Introduction to GIS
Computer Science in Modern Biology Workshop 2021
Gwendolyn Lloyd (lloydg@miamioh.edu)
TA(s): Kallie Koon (koonkr@miamioh.edu)

Vegetation along roads within Great Smoky Mountain National Park: Day 2

Task: Is there a difference in understory vegetation present near (**within 50 meters**) of **primary** roads in the Great Smoky Mountains compared to the vegetation types **within 50 meters** of **local** roads? What is the composition of the vegetation in both these locations?

Background: Vegetation surveys are a huge component of monitoring and conservation, particularly for locations such as National Parks. By doing a vegetation survey, researchers are able to learn what plants are growing in particular areas, and can then be used to draw conclusions about the animals as well as the soil in these areas. Man-made construction, such as roads may negatively impact diversity in vegetation due to chemical or physical influences, such as exhaust or loose asphalt.

For more background information:

https://www.eco-action.org/dt/roads.html

Part 1: Create a workflow

Before we begin we need to create a workflow or a list of methods required for this project.

1. First we need to figure out which variables to use! We want one variable of different vegetation types present in the Smoky Mountains, as well as a road variable that includes primary and local road types. In addition, we want a variable we can use to measure vegetation! We could compare the *types* of vegetation present in an area, but a numerical value that represents the vegetation present in the area would allow us to compare between the road types more easily. The vegetation attribute table is extremely large, so using the online version may be better:

   https://public-nps.opendata.arcgis.com/datasets/understory-vegetation/data

Variable	Shapefile	Information it contains

2. Determine which tools to use for this project. We will need to create a 50 meter buffer surrounding the roads, and then use that as a template to 'cut out' the vegetation layer. To do so, we will need to use some of the geoprocessing tools (Vector → Geoprocessing Tools). Their functions are described below:

Geoprocessing tools

Icon	Tool	Purpose
	Convex hull(s)	Create minimum convex hull(s) for an input layer, or based on an ID field.
	Buffer with * fixed distance * distance field	Create buffer(s) around features * based on fixed distance * based on distance field
	Intersect	Overlay layers such that output contains areas where both layers intersect.
	Union	Overlay layers such that output contains intersecting and non-intersecting areas.
	Symmetrical difference	Overlay layers such that output contains those areas of the input and difference layers that do not intersect.
	Clip	Overlay layers such that output contains areas that intersect the clip layer.
	Difference	Overlay layers such that output contains areas not intersecting the clip layer.
	Dissolve	Merge features based on input field. All features with identical input values are combined to form one single feature.
	Eliminate sliver polygons	Merges selected features with the neighboring polygon with the largest area or largest common boundary.

Also remember that you want to get information for vegetation both near primary and local roads. The best way to do this is to have two layers, one of vegetation around primary roads and one of vegetation near local roads.

3. Once you have determined your variables and tools, create a **workflow**. A workflow is a flowchart or list of steps and necessary information you need for each step. I would recommend drawing this out on paper or in a separate document. (Don't worry! We will talk through this together before moving on!)

Part 2: Geoprocessing

Once you feel comfortable with your workflow, you can follow it to create new data files! Make sure to change the appearance of your data and label the shapefiles so that you can make sure any analyses you are doing makes sense!

Part 3: Analyses

We will be learning a new technique for the analyses today because we are dealing with categorical values as well as numerical values, while yesterday we only used numerical values, so we could subset by creating cut-offs in our data.

We could do the same today, and you are welcome to, but this should provide a more straightforward end result.

1. This method will involve creating a database and using the database manager (Database → DB Manager)
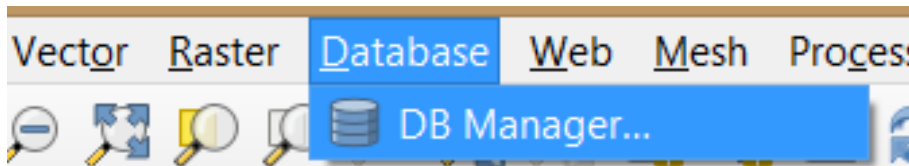
2. Then under the virtual layers database we will then select the layer that we want to get vegetation information for. A virtual layer creates a temporary file that does the calculations from another shapefile.
   Next, click on the second button from the left to open an SQL Query window.
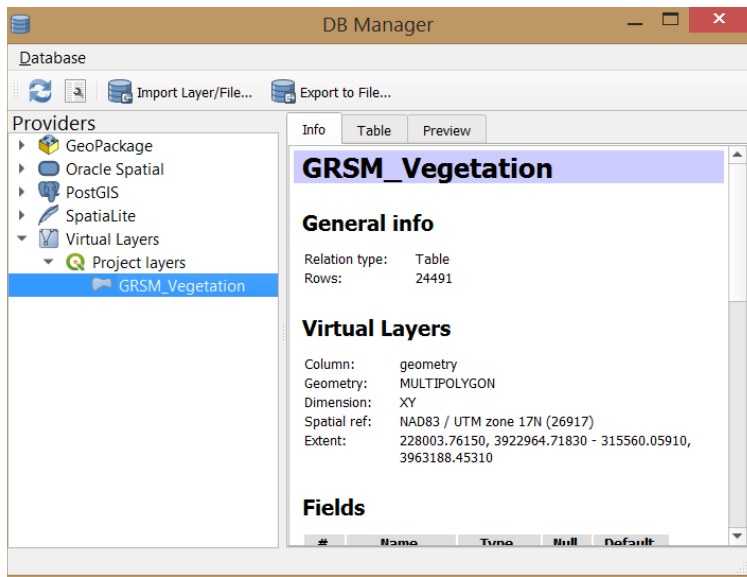
3. Once open, you should type the code as written below. The second row, FROM " ", is the layer which you are getting the calculations from, so that should change.

Otherwise what it means is that it is selecting the two variables, Class and SHAPE_Area from the herbaceous data, and taking the sum of the area and grouping it by the categorical variable Class.

SELECT Class, Sum(SHAPE_Area)
FROM "*LAYER OF INTEREST HERE*"
GROUP BY Class
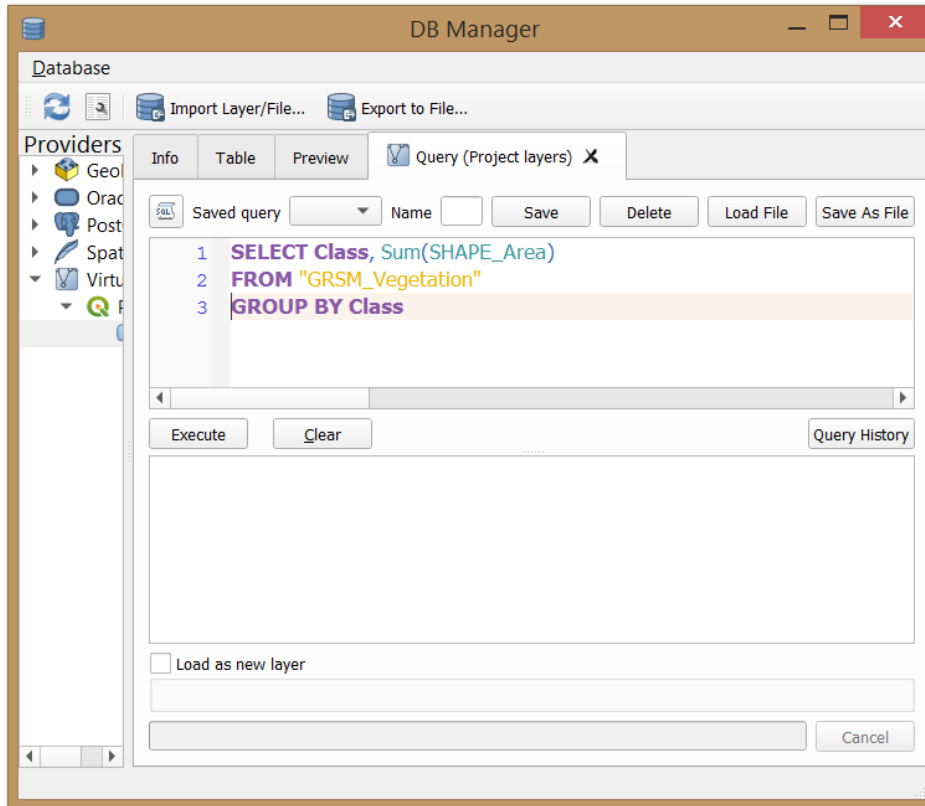
After you have written your code click execute.

4. Your screen should look similar to the image below, with a table containing the total area covered by each vegetation class. These values are in meters squared, due to the original data.

To open this table as an attribute table under the layer menu, check the box that states 'Load as new layer' and then click load in the window.
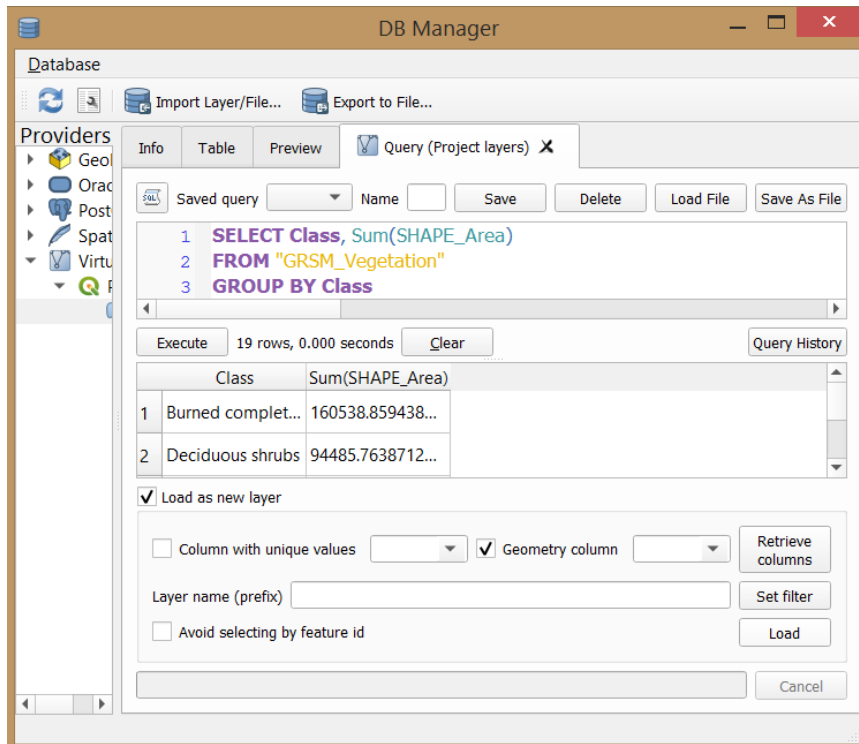
5. Once there, you can repeat the process for the other layer file (primary or local roads) and compare the values. Are the vegetation types the same? Is the area they cover the same? Do different types of road make a difference? What is different about these two road types?

Once complete, you have both data management and geoprocessing skills as well as aesthetics. Take the remaining time to work do the following options:
- Display your results from today yesterday on one map
- Try to answer a different question from today's data.

Data from Great Smoky Mountains National Park. If you want to play with more of their data and from other national parks use the link below:
https://public-nps.opendata.arcgis.com/search?owner=GRSM_GIS