**Lab 3: Tornado chasing**

*Geog2011: Introduction to GIScience*

**Introduction**

In this lab, you’ll be analyzing the locations of tornadoes in Georgia from 1980-2018 using data from the National Land Cover Database (NLCD). The NLCD uses satellite data to identify land use types across the country. [See this page](https://www.mrlc.gov/data/legends/national-land-cover-database-2016-nlcd2016-legend) for more information on the NLCD classifications. You will also identify which Atlanta county had the most tornadoes during this time.

This lab will also follow a different format than your first two. Rather than providing step by step instructions, you’ll get tasks to complete and resources to help you do so. This better reflects the ways most professionals learn GIS software and gives you practice working independently with the software. Sites like [QGIS Tutorials](https://www.qgistutorials.com/en/index.html) are crucial for most folks learning new software.

Frustration is a normal part of this process! Just be patient with yourself as you work through these tasks. You’re also encouraged to Google for solutions when you encounter problems, as this is what even professionals do on a regular basis.

**Preparation: Software, county data, and NHGIS registration**

Download the zipped data file from the Lab 3 folder. It includes the following files:

* **nlcd2016\_ga\_updated**: This is a geotiff of the land cover data. As discussed in the class on data types, geotiff is a common format for raster data. This is a LARGE file (> 1GB) that has been compressed in a zip folder. You will need to click on the downloaded folder and then drag the geotiff file to your working folder.
* **tornadoes\_ga\_1980\_2018**: A CSV file with data on all tornadoes in Georgia from 1980 to 2018. This includes starting and ending coordinates, severity, and both injuries and fatalities. [Data source](https://www.spc.noaa.gov/wcm/#data)
* **Atlanta\_msa\_counties**: A geopackage file with counties in the Atlanta [metropolitan statistical area](https://www.census.gov/programs-surveys/metro-micro.html).

Download these files and save to your working folder.

**Lab structure**

Your goal in this lab is to calculate the percentage of tornadoes in Georgia that originated in urban (developed) areas during the study time period. To do so, you’ll need to do the following:

1. Load the NLCD dataset and tornado data
2. Use the Sample raster values tool to identify the land cover type for each tornado’s origin point
3. Calculate the percentage of tornadoes in areas designated as developed in the NLCD
4. Count the number of tornadoes in each Atlanta county using a spatial join
5. Make a choropleth map showing the number of tornadoes per county

**1) Load the NLCD and tornado data**

When you open QGIS, make sure you are choosing QGIS with GRASS. You learned how to use the add layers tool in QGIS in Lab 2. For this lab, start by adding the GeoTiff you downloaded.

* [QGIS Tutorials support page on raster styling](https://www.qgistutorials.com/en/docs/3/raster_styling_and_analysis.html)
* [Duke University’s guide to adding data, including raster data](https://guides.library.duke.edu/QGIS/ImportData)

You should now see a map of the whole state colored by land use type, something like this:

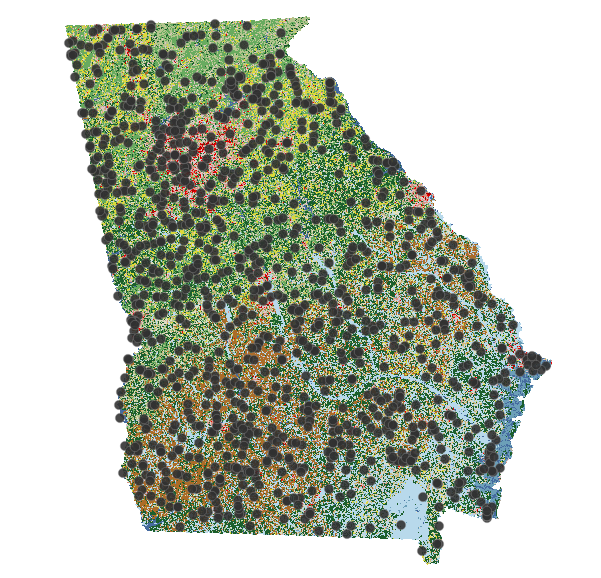


You can click on the identify tool (circled in red below) and click on any one of the pixels to see it’s value. [See this page](https://www.mrlc.gov/data/legends/national-land-cover-database-2016-nlcd2016-legend) for more information on how to interpret these numbers. For example, 22 is Developed, low intensity. **In this lab, we’re interested in codes 21-24, which refer to low to high intensity developed (urban) land**.



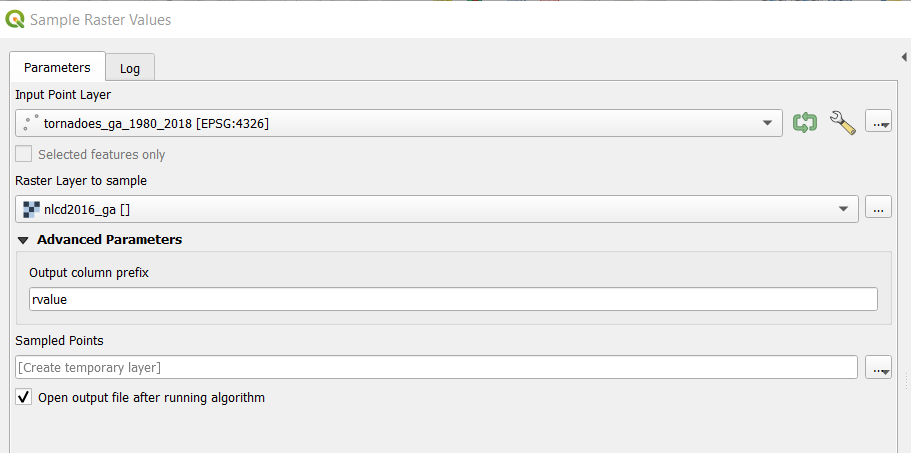
The tornado data is a CSV file, which you can load as Delimited Text, as we did your residence points in Lab 2. Use the slon and slat columns for the X and Y values. These are the points at which each tornado started. The elat and elon columns give an end point (if there is one available).

Change your projection to UTM zone 16N (crs 32616)--see lab 2 for more information on how to do so. You should now have both the land use map and points for the tornado origins loaded on your map, something like this:



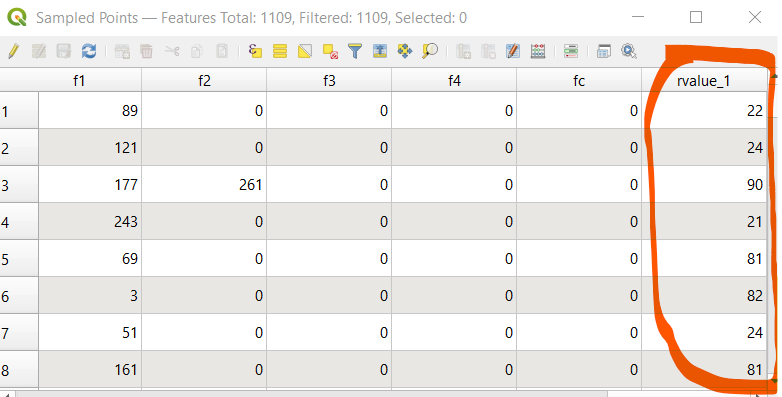
**2) Use the Sample raster values tool to identify the land cover type for each tornado’s origin point**

To identify what land type is present at each tornado’s origin, you will need to sample at each point. The tool to do this in QGIS is ***Sample raster values***. This will detect what the value of the NLCD data layer is at each point location and join that value to the point data. You can find it in the Processing Toolbox. If this is not visible, you can find it under the Processing drop down on the top menu bar. You’ll get a dialog box that looks like this:

/

If you’d like to save the output file (which will have the NLCD land type added in the “rvalue” column), you can do so in the lowest box under Sampled Points. Otherwise, it will be added to your project as a temporary layer--one that is deleted when you close QGIS.

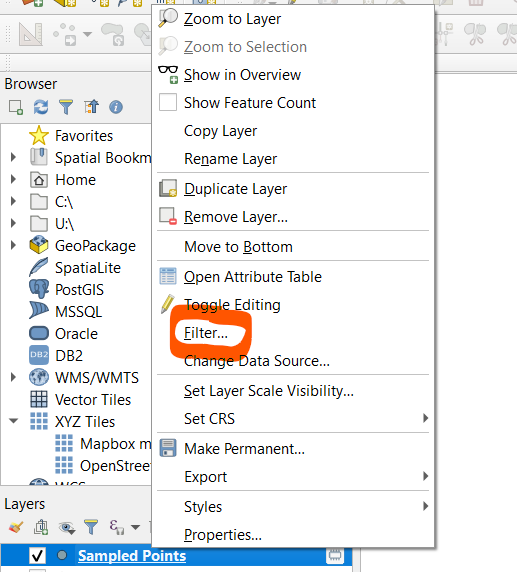
Once you’re done, open the attribute table for the new “Sampled points” layer. You should see a list of NLCD land use codes under the rvalue column--the last one in your dataset.



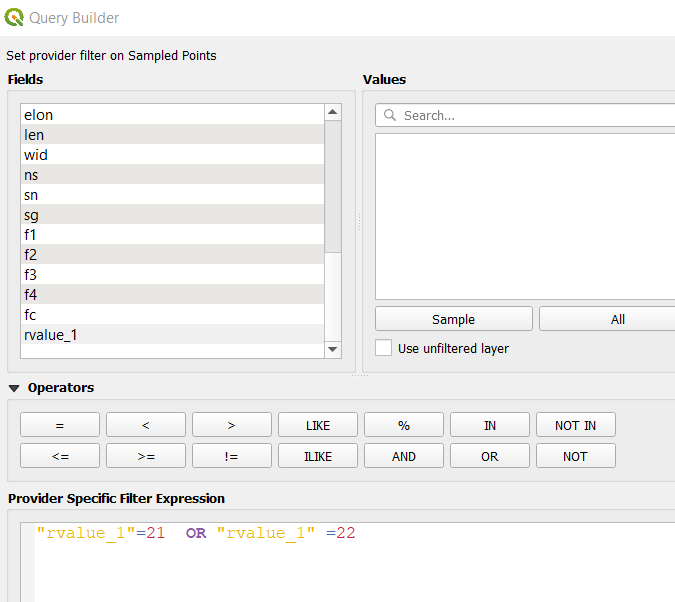
**3) Calculate the percentage of tornadoes in areas designated as developed in the NLCD**

How many of these tornados are in developed areas? To calculate this, you’ll need two numbers: the total number of tornados and the count of tornados with codes 21-24. The first number is easy. It’s the total count of observations in your tornado dataset. The second can be calculated in multiple ways, but in this case you can use a ***filter***. This tool removes all rows that don’t fit a given set of criteria.

Right-click on the Sampled Points layer in QGIS and choose filter.



This will open the Query Builder window, which allows you to set the filtering criteria using the SQL programming language. This isn’t too complicated. For example, the entry below filters for only tornados with codes 21 and 22.



Adapt this code so that you include codes 21, 22, 23, and 24 in your filter. Count the number of points in this filtered dataset, and you’ll be able to say what percentage of all points have these values.

For reference, here’s a count of pixel values for the entire NLCD file:

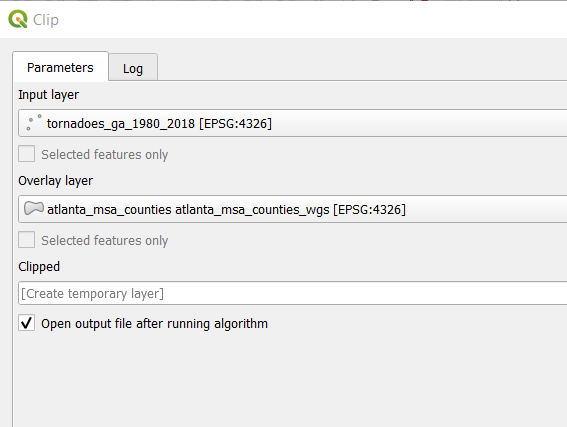
| Pixel value | Count of pixels |  | Pixel value | Count of pixels |
| --- | --- | --- | --- | --- |
| 11 | 3,316,130 |  | 43 | 11,402,674 |
| 21 | 9,566,652 |  | 52 | 5,867,867 |
| 22 | 4,800,590 |  | 71 | 7,919,851 |
| 23 | 1,557,884 |  | 81 | 11,843,231 |
| 24 | 716,061 |  | 82 | 20,210,598 |
| 31 | 421,631 |  | 90 | 26,776,414 |
| 41 | 22,516,932 |  | 95 | 2,750,118 |
| 42 | 39,873,744 |  | Total | 169,540,377 |

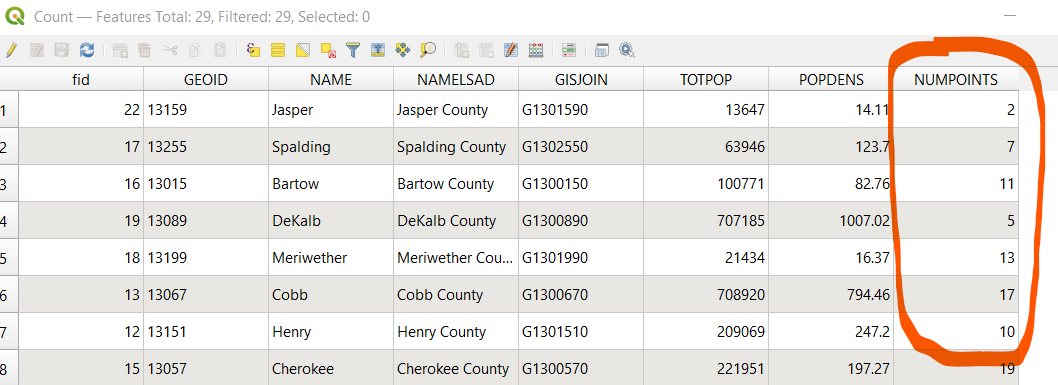
**4) Count the number of tornadoes in each Atlanta county using a spatial join**

In the second half of this lab, you will focus specifically on tornadoes in the Atlanta metro. Load the geopackage you downloaded with the Atlanta metro counties. This will be very similar to Lab 2.

To count tornadoes within each county, you will need to do a ***spatial join***. You did an attribute join in Lab 2, but a spatial join connects two layers that have a spatial relationship, such as points within polygons. There’s a few different tools for doing this in QGIS, but you want the ***Count points in polygon*** tool in the Processing toolbox. The tool is straightforward--indicate the polygons you want to use (Atlanta counties) and the points you’re counting within those polygons (the **original** tornado dataset). Leave all the other fields the same, unless you choose to save the result on your local computer.

When this tool is done, open the attribute table of the new county layer. You should see a NUMPOINTS field with a count of tornadoes by county.





**5) Make a choropleth map showing the number of tornadoes per county**

You’re almost done. The last task is to create a map with the results of step 4. You learned the basics of Print Composer in QGIS in Lab 2. Create a choropleth map of Atlanta counties that includes the following:

* The counties themselves, shaded by the number of tornadoes starting within it. Use a classification scheme that you think best fits the data.
* The origin points of tornadoes within these counties. You can subset these by using the ***Clip*** tool, using the original tornado dataset as your input layer and the county polygons as the Overlay layer (see below)

Your map should also include a suitable title, legend, and scale bar. Include credits that give your name and the source of the tornado data (NOAA). Make sure it is in the UTM 16N projection listed earlier in the lab. Do NOT include the NLCD data for this map.

**Lab deliverables**

Upload the map image to ELC when you submit your assignment. You’ll receive up to **15 points** depending on how closely it matches the specifications outlined above.

In addition, submit a Word document with answers to the following questions:

1. (5 points) According to this [optional reading](https://www.e-education.psu.edu/natureofgeoinfo/node/1892) from DiBiase’s text on remote sensing, the NLCD went from using unsupervised to supervised classification methods for the 2001 dataset. Based on that reading and our class lecture, explain **in your own words** what this means.
2. (4 points) According to [the metadata](https://www.mrlc.gov/data/legends/national-land-cover-database-2016-nlcd2016-legend), developed land is classified primarily by the amount of impervious surfaces it contains. Pick one other class of land use included in the NLCD (water, barren, wetland, etc.). What criteria are used to place land areas in that classification type?
3. (4 points) In step 3, you calculated the percentage of tornadoes originating in developed areas. What was that percentage? How does it compare to the percentage of pixels with that value for the whole state (see the end of section 3 above)?
4. (5 points) Pick one of the tornadoes with an origin point in a developed area and find an article that provides more information on it. You may have the best luck with tornadoes that either resulted in injuries (the inj variable), fatalities (the fat variable), or property damage (the loss variable). Provide a link to that article and briefly summarise anything additional you learned from it.
5. (4 points) Write a brief summary of the map you created in step 5. What spatial pattern (if any) do you see in this map? Be specific--give county names and/or exact point counts.
6. (3 points) Thinking of the modifiable areal unit problem discussed in class lectures, do counties seem appropriate for the trend you see in the point data? Would differently shaped or sized polygons potentially be a better fit? Give one specific justification for your answer.

For ***2 points extra credit***, submit questions 4 and 5 as a 1-2 minute video summary with your lab.