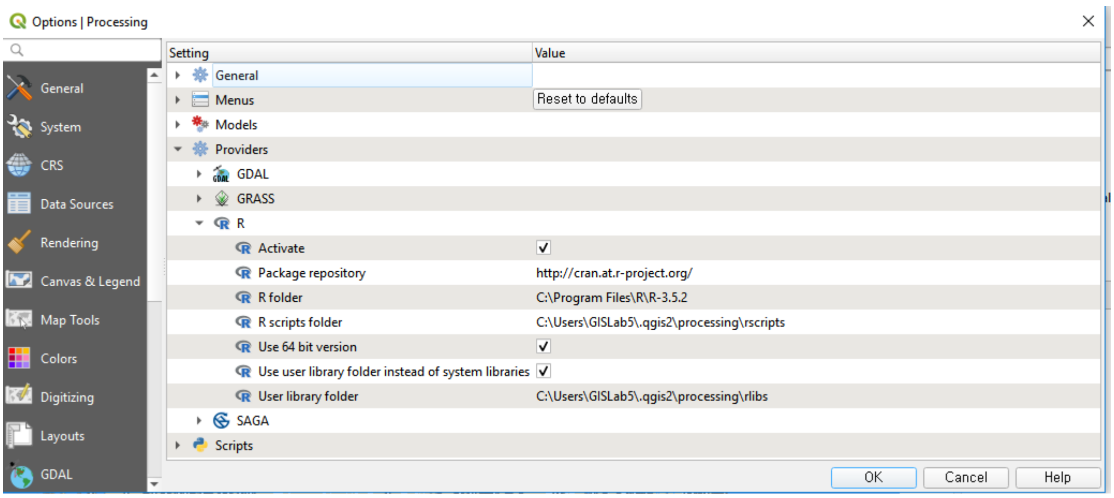


Fig. 8.10. Setting up R in QGIS 3.4.

Once you have satisfactory R environment established, you are ready to get started. The first tests we will generate will be a G function followed by an F function spatial domain test. We will also perform a K test, which is called “Monte Carlo Spatial Randomness” to R. Before you do this, you need to get the R scripts in the Processing Toolbox options (Figure 8.11). Once you have selected this option, you will find the options under Point Pattern Analysis options.

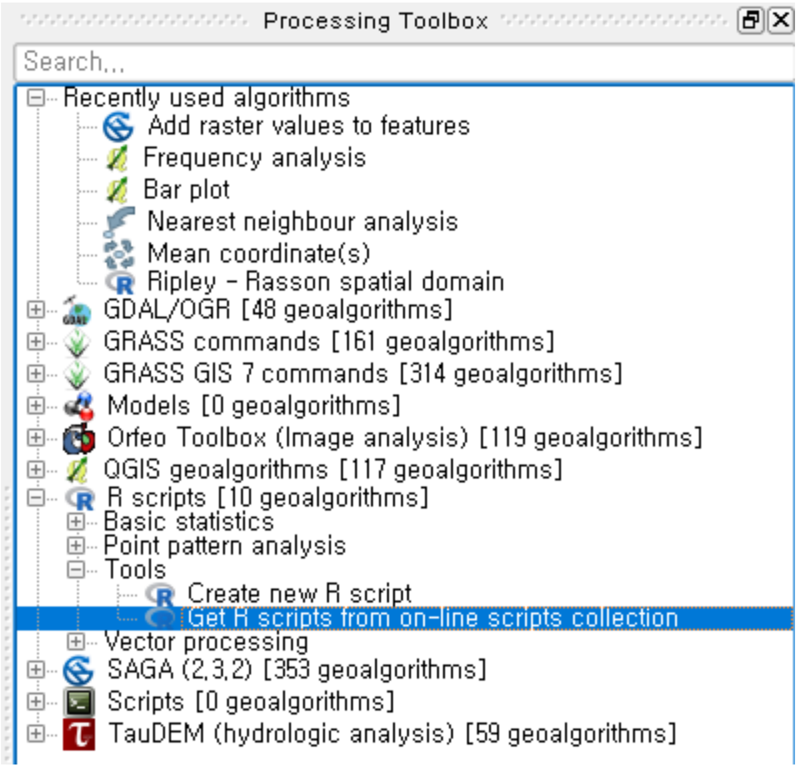


Fig. 8.11. Get R scripts from the online scripts collection.

These tests require two dependencies: maptools and spatstat. Open your RGui program and click Packages>Set CRAN mirror… and pick Seoul. Next, Select repositories… and pick CRAN. Next select Install package(s)… and select maptools and spatstat (Figure 8.12).

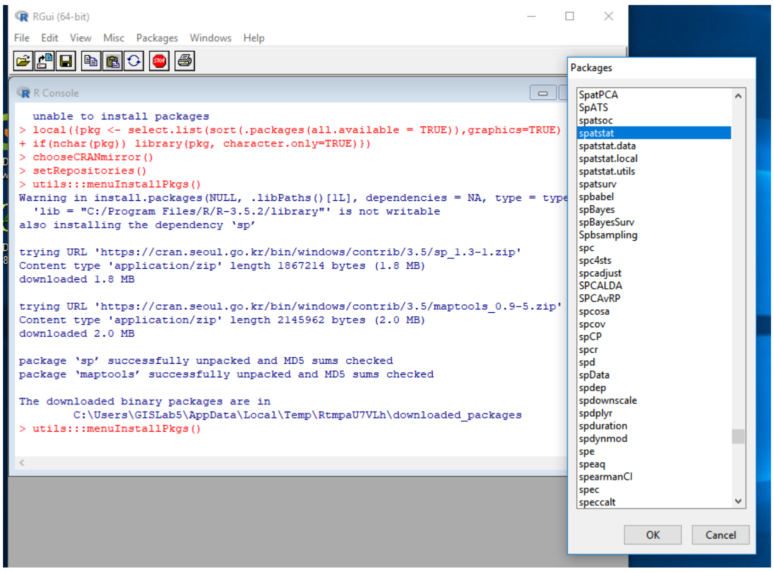


Fig. 8.12. Install maptools and spatstat dependencies in R.

We will now run G and F functions. These tests measure the distribution of points in a spatial domain against permutations of random points at different distances.[17] In the nearest neighbor analysis we ran above, we generally determined that our data was slightly clustered, but "clustering" and "dispersion" are always a question of scale. Cows in a pasture viewed from outer space may appear clustered, but from a drone that is hovering 100 m above the field, they may appear dispersed. G and F functions use nearest neighbor statistics to determine the expected distances of points from a theoretical neighbor (G function) or all neighbors (F function). The F Function is often called the "empty space" function or "point-to-event" function. The G function is an event-to-event test.[18] In these tests *r* is includes all values against which the summary functions G(*r*) or F(*r*) are evaluated. The statistic generates a cumulative frequency distribution based on distance thresholds that are automatically established by the program.

Double click G function and use the TrainingData_UTM_52N shapefile, allow it to run 10 simulations, click OK. Repeat the same for the F function. Go to Processing>Results Viewer… and you will see two R Plots that show a theoretical distribution of 16101 points within the spatial domain of Korea (G/F_{theo} (r)) a confidence envelope in gray (hi and lo) and then the observed distribution (G_{obs} (r)) (Figure 8.13a, b).

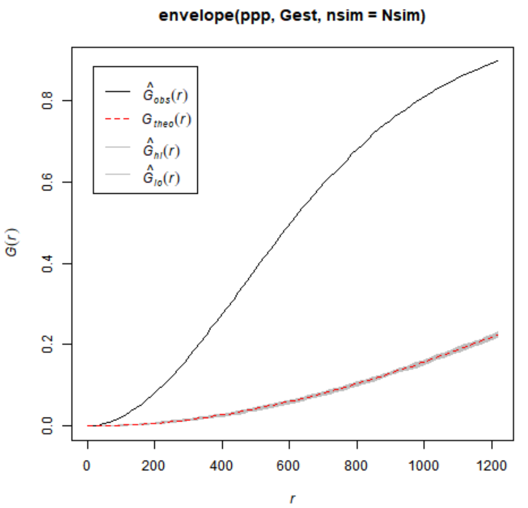


Fig. 8.13a (left). G statistic output.

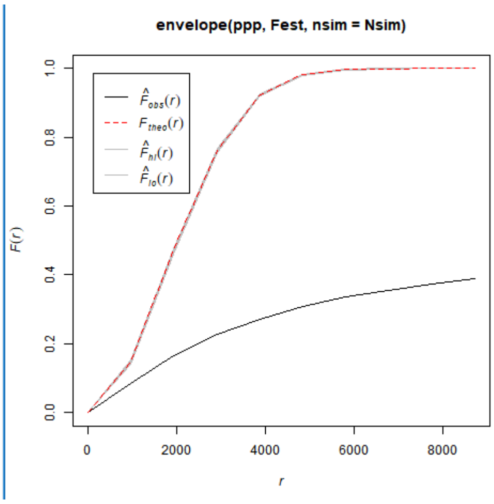


Fig. 8.13b (right). F statistic output

Now, run the K statistic (called "Monte Carlo spatial randomness"). The K statistic looks at the expected number of random points surrounding a given point. It can take a while to run this test because it creates random permutations of points at increasing distances from each one of the points in the analysis. Since we have 16,101 points, it takes a while.

---

[17] A clear summary of what the tests tell you can be found in this pdf.

[18] See Baddeley, A. J., & Turner, R. (2004). Spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software* 12(6):1-42.

The interpretation of these statistical tests is that the points are significantly clustered at low thresholds (within several meters) relative to a completely random distribution of points. If the data were completely random, G(*r*) would increase slowly up to distance where most events spaced, then increase rapidly, and F(*r*) would rise rapidly at first, then slowly at longer distances. K(*r*) would follow the confidence envelope closely. This is not surprising given how the data were generated. We would expect similar distributions for our real data since our random points were generated from the same general spatial domains.[19]

So, we now know the general tendency of our data, but we want to get deeper into the analysis and see if time is a factor in the relative correlations of data clusters. We are going to run a Moran's I analysis. This analysis measures the interdependency of attributes in spatial relation to each other, which is called "autocorrelation." Moran's I performs Monte Carlo (Bayesian) simulations of data points in the range of attributes identical to the data provided. In our case, there are 16,101 points with attributes between 150 and 54,850 that are bounded within the country of South Korea (the statistic creates a minimum bounding box). If the spatial distribution of the attributes are perfectly clustered, the Moran's I statistic will be 1. If they are perfectly dispersed, the statistic will be -1. If the spatial correlations between the attributes are completely random, the Moran's I will be 0 (see also Conolly and Lake 2006:158).

Click the Autocor spatiale R tool and load in your TrainingData_UTM_52N shapefile and the field will be BP (Figure 8.14). Keep the number of simulations neighborhoods at the default levels of 100 and 10, respectively. Click Run.

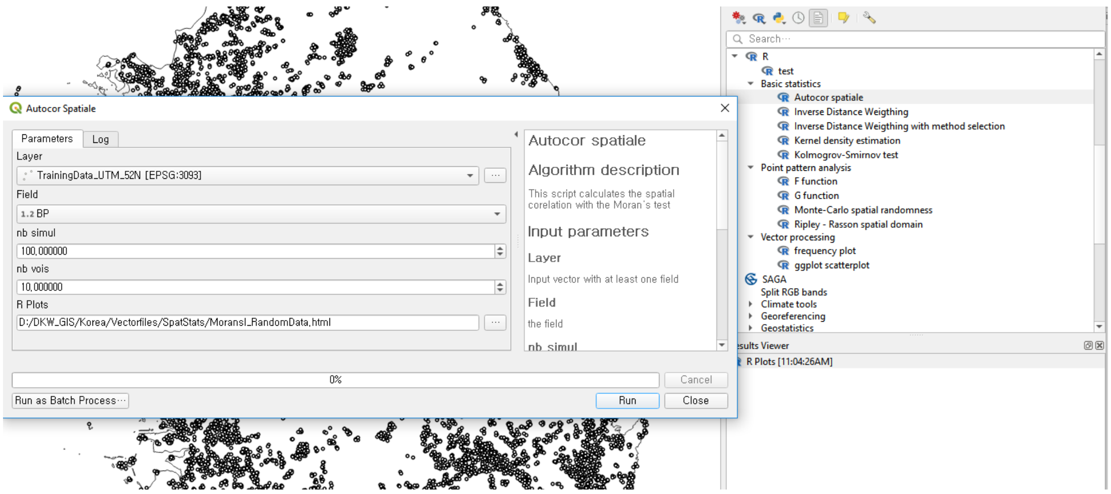
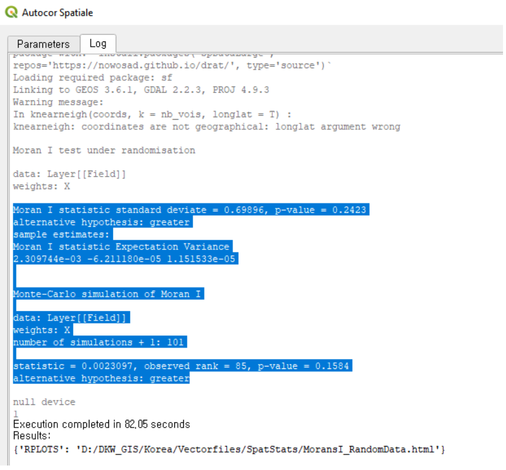


Fig. 8.14. Run the Moran's I analysis of the site distributions.

The tool make take a minute or two to complete the analysis, but when it is finished, you will get an output similar (though not identical) to Figure 8.15a, b. The important factors are the fields "statistic" and "p-value." You will have a statistic pretty close to 0 with a p-value between 0.1 and 0.3. This means that the

[19] G does not work with the real data because so many of the points do not have neighbors in the way they are plotted. This can be overcome by adding decimals to separate the points at a submeter scale. This would allow the statistic to function, but the results would be meaningless because clustering would be so significant at small scales that it would overwhelm the interpretation of clustering or uniformity of the data.

attribute data are randomly distributed (the null hypothesis of spatial randomness *cannot* be rejected).



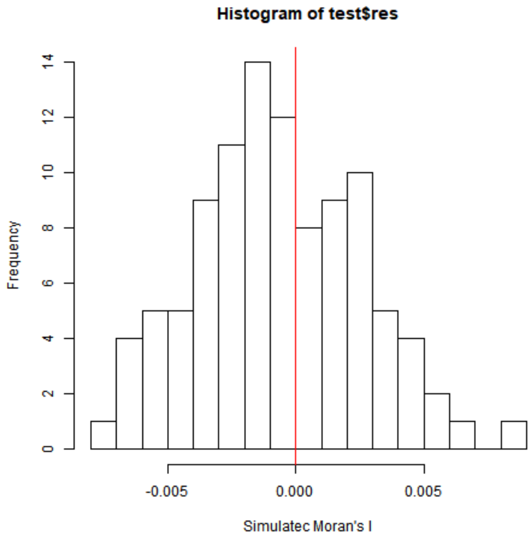


Fig. 8.15a (left). Moran's I statistic output.   Fig. 8.13b (right). Moran's I graphic.

A Global Moran's I can also be performed for raster data in the SAGA toolbox. Note that if you are a Mac user, you need to enable all of the provider options in the SAGA Options | Processing dialogue, otherwise you will get errors for most of the functions.

To this point, we have learned that our data points are relatively clustered (due to how the random points were generated), but there is no spatial correlation between the dates (BP) of those points. Our final spatial analysis will be to generate a type of raster map called a "heatmap." Heatmaps use a Kernel Density Estimation (KDE) to graphically represent hotspots and coldspots of point data. There are several ways to create heatmaps in QGIS: R tools have a KDE creator, you can use the interpolation tools (SAGA or GRASS) we used to create surface models in Chapter 6 if you want to create weighted heatmaps and there is GRASS tool called v.kernel that creates either a raster or vector KDE just from the points. There is also a Heatmap option in the symbology editor of the Properties dialogue that allows you to simply adjust the color scheme based on attribute values of any given column. However, my favorite tool is the "Heatmap" plugin (*NOTE: This tool is not available to Mac users.*) Download the plugin from the repository and double click it to open the interface.

Once the plugin is installed, select Raster>Heatmap>Heatmap… and a dialogue box will appear (Figure 8.16). Create an output raster folder called Heatmap_Sites and save your tiff in that folder. The default radius of 20,000 (m) is fine. Click OK and the computer will calculate your heatmap.

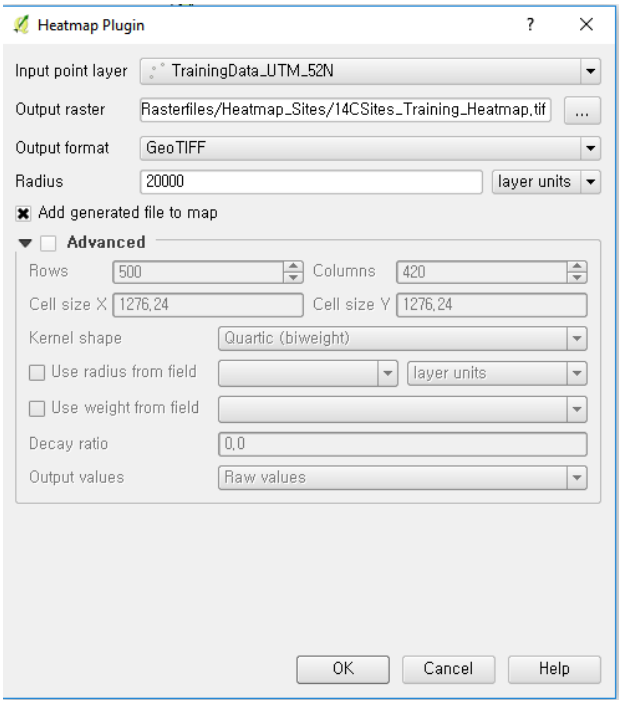


Fig. 8.16. Heatmap plugin setup.

Once your raster is on the screen, clip the raster (Raster>Extraction>Clipper…) to your KOR_adm shapefile dimensions and reset the color parameters to Inferno (or whatever you like) and you will have a heatmap that looks similar to Figure 8.17. In this map, the yellow areas have the highest density of (randomly placed) archaeological sites while the black areas have the lowest density. Repeat this exercise using the v.kernel.rast tool (Figure 8.18).

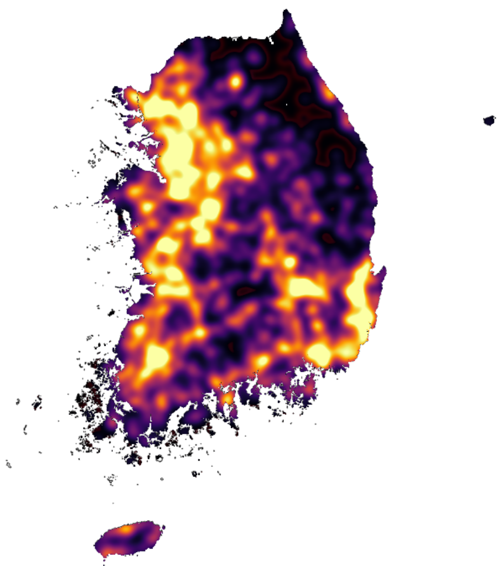


Fig. 8.17 (left). Heatmap plugin heatmap of site distribution in South Korea. Fig. 8.18 (right. v.kernel.rast GRASS tool heatmap of site distribution in South Korea.

8.4. EXERCISE

You have a set of real radiocarbon data from Korea. In that file, there is a classification of features (유구종류-정리본) called pit-houses (주거지) that range from 150 to 35,550 cal. yr. BP. Most of the data are ≤5000 years old. I want you to separate the data ≤5000 years BP into 10 bins (each bin will be 485 years). Select **three** of those bins and identify the mean coordinate, run a nearest neighbor, followed by F, G and K statistics on them. Also calculate a Moran's I on each of these data bins. Write two to three paragraphs to describe what the statistics tell you about human settlement in Korea from these time periods. In what ways are these outputs similar or different from the tests you ran on the training data drawn from random points located within 1 km of known archaeological survey boundaries? Include figures of the graphical outputs similar to figures 8.13 and 8.15 to assist in your discussion.

Next, create heatmaps of the 10 bins of data. You will create .png files of the maps (with a scale, north arrow and title that includes the years included in the particular bin you are analyzing—e.g., Pithouse density in S. Korea 5000-4515 cal years BP). Create a gif file (here is a link to a free online gif maker if you don't have any other way to do it). This process can be a little difficult because in order for the gif to look good, you cannot have any variability in the scale of the maps you create. Be sure to use the Scale tool (set it to 3000000) in the Composer.

Also, I am also curious to know if you were to create a logistical regression model similar to what Conolly and Lake (2006:179-183) discuss, what kinds of variables would you want to include in the Korean context and why (give me five variables and reasons)? To help you, watch all four of the videos that begin here. The final video gives and example of how the execution of the logistic regression algorithm can be executed in R; but in SAGA, one of the Multiple Regression Analysis tools could easily substitute for this stage of the analysis.

Provide all of the results as *YOURNAME*_Ch8.docx and *YOURNAME*_Ch8.gif and **upload it to the eTL by May 6 at 9:00 am.** (10 points for completing the assignment correctly and on time).

Chapter 9: Map Algebra, Surface Derivatives and Spatial Processes

9.1. Principles of Map Algebra

Map algebra is a raster-based processed whereby values from one raster are added, subtracted, multiplied, divided, squared, logically-derived, etc. by values from another raster or an equation/algorithm to produce a new set of raster values. These processes can take place on one of four levels: local, focal, zonal or global.[20] In some sense, we have already undertaken some map algebraic functions when we use bicubic interpolation of DEMs. Bicubic interpolation is a focal (or neighborhood) resampling of the raster to accommodate the values of the surrounding raster cells. Similarly, a 3-arc second DEM can be created using zonal operations on a 1-arc second DEM. Raster regions are resampled from ~30-m resolution into ~90-m resolution by averaging all nine of the pixels within the 90-x-90 m square.

QGIS has a powerful tool for performing map algebra called the "Raster Calculator." This allows you to resample rasters according to any set of variables you can conceive. Raster calculations are typically done with Boolean operations, but there are many automated resampling tools as well, some of which you have already used. This exercise will present you with a real-world scenario typical of cultural resource management and you will use map algebraic functions to solve the problem. We will use QGIS 3.4 (Madeira) for this exercise.

9.2. Map Algebra to Select Areas

Using Map Algebra, we can use rasters to select specific areas. This can be useful in cultural resource management applications. When developers want to build infrastructure, they normally have multiple criteria for selecting locations. While this can be performed using paper and pencil, it is far more efficient (and accurate) to use GIS to sift through areas suitable and not suitable for development.

In this exercise[21], you have a developer who wants to locate a new museum in a cultivated or urban area with a population density >750 people/km^2 in Gyeonggi-do, Incheon or Seoul within 1 km of a major highway and not within 250 m of an archaeological site. She also wants to put the museum on a low slope area (less than 5% slope) and the site must have 20 ha (20,000 m^2) of contiguous space in which a museum site can be located.

---

[20] A really nice summary of the different scales of map algebra can be found at this website: https://mgimond.github.io/Spatial/map-algebra.html.

[21] This exercise was adapted from a workflow published on the YouTube channel of Hans van der Kwast (https://www.youtube.com/watch?v=IfZSyBf6o84). I have adapted the data model to fit this course and added components to deal with complicated raster functions commonly needed in cultural heritage impact assessments.

The first task is to gather all of the data you will need in to your GIS. You have already used the GLCC map in Chapter 3 and that contains the land use categories that include categories of "artificial surfaces" and "cultivated." Add the population estimates KOR_msk_pop shapefile from DIVA.[22] Add the KOR_roads shapefile from DIVA as well. Use the regional DEM we created in Chapter 7 from which you will calculate the slope. Reproject all files into UTM 52N. We will use the TrainingData_UTM52N shapefile (use the one that has the administrative categories already coded into the attribute table) for the locations of our archaeological sites (Figure 9.1). Be sure all of your data are projected into UTM 52N (EPSG:3093).

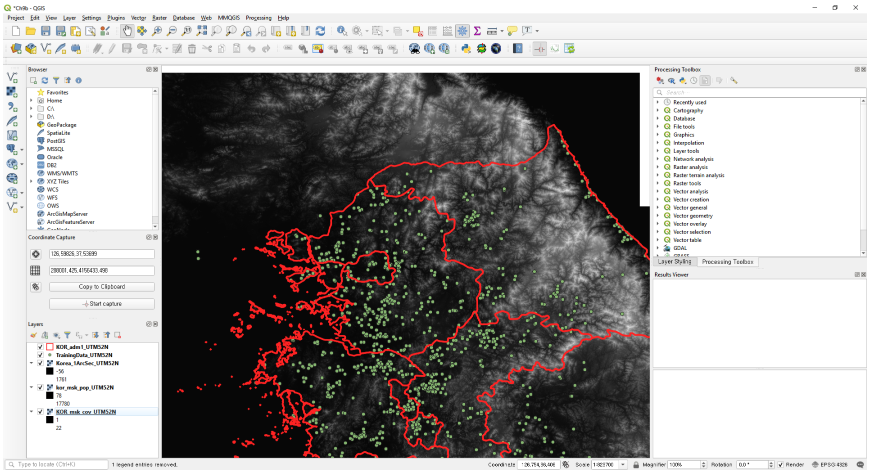


Fig. 9.1. Add the needed files to the GIS.

The second task is to clip the rasters to the project area. Since your developer wants to build in Gyeonggi, Incheon or Seoul, you can select those administrative areas from the KOR_adm1 shapefile. Once they are selected, clip the KOR_msk_cov, KOR_msk_pop and DEM by the selected features (derived from NAME1) from the KOR_adm1 file (Figure 9.2).

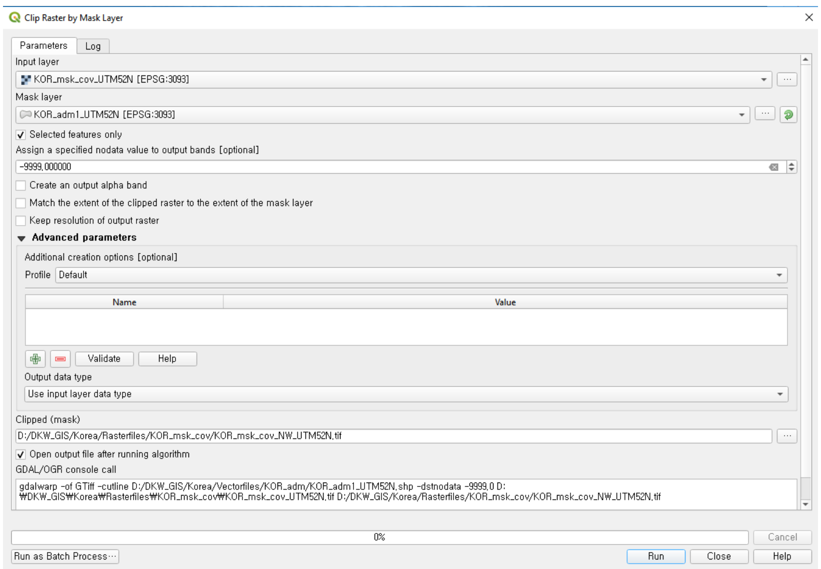


Fig. 9.2. Select Gyeonggi, Incheon, Seoul from the KOR_adm1 attribute table. Clip Raster by Mask Layer (Select Features only). Repeat this for the other rasters. Rename the rasters (can be easily done in Properties>Source or just by clicking on the name in the Layers Panel).

[22] Note that the best statistical data comes from www.kosis.kr, but for the purpose of training, we will use the old, simplified population model.

Next, you will use the Select by Expression calculator to isolate archaeological sites only within the provinces under consideration for constructing the museum. Click the Select by Expression button in the attribute table and use the Fields and Values and Operators to build an expression to select the sites you are interested in displaying (Figure 9.3). Save the selected features as a separate shapefile.

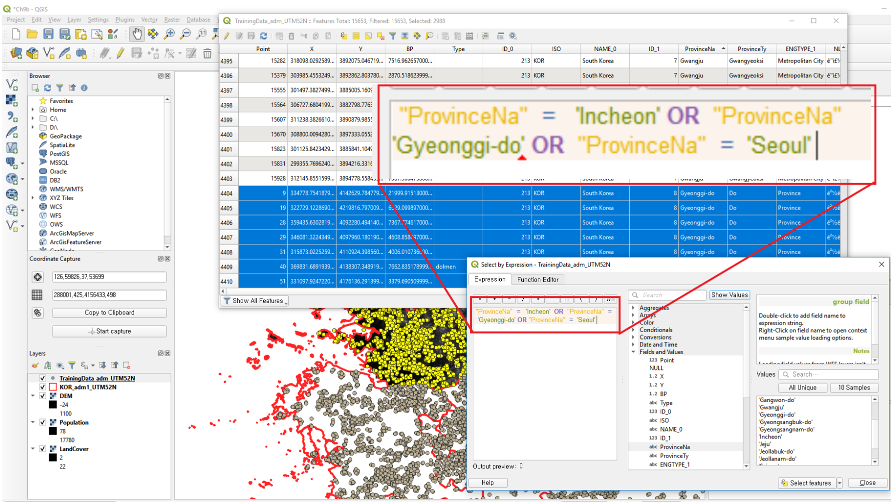


Fig. 9.3. You can use the TrainingData_adm_UTM_52N file we created in Chapter 2. Use the expression calculator to select the site you want.

In order to establish the required distance to roads the developer requires, clip the KOR_roads shapefile (Vector>Geoprocessing Tools>Clip…) according to the geographic boundaries used to clip the other files (Figure 9.4). Create Roads for just the NW corner of Korea. Use the buffer tool (Vector>Geoprocessing Tools>Buffer…) to create 1000 m buffers around the roads and 250 m around archaeological sites. Select the "Dissolve result" option when creating the buffers as we did in Chapter 7.[23]

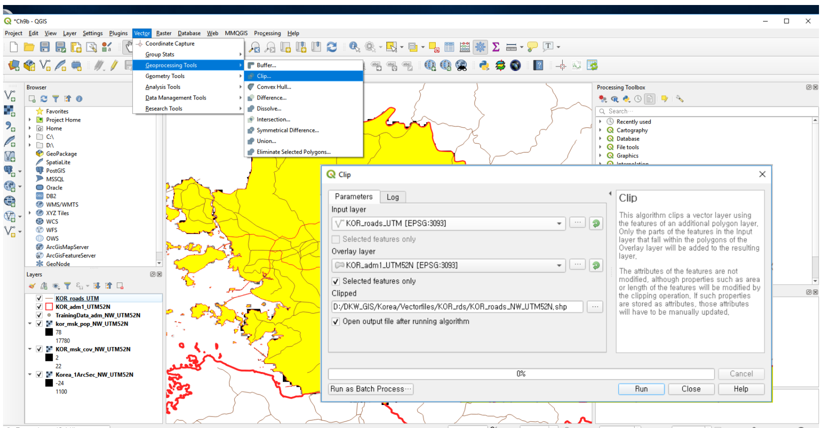


Fig. 9.5. Clip the KOR_roads shapefile. Be sure the coordinate system is in UTM 52N.

Once your polygons are clipped, you will need to rasterize the results. Type "rasterize" into the Processing Toolbox search bar and select the SAGA Rasterize

23 Note that there is a small section of road in the SE of the country that is included. Delete this before proceeding.

tool. You are going to create a data/nodata file (Figure 9.6). You will want to zoom in on the area of the road network and set the output extent to that view. Make a MuseumSites folder in your Rasterfiles folder from which to work. Your output should look similar to Figure 9.7.

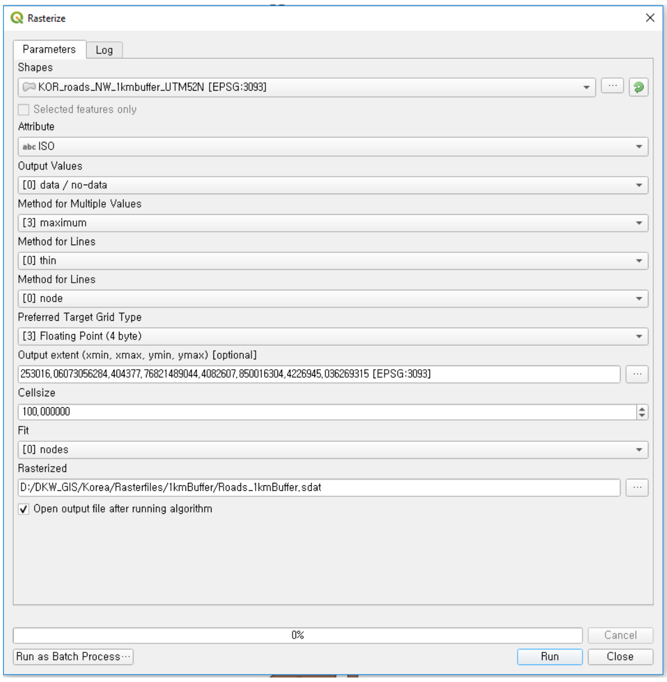


Fig. 9.6. Rasterize polygons. SAGA>Raster creations tools>Rasterize.

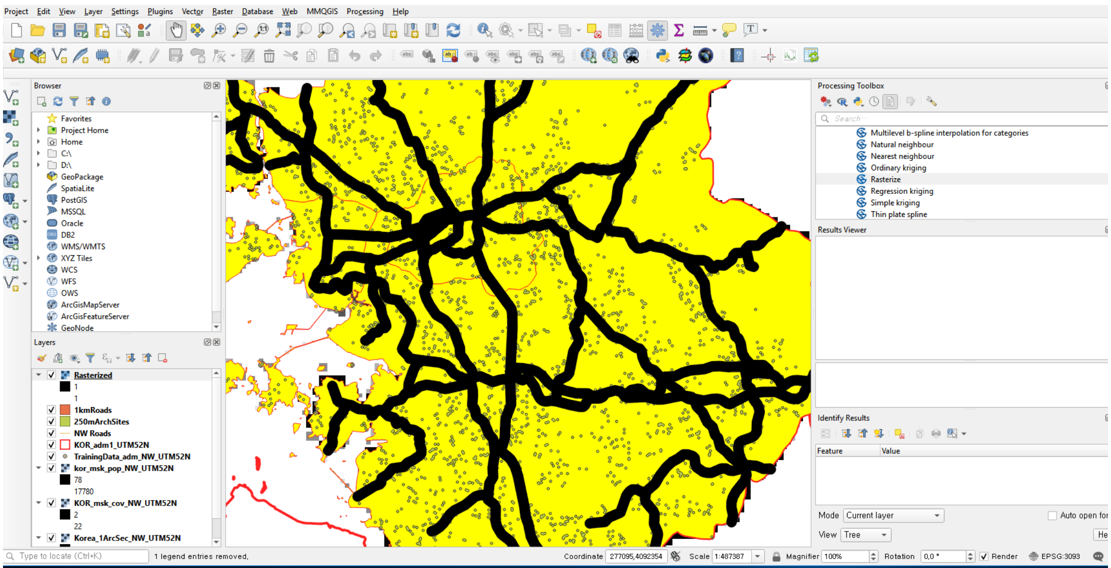


Fig. 9.7. Rasterized road network in Gyeonggi, Incheon and Seoul.

We are constructing a computer-based logic operation in which 1s on a raster maps are the areas suitable to build in (can-build) and the 0s are not suitable (no-build) according to the developer's parameters. However, in the 250-m buffer map we created, the data (1s) are the areas we need to avoid. Therefore, after you create your 250-m buffers and rasterize the archaeological site map, you need to invert the values so that the areas within 250 m of archaeological sites have a no data value and those areas outside 250 m are classified as 1. The SAGA Raster tool Invert data/no-data is perfect for this task (Figure 9.8).

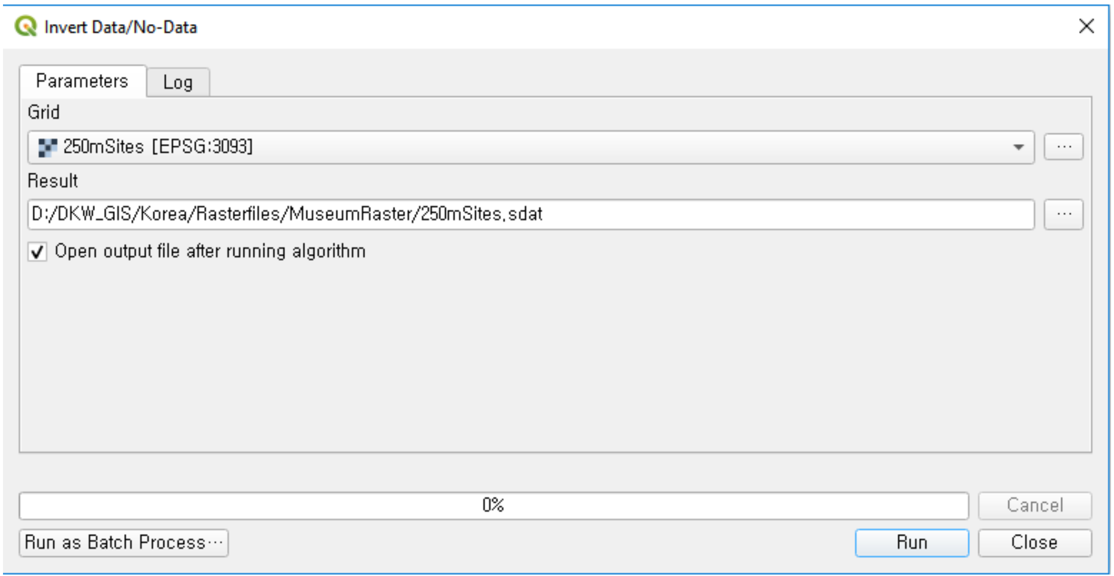


Fig. 9.8. Invert the data/no-data parameters of the buffered archaeological site raster.

Once your rasters are created, you are going to want to create a color-based classification scheme to keep everything conceptually organized in your mind. For the purpose of this exercise, we will make all areas in which we can build the museum coded as green and areas where we cannot build the museum will be coded as red (if you are colorblind you will want a different color scheme). Most importantly, in our raster classification scheme, 0s will be the red (or no-build) locations and 1s will be the green (or can-build) locations (e.g., Figure 9.9).

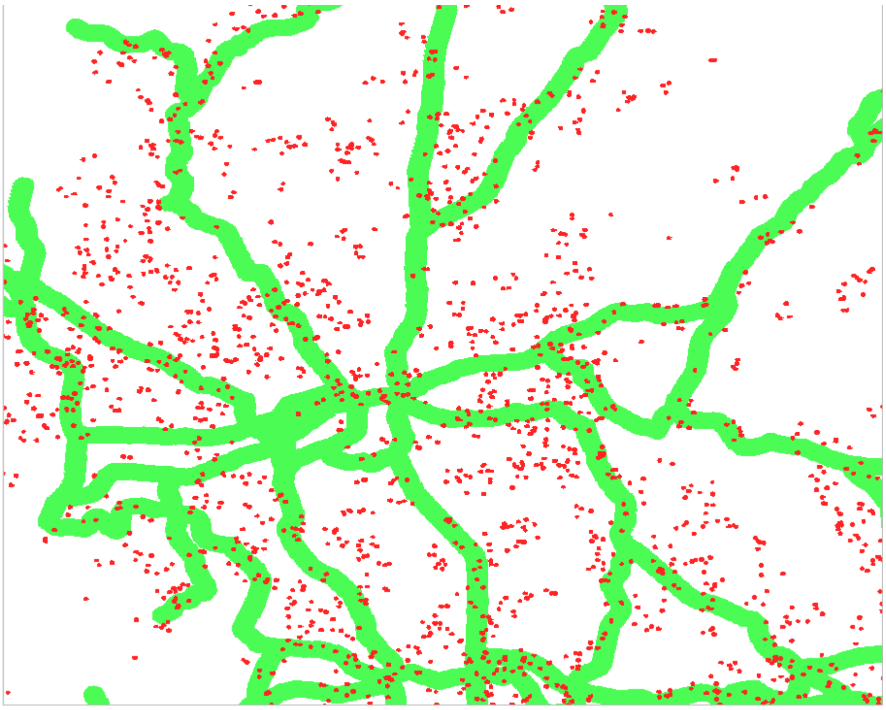


Fig. 9.9. Create a color-based classification scheme for can-build and no-build locales.

From Figure 9.9, you can notice that there is missing data: no-build locations within the polygonized 1-km buffered roads file are not displayed and can-build locations in the polygonized 250-m buffered sites file are not displayed. You can give them attributes using the r.null GRASS tool (Figure 9.10).

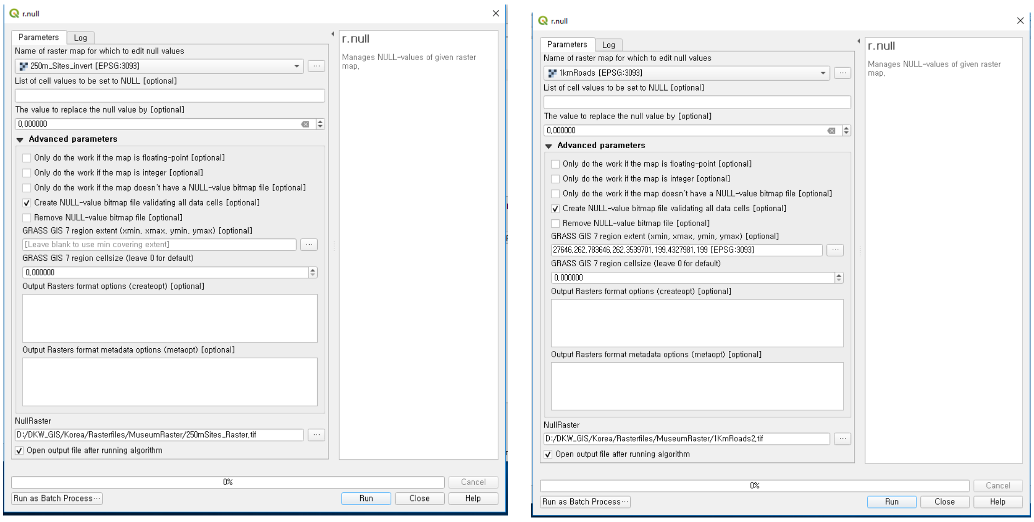


Fig. 9.10. Use the r.null tool to classify the nodata portions of your rasters.

Once you have finished this operation, you can reclassify the symbology of your new rasters according to the can-build/no-build scheme above (Figure 9.11). This raster will need to be clipped to the shape of Gyeonggi, Incheon, Seoul as explained above.

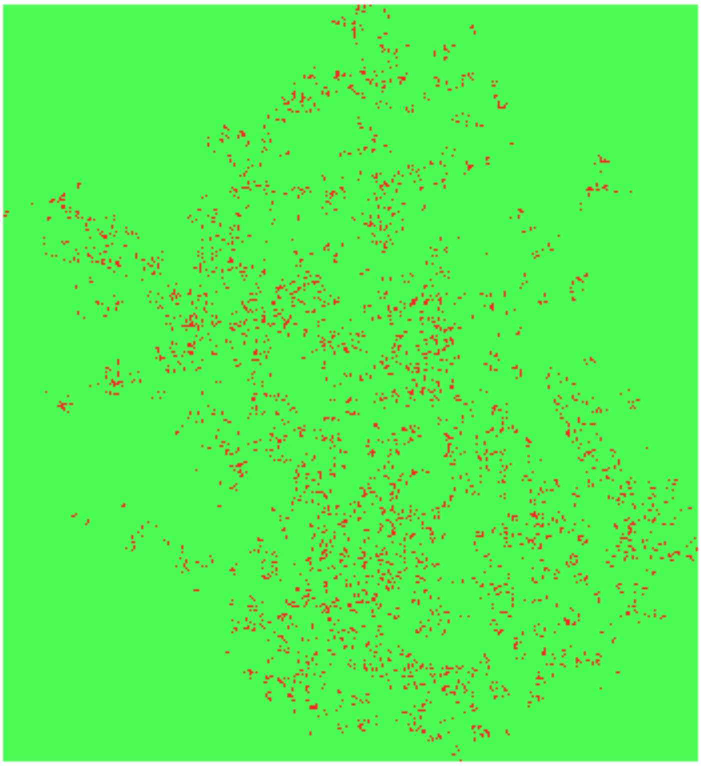


Fig. 9.11. Buffered archaeological sites raster after r.null and changed symbology.

Now, classify the slope to the developer's criterion. The first step is to convert the DEM into a slope map using the GDAL tool found in the Processing Toolbox. Be sure to categorize this map in percent instead of degrees, as you did the DEM in Chapter 7 (Figure 9.12). Create a new tif file in your DEM folder.

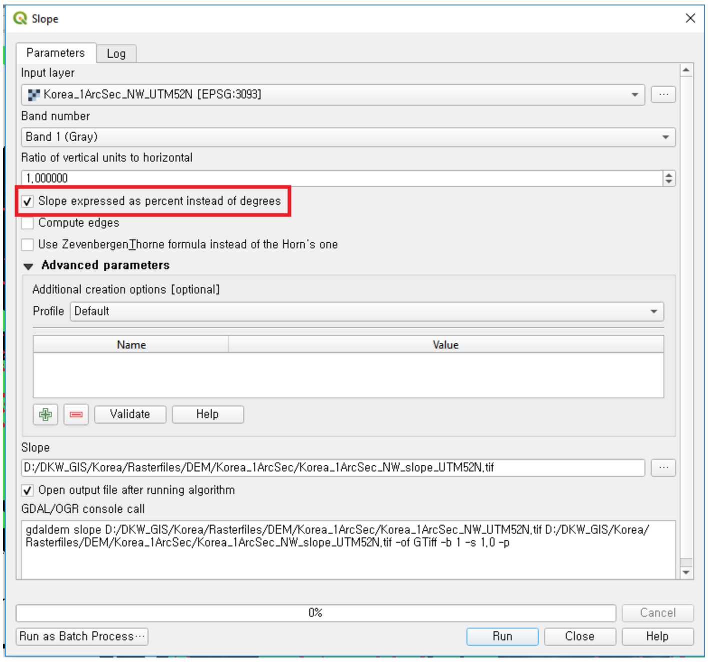


Fig. 9.12. Create a slope map in percents from the DEM.

Use the Raster Calculator to create a new raster that reclassifies the values on the basis of whether or not they are greater than 5% slope (Figure 9.13). Store the output in your MuseumRaster folder.

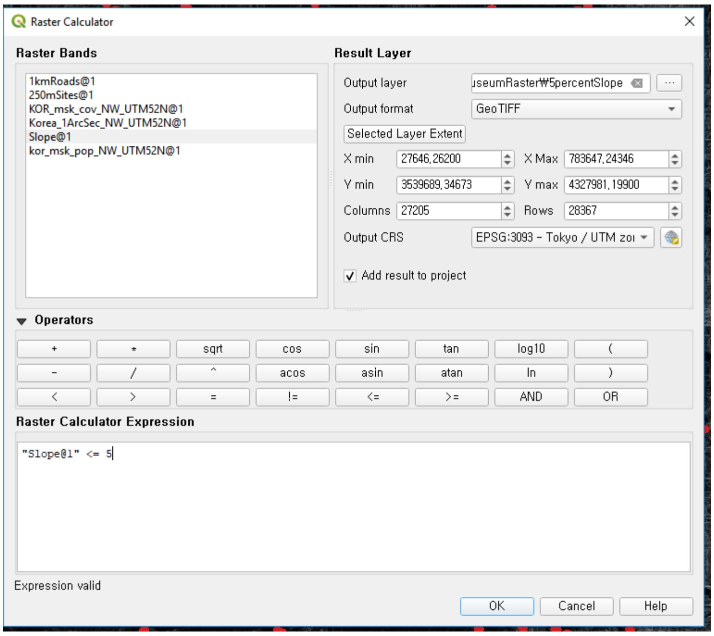


Fig. 9.13. Recalculate the raster into binary variables. In the new file, cells that are >5 will be reclassified as 0 (no-build) and those that are ≤5 will be 1 (can-build).

Change the output map into the no-build/can-build binary color scheme (Figure 9.14).

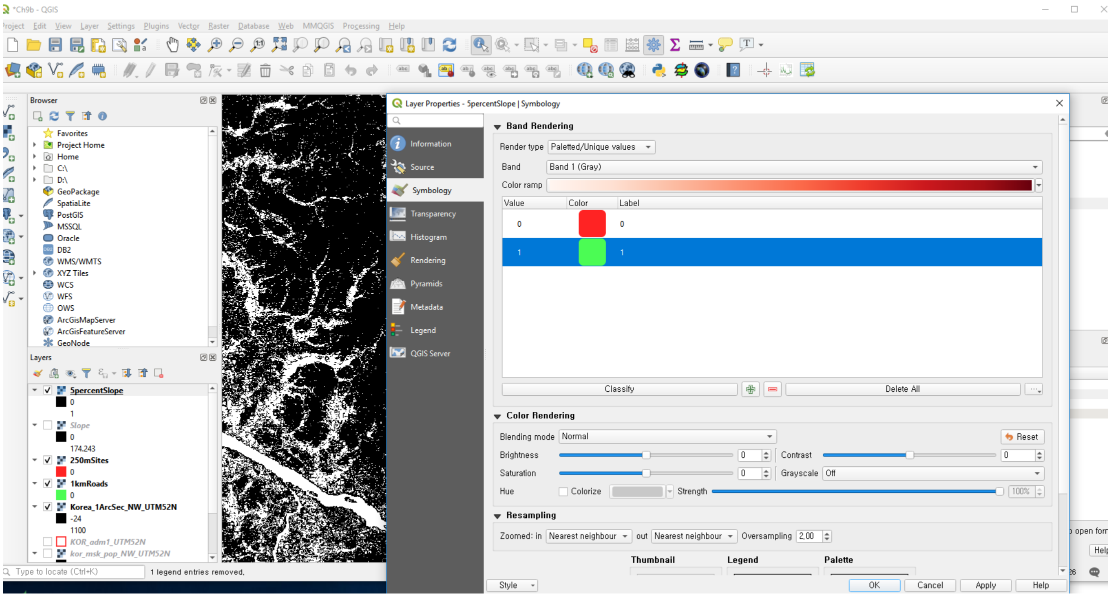


Fig. 9.14. Reclassify the raster map into the binary color scheme in the symbology.

Using the KOR_mask_pop shapefile clipped to Gyeonggi, Incheon and Seoul, you can make a binary population map of areas ≥750 people/km² and <750 people/km². Use the Raster Calculator to build an expression as shown in Figure 9.15. You will have created a new raster map in which the 0 value represents <750 people/km² and the 1 value is ≥750 people/km². Reclassify the symbology into red (0) and green (1) binaries.

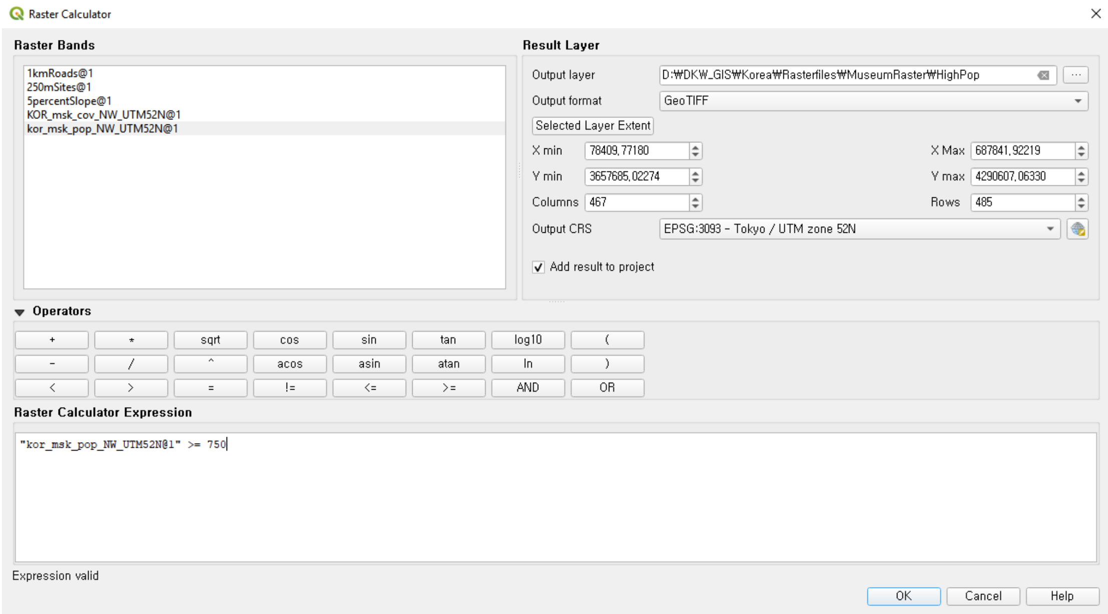


Fig. 9.15. Select population >750/km².

You will perform a similar operation on the GLCC map that has been clipped to include only Gyeonggi, Incheon and Seoul, but this time you will classify the raster according to the land categories 16 (Cultivated and managed areas) and 22 (Artificial surfaces and associated areas, which we will use as a proxy for the developer's criterion of "urban"). Once again, this operation can be performed simply using the Raster Calculator (Figure 9.16). Remember, the "OR" operator expands the Boolean search to include all of the categories you include.

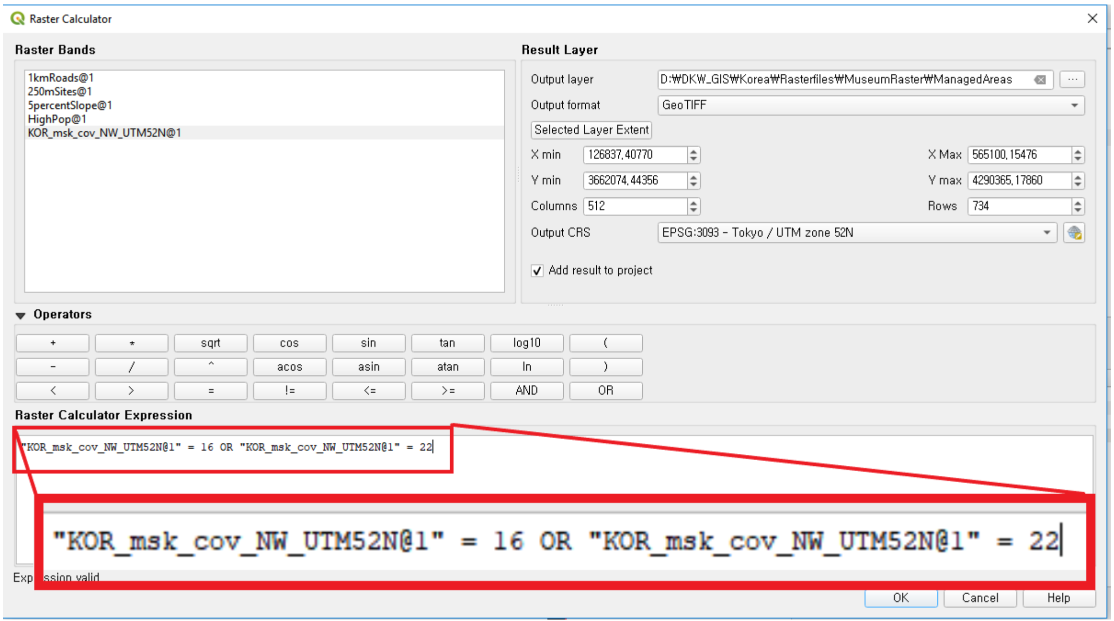


Fig. 9.16. Reclassify the GLCC land cover bare/cultivated areas.

At this point, all of your rasters should be set to meet each of the developer's demands. Remove any raster map that is not the final product of binary operators of can-build/no-build. You now want to combine all of the remaining rasters using the Raster Calculator limiting the output to only those parameters in which satisfy all of the demands. In this instance, the "AND" operator will identify all of the rasters that (1) fall within 1 km of a major highway, (2) are more than 250 m from a known archaeological site, (3) are situated on an area <5% slope, (4) are in a population cluster of more than 750 people/km^2 and (5) are in a GLCC category of either cultivated or artificial surface (Figure 9.17).

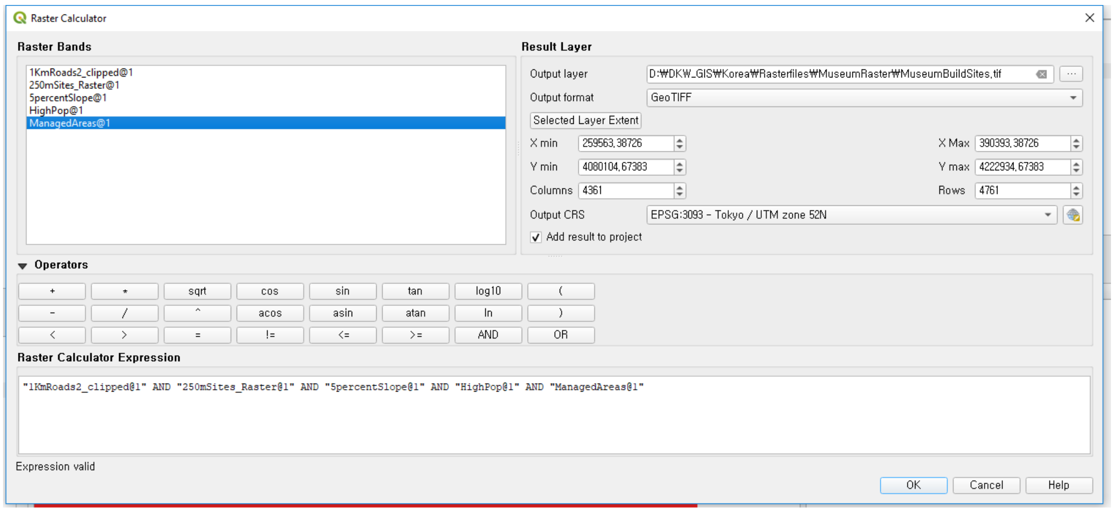


Fig. 9.17. Use the Raster Calculator to produce the final raster view of buildable area in Gyeonggi, Incheon and Seoul (Figure 9.18).

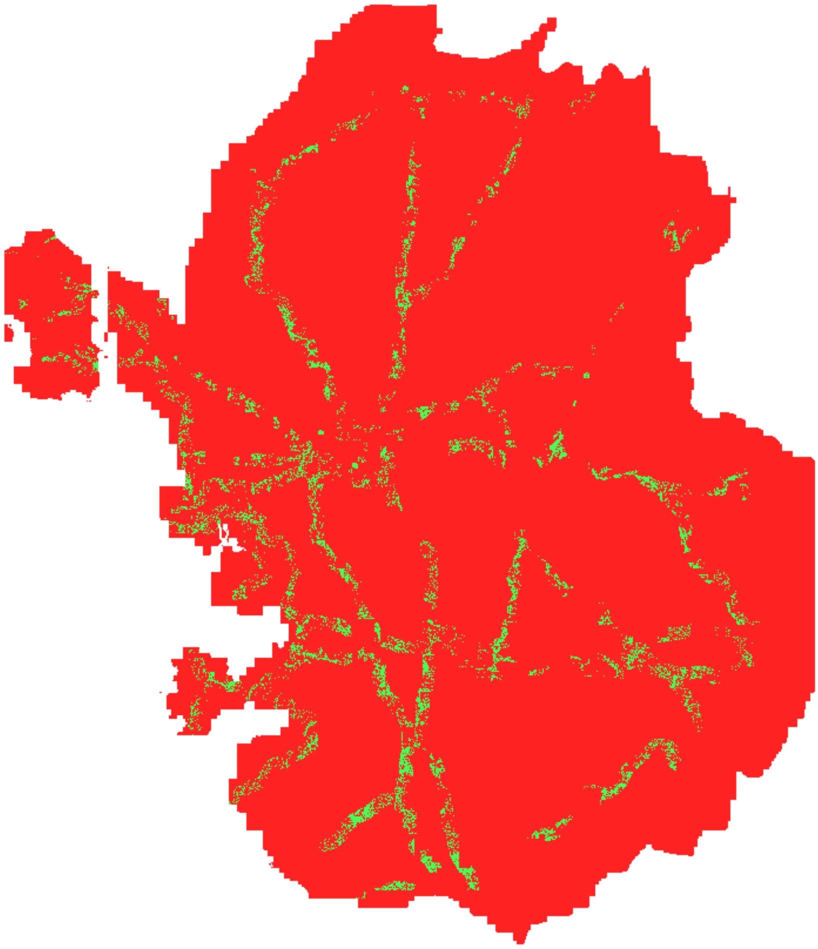


Fig. 9.18. Final calculated raster map of potential sites for the new museum.

To get a sense of the scale, you can reclassify the symbology of the raster to give the no-build areas a transparent symbology (Figure 9.19).

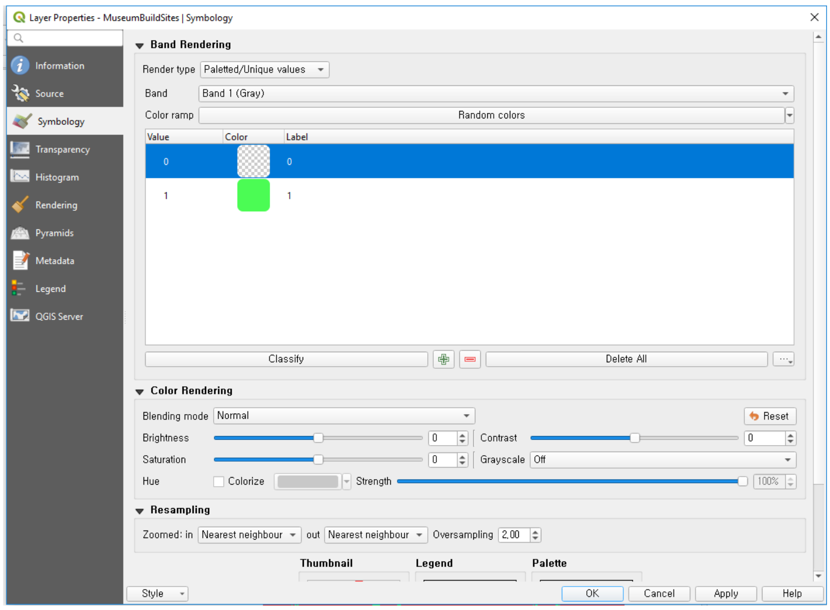


Fig. 9.19. Change the symbology of the can-build/no-build raster map.

If you open the “QuickMapServices” plugin, you will have an option to add an OSM (Open Street Map) background. This gives you a sense of where your preselected site locations are found in real space (Figure 9.20).

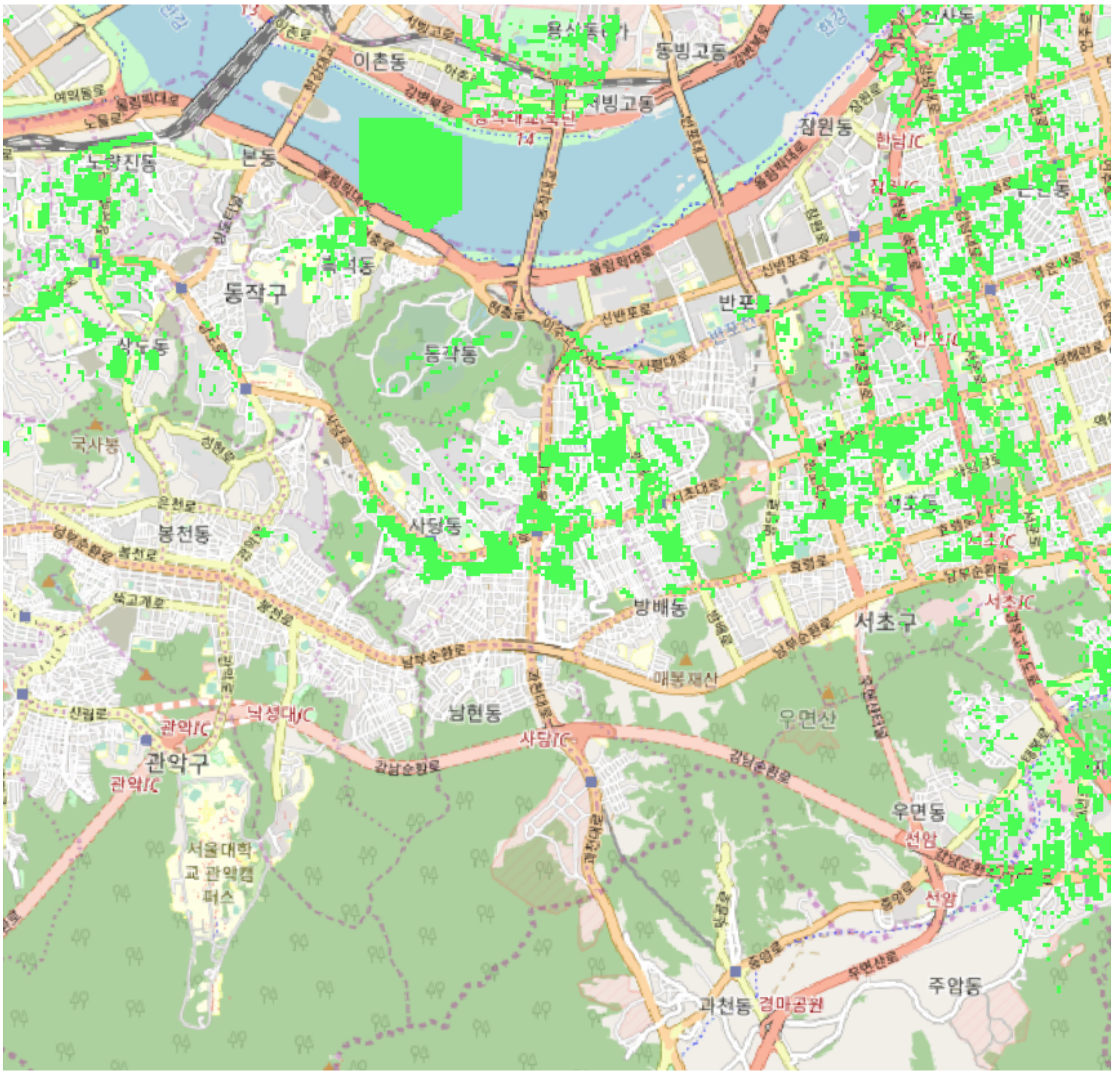


Fig. 9.20. Potential museum can-build sites on the OSM.

We still are not done, though. You should remember that the developer also wanted a site that has more than 20 ha of contiguous land because the new museum is going to be large. This requirement will take a few more steps. First, you will want to create a new no-data band from the 0 parameter. The easiest way to do this is to use the Clip Raster by Extent tool and be sure to assign 0 as the no-data value (Figure 9.21). Call your new file “IsMuseumSite” because all of the data that will be output are potential sites for the new museum.

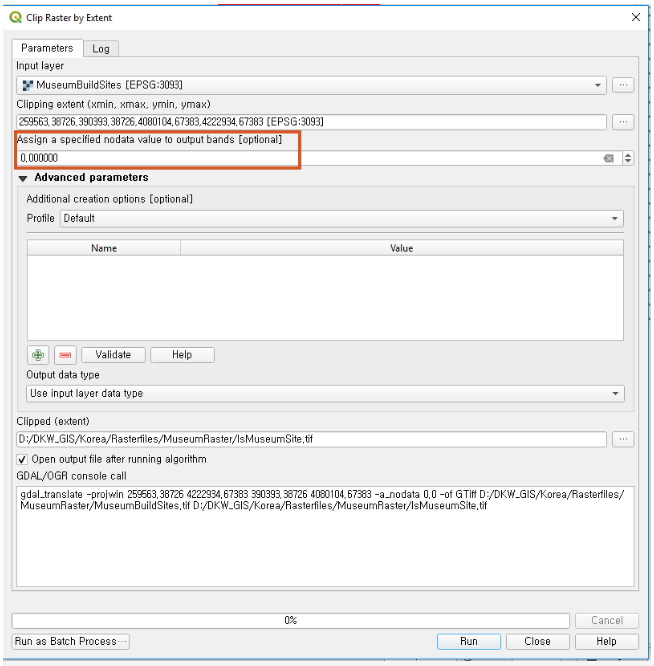
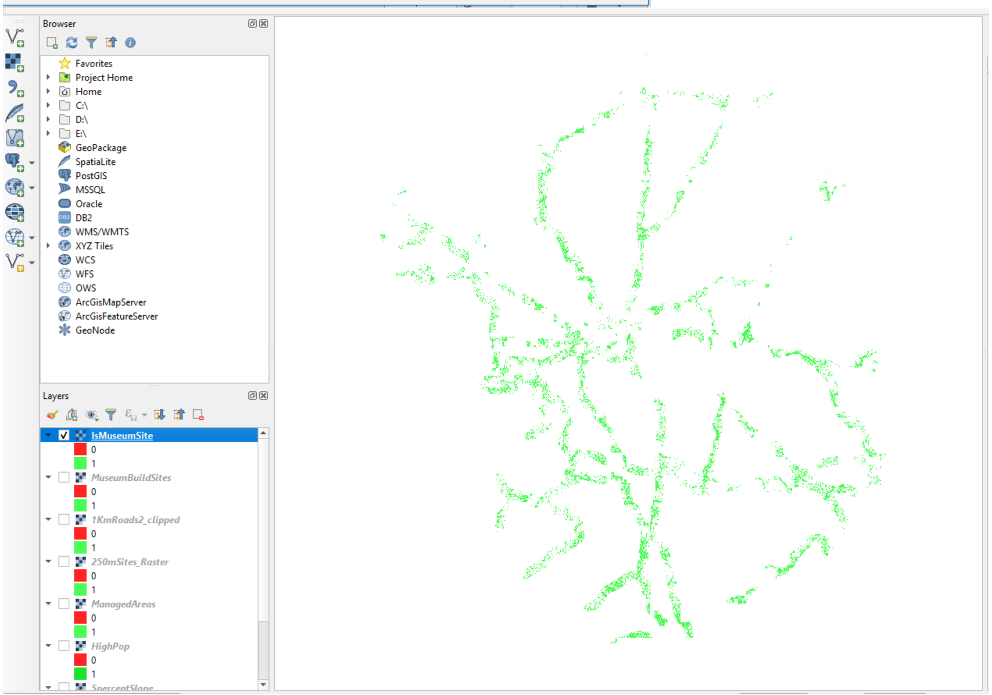


Fig. 9.21 (left). Clip the Raster by extent and reclassify 0 as "nodata." Be sure that your extent covers all of your project area.

Fig. 9.22 (below). The derived, reclassified raster in which the 0s have been eliminated.



Next, we will make "clumps" out of the new raster. Naturally, there is a GRASS tool called r.clump to perform this function. You can find the tool in the Processing Toolbox by searching for it. Just add the IsMuseumSite tif file you just created to the dialogue and create a new tif IsMuseumSite_clump.tif (Figure 9.23). The created raster include only those cells which have contiguous cells with no diagonal connections. Change the symbology of the new raster to have random colors.

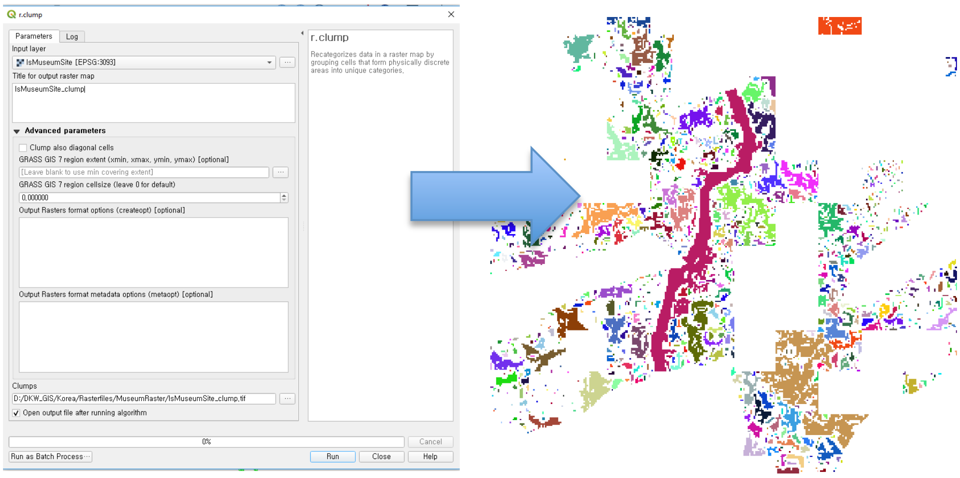


Fig. 9.23. Use r.clump to identify contiguous cells in your raster. Change the symbology to random colors so you can identify the blocks of contiguous cells.

Next, you will change this raster into a shapefile file using a Raster>Conversion> Polygonize (Raster to Vector…) function (Figure 9.24).

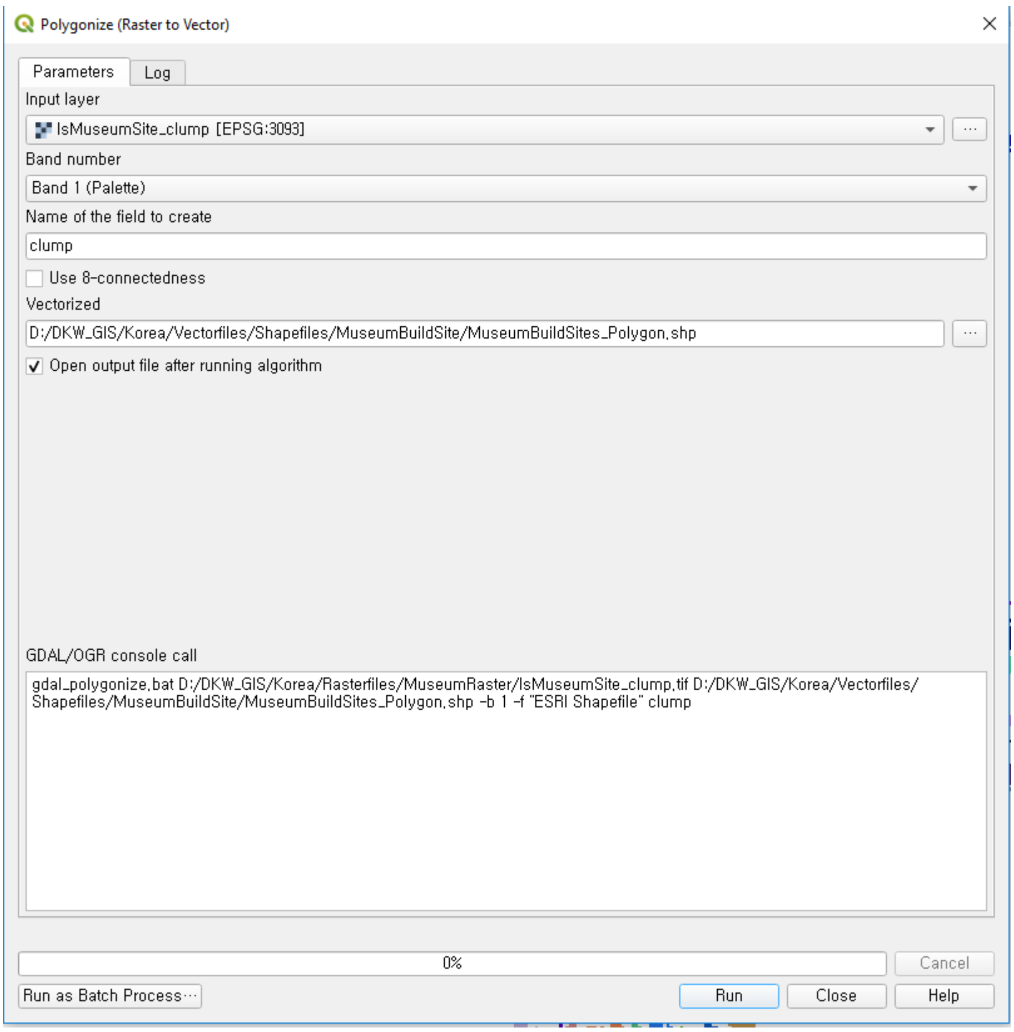


Fig. 9.24. Polygonize the clumped raster map.

Now that you have a polygon shapefile, you can easily identify those geometries that meet the developer's requirements. From the attribute table, you can Select by Expression: `$area >= 20000`, which are polygons >20 ha or 20,000 m^2 (Figure 9.25).

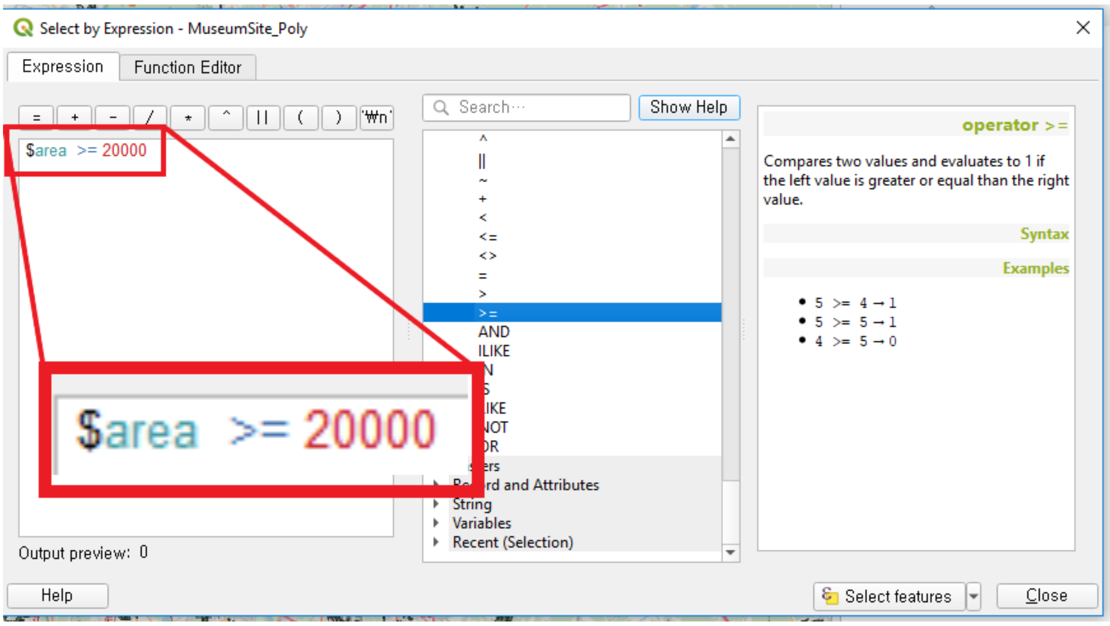


Fig. 9.25. Select potential museum build sites by area.

Finally, once those sites are selected, save the selection into a new shapefile (Figure 9.26). This is the shapefile you will give to the developer that identifies areas in which she can locate the new museum.

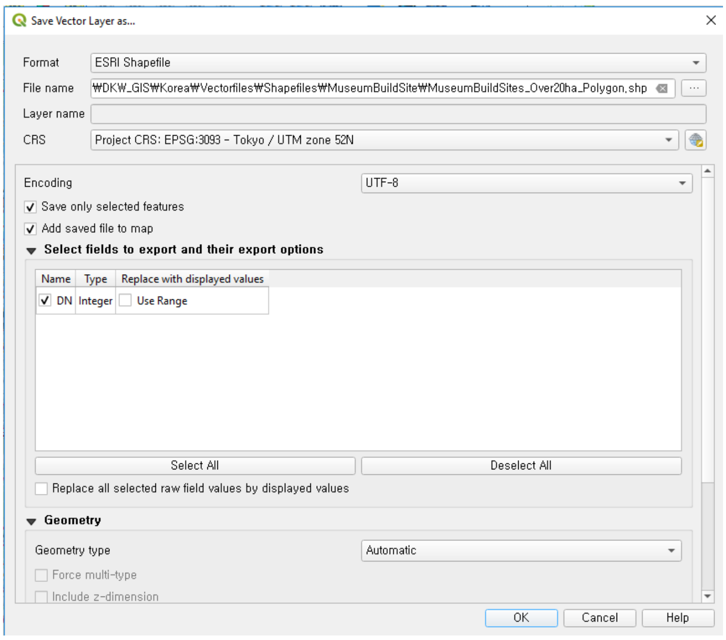


Fig. 9.26. Save the selected >20ha sites as a separate shapefile.

Often, developers approach cultural resource managers with multiple criteria for an infrastructure project. Using map algebra, you can provide a detailed set of parameters in a short amount of time that meet your client's needs (Figure 9.27).

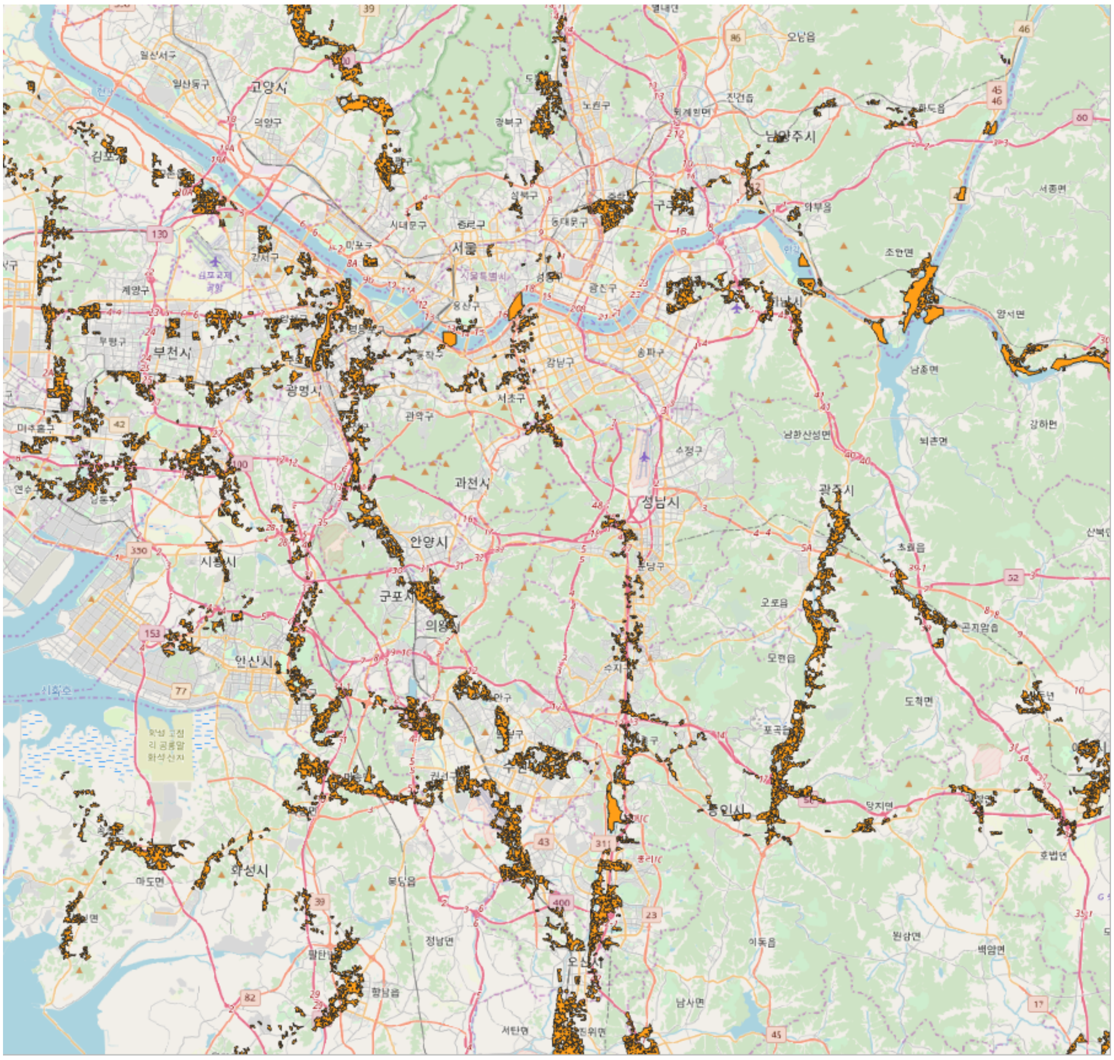


Fig. 9.27. The output of the >20 ha locations that satisfy the developer's parameters for a potential build location for the new museum with an OSM background.

9.3. EXERCISE

Your professor is convinced that the best place to look for sites that have evidence for early agriculture are in river valleys north of 37°N that occur within 10 km of known archaeological sites that date to between 2000 and 4000 cal. years BP. Your professor has a hypothesis: these sites are likely to occur on a soil that contains between 30% and 40% silt fraction at 5 cm depth. Make a prediction map that combines these correlates and identifies all of the potential locations where early agricultural sites might be found. Be sure your map has a north arrow, scale bar in kilometers, a title, comprehensible legend and UTM-marked zebra frame around the outer border.

For this exercise, you can access the waterbodies shapefile from diva-gis.org and set a buffer of 2 km. You can create the "valley" characteristic based on slope that is <5%. You can get the soil fraction data from soilgrids.org. Finally, use the real radiocarbon dated data to make your real prediction model.

Explain the results of your analysis in two or three paragraphs. Provide a final calculation table of the square kilometers of potential sites within each of the major Korean administrative districts (-도/서울) that fall north of 37°N. Where are the major locations of contiguous ground that meet the requirements? What are the basic statistical parameters of the target locations (mean, median, standard deviation, minimum, maximum area)? How do the target locations relate to known important archaeological sites? You will receive 5 points for completing the mapping portion of the assignment correctly and on time and thoroughly answering the questions I have asked above.

STUDY QUESTION (1): Conolly and Lake (2006:198-201) discuss different types of filters to decrease noise or sharpen the resolution of an image. In what ways could these filters improve the resolution of your image at different scales? In other words, how and why would you apply a high-pass filter and how and why would you apply a low-pass filter? (3 points)

STUDY QUESTION (2): Conolly and Lake (2006) discuss the how soil erosion models are constructed and for what purpose. In one paragraph, explain to me why archaeologists should construction of soil erosion models in the Korean context. What usefulness would they have in detected and preserving archaeological sites? (2 points)

Provide all of the results as *YOURNAME*_Ch9.docx and *YOURNAME*_Ch9.png and **upload it to the eTL by May 13 at 9:00 am.**

Chapter 10: Regions and Territories

10.1. Principles of Creating Bounded Regions

Humans like to create boundaries to conceptualize classes of space. Think about the border between North and South Korea. There are many social and historical reasons the border exists as it does, but there are very few natural or geographic reasons for the border to be located where it is. This is not always the case, though. Often social boundaries correspond to physical boundaries such as mountains, rivers and oceans. But, by making distinctions between places using boundaries, humans create zones in which the identify more with the people within their zone than outside of it. However, such boundaries are fluid and they can be difficult to identify in prehistoric contexts.

The viewshed is a space in which the landscape is bounded by the distance the human eye can see. Much has been written about this topic (see, for example, a recent article by Mark Gillings that reviews the topic from the site of Exmoor, England). In an opposite case study to Gillings's study, I published an article about how special stone structures were constructed on the tops of mountains to facilitate intervisibility between the locations. These examples use GIS statistical techniques to calculate and analyze viewsheds over large areas.

The implication with boundaries and viewsheds in archaeology is that there are connections forged between them by human communities. Distance traveled between locations can be measured in a path model in which geographic barriers serve as "friction" to repel human travel. However, it is often the case that roads and networks do not follow the easiest (most frictionless) path. For example, a Least Cost Path analysis in the Mimbres region of the southwest US found that obsidian trade networks changed over time and not necessarily following the easiest routes. We are left with cultural explanations for why such disparities arise in prehistory.

This week, you are going to learn basic boundary creation techniques and build a path model between sites. I will provide some bounded data for you (which you will create for yourself again next week). You will then look at the relationships between areas by creating a Least Cost Path model and evaluating population distributions by region according to that model.

10.2. Tessellating a Map

In this portion of the training, you are going to make some simple boundaries using tessellation. The first step is to add your TrainingData_UTM_52N file to the GIS (Figure 10.1). I have also added the KOR_adm0 shapefile for use later in the exercise.

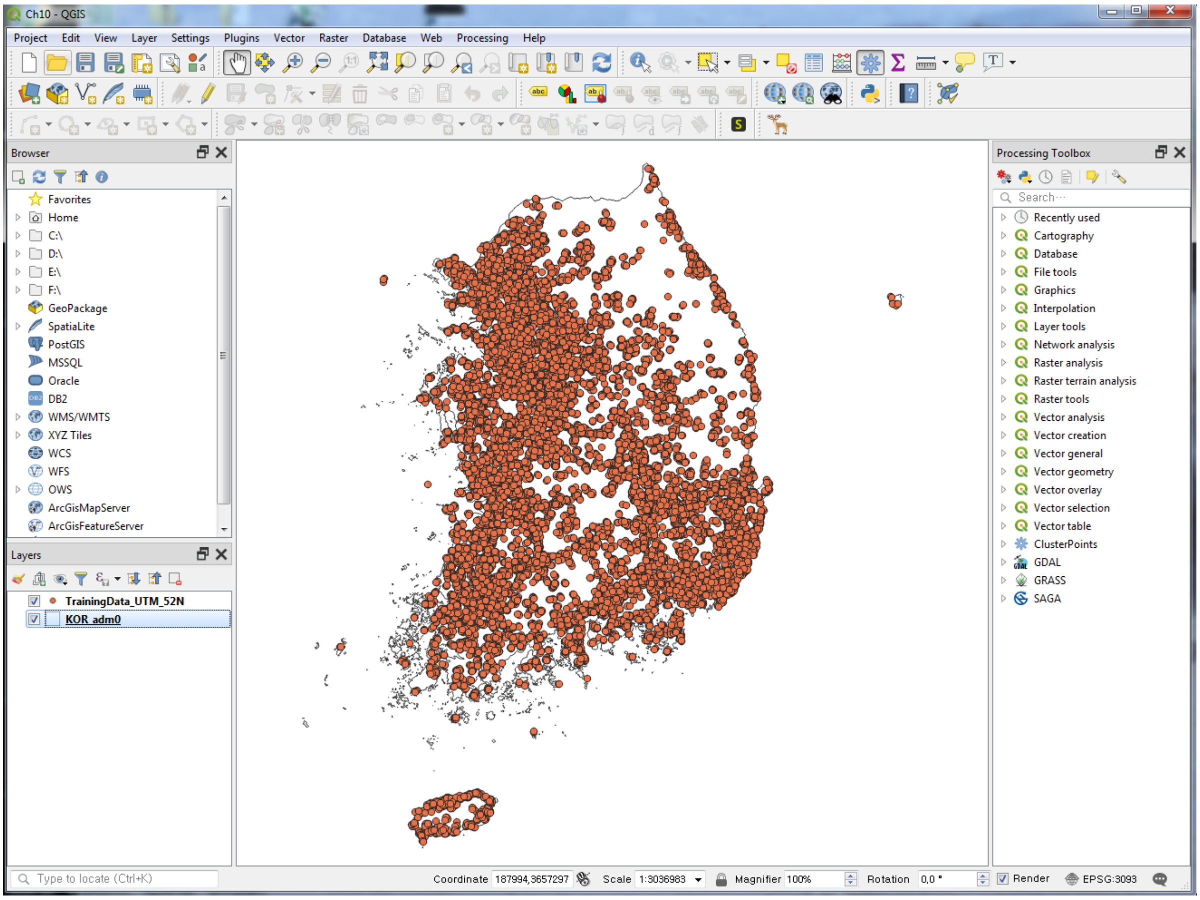


Fig 10.1. Add the training data to QGIS.

The next thing you need to do is select sites between 2500 and 5000 years BP using the Select features using an expression tool ( ) in the attribute table. If you don't remember how to use this tool, go back to Chapter 3 of the Training Manual. Once you have set it up correctly, your Map Window should look like Figure 10.2.

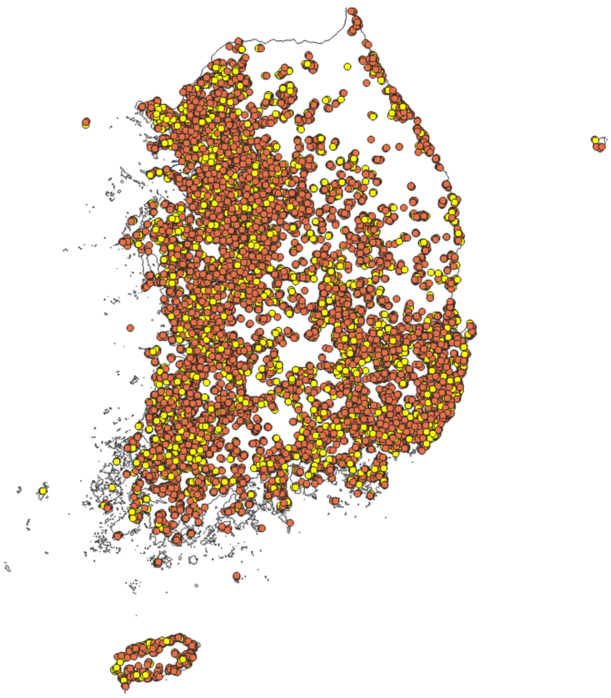


Fig. 10.2. Sites between 2500 and 5000 years BP selected using the selection tool in the training data.

For the next steps, it will be easier for you to create a separate shapefile from your 2500 to 5000 year BP data. You should already know how to do this as well (Export>Save Selected Features as…).

Next, you are going to tessellate your sites. You are going to create Voronoi polygons, which treat your points (sites) as centroids when it creates polygons. There are a couple of ways to find your tool. In Figure 10.3, I have used the GRASS tool (access it from the Processing Toolbox). The reason is that the GRASS tool allows you to set a boundary creation area (GRASS 7 region extent). In this example, I selected the extent from the canvas.

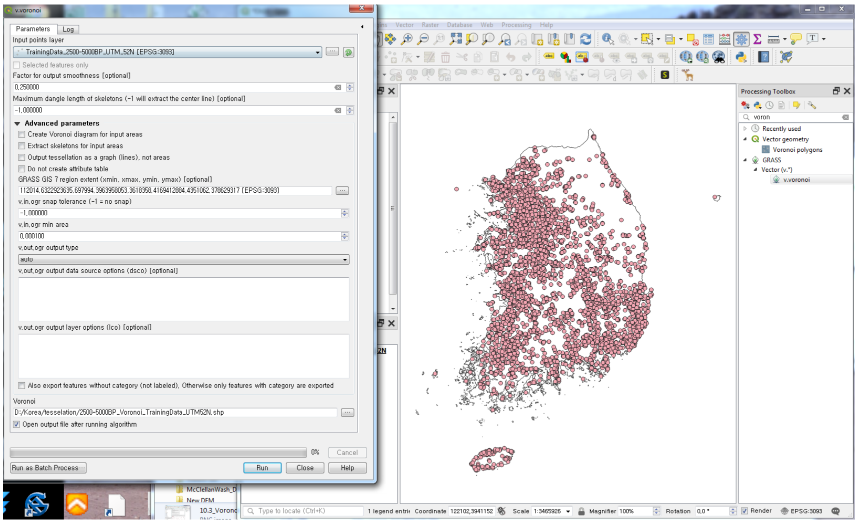


Fig. 10.3. Create Voronoi polygons. Be sure your buffer is set to 0% of the extent.

If you look closely at the polygons, you have an "edge effect." This means that the points found along the edges of the analytical frame have tessellated regions that extend into the ocean. You can fix this by clipping (Vector>Geoprocessing Tools>Clip…) the Voronoi polygon using the KOR_adm0 shapefile (Figure 10.4).

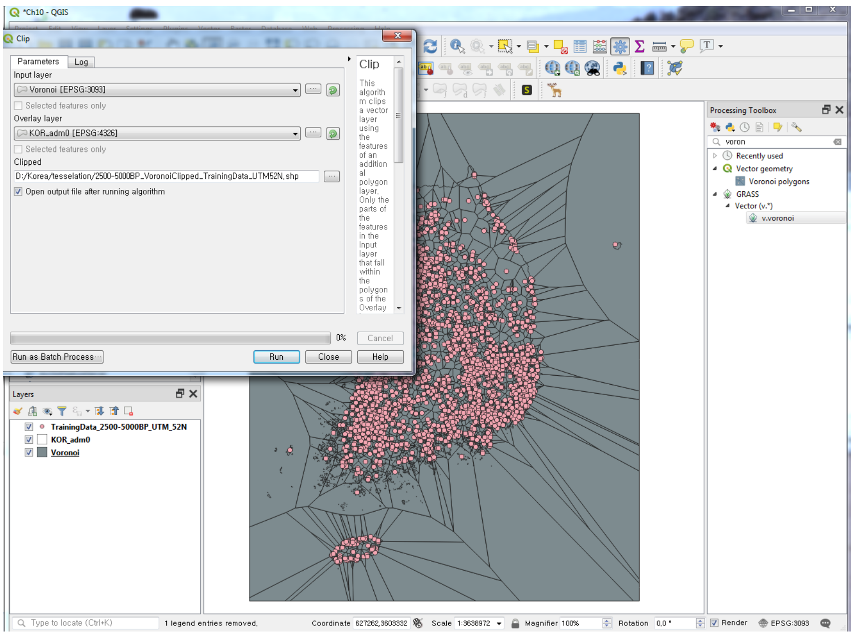


Fig. 10.4. Clip your Voronoi polygon using the KOR_adm0 shapefile.

When you have finished, you will have a map similar to what I have created in Figure 10.5.

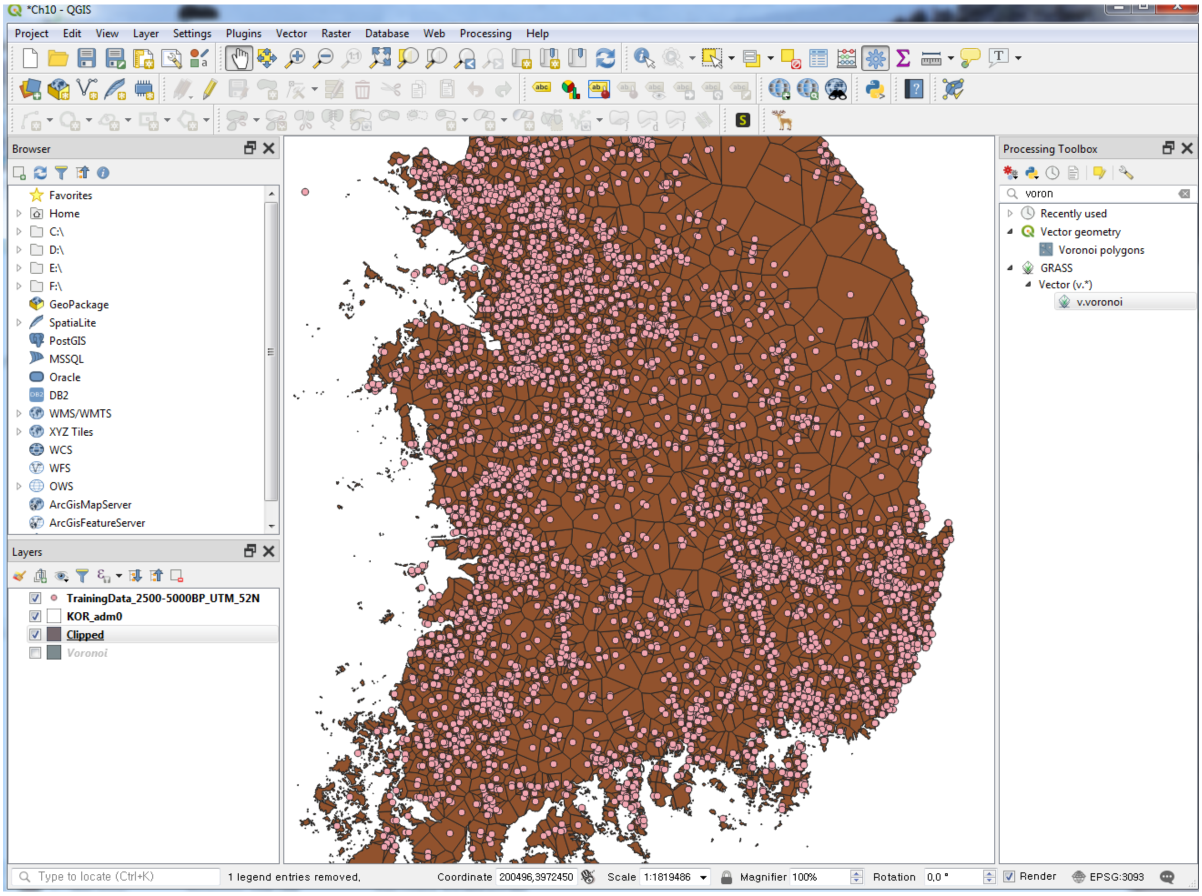


Fig. 10.5. Voronoi polygons of randomized sites between 2500 and 5000 years BP from the training data.

10.3. Advanced Boundary Creation

Next, we want to make clusters of our data. In QGIS 2.18, there is a plugin called "ClusterPoints" that can be downloaded and used to build clusters of points. In QGIS 3.4, there is a native operation in the Vector Analysis toolkit in the Processing Toolbox that does the same thing as ClusterPoints for what we need to do today. You should know that the 2.18 plugin has many more mathematical operations than the clustering tool in 3.4, so if you are keen to do more with clusters of points, you should try using the 2.18 plugin.[24]

Call up the K-Means Cluster tool and add your Training Data (2500-5000 years BP) file (Figure 10.6). Note that you can call up a selection of the data for clustering as well. We are going to make 100 clusters from the data and create a new field in the attribute table called "CLUSTER_ID." Note that by establishing 100 clusters, you are making an *a priori* assumption about the nature of your data. Once you have run the operation, you will see that there are 100 separate clusters (0-99) in your attribute table.

---

[24] You should also look at the DBSCAN Clustering tool in QGIS 3.4. This tool has a bit more power to control your cluster sizes.

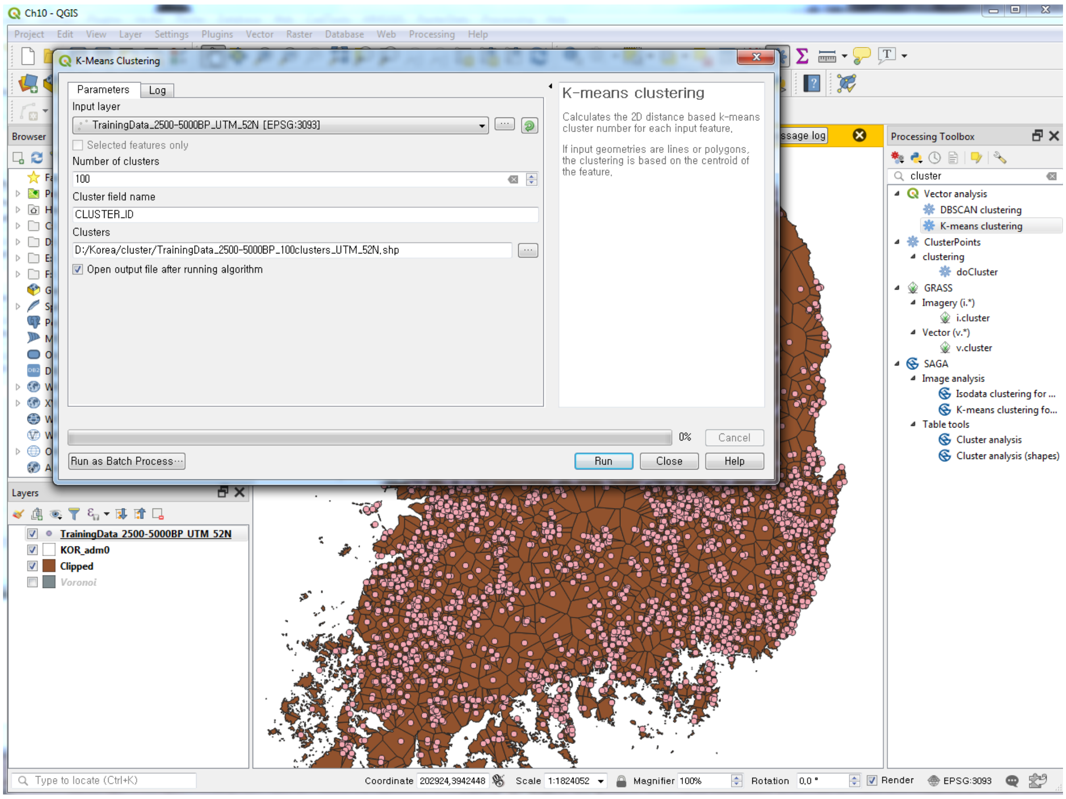


Fig. 10.6. Create 100 K-Means Clusters from the Training Data (2500-5000 BP).

Now, we want to create a new set of boundaries to encircle the clusters. Within the Vector Geometry toolkit in the Processing Toolbox, there is a tool called Minimum Bounding Geometry, which allows you to create polygons based on attributes (Figure 10.7). Select your shapefile that has the cluster values, and set the cluster field to CLUSTER_ID. You will create convex hull polygons, which encompass the points within a cluster analysis.

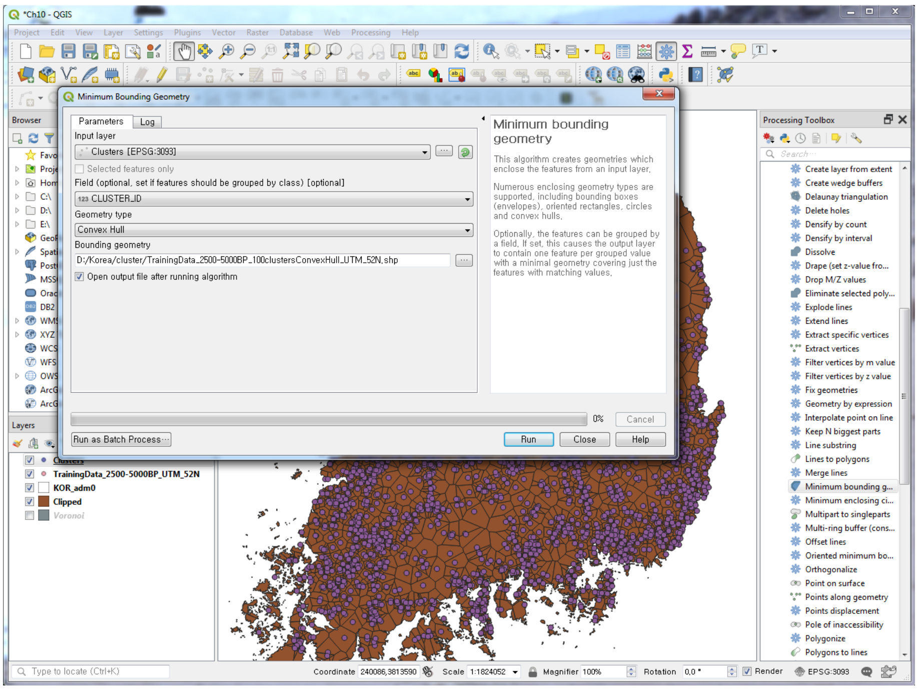


Fig. 10.7. Create convex hull polygons of your clusters.

After you run the operation, you will notice that some points were not polygonised. This is because the cluster was comprised of only one or two points, which cannot receive a convex hull.

You can now create a Join on your Voronoi polygons in the Properties (Figure 10.8). Add a Join by clicking the green plus arrow at the bottom of the Join window (). To make matters easy, use 

as both the Join and Target field. Unclick Create attribute index on join field. Click OK. When you open up your attribute table, you will now see that you have a CLUSTER_ID attribute.

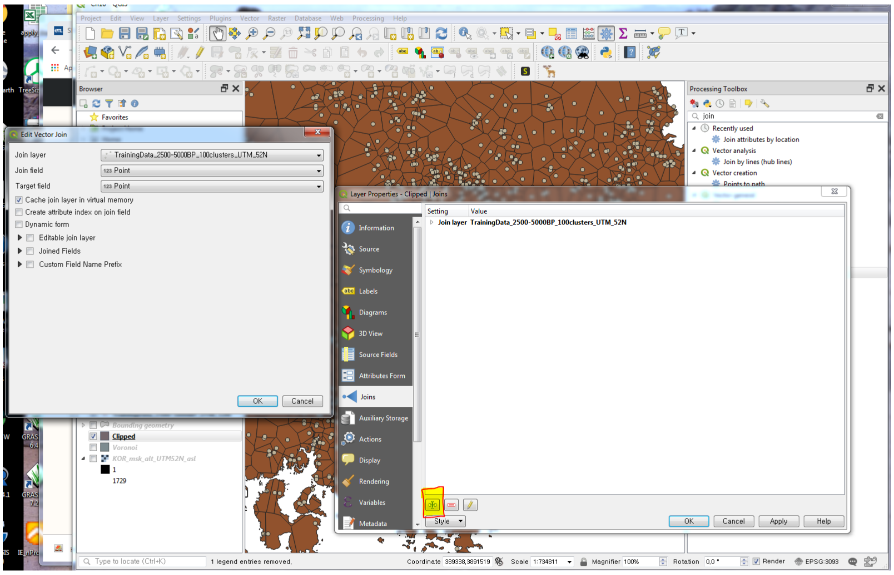


Fig. 10.8. Create a join.

Next, in the Processing Toolbox, select Vector Geometry>Dissolve. Add your Clipped layer, select CLUSTER_ID as the dissolve field and create a new shapefile (Figure 10.9). Click OK and Run the operation.

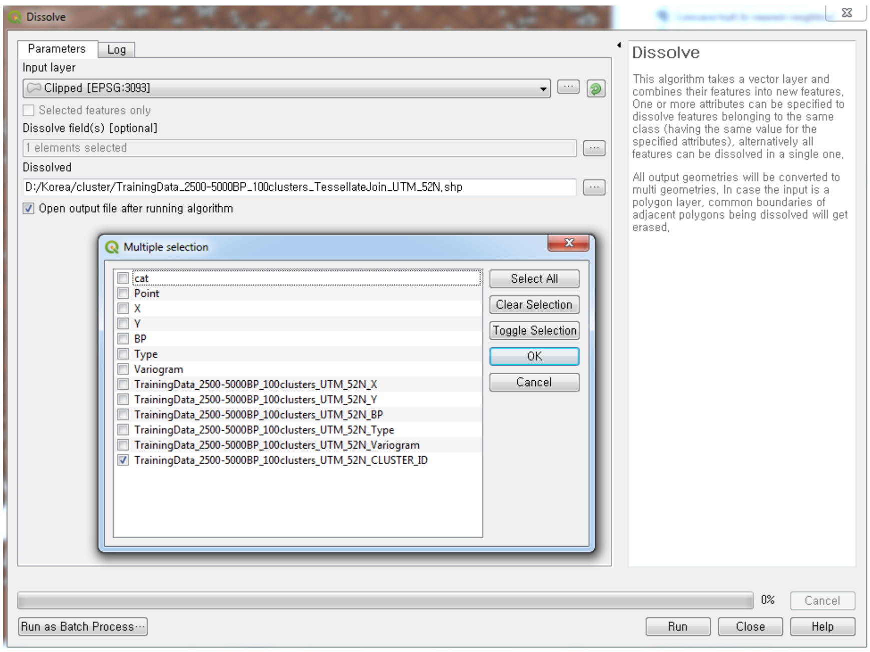


Fig. 10.9. Dissolve the clipped polygons based on the clusters to which they belong.

In your new Dissolved shapefile, you can remove fields in the Layer Properties with the editor enabled (Figure 10.10). Delete the extra fields you don't need and rename the Training5_ field as CLUSTER_ID. You have created a type of polity map based on clusters of points.

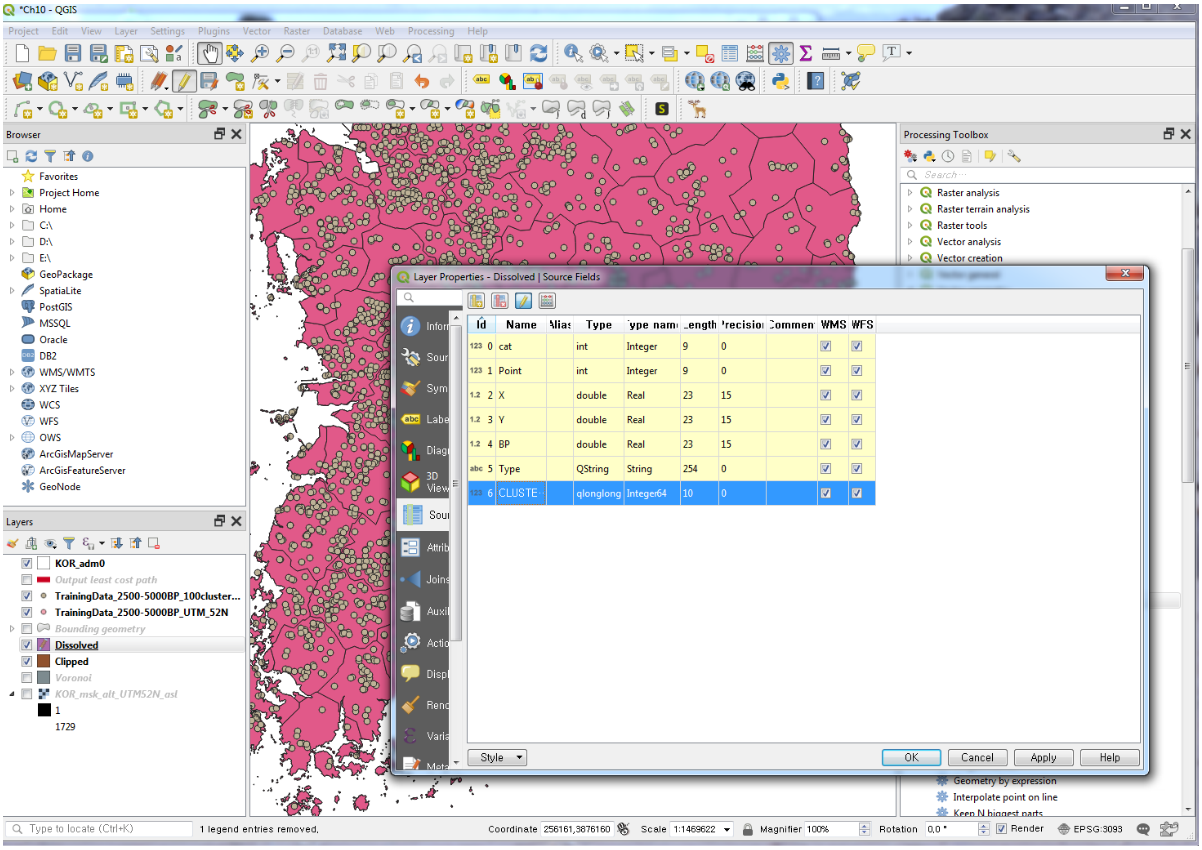


Fig. 10.10. Edit the attribute fields of your new tessellated map.

10.4. Least Cost Path

We are going to calculate a Least Cost Path (LCP) model between two of the clusters we created in Section 10.3. As introduced above, an LCP provides a friction model for overland distances between points.[25] LCP models are effective ways to understand optimal movement paths between locations. Deviations from the optimal path raise questions of why humans would choose to not take the optimal route.

The primary basis for such calculations is elevation, a file of which can be found on the KOR_msk_alt_UTM52N DEM I have placed in the eTL. A latitude/longitude version of this file is downloadable at DIVA-GIS, however this website has not been working well lately, so I have provided a reprojected version for you. However, there are many other ways of creating "costs" for an LCP. For example, areas with tree cover or river crossings are typically more difficult than grassfields to traverse. So, if you were to build an LCP for your final class project, I would want you to use the Raster Calculator

---

[25] For a relatively recent overview of the application of LCPs (via Least Cost Analyses), see White, D. (2015). The Basics of Least Cost Analysis for Archaeological Applications. *Advances in Archaeological Practice*, 3(4), 407-414. doi:10.7183/2326-3768.3.4.407.

(Chapter 9) to structure in a more realistic cost model than just using elevation. However, for this illustration, we will keep it simple.

When you load the DEM into your Layers panel, you will see that some of the elevations are below sea level. This will make it impossible to calculate a friction model, so you need to use the Raster Calculator to remove the negative values using the general formula: *@YOUR_RASTER*/(*@YOUR_RASTER* > 0).

Once you have done this, you need to select two points in your clusters between which you will calculate a LCP. It doesn't matter where you perform this operation in the training exercise, but I will select points 14912 and 12429 using the attribute tables of the two shapefiles you have loaded (TrainingData_2500-5000BP_UTM_52N and TrainingData_2500-5000BP_100clusters_UTM_52N). In the Processing Toolbox, type "cost" and you will see several options appear for a LCP. In this example, I am using the Least Cost Path tool (Figure 10.11).[26] The cost raster layer is the DEM, and I selected point 14912 from the 100 clusters shapefile and point 12429 from the Training Data shapefile. Save your path into a separate folder. I have created a thick red line shown in Figure 10.11 for the LCP generated.

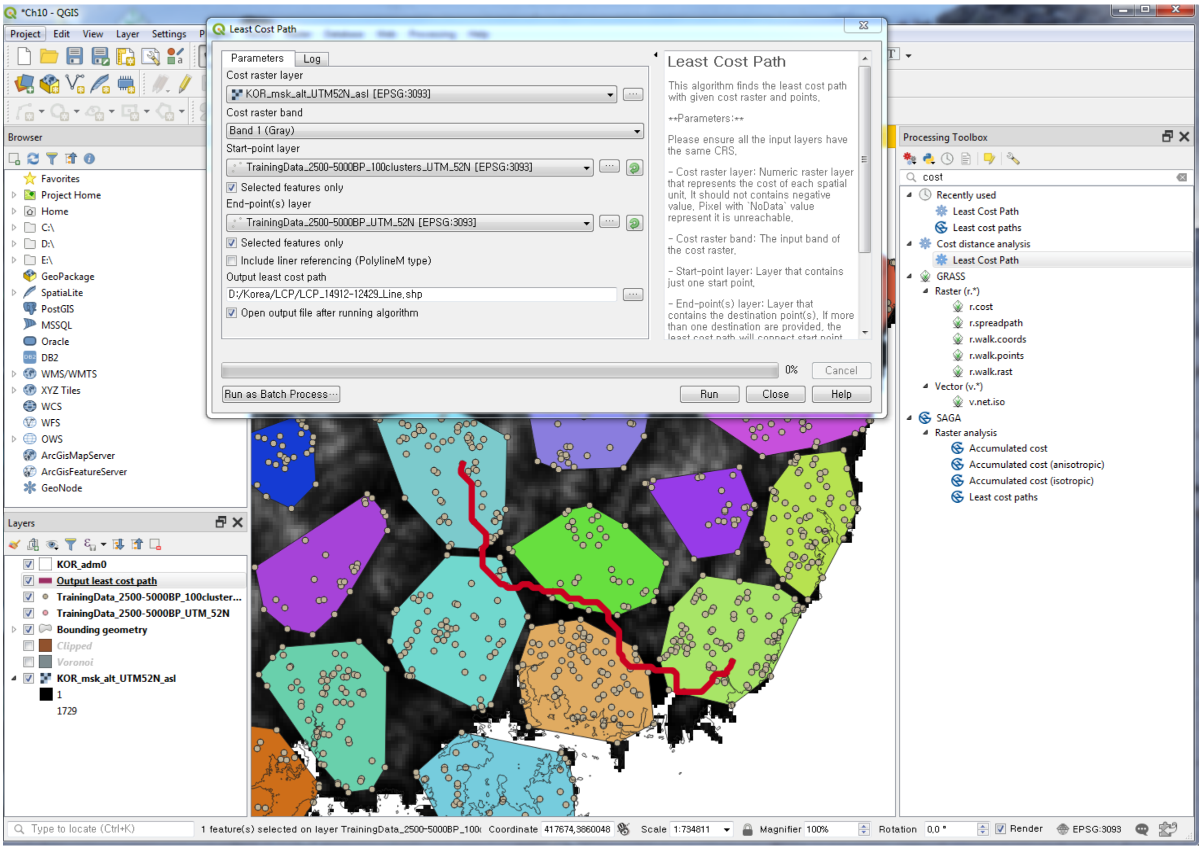


Fig. 10.11. Least Cost Path created using QGIS native tool. The path generated is a bold red line.

[26] The r.cost tool does the same essential function, except you can provide coordinates, which may be easier in certain circumstances than selecting points from separate shapefiles. You can also map out movement directionality in r.cost. Note when you add your cost raster that the

In the examples above, you have learned how to make different bounded regions either through tessellation or creating convex hulls around clusters. You have also learned how to calculate a LCP to connect specific centers to each other across bounded areas.

10.5. EXERCISE

I have placed a shapefile in the eTL called "CapitolSites." These include the ancient capitol cities of Seorabeol and Ungjin.

Create a LCP between these two locations using the DEM and a recalculated Strahler Order map. I have given you a raster file that includes stream order data, called Strahler Order (a modified version of what you will create in Chapter 11). The biggest waterways in Korea are coded as 7 and flat land is coded as 1 (the mouth of the Han or Geum river is a 7). Our assumption is that the bigger the stream, the more difficult it will be to cross. Your LCP will need to multiply the DEM values by the Strahler Order values, then square root the results. Review Chapter 9 of this training manual if you don't know how to use the Raster Calculator.

Next, buffer the LCP route by 25 km. Review Chapter 7 of this training manual if you do not remember how to create a buffer.

Use the real ^{14}C data to plot points occurring between 1600 and 1800 BP. Sum the ^{14}C dates by site (i.e., overlapping point geometries should not have multiple points or the tessellation you will do next will crash). *BIG HINT:* Chapter 7 of this training manual has the method for setting up a pivot table. Your rows need to be your coordinates and your values need to be counts of the "BP" field.

Create Voronoi polygons within the buffer zone (you will need to clip the polygon using the buffer). Create a graduated color symbology in which the polygons with fewer radiocarbon dates are a lighter color than polygons with more radiocarbon dates.

Make a map that shows the relationship between your Ungjin-to-Seorabeol LCP and the total number of radiocarbon ages within the Voronoi polygons tessellated to within 25 km of the LCP. Be sure your map has a north arrow, scale bar in kilometers, a title, comprehensible legend and UTM-marked zebra frame around the outer border. You will receive 5 points for completing the mapping portion of the assignment correctly and on time.

STUDY QUESTION (1): What are the sites with the most amount of radiocarbon dates? Assume radiocarbon dates are population proxies, discuss the relative importance of the two major population centers in the network relative to the archaeological record. Were these important archaeological sites between the two capitol cities? Or, were they something else not connected to a network? Cite literature (one or two references) to support your argument. (3 points)

STUDY QUESTION (2): Conolly and Lake (2006: Ch. 10) discuss different types of costs associated with modeling movement across landscapes (e.g., accumulated, isotropic and anisotropic costs). Of the various options presented, explain what type of model best fits the type of model you created for this assignment. Why is the LCP you made this type of model as opposed to any of the other available options? Select one of the models you did not create and briefly tell me how you would create a such a model out of the radiocarbon data you used in this assignment.

Provide all of the results as *YOURNAME*_Ch10.docx and *YOURNAME*_Ch10.png and **upload it to the eTL by May 20 at 9:00 am.**

Chapter 11: Network and Path Analyses

11.1. Principles of Network Analyses

Networks are the ways in which things connect to each other. Typically, networks are created using the shortest or easiest paths possible to make those connections. Archaeologists with lots of fieldwork experience are well aware of the principles **distance parity** whereby artifact (or feature) occurrences tend to be more similar to each other the closer in space and time they are found. This is because as distance (and time) increases the amount of work needed to connect the network also increases. This is also called a **cost surface**. Generally speaking, unless there is some good reason, people (and objects) will not violate the principles of distance parity. Network analysis concerns itself with determining the relative ease of forging connections between spatial situations.

In QGIS, network analyses are normally constructed using open source maps. ORS tools (formally [OSM tools]) is the most common plugin and it can make nice path models to establish a network. However, this tool (and most QGIS plugins that analyze networks) are really only designed to evaluate them within the context of modern infrastructure. There is a Networks plugin that can perform a network analysis in ways that would be useful for reconstructing prehistoric human trade networks, but it is not intuitive. There are also a native and GRASS network analysis tools in the Processing toolbox that determine the shortest paths between points, lines and polygons. These tools require you to have a priori knowledge of your network, which you may ascertain by using historical maps or archaeological data. The GRASS tool r.walk.rast is probably the best way to develop a priori knowledge in the absence of extraneous documentation. If you are interested in making a network path model for your final assignment in this class, let's discuss your ideas on an individual basis.

This week, we will concentrate on building a hydrology model in order to create basin **catchments**. We can think of these as "water flow networks," but they also tend to be important regions in which archaeological activity is encapsulated. This is because while the linear distance between two points in different catchments can be shorter than the distances between points within a catchment, the cost to travel between those more distant points is normally less. Travel friction increases as one traverses difficult terrain such as mountains, coastlines affected by strong tides or (as in the case of western Korea) tidal mudflats.

11.2. Make a Watershed Map

This week, we are going start simply (quick, but not very precise) and work up to a more complex (laborious, but very precise) workflow for creating watersheds. QGIS has a nice GRASS tool to simply construct watersheds called r.watershed. This tool gives you a powerful way to model catchments, which are typically the primary focus of prehistoric social and subsistence life. For the purposes of learning, we will resample the data using GDAL and SAGA tools in Madeira (QGIS 3.4) this week, so we will do all of the analysis in that version of QGIS.

After opening Madeira, add the KOR_msk_alt (from www.diva-gis.org) DEM to your Layers panel. For this portion of the exercise, it is fine to leave the coordinate system in EPSG:4326 (Lat/long, WGS84). In the Processing Toolbox dialogue box, type r.watershed and select the tool. Accept all of the default outputs (including the creation of temporary files) with one exception: for the Minimum size of exterior watershed basin, select 2000 cells (Figure 11.1).

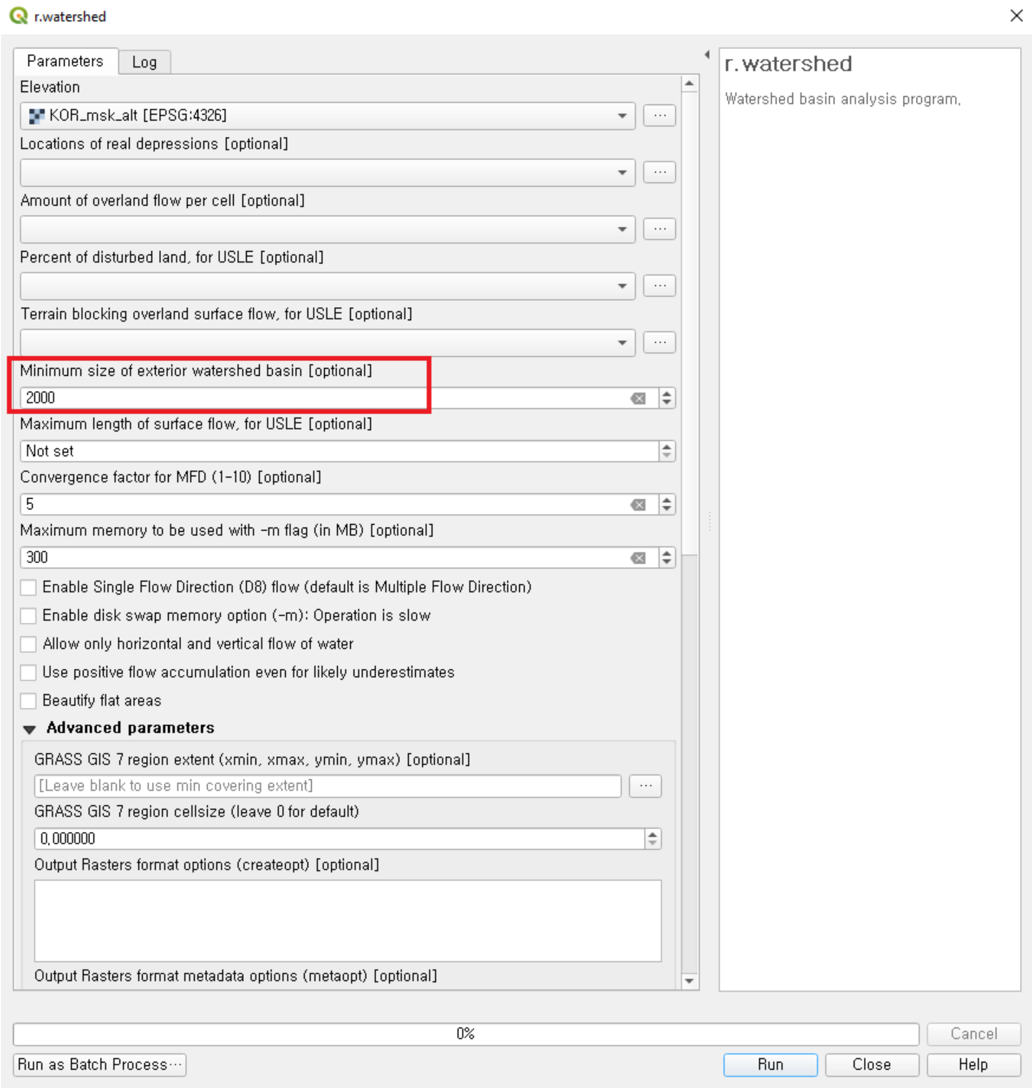


Fig. 11.1. Set the r.watershed output parameters.

The calculations are fast and you will see a bunch of files in your Layers panel (Figure 11.2). *Stream segments* are useful rasters that tend to follow the trajectory of the major rivers (you can test the accuracy of this model in Section 11.3). The *topographic index* is a measure of roughness with the lighter areas being less rough than the darker areas. *Stream power index* can be useful in some regions, but in Korea, it isn't so useful. If you adjust the symbology of this layer you can see that it identifies the Han, Geum and Nakdong rivers pretty well, but not much else. But, if you are interested in just identifying major water bodies,

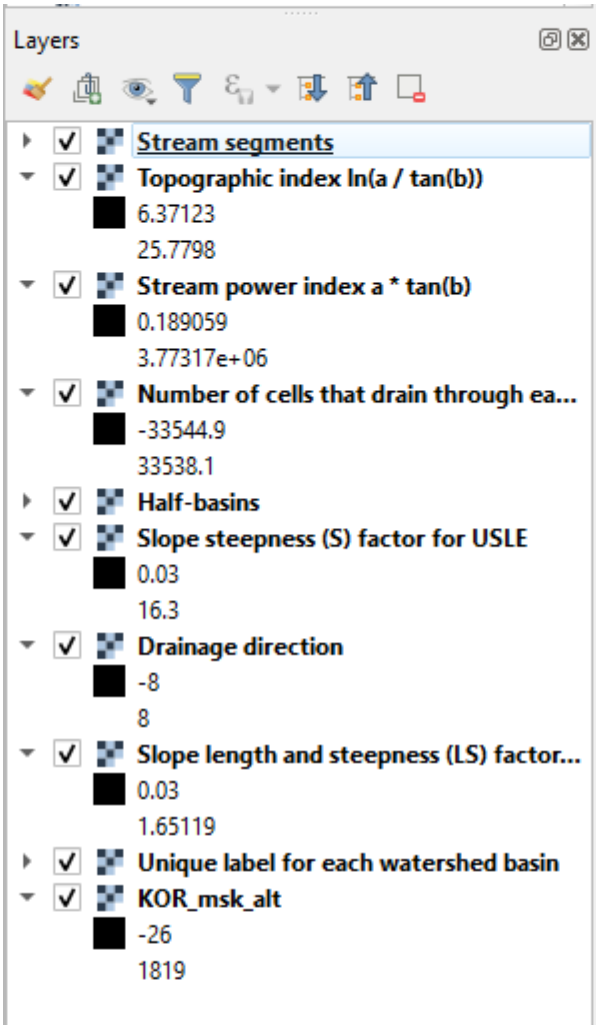


Fig. 11.2. Output of r.watershed GRASS tool.

this is an easy way to do that (and in Chapter 9, you learned how to use the Raster Calculator to sift the data you want from which you can create a vector from the raster). The *number of cells that drain through each cell* gives you more drainageway construction options. You can interpret these as measures of net inflow or net outflow. *Half basins* and *unique label for each watershed basin* are different sensitivities to the model, with half basins being at least 1000 cells (in our present model) and full basins are 2000 cells. Save the “full basin” shapefile as a vector somewhere in your Rasterfiles folder for later (save it as UTM 52N). *Slope steepness (S) factor for USLE* (universal soil loss equation) is a gradient factor used to calculate the potential for the susceptibility of an area to have soil loss measured in metric tonnes/hectare/year. For some reason (unknown to me). there are only two values calculated from the KOR_msk_alt DEM, which are 0.03 and 16.3406. *Slope length and steepness (LS) factor for USLE* is similar to the *S* factor, except that the length of the gradient is also factored. *Drainage direction* is complicated, but negative values indicate that the surface runoff is leaving the map and values 1 through 8 correspond to the cardinal directions (1 is north, 4 is west, 6 is south, 8 is east). More details for these measures can be located by following this link.

Rerun the analysis this time but this time change the *minimum size of exterior watershed basin* to 500. All of the outputs will be the same except for the *Unique label for each watershed basin* output. If you compare the two outputs (Figure 11.3), you will notice that the 500-cell output creates more basins. Now run it for 5000 and you will notice significantly larger drainage basins were constructed.

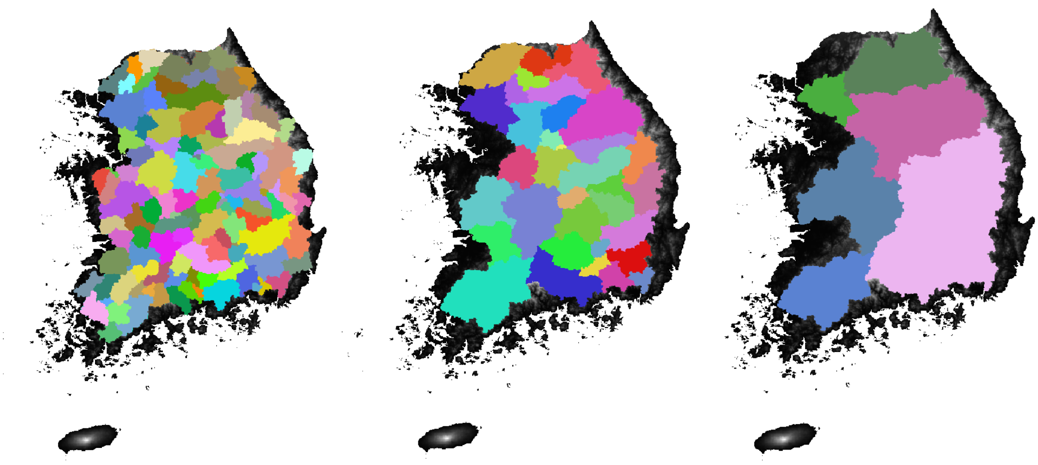


Fig. 11.3. Outputs of r.watershed basins for 500 cells (left), 2000 cells (center) and 5000 cells (right). Increasing the sensitivity of the basin model increases the precision of the catchment boundaries, but complicates interpretations by increasing the perceived noise.

You can notice some of the limitations in these models, though. There are missing watersheds in this model. You can now tell that if we wanted to get these watersheds adjacent to the coast, we could easily do so by decreasing the number of cells required to make a basin, but this would mean that we would have so many catchment basins in the middle of the country that it would be difficult to interpret the results. In other words, this is a good first start, but we need to learn more precise ways to construct hydrological maps.

11.3. Making a Detailed Drainage Basin Map

We have already downloaded a DEM and reprojected it in Chapter 7. However, we have data at 1-arc second resolution, which is too detailed for the construction of a useful countrywide basin map. Therefore, we will either need to resample the data or download coarser data from USGS. Be sure your coordinate system is set to EPSG:3093 (UTM 52N).

First, for orientation purposes, open the QMS plugin and add an OSM map (Figure 11.4).

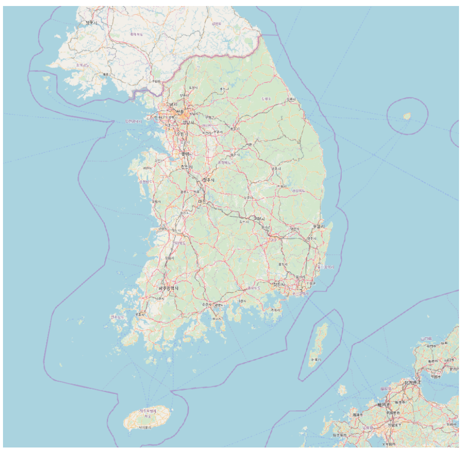


Fig. 11.4. From the QMS plugin, add an OSM map.

You will now downsample the spatial resolution of the DEM to 100 m. This is most easily done through the warp GDAL tool located in the Raster tools (Figure 11.5). Save the new resampled image as a tif in your Rasterfiles folder.

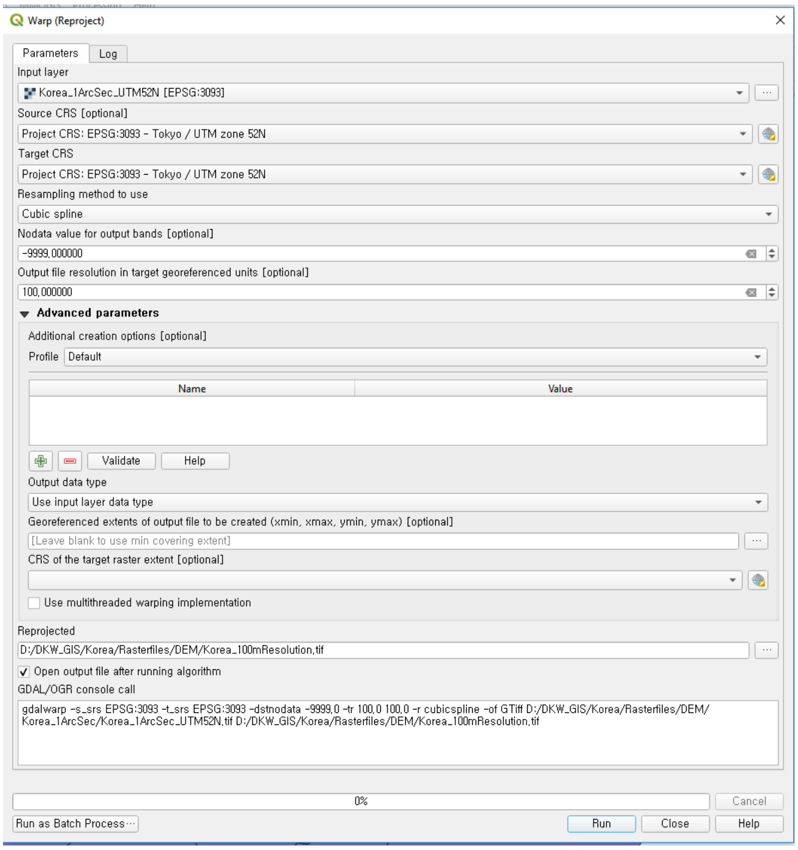


Fig. 11.5. The 1-arc second DEM needs to be resampled in GDAL to a more manageable size.

DEMs invariably have “sinks” in them. You can think of a sink as a small lake or pond in which a small number of cells drain into them, but they readily fill and spill into major drainageways. Think of 자하연 on campus…that is a sink that would need to be filled before constructing a drainage basin map. To do this, use the SAGA “Fill sinks (wang & lui xxl)” tool (Figure 11.6).

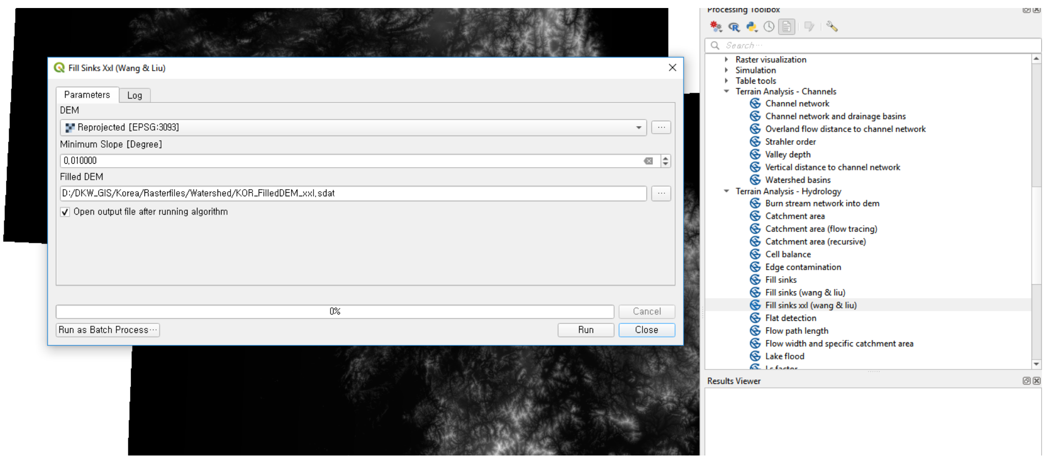


Fig. 11.6. Fill sinks (wang & lui xxl) of reprojected file.

The next step is to calculate the Strahler Order of drainages. We did this in the r.watershed tool, but the SAGA tool is a bit more sensitive. The Strahler Order is the stream order with 10 being the largest order drainageway (like the Han or Geum River) and 1 being a small stream similar to what you see outside of Building 5. You will use your filled DEM as the basis of the calculation of the Strahler Order (Figure 11.7).

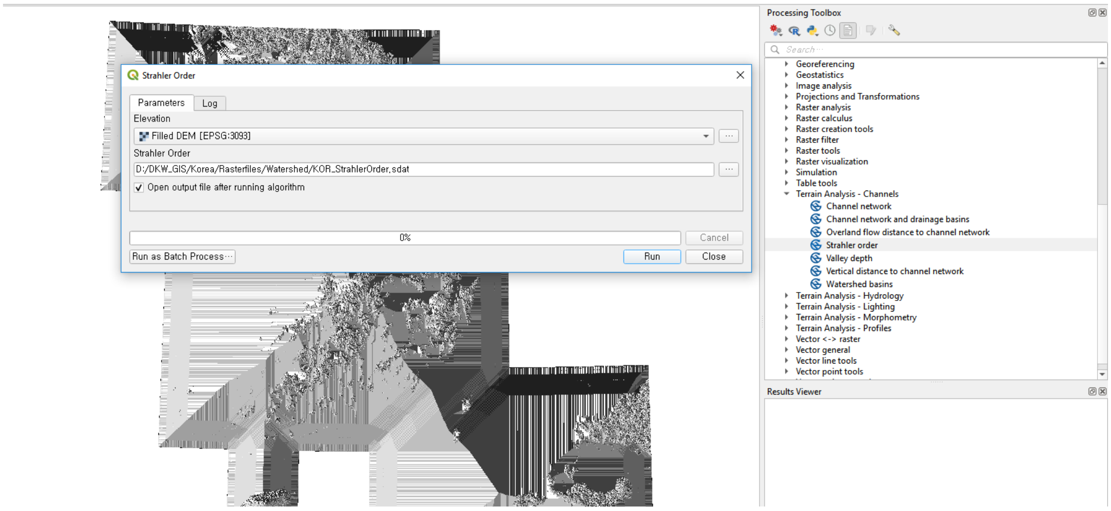
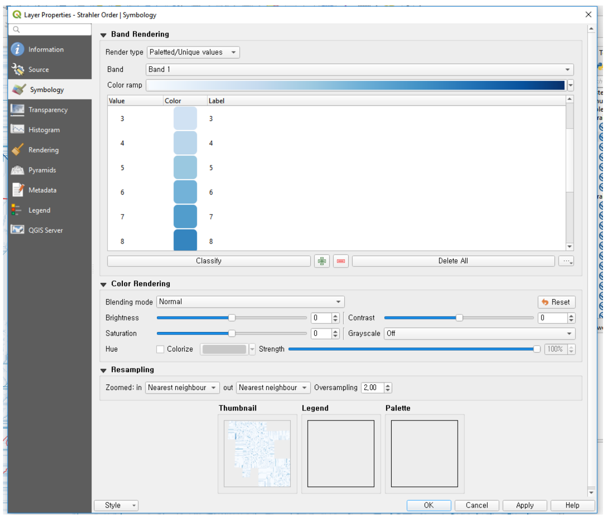


Fig. 11.7. Calculate the Strahler Order from the filled DEM.

In the next step, you will change the symbology of the created Strahler Order map so that you can see how many high-order drainageways there are. Code the colors so that the highest order streams appear in dark. In a normal situation if you were working independently, you would use the color scheme to make a judgment on which Strahler Order of streams were the important ones to construct the catchment model.



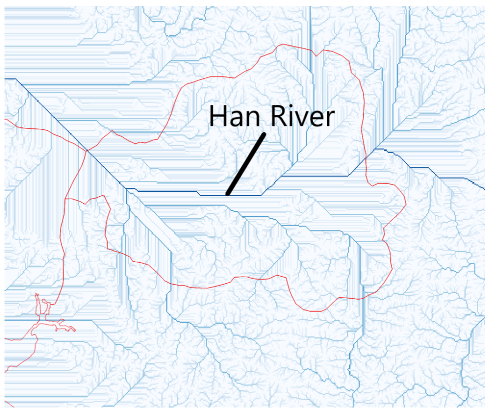


Fig. 11.8a. (left) Reclassify the symbology of the Strahler order map. Fig. 11.8b. (right) Reclassified symbology results

Based on the color scheme and the numbers of high-order drainageways in your model, we can set the threshold to Strahler Order 8. That means that streams <8 will not be calculated as independent drainages while Strahler Order ≥8 is part of the model. Based on the results from the r.watershed, if we were making a basin map on the east coast of Korea, we would need to use a much smaller order Strahler value (like 5 or something) in order to calculate a basin.

We will use the Raster Calculator to make a binary raster of ≥8 and <8 (Figure 11.9). The expression will be `"Strahler Order@1" >= 8`. Save the file as a new raster (tif) file.

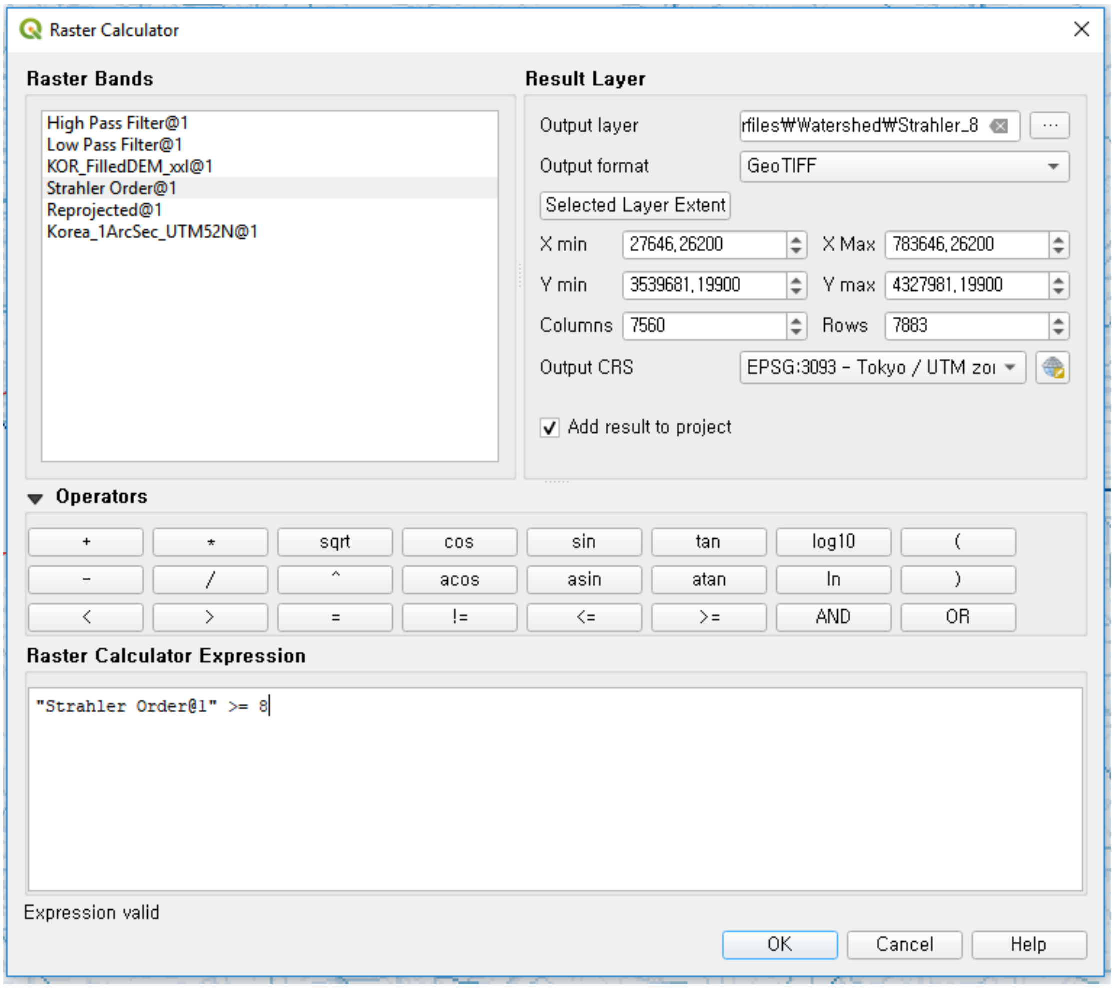


Fig. 11.9. Recalculate the raster into binaries based on Strahler values ≥8.

In order to be able to see the raster map, we will perform a little visual trick in the Layer Properties by making 0 values as "no data" in the transparency options (which will give it a transparent appearance) and classify the 1 values as blue in the symbology options (Figure 11.10). You can now notice that there is a small problem with the model in which the Han River drainage appears to primarily empty into the West Sea near Incheon in the Siheung area, but from the OMS view, it is clear that is not the major outlet. It is also clear that the location of the channel is difficult to model properly as the channel is a few km west of its real location. This is due to the heavy channelization of the Han River and flat estuary with low relief. There isn't much we can do to change this situation. In low-relief settings that have been heavily modified, establishing clear boundaries between basins is an impossible task.

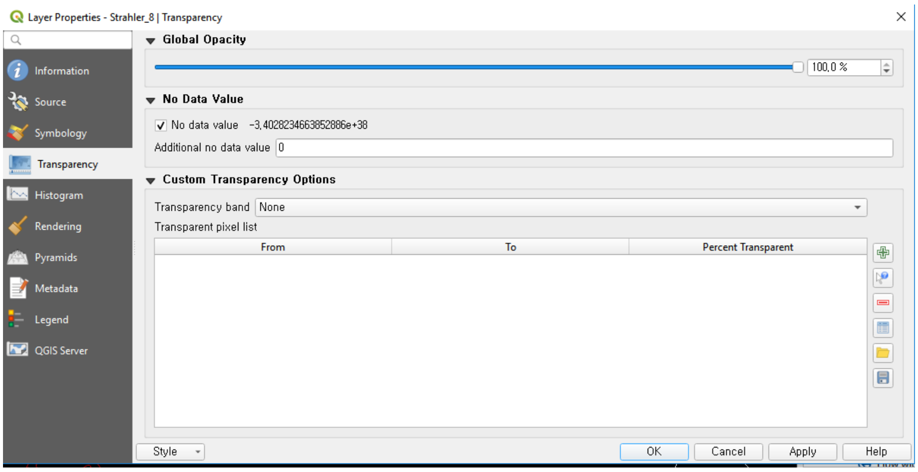


Fig. 11.10a. Make 0 a "no data" operator in the reclassified Strahler order map.

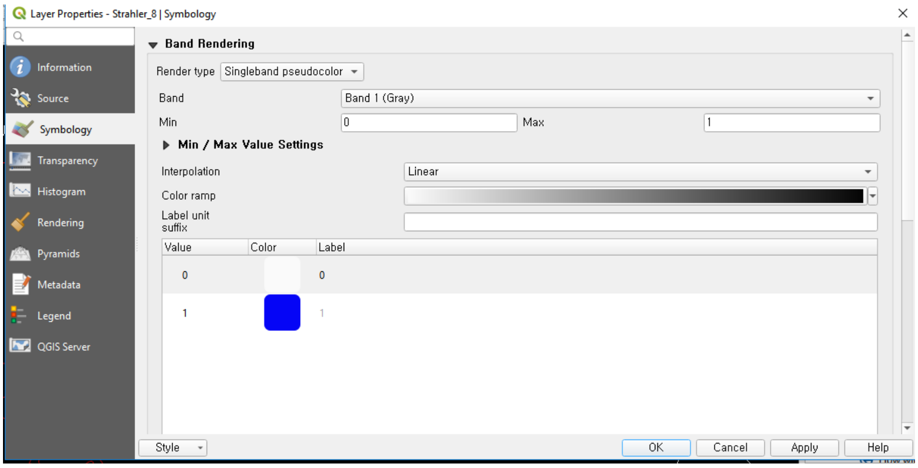


Fig. 11.10b. Color the ≥8 Strahler value in blue.

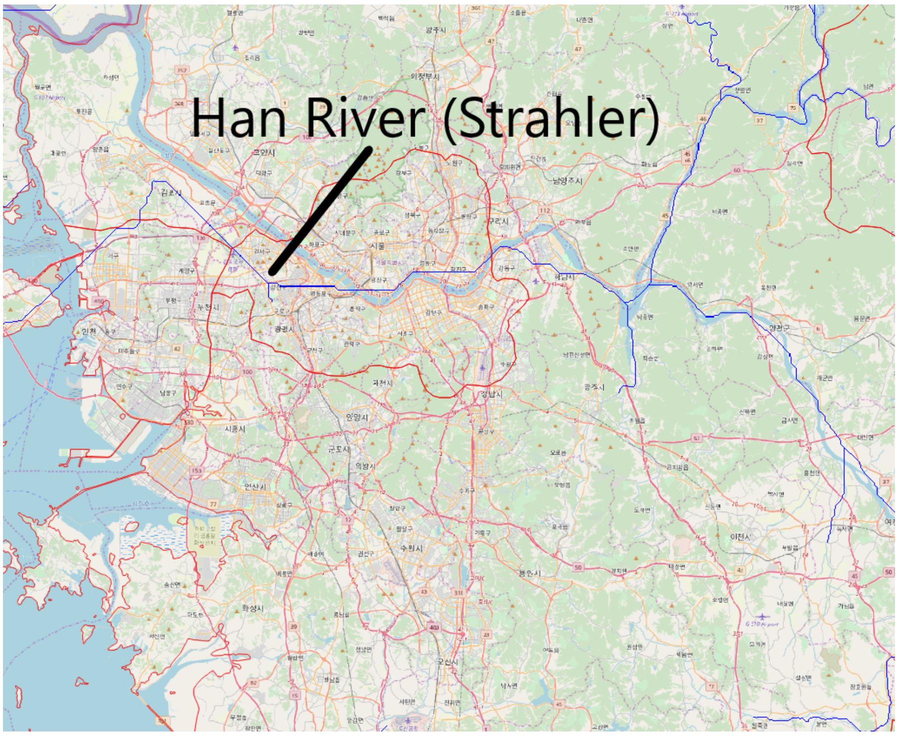


Fig. 11.10c. Output of the new symbology.

Next, use the “Channel Network and Drainage Basins” SAGA tool. Only select the “Channels” and “Basins” options, and save the files as shapefiles in your Vectorfiles folder (Figure 11.11).

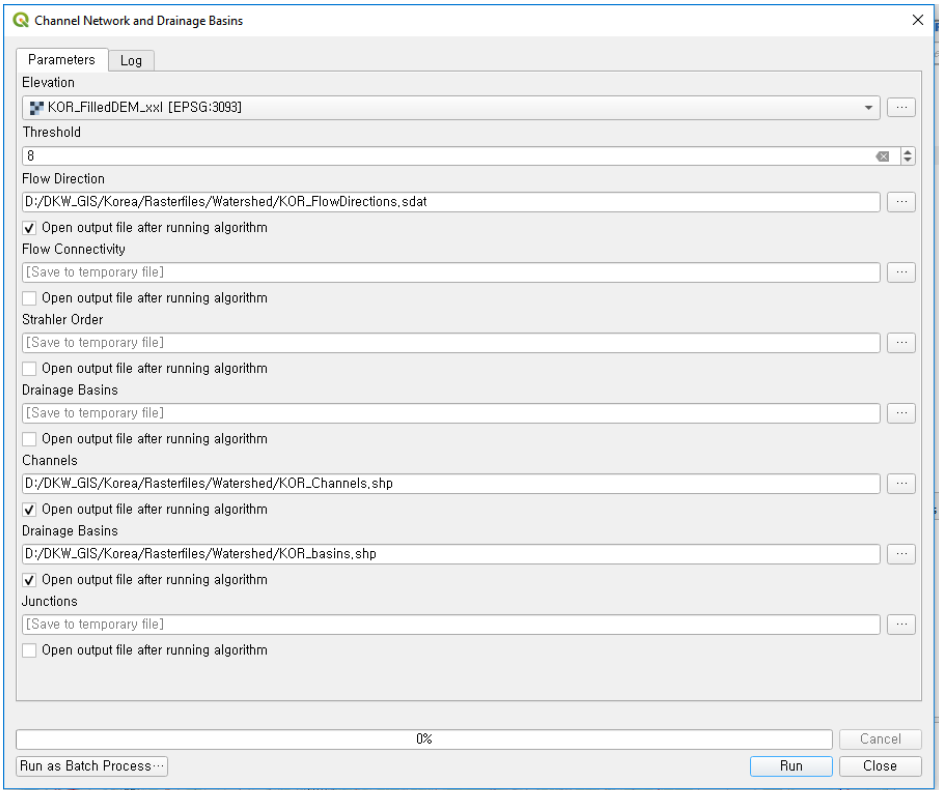


Fig. 11.11. Calculate the drainage basins and save as a vector file.

The next operation is to use the Coordinate Capture plugin tool to identify the location of the “pour point,” which is the location where the channel meets the sea. Capture the coordinate using the tool and copy the coordinates to the clipboard (Figure 11.12).

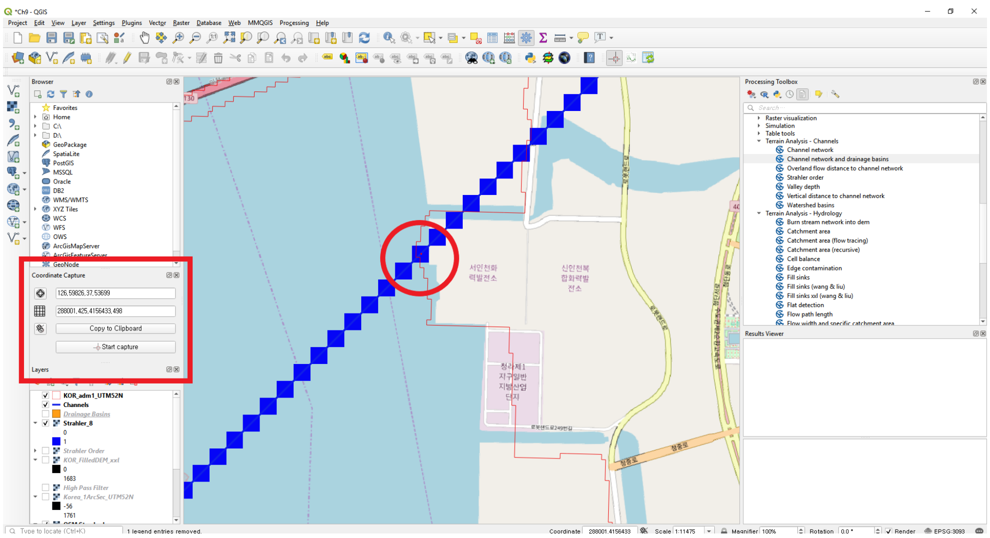


Fig. 11.12. Use the coordinate capture tool to identify the “pour point” of your drainage channel.

Calculate the Upslope Area catchment using the SAGA tool of the same name. Paste the coordinate into the target X and Y areas (Figure 11.13). This can be done as a bulk process if you are attempting to calculate multiple drainageways simultaneously. The resulting sdat raster will have values of 0 and 100.

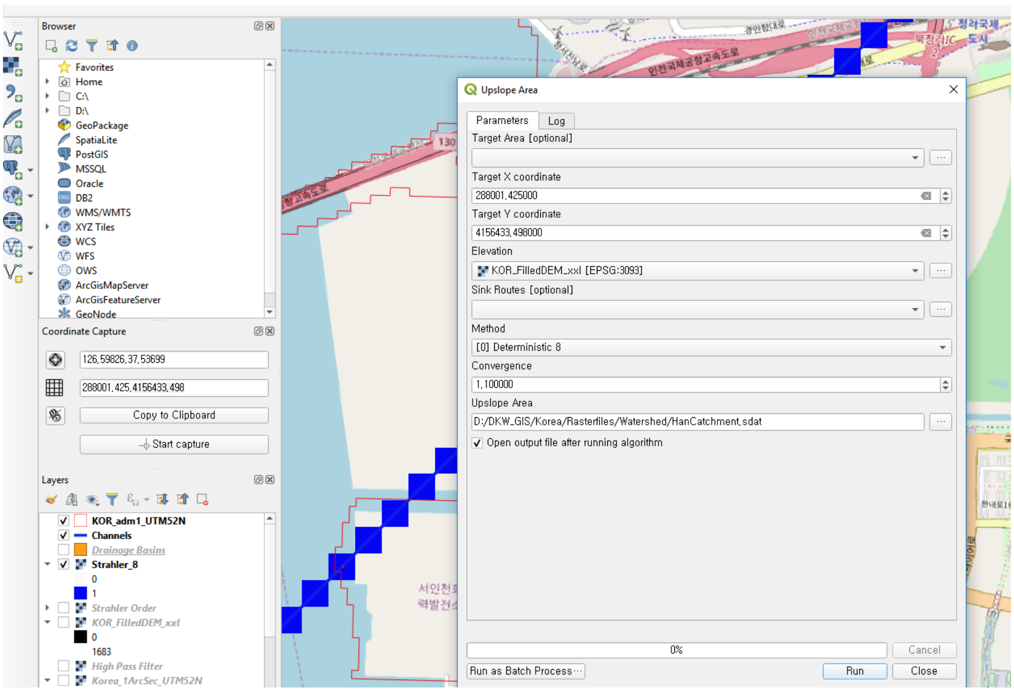


Fig. 11.13. Calculate the upslope area relative to your pour point.

The resulting raster will look something similar to Figure 11.14. The red lines are the KOR_adm1 shapfiles (just for your reference).

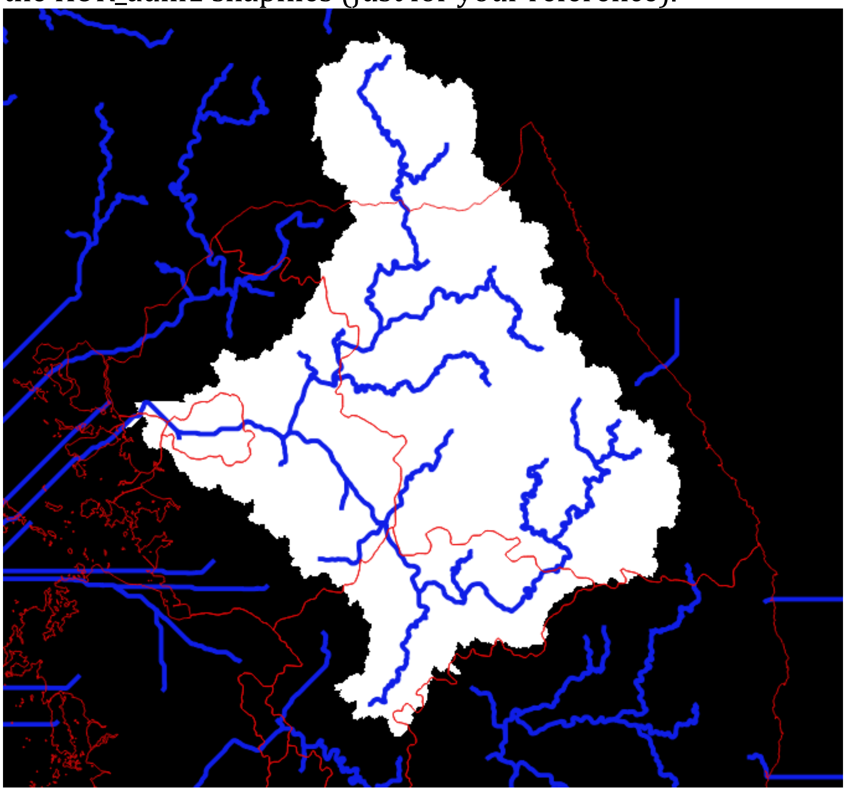


Fig. 11.14. The resulting Han River catchment raster

Convert the raster into a vector using Raster>Conversion>Polygonize (Raster to Vector)... (Figure 11.15).

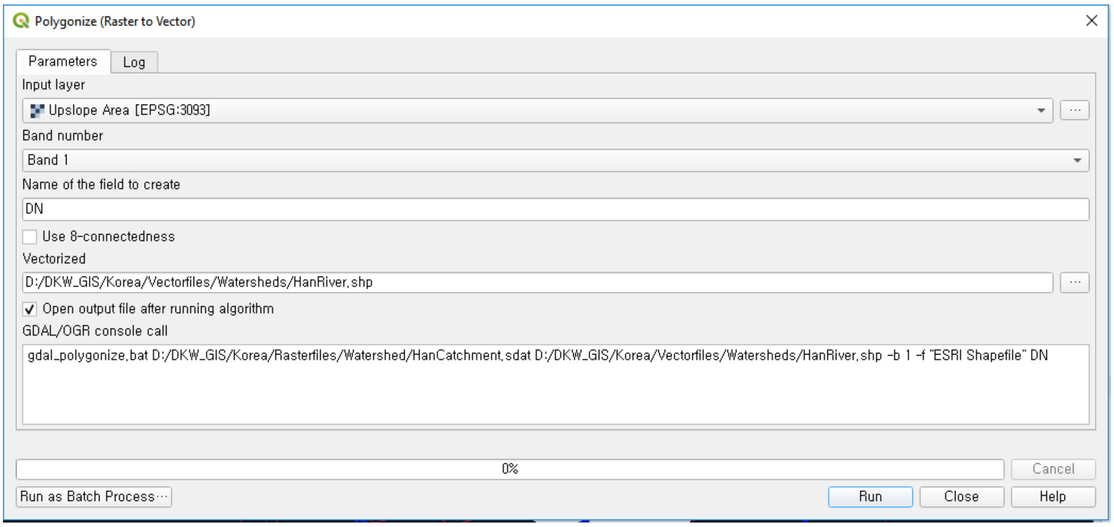


Fig. 11.15. Convert the catchment raster into a vector.

Once you have the shapefile created, then delete the 0 value from the attribute table of the vector file (Figure 11.16).

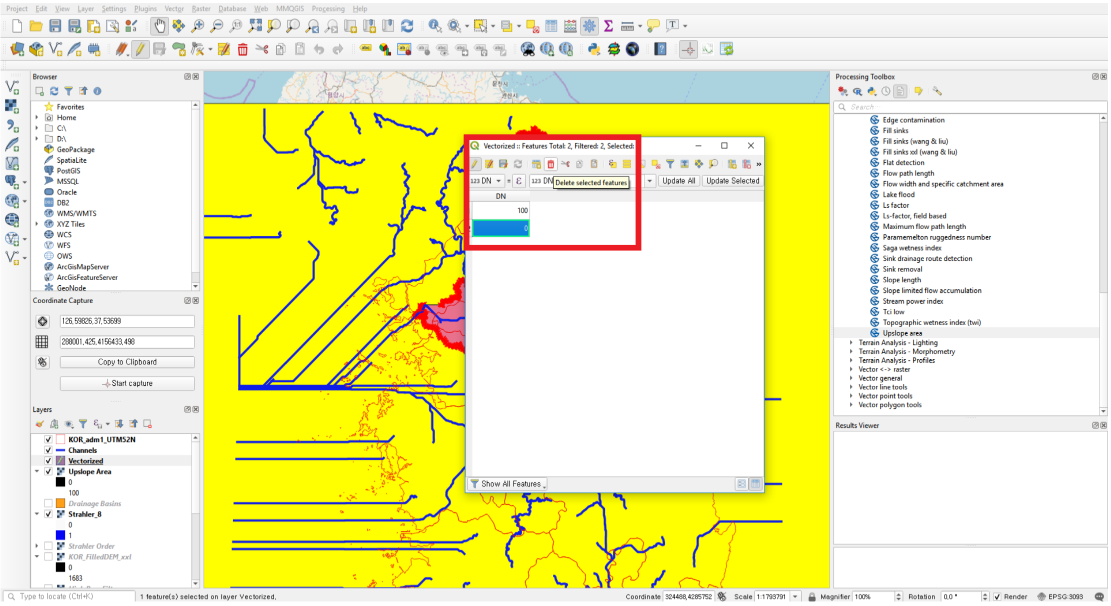


Fig. 11.16. Delete the 0 attribute from the attribute table.

You will notice that there are a lot of long lines extending out into the ocean. This is an artifact of your DEM and they can be deleted relatively simply. First, using one of the vector clipping tools (GDAL, SAGA or native algorithm), clip the newly created streamway vector by the KOR_adm shapefile (Figure 11.17). This will get rid of most of the lines that are not streamways (Figure 11.18).

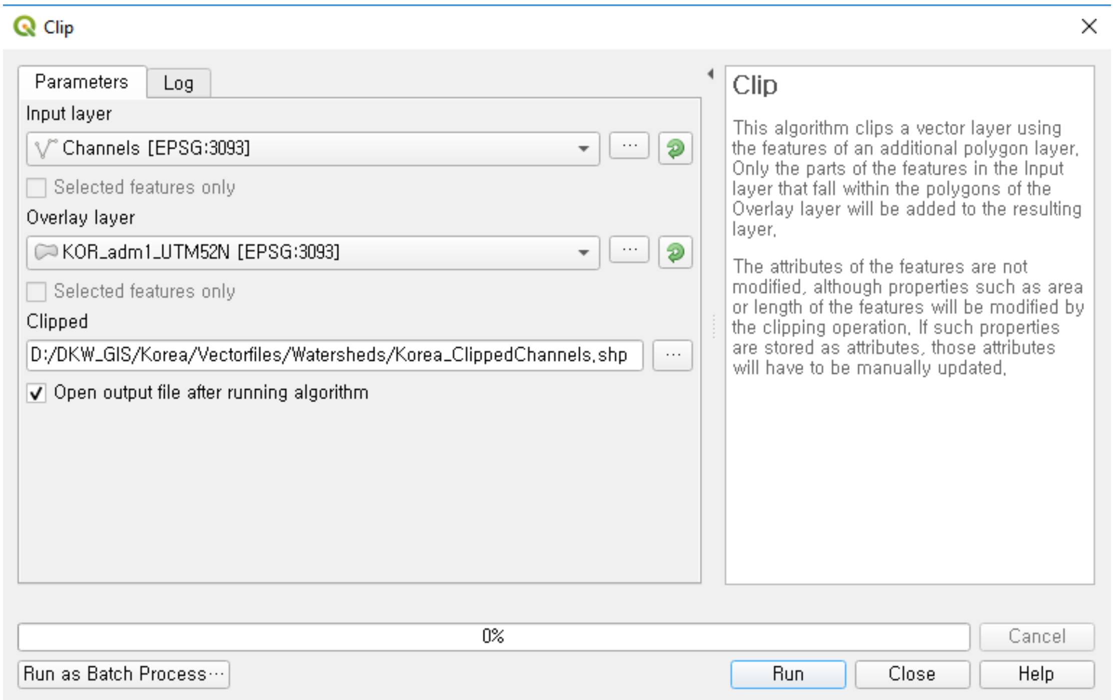


Fig. 11.17. Clip the stream vectors using the KOR_adm shapefile.

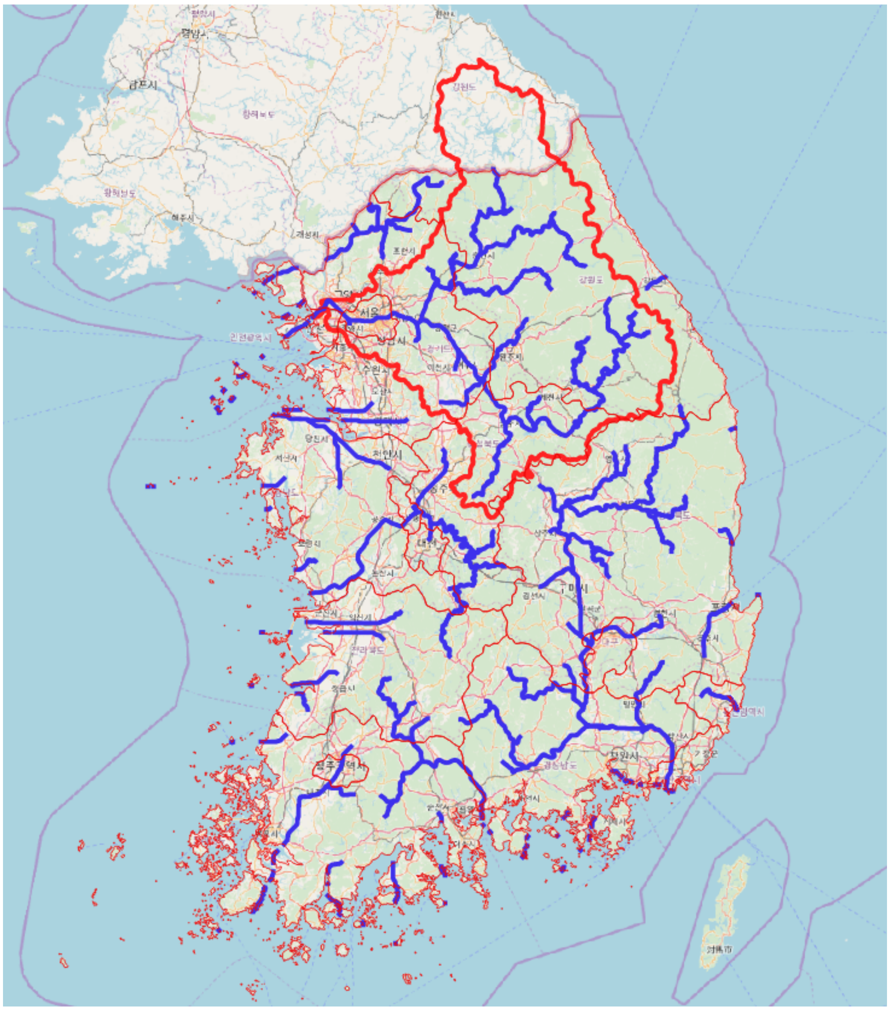


Fig. 11.18. The resulting stream map of South Korea

Next, you will clean up the bad vectors that run across the islands. Enable the Toggle Editing button. Select Vertex Tool (Current Layer) (Figure 11.19). Left click and drag over the area you want to delete and then click the delete button. You will have to repeat this operation several times to get rid of all of the vertexes that are inaccurate.

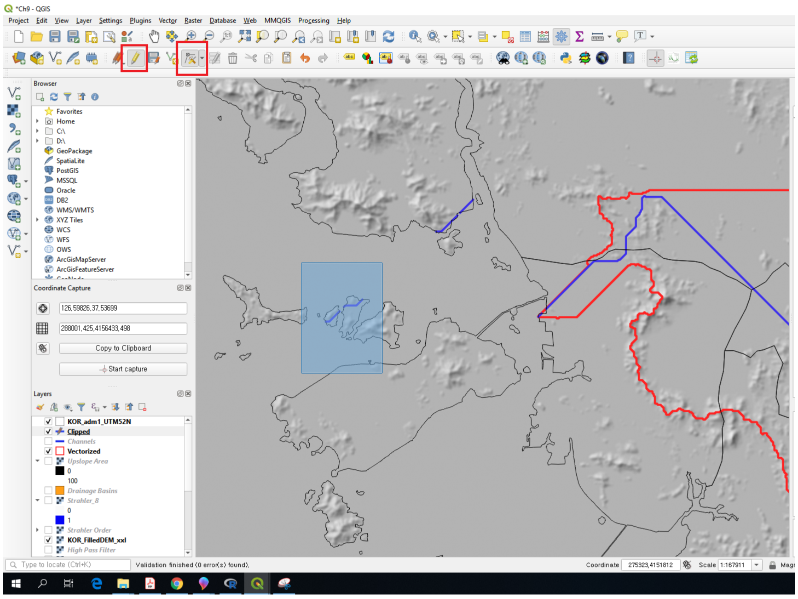


Fig. 11.19a. Clean up the stream vectors that don't belong.

In the end, you can have a nice map of the major streamways (Strahler Order ≥8) in South Korea as well as a drainage basin map of the Han River (Figure 11.20).

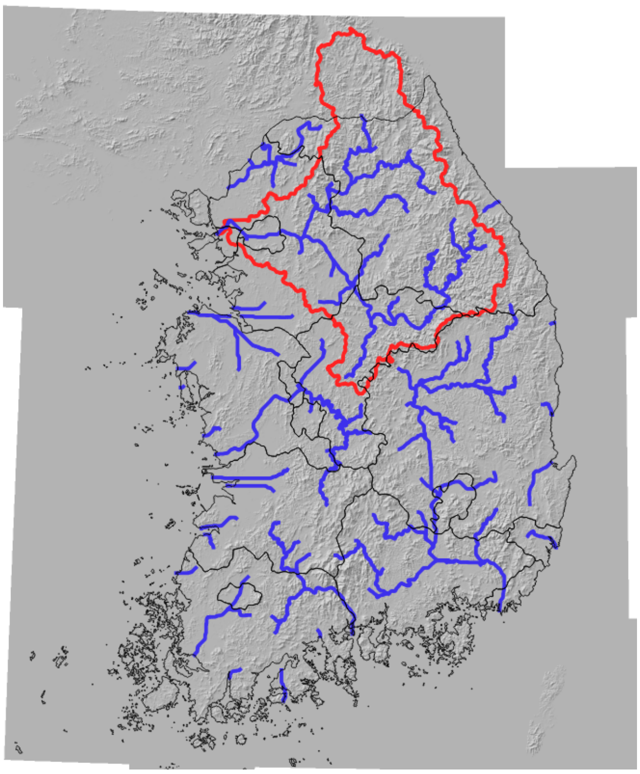


Fig. 11.20. The resulting stream catchment map.

If you compare the results of the r.watershed map to what we just created, you can see that r.watershed did a pretty good job with matching the contours of the basin (note some small disparities as in Figure 11.21b), but it overpredicted the number of basins relative to the pour point, even when we set the parameters at 5000 cells. So, r.watershed can be a good tool for giving you general catchments over large areas, but it lacks precision in mapping total drainageways. If you want to develop a true hydrological model, you have to spend the time performing calculations on each basin.

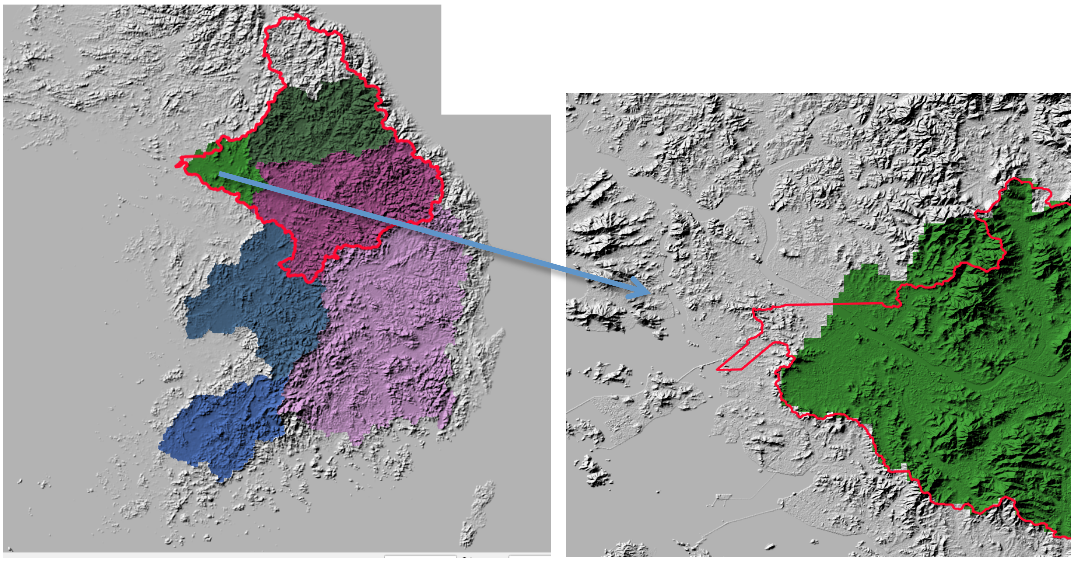


Fig. 11.21. Compare the basin construction models between r.watershed and the multistep process you just undertook.

11.4. EXERCISE

Create a catchment basin for the Geum River (금강). Plot the Neolithic archaeological sites (5800–3600 cal. years BP) that fall within this basin twice: 5800–5200 and 5200–3600 years BP). Do the same for the Han River (you already plotted this basin in the exercise, but be sure that you plot the Incheon sites in the Han River catchment). From the CW_chulmun_ceramics.xlsx file, calculate the median rim diameters from the sites that have ceramic rims. Use a graduated symbol size in the symbology to express median rim sizes on a scale from largest to smallest. Under Method you need to change from color to size. For the points that do not have ceramics, you will not make a graduated scale—make those points a different color than your sites with ceramics. Provide a png file of two maps (early phase, late phase) that has a scale bar, north arrow and **gradient ticks** of UTM coordinates on the outside of the map. In addition to titles, north arrows and scale bars, the maps need to have legends as well identifying the elements as well as what the size gradients represent in relation to median rim sizes.

In two or three paragraphs, explain the similarities and differences between the distribution of ceramic rim sizes in space and time in your study region. Archaeologists often study catchments because they detect significant differences between them in human behavior. Did this occur in the Neolithic of the Geum and Han River basins? (5 points for completing the mapping portion of the assignment correctly and on time and thoroughly answering the questions I have asked above.)

STUDY QUESTION (1): Conolly and Lake (2006: Ch. 11) discuss different types of ways to measure connectivity between communities. How could you use the data you have mapped in this exercise to develop a network analysis within and beyond Korea? (2 points)

STUDY QUESTION (2): Reference an academic paper (provide the full citation in the same style as *American Antiquity*—see pg. 3 of linked pdf) that discusses network analysis of some kind in relation to the final paper you are writing for this class. Even if your final paper is not about network analysis, there is always some kind of research on how communities or objects connect to other places. What are the relative strengths and weaknesses of this paper based on the GIS you compiled this week? (3 points)

Provide all of the results as *YOURNAME*_Ch11.docx and *YOURNAME*_ Ch11.png and **upload it to the eTL by May 27 at 9:00 am.**

Appendix A: List of excellent sources of free data

Aggregators

- Free GIS data (https://freegisdata.rtwilson.com)
- Wikipedia (https://en.wikipedia.org/wiki/List_of_GIS_data_sources): Surprisingly, a collection of GIS data sources are available on Wikipedia. Some of the links may be dead, but it is worth checking here if you need vector data.
- ESRI Open Data (http://opendata.arcgis.com): A very comprehensive assemblage of datasets from all over the world, including many parts of Kenya.
- Geospatial Data Gateway (https://gdg.sc.egov.usda.gov): Aggregator of geospatial data available from the US government.
- Stanford Libraries (https://library.stanford.edu/research/stanford-geospatial-center/data): A metadatabase of geospatial information.
- Harvard AfricaMap (https://worldmap.harvard.edu/africamap/): A project aggregating GIS data from across Africa. Includes language, ethnic, religion and historical boundary maps.
- Harvard WorldMap (http://worldmap.harvard.edu/maps/search?sort=last_modified&dir=DESC): Spawned from AfricaMap, Harvard hosts a repository of historic maps from all over the world.

Primary datasets

- Earth Explorer (https://earthexplorer.usgs.gov): Remote sensing data from the United States Geological Survey.
- ETOPO1 (https://www.ngdc.noaa.gov/mgg/global/global.html): A 1 arc-minute DEM of the world.
- UNEP Environmental Data (http://geodata.grid.unep.ch): Lots of environmental datasets from across the world.
- ILRI data hub (http://192.156.137.110/gis/search.asp): A comprehensive database of environmental and administrative data from across the tropics.
- Copernicus Open Data Hub (https://scihub.copernicus.eu/dhus/#/home): Data from the European Space Agency. Includes the highest resolution orthographic maps currently available for free online.
- Food and Agriculture Organization (http://www.fao.org/geonetwork/srv/en/main.home): European dataset portal for various climatic and ecological data.
- Diva GIS (http://www.diva-gis.org): Vector datasets available on a country-by-country basis.
- Natural Earth (http://www.naturalearthdata.com): Topographic and other raster datasets available on global scales.
- ISRIC Library (http://library.wur.nl/WebQuery/isric/start): An excellent assemblage of historical maps. These maps need to be georectified in a GIS.
- World Resources Institute (http://datasets.wri.org/dataset?_tags_limit=0): Land cover and biodiversity data throughout the tropics.
- European Digital Archive of Soil Maps (http://esdac.jrc.ec.europa.eu/ESDB_Archive/EuDASM/Africa/index.htm): A

collection of old soil and geologic maps from across Africa. These need to be georectified to use in a GIS.

- Mineral resources online USGS (https://mrdata.usgs.gov/general/global.html): A repository of geologic and minerologic data on a global scale
- Canadian Archaeological Research Database (http://www.canadianarchaeology.ca): Includes locations of radiocarbon dates from around the world.

QGIS Information/Resources

- QGIS home page (http://www.qgis.org/en/site/): Download the QGIS program and get information on the latest (and old) software releases.
- QGIS on Twitter (https://twitter.com/qgis?lang=en): Updates from power users and programmers on software features.
- QGIS Tutorials and Tips (www.qgistutorials.com): Useful tutorials on how to execute various objectives in QGIS.
- QGIS for Absolute Beginners (https://www.youtube.com/watch?v=aLmMovuydqI&vl=en): This is a great rundown on the basic features of QGIS.

Open Source Data Forums

- OSGEO (http://osgeo-org.1560.x6.nabble.com): An online community where you can post questions and receive responses about problems you are experiencing with the software.
- GIS Stack Exchange (https://gis.stackexchange.com): A question and answer forum where you can search for or post questions to receive answers to vexing questions you might be facing.
- QGIS Wiki (http://wiki.openstreetmap.org/wiki/QGIS): A wiki for obtaining information and tips on how to use QGIS.
- Geoacademy (https://fossgeo.org): Free online training courses in open-source software following registration. A certificate can be obtained for US$50 following successful completion of the course.
- Geospatial Data Abstraction Library (http://www.gdal.org): A translator library for raster and vector geospatial data formats. If you don't know what your file format means (or how to open it in a GIS), visit this webpage.

Appendix B: R Code for the creation of modeled radiocarbon dates on an exponential growth curve

```
set.seed(35)          # for sake of reproducibility
n <- 16101            # 16,101 data points
mr <- c(54850, 150)   # dates between 150-54850 cal BP

y0 <- rexp(n, 1/3.4) # simulate exp. dist.
y <- y0[order(-y0)]  # sort

# f(x) = (b-a)(x - min(x))/(max(x)-min(x)) + a
y.scaled <- (mr[1] - mr[2]) * (y - min(y)) / (max(y) - min(y))
+ mr[2]

range(y.scaled)
print(y.scaled)

plot(y.scaled, type="l", ylab="cal years BP", xlab="# of
dates")

write.csv(y.scaled, "~/Documents/SNU
teaching/@2019Spring_GIS/14C_data/14CTrainingSet2_expdates.csv
")
```

Resulting image showing the exponential plot of numbers (arbitrary radiocarbon dates) between 150 and 54,850, which mirrors the real distribution of our data: