CUNY FemSTEM QGIS Hack: Subways and population in NYC – PART I

Set Up

Make sure that you have QGIS 2.18 installed on your computer. If you do not, please refer to the installation instructions (https://github.com/michellejm/Intro-QGIS-CUNY-FemSTEM/Install-QGIS.md). If you have 2.14, please update now by opening QGIS and clicking on the update link at the bottom. **Do NOT install 3.0** – that is the Early Adopter release, and it is not stable yet.

Download the GitHub repository for this course. Using the green button here (https://github.com/michellejm/Intro-QGIS-CUNY-FemSTEM), select `Download ZIP`. The Data folder has all of the datasets you need for our hack today.

Premise

For this hack, we are going to estimate how many people live within a 1/2 mile radius of subway stations in New York City. We will use our findings to identify "transportation deserts", as well as "transportation hot spots". We will represent our findings with a heatmap.

Data

For this hack, we will need four different datasets:

- A shapefile of New York City census tracts (provided)
- A shapefile of New York City neighborhoods (from NYC Open Data Portal)
- A shapefile of the subway stations (provided)
- Population data (from the Census) that can be used with our shapefile of census tracts (provided)

Notes on data

When working with census data from the US Census, there are a few levels of granularity that you can look at: State, County, Census Tracts, and Block Group. Some areas have intermediate levels that may be tabulated or reported by the census, such as Neighborhood Tabulation Areas, or Metropolitan Regions. Usually, these are the result of combining smaller areas to make intermediate areas in regions of high population density.

Census Tracts are areas created by the census for analytic purposes, and Block Groups are, just as described, groups of city blocks, smaller than census tracts. When trying to make population estimates of any kind, it is best practice to use the smallest area available for the data you are interested in (to protect privacy, some data is only available at the Tract or County level). However, even though population data is available at the block group level, we are going to use census tracts because this project comes out of a larger project on languages, education, and income in New York City, and those datasets are only available at the census tract level.

Finding Data

Census data is published by the US Government in their annual American Community Survey Report. You can find historic census data by doing a data query at NHGIS. However, because we are really only interested in New York City, we are able to use the NYC Open Data Portal. In recent

years, many municipalities have established these data portals to allow students, researchers, and enthusiasts easy acces to relatively (sometimes very) clean data.

Set Up QGIS

Launch QGIS. Your new blank map project will look like this:

Begin to familiarize yourself with the interface. For more information, refer to this [brief description (https://github.com/CenterForSpatialResearch/MappingForTheUrbanHumanities/blob/master/Reso](https://github.com/CenterForSpatialResearch/MappingForTheUrbanHumanities/blob/master/Reso) of the elements of the interface.

Adding Layers

In order to construct our map within QGIS we will need to add our data layers to the map project. There are several ways to accomplish this. We will use the `Add Vector Layer` button. Click on `Add Vector Layer`

Navigate to the Data Folder. There are a number of different file extensions that may be unfamiliar. The files outlined in magenta are elements of the nyct2010 shapefile (which represents the 2010 census tracts. The files outlined in blue are components of the nynta shapefile (which represents the neighborhood tabulation areas). It is very important that all of these files stay together in the same folder otherwise QGIS will not be able to load the layer.

- .shp – The main file that stores the feature geometry (required).
- .shx – The index file that stores the index of the feature geometry (required).
- .dbf – The dBASE table that stores the attribute information of features (required).
- .sbn and .sbx – The files that store the spatial index of features.
- .prj – The file that stores the coordinate system information.
- For more information on these extensions and others see [this explanation by ESRI (http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Shapefile_file_extensions)](http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Shapefile_file_extensions).

Select the `nyc_2010.shp` file. A map should appear in the data window of QGIS (yours will likely be a different color)

SAVE YOUR PROJECT!! Make a Project Folder in the FemSTEM Folder that you downloaded. QGIS saves *relative* file paths, so if you move a file, it will break the project. BUT if you move the *Entire* FemSTEM Folder, it will be ok.

Projections

Our map looks a little skewed – this isn't quite how we are used to looking at New York City. There is an issue with the projection. Projections are sets of mathematical transformations that turn a bumpy 3D figure (usually the Earth) into a 2D map. Something is always sacrificed in this

transformation, but we can limit the distortion by selecting a projection that is optimized to the region we are concerned with.

We want to optimize our map by selecting a projection best suited to New York City. For a detailed discussion of projections, see here (http://www.geo.hunter.cuny.edu/~jochen/gtech201/lectures/lec6concepts/map%20coordinate%2

Navigate to Project >> Project Properties

Click on CRS at the left. Make sure that `Enable 'on the fly' CRS transformation` is enabled. Search for NAD_1983_StatePlane_New_York_East_FIPS_3101_Feet (ESPG: 102715). **Click Apply and then OK**

Notice that the number in the bottom right hand corner reflects the new CRS.

Now we will add all of the other shapefiles.

Again, **click** on the `Add Vector Layer` button, and add the

- `nyc_nta.shp,`
- `mta_stations,`
- `mta_lines, and`
- `nycwater.`

Once you add all of the files, your map should look something like this. If all 5 layers appear in the Layers Panel, but you can't see the subway lines or the stations, rearrange the order by dragging and dropping them in the Layers Panel.

Now we want to add the population data.

Click on the `Add Delimited Text Layer` button, and make the following selections:

- File Name: FemSTEM/Data/Tabular/nyc_population-by-tract_2016.csv
- File format: CSV
- Geometry definition: No geometry
- OK

It should appear in the layers panel, but not in the map since there is no geometry associated with it.

SAVE YOUR PROJECT

Joins

Our census tracts don't have any population data, so we want to **join** population data **to** the census tracts. **Right click** on the acs_shape layer and open the Attribute Table. Notice the GEOID column – it starts with 14000000US and has some other numbers. Remember this.

Now open the Attribute Table for the nyc_population_tracts_2016. Notice the GEO.id column. It is the same code. This is a unique identifier that identifies each tract. We can use this to our advantage. Close the Attribute Table.

Right Click on the `acs_shape` file again and select `Properties`

- Select **Joins**
- Click the '+' button at the bottom
- Make the following selections:

    - Join layer: `nyc_population-by-tract_2016`
    - Join field: GEO.id
    - Target field: GEO_ID
    - Custom Field name prefix: delete contents

- Click `Apply` then OK

Open the attribute table again and you should see the population data joined in the far right column. This join is only saved locally, so if we wanted to use this file again, we will have to save it.

Right click on the acs_shape layer and select `Save As`. Save your layer as and ESRI Shape File named 'nyc_tracts_popJoin`.

We can now "turn off" the acs_shape layer by unselecting its box.

SAVE YOUR PROJECT

Visualization

We want to make a simple chloropleth map illustrating population in New York City. We will do this with Syles. In the process, we will clean up some of our layers.

- Make sure that the nycwater layer is on the top. The purpose of this layer is to essentially hide the fact that our census tracts extend into the water (where no people live), and makes our map look a little odd. We could clip to the shoreline, but that is unnecessary for this hack.

- Move the two subway layers under the nyc water layer.

- Move the nyc_nta layer under the transportation layers. (the purpose of the nta layer is to give us a visual reference to NYC – something to help ground us in what we are looking at)

- Move the nyc_tracts_popJoin layer under the nyc_nta layer

- **Right Click** on the nycwater layer and select Properties and then the 'Style' Tab. Click on `Simple Fill` and change the color to something a little more water-like, and adjust the outline color if oyu like (you can also make this white). Click `Apply` and `OK`

- **Right Click** on the nyc_nta layer and select Properties and then the 'Style' Tab. Click on `Simple Fill` and change the style to `Simple Line`.

- **Right Click** on the nyc_tracts_popJoin layer and select Properties and then the 'Style' Tab. Click on `Single symbol` at the top, and change it to `Graduated Symbol`.

  - Make the Column: Population
  - Pick your color ramp
  - Select the Mode: Jenks
  - **Click** `Classify`
  - **Click** `Apply` and `OK`

Designing a map

To present this map, we will now compose a map layout using the QGIS map composer. The map composer allows you to add a legend, north arrow and scale bar as well as to export our work.

YOu can move your map around, reorder the layers and change the colors. Once you are pleased with your map composition we will create a new print composer. When prompted you can either name the print composer or not.

To add a new map to the composer select the *add new map* button. Then click once and drag a rectangle over the area that you would like the map to occupy and click again to stop. Whatever is showing in your QGIS map project window when you place this map in the print composer is what will appear in the new map.

Next, add a legend. Select Add New Legend, and again click to draw a rectangle where you would like to place the legend. An unformatted legend that matches the information from the Layers panel will appear. You can use the options in the Item Properties tab (highlighted in blue) to change which layers are represented in the legend and to change the labeling of the layers in the legend. Scroll within this tab to familiarize yourself with which properties about the legend you can change.

Format the legend and change the titles of each dataset so they are more descriptive. Un-click "Auto update" to make changes, then change the layer names by clicking the "legend item properties" button.

Next we will add a scale bar. Click the add new scale bar button. Again you will be able to change the properties of the scale bar, including the style, number of segments and size.

Last we will add two text boxes, one with a title for the map and another with abbreviated citations for our data sources. Click the add new label button then use the Main properties field to add the text you want, and use the Font button to change the text size and font.

Finally use one of the export options circled in blue above to save the map composition as an image file, PDF, or SVG.

You can modify the subway lines and stations as you wish. In the next half of the tutorial, we will take this Project and use it to figure out how many people live within 1/2 mile of the subways.

Tutorial written by Michelle McSweeney, PhD, a Research Fellow in the Center for Spatial Research (http://cunywomeninstem.org) at Columbia University. Taught at *CUNY FemSTEM HackDay*, one day hack-a-thon by the CUNY Women in STEM (http://cunywomeninstem.org). More information about the hack day is available here (http://cunywomeninstem.org/hackday/).