Spatial Analysis Using QGIS

Last Modified: February 2016



This reference and training manual was produced from material provided by:

University of Maryland Libraries
U.S. Government Information, Maps & GIS Services
McKeldin Library, Room 4118
College Park, MD 20742-7011
http://www.lib.umd.edu/gov-info-gis/maps-and-gis/gis

Modified by

University Library System, University of Pittsburgh Geographic & Spatial Data Services, Hillman Library, G-49 3960 Forbes Ave Pittsburgh, PA 15260 http://library.pitt.edu/GIS

Permission to reproduce this manual or any of its parts for non-commercial, educational purposes is granted. Appropriate citation is appreciated.

Introduction

A GIS (Geographic Information System) is a powerful tool used for computerized mapping and spatial analysis. A GIS provides functionality to capture, store, query, analyze, display and output geographic information.

For this seminar we will be using QGIS Pisa, the newest version of a popular open source GIS software. This course is meant to teach some fundamental GIS operations using QGIS. It is not meant to be a comprehensive course in GIS or QGIS. However, we hope this seminar will get you started using GIS and excited about learning more.

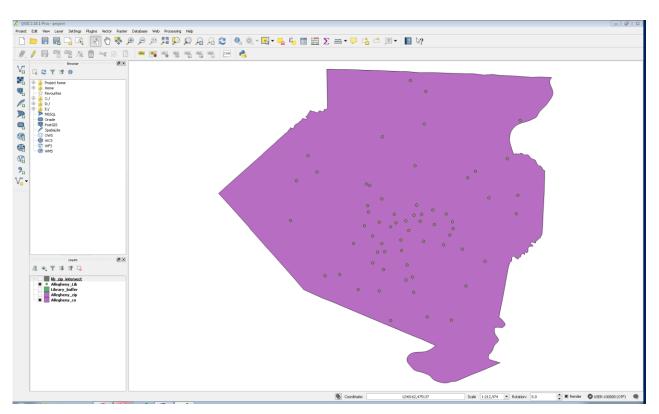
Exercise: Five Common QGIS Tools

1. Buffer

Our first geoprocessing task will be creating a new area around an already existing layer. This is called a **Buffer**. Buffers can be created around any of the vector file types—points, lines, and/or polygons. We will be using the point file of libraries in Allegheny County named Allegheny_Lib.

If you do not already have the file available in the table of contents, do the following:

- 1. Open QGIS and a new blank map. Click on Add Vector Layer on the left side menu.
- 2. Navigate to the workshop folder and select the Allegheny_Lib.shp file. Click open twice.
- 3. Also add Allegheny_Co.shp. Rearrange layers so the points appear on top by dragging this layer to the bottom of the layers list in the lower left of the screen.



Let's pretend that we are working for Allegheny County and are looking into building a new library. We'd like to know who currently has library access and how we can improve service to those who don't. We have been told by experts that we should be looking for locations that are outside a 2 mile radius of current libraries. This is a great situation to create a **buffer**. Buffers are new areas of a certain distance that can be created around points, lines, or polygons. We will make a 2 mile buffer around the libraries to see the areas that currently

have library access.

We will use Geoprocessing tools to create our buffers.

- 4. Navigate to Vector > Geoprocessing tools > Buffers
- 5. Set the input layer to Allegheny_Lib with a buffer distance of 10560 (number of feet in two miles)
- 6. Click browse and save the output shapefile to your workshop folder as "Library_buffer".

🌠 Buffer(s)

Input vector layer Allegheny_Lib

Use only selected features

Segments to approximate

Buffer distance field

Dissolve buffer results

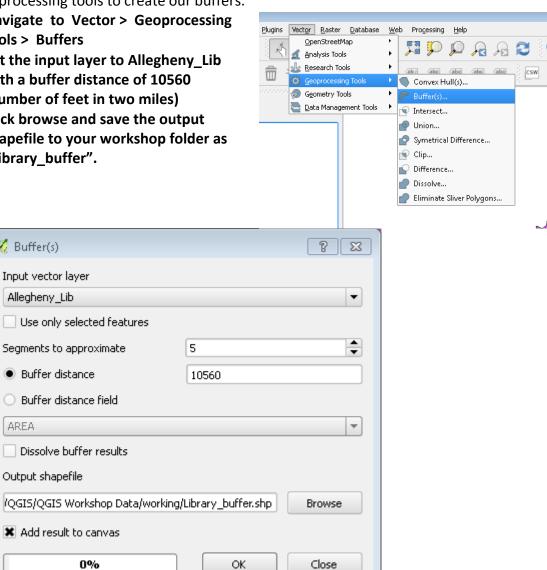
X Add result to canvas

0%

Output shapefile

Buffer distance

AREA

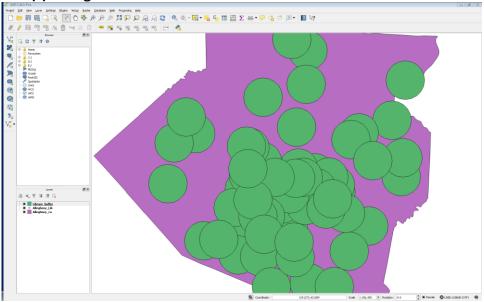


5

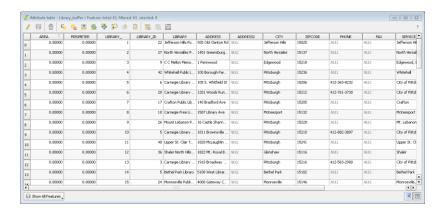
10560

The new shapefile, Library_Buffer, now appears in the table of contents and in the map display.

7. Close the buffer(s) dialog.

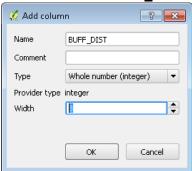


8. Let's look at the attribute table of the buffer file. In the table of contents, **right click on** Library_Buffer > Open Attribute Table.

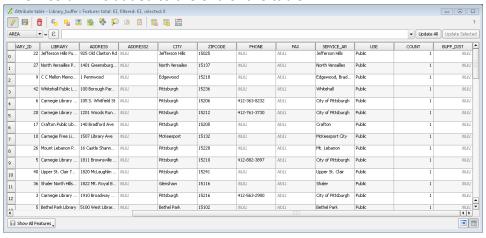


The table shows data describing each library but no field showing the buffer distance. Let's add that.

- 9. Enable editing mode
- 10. Click 'New Column'
- 11. Set the name to BUFF_DIST with a width of 8. Click OK.

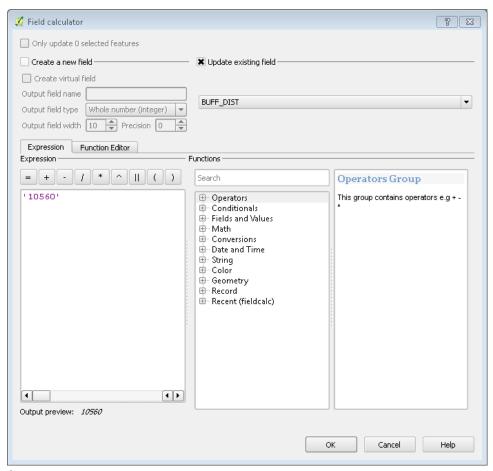


A new column is added to the end of the table.



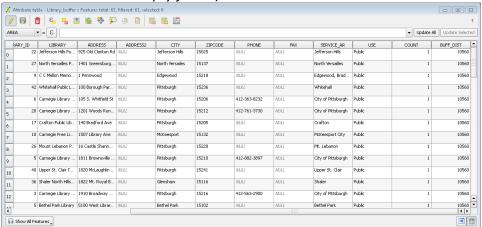
Let's populate this field with the buffer distance of 10560 feet.

- 12. Open the field calculator
- 13. Check the 'Update existing field' box and set the dropdown menu to 'BUFF_DIST'
- **14.** In Expression type: <u>'10560'</u> (*Include apostrophes!*). The box should match the image on the next page. Your field calculator might look different if you're using QGIS on a Mac.



15. Click OK.

The field BUFF-DIST is now populated with the distance 10560, which is a two miles in feet. (You alternatively could have used the field calculator to create a new field with the output name BUFF_DIST and the same expression. This method was used to show how to add empty fields.)



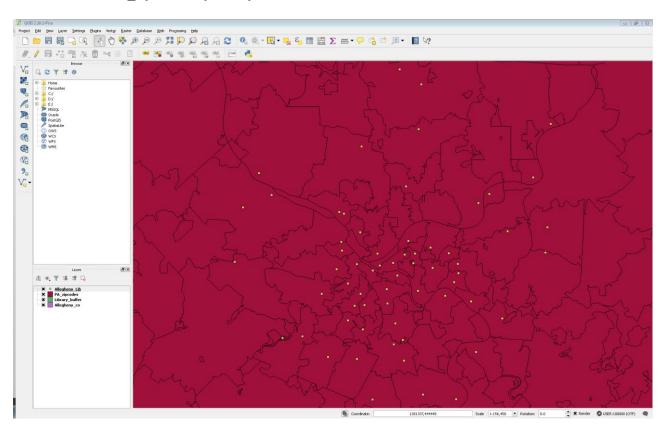
16. End the edit session by clicking the edit button again and save the changes and close the attribute table.

Clip Operation

Now that we have a 2 mile buffer of the areas that are within 2 miles of a library, let's look at **who** is nearby. We will use zip codes as a way to identify areas. The only zip code file that we could find is for the entire country. We'd like it to be for only Allegheny County. This is a perfect opportunity to use the **clip** function.

Usually associated with the "within" condition, the clip function works like a cookie cutter. In our case we will be using the US zip code file as the input layer (the cookie dough) and cutting it with Allegheny County (the cookie cutter). We'll end up with the shape of Allegheny County and the attributes of the zip codes.

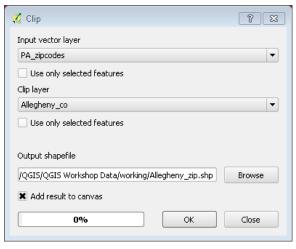
- 1. Click on the add vector data icon.
- 2. Browse to the workshop folder.
- 3. Select PA_Zipcodes.shp and open it.



A layer for the zip codes in the US from 2014 will be added.

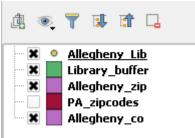
Now, let's clip the zip codes to Allegheny County.

- 4. Navigate to Vector > Geoprocessing tools > Clip...
- Set 'PA_Zipcodes' as the input vector layer and Allegheny_co as the clip layer.
- 6. Click browse under output shapefile and navigate to your workshop folder.
- 7. Name the file 'Allegheny_zip'.
- 8. Click OK and close the dialog.

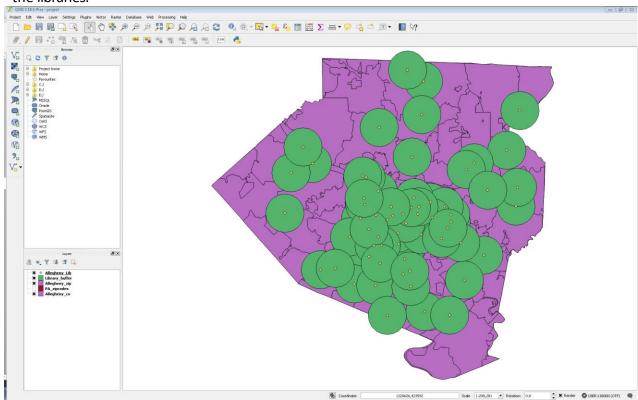


Let's reorganize our layers so that we can see everything.

- 9. Turn off PA_zipcodes by unchecking the box next to it...
- 10. Drag Library_Buffer under Allegheny_Lib and Allegheny_zip under Library_buffer.

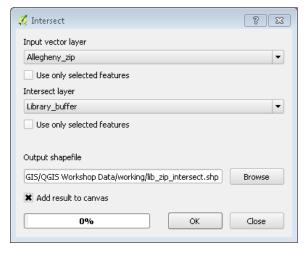


You should see the zip codes for Allegheny County as well as the libraries and buffers around the libraries.



Intersect

Now that we have the zip codes in Allegheny County, we can use that with the buffer of the libraries that we created. We can visually see where the overlap between the two files is, but we can take a closer look by using the **intersection** operation.



From the Geoprocessing tools menu we used in the last two sections, navigate to Intersect...:

- 1. Navigate to Vector > Geoprocessing tools > Intersect...
- 2. Set the input layer as 'Allegheny zip'
- 3.Set the intersect layer to 'library buffer'
- 4.Browse to your workshop folder and save the file as lib_zip_intersect
- 5.Click OK

6. Turn off all layers EXCEPT:

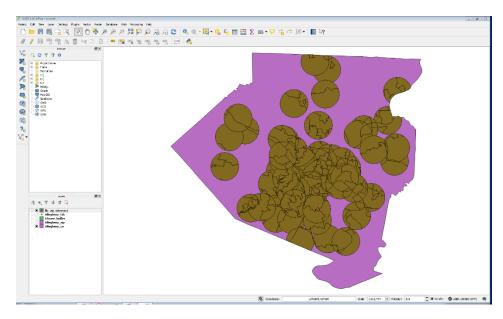
- the newly created lib_zip_Intersect file and
- Allegheny_Co

In the map display we can visually see the zip code areas in the 2 mile buffer.

7. Right click on lib_zip_Intersect and open the attribute table.

The attribute table has attribute information of **both** the zip codes and the libraries that are in a 2 mile buffer of the library. The shape of the file is of the shared areas.

8. Close the attribute table.



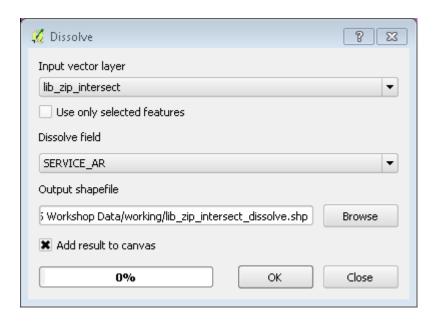
Dissolve

Still pretending to be Allegheny County, we have decided that it would be nice to have a map and table of the city located in each zip code. We'd like to make sure we don't build a new library in a service area that already has coverage. As the larger areas have multiple zip codes that belong to one city, we'd like to combine the areas both visually and in the attribute table. The dissolve operation is ideal for this. As the dissolve operation allows you the option of adding statistical fields, we use the layer we just created (lib_zip_intersect) and add a column with the total area of the library service areas combined.

1. Navigate to Vector > Geoprocessing Tools > Dissolve

The Dissolve window will open.

- 2. In the drop down box for 'Input Features' select lib zip Intersect.
- 3. Set "SERVICE_AR" in the 'Dissolve_Field(s) box. This allows us to dissolve the lines that connect zip codes that are shared by the same city name.
- 4. Set the output shapefile to 'lib_zip_intersect_dissolve' in your workshop folder.
- 5. Click "OK."

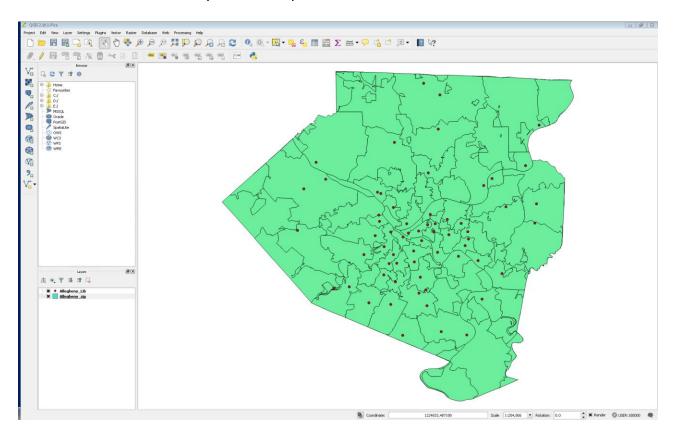


In the map display you can see that there are now fewer areas. Open the attribute table by **right clicking on lib_zip_Intersect_Dissolve.** You can see that all 63 libraries are located within forty-two service areas in the SERVICE_AR field.

Join attributes by location (spatial join)

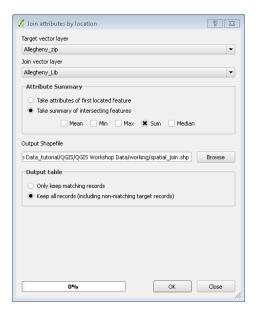
Let's investigate another option in the toolbox. In a **spatial join,** fields from one layer's attribute table are appended to another layer's attribute table based on the relative locations of the features in the two layers. Let's take a look at how it works.

We will be using the Allegheny_lib and Allegheny_zip shapefiles. Open a new map in QGIS and add these two files from your workshop folder.



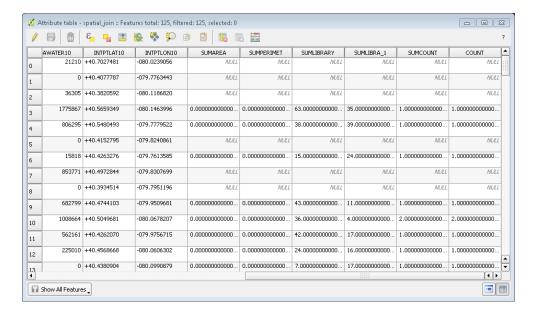
A spatial join can be useful, as in this example, for determining how many points fall within polygons.

- Navigate to Vector > Data Management Tools > Join Attributes by Location...
- 7. Set the target vector layer to Allegheny_zip and the join vector layer to Allegheny_lib
- 8. In the Attribute Summary box, click the radio button for 'Take summary of intersecting features' and select SUM. Deselect any other summaries.
- 9. Name the output feature 'spatial_join' in your workshop folder.
- 10. For output table, select 'Keep all records'
- 11. Click OK
- 12. If prompted, add the new file to the map.
- 13. Close the dialog.



A new shapefile will be added to the map and table of contents. As you can see, the shape matches the shape of the Allegheny_zip shapefile. Let's check the attribute table.

Scroll through and see how this spatial join added the attributes for Allegheny_lib to the end of the Allegheny_zip file. Note how the COUNT field keeps track of which ZIP codes have a library with a count of 1 and how some zip codes might even have 2 libraries.



Conclusion

QGIS has several geoprocessing tools. In this tutorial, we've seen five of the most commonly used Geoprocessing tools:

Operation	Description	Attribute table changes	New Shape is:	Attribute Table
Buffer	Creates a new polygon at a buffered distance around a point, line or another polygon.	None	New polygon of distance	Attributes of layer buffered.
Clip	Cuts out a piece of one theme using another theme as a "cookie cutter."	None	Overlay layer	Input layer
Dissolve	Removes boundaries or nodes between adjacent polygons or lines that have the same values for a specified attribute.	Yes	Polygon layer of areas that share the specified value.	Only contains values that the dissolve affects. Lose all other attributes.
Intersect	Integrates two spatial data sets while preserving only those features falling within the spatial extent common to both themes (similar to Boolean AND).	Yes	Just the areas in common between the layers.	Combined attribute table of all layers used in operation.
Spatial Join	Creates a table join in which fields from one layer's attribute table are appended to another layer's attribute table based on the relative locations of the features in the two layers.	Yes	Shape of target features layer (first one input)	Target layer attribute table with join features attributes at end.
Union	Joins two layers together visually with the new attribute table consisting of shared/overlapping areas. NOT shown in this class.	Yes	Both layers used in operation	Attributes of shared areas

Exercise: Spatial Analysis with Raster Data

Background

The analytical functions we have just performed in QGIS are based on a vector data structure of points, lines and polygons. As demonstrated, this data model is ideal for elements that have discrete boundaries like airports, highways and urban areas. But what if our data does not have discrete boundaries and is *continuous* over space (for example, elevation, slope, aspect or soil type)? The *raster* data model can be particularly beneficial in analyzing data that is continuous. As stated earlier, raster data is composed of a two-dimensional matrix of grid cells, with each cell assigned a numerical value.

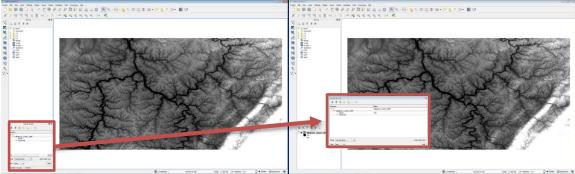
In the following section, we will work through some analyses that may be performed on raster data. Bear in mind that with raster data, even more than with vector data, for a successful outcome it is essential to understand the parameters of the data you're working with and what you hope to achieve in your analysis. The exercises below are simply to introduce you to where the tools are located and to show you some of the mechanics—for true analysis, you will have to make the decisions!

Examine raster data

- 1. Open a new map.
- 2. Add an elevation data to QGIS by clicking on the Add Raster Layer icon

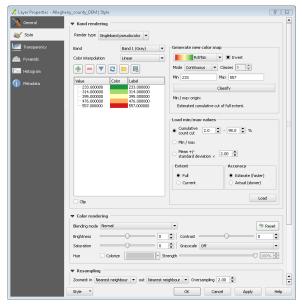


- 3. Navigate to 'Allegheny_county_DEM.tif' in your workshop folder
- 4. Use the 'identify' tool to get each cell's value, it will be displayed in the Identify Results panel (which might appear under the browser panel on the left side, you can drag this to a window for easier viewing, see following image.)

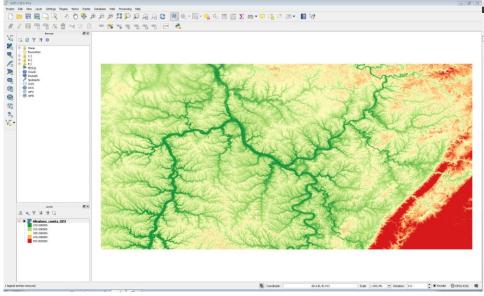


5. Right click on the elevation layer and go to 'Properties'.

- a. Look at the "Metadata" tab to find the size of the cells, the extent of the data, what spatial reference it is using, and other good information under the Properties field.
- b. Look at the "Style" tab to see some color ramp options for display. In the 'Render type' dropdown menu, select 'Singleband pseudocolor'.
 - In the "generate new color map" box, select RdYlGn in the dropdown menu and check the Invert box.
 - Click the Classify Button



c. Click "Apply" to view the image with the new color ramp. Lower elevations appear in green and higher elevations are red. Click OK



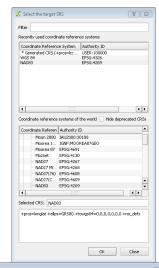
Surface Analysis

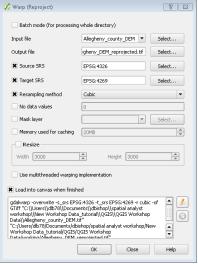
Surface analysis produces a new dataset—this can help you identify or derive patterns within the original dataset that may not have been evident. We will look at two examples: **contour** and **slope**. Contour will produce a vector output, while slope will produce a raster output. The raster output you get from spatial analysis on grid data can be either grid with unique values or Boolean values.

The coordinate system used for the elevation raster uses Decimal Degrees for latitude and longitude for X and Y and meters for the elevations (Z). Most of the analysis we want to do using elevation datasets – such as creating a hill shade raster or creating contour intervals requires us to transform the coordinate system of the raster so that X, Y and Z are all in the same units. We should choose the projected coordinate system that is appropriate for the area of geographic coverage.

Warp (reproject)

- 1. Navigate to: Raster > Projections > Warp (reproject)...
- 2. Keep the Input Raster as the Allegheny county DEM file.
- 3. In the Output Raster Dataset blank, chose a name and file location (workshop folder on your desktop). Rename the file 'Allegheny_DEM_reprojected' and keep the file type as GeoTIFF. Click Save.
- Check the Target SRS box and then the Select... button next to it, select EPSG: 3724 (NSRS2007 UTM 17N) as your coordinate reference system. Click OK in the Select Target SRS window.
- 5. Check the 'No data values' box and keep the default value 0.
- 6. Finally, check the Resampling Method box and choose Cubic
- 7. Your Warp window should match the image below, click OK. Then close the window. A new raster will be added to your project.





Resampling

When a raster dataset is projected from a plain geographic coordinate system to a projected one, the size and shape of the cells is transformed. You can think of Resampling as creating a new grid of cells and assigning values to the new cells based on the values of the old cells. The default method (Nearest Neighbor Resampling) simply assigns a value to the new cell based on the value of the cell in the old raster that is nearest to the center of the new one. This is the right thing to do with rasters whose values represent categories. For rasters that represent a continuous surface the **Nearest Neighbor** method can create stair-stepped striations. For continuous value rasters, like elevation models, the **Cubic** resampling method is best. (Source: http://www.gsd.harvard.edu/gis/manual/dem/)

Contour (Vector)

Why might you want to see contours? This is a very good way to look at the overall gradation of the land, and is familiar to the eye. Using contour as an example also illustrates how you can get a vector result from a raster original.

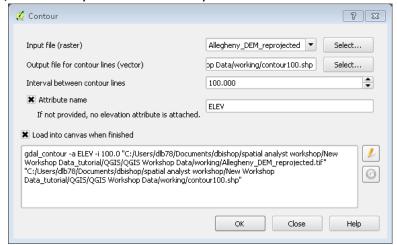
1. Navigate to Raster > Extraction > Contour

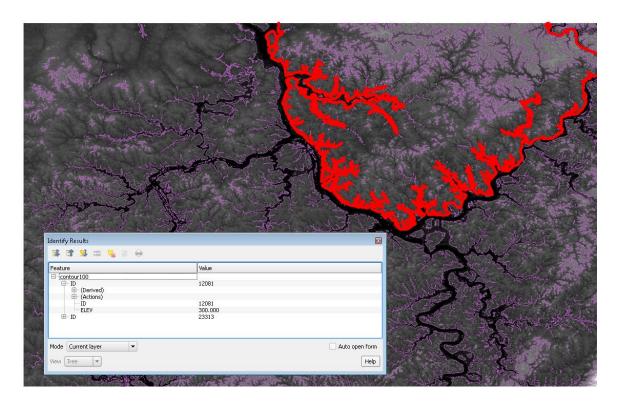
In the 'Contour' dialog,

- Keep "Allegheny DEM reprojected" in the drop down box under 'Input file (raster)'.
- 3. Navigate to the workshop folder on the desktop in the 'Output file for contour lines' box using the 'Select...' button.
- 4. Name the file contour 100.
- 5. Put '100' (or your own choice) in 'Interval between contour lines' field.
- Check the attribute name box, otherwise you won't have any elevation info!
- 7. Click 'OK', close the dialog

Elevation raster data is converted into vector data—contour lines.

8. Click on a line with the identify tool and notice that the line on the map is highlighted. The value of the contour line is listed under the ELEV field.



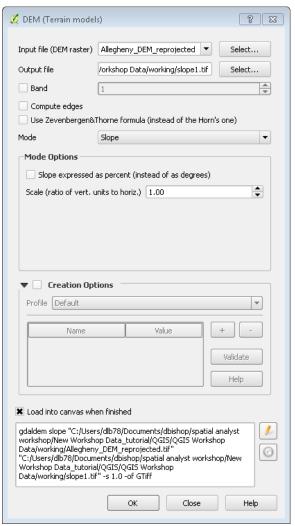


Let's try another function from the surface menu.

Slope (Grid)

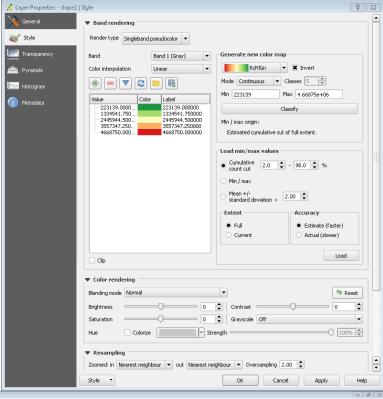
Examining slope tells you how steep the terrain is—this kind of output can then be analyzed and used for a variety of determinations such as likelihood of flooding, best places to locate buildings, etc. What the software is doing in this case is calculating the maximum rate of change between each raster cell and its neighbors. You can calculate your slope output either as percent slope or degree of slope. We will use percent in this example, and our output will be a raster.

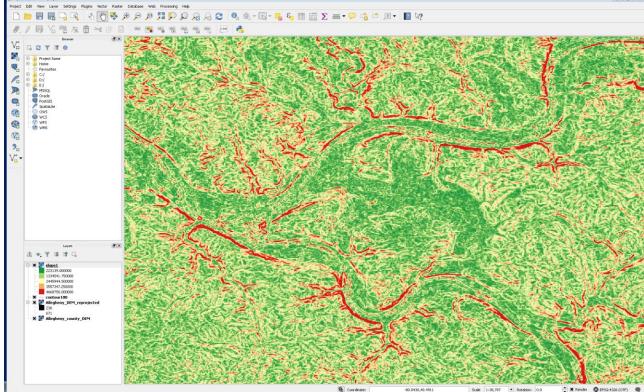
- Navigate to Raster > Analysis > DEM (Terrain Models)
- 2. Keep "Allegheny_DEM_reprojected" in the drop down box under 'Input file'.
- 3. Navigate to the workshop folder on the desktop in the 'Output file' box using the 'Select...' button.
- 4. Name the file slope1.tif.
- 5. Change the mode to Slope
- 6. Keep the defaults for the remaining boxes.
- 7. Click 'OK', close the dialog



The resulting output is a slope map of the elevation data. The grid cells have been given new values based on the difference between their elevation value and that of their neighbors. In the example the dark areas are those with the steepest slopes. Zoom in to view image details.

8. To adjust the classification, right click on the slope1 layer and select properties. Change the render type to Singleband Pseudocolor and apply an inverted RdYIGn ramp.





Conclusions

The examples provided here are very simple ones. We could have built a query that was far more complex. Other areas such as wetlands, parks, etc. could have been incorporated into the model and eliminated from consideration.

Grid analysis can be very complex. For example, you can use the themes of slopes, wetland areas, soils, forested areas, areas served by water/sewer utilities, proximity to transit, etc. to determine the best sites for building. The accuracy of the analysis depends on several things:

- The accuracy and currency of your data the old saying "garbage in, garbage out" holds true here. You want to have data that is current and accurate for the purpose of your analysis. QGIS allows vector data to be converted to grid data so any investment in vector data is not lost.
- The resolution of your data the finer the resolution, the more detailed your analysis. You can set a resolution size so high that the grid data approaches vector data in appearance. The down side of very high resolution data is the amount of storage space that it uses. A resolution should be chosen appropriate to the scale for the data and the task at hand.
- The values you assign to the cells and the appropriateness of your mathematical formula spatial analysis is ultimately a judgement call of what is deemed "valuable" for some purpose, and your formula should be logical in its structure and sequence. The reasoning behind the value assignments and the formula used must be defensible.

Please take the last few minutes to complete the Workshop Evaluation Form http://bit.ly/ULS_assess