GIS Level 2 Exercise using QGIS

MIT GIS Services: libguides.mit.edu/gis

Introduction

Covered in this workshop are a range of tools that are useful in a variety of applications and disciplines. This exercise will be partly performed using the Processing Modeler tool in QGIS, which allows users to develop a workflow of their methodology. This workflow can be saved and shared on different machines and with other users so that a process that would normally comprise multiple steps is collapsed into one tool.

Tools this workshop covers include:
- Query Builder
- Create a new field in the attribute table and use the Field Calculator to create an expression
- Add XY Data and Projections
- Creating a Model and Clip, Buffer, and Intersect
- Dissolve
- Select by Location and Select by Attributes
- Spatial Join
- Create and Edit Features
- Optional bonus exercise: Geocoding with QGIS

Getting Started

Move Data to your Computer
1. Create a folder on your desktop and name it your last name.
2. Download the example data.

Add Data

GIS is commonly used for site selection analysis. In this exercise we will use tools to select a site for a new apartment building in New York City.

Much of the data used in this workshop is available for download on the MIT Geodata Repository (http://arrowsmith.mit.edu/mitogp) We are also going to use some zoning data from the New York City Department of City Planning website. To save time we have already downloaded this free data and placed it the folder that you've copied onto your desktop. All the data is in a specific projection that minimizes distortion of shapes and areas. The projection can be found in the metadata, which can be read in the MIT Geodata Repository. If you search for the layers we will be using, you will find they are in NAD 1983 StatePlane New York Long Isl FIPS 3104 (US feet), a projection well suited for the area we are studying. In QGIS this projection should be set before the data is added.

1. From the Start Menu, All Programs, open QGIS Desktop 2.12.3.
2. In the top menu open Project > Project Properties and click on the CRS tab.
   - Enable 'on the fly' CRS transformation.
   - In the Filter box, type NAD stateplane new York.
   - From the coordinate reference systems of the world, select NAD_1983_StatePlane_New_York_Long_Island_FIPS_3104_Feet EPSG:102718
   - It should show up in the Selected CRS: box.
- Click OK.
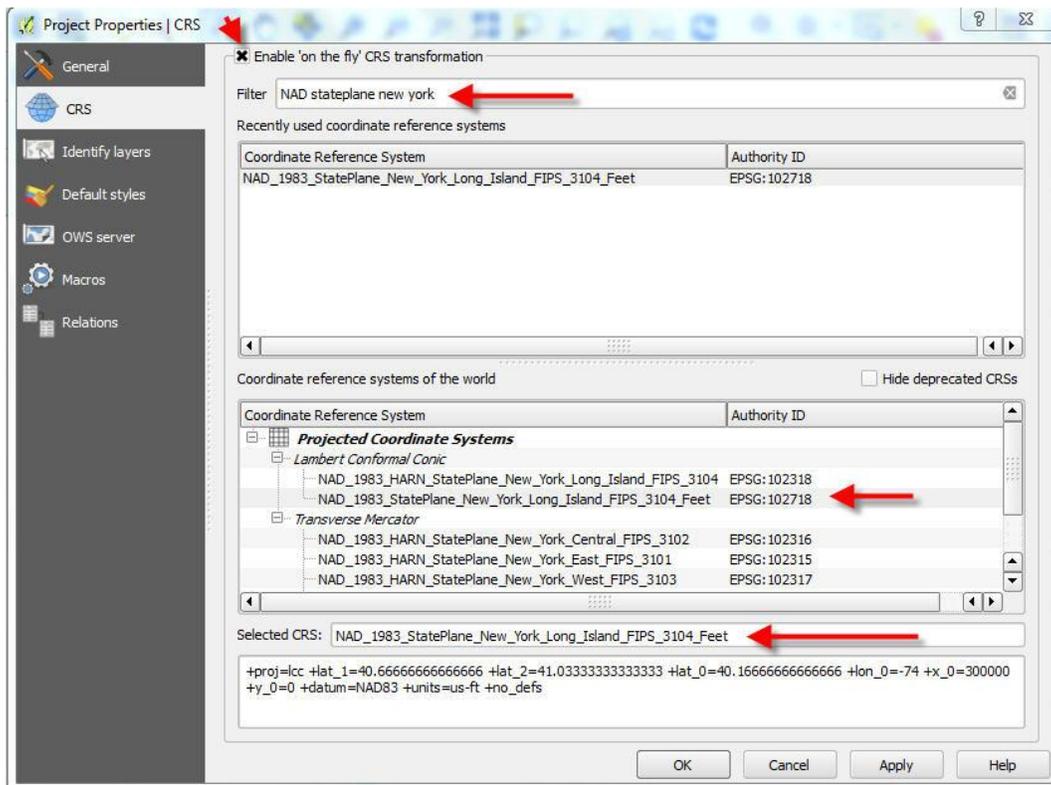
3. Click the Add Vector data button ( ) in QGIS.
4. In the Add vector layer dialog box, click on the Browse button and navigate to your folder on the desktop.

5. Add from the following to your map (these may be in 2 separate folders):
   - TXLOTS.shp
   - SCHOOLDIST.shp
   - BOROBNDRYS.shp
   - nyzd.shp

These shapefiles represent Manhattan tax lot boundaries, New York City school district boundaries, New York City borough boundaries, and New York City zoning data, respectively.

QGIS doesn't see the projection information contained in the ESRI shapefile .prj file when these layers are added. We know from metadata that these layers are in a projection for New York City and Long Island. It is the NAD_1983_StatePlane_New_York_Long_Island_FIPS_3104_Feet we just chose within the Project Properties. We want to work with these layers together, so we will select this projection for each layer.

6. Right click on the nyzd layer name and choose Set Layer CRS (Coordinate Reference System). In the Filter box, search for NAD 1983 New York. Choose NAD_1983_StatePlane_New_York_Long_Island_FIPS_3104_Feet.
7. Repeat this step for each of the other three layers.
8. Set each layer's Coordinate Reference System to the proper NAD_1983_StatePlane_New_York_Long_Island_FIPS_3104_Feet. Save your map document in your working directory (Project→ Save As). Save it in the folder you created on the desktop. Whenever working in GIS, make sure to save early and often.

Query Builder

A layer's query builder provides the option to display and interact with a subset of a layer based on an expression. The layer's query builder makes no permanent changes to your underlying data nor does it create a new layer.

We want to find vacant land for our apartment building. Vacant land is a land use category used by the New York City planning department.

Every data layer in the MIT Geodata Repository (and typically GIS data from other sources as well) includes metadata. Metadata provides information about the data layer such as what any attributes and codes mean, who created the file and when, information to use when you cite the data, and whether there are use constraints for the file. To explore the metadata for the tax lots file, you can navigate to this direct link:
https://arrowsmith.mit.edu/mitogp/layer/MIT.SDE_DATA.US_NY_NYC_MN_G47TXLOTS_2011/ . If you scroll down and expand the entity and attribute information section, you will find in this metadata that a land use (LANDUSE) value of 11 represents vacant land.

1. Move your tax lot layer (the layer named TXLOTS) to the top of the drawing order (click and drag in the Table of Contents). Turn off the other three layers so only TXLOTS has an x next to it. This allows you to see the Tax lot layer alone.
2. Double click on the layer name of your tax lots to open the layer properties.
3. Click on the General tab and click on Query Builder near the bottom right side of the dialog box. You may have to scroll down to see the button.

4. Double click on LANDUSE in the list of fields. The field name will be added to the expression below.
5. Add an equal sign (=) to the expression.
6. Now click on All and double click 11 to add this to the expression.
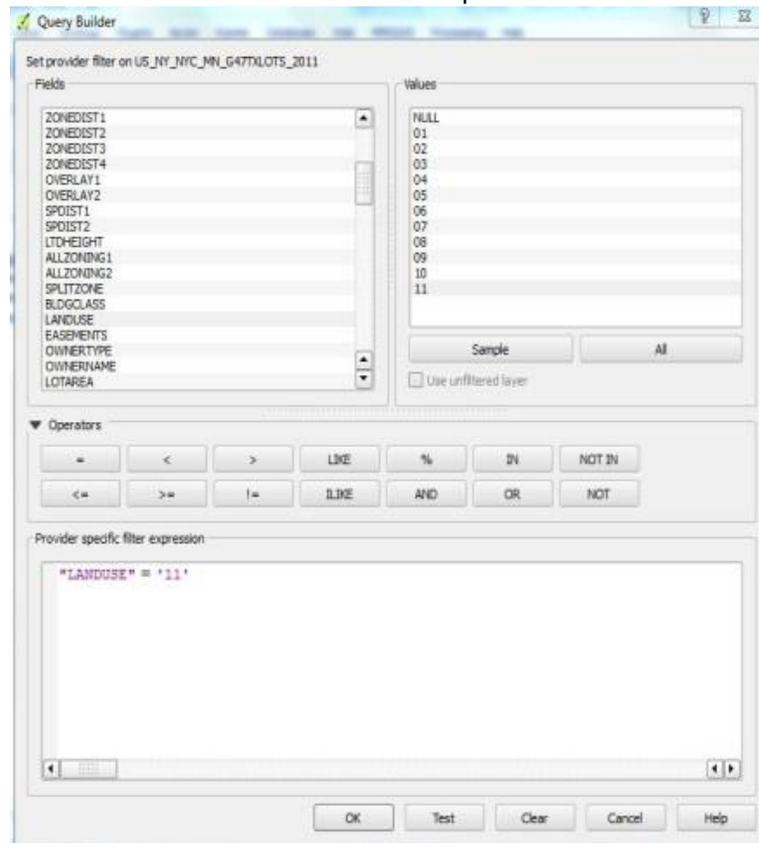
7. Click OK > OK. As you can see, the only parcels that remain are ones that are currently vacant. If you have trouble seeing the remaining parcels, double click on the color under the layer name and choose a brighter color, such as yellow.
8. Turn on the borough boundaries layer by clicking on the box next to the layer name (BOROBNDRYS). Using the same process as above, restrict the boroughs to Manhattan using the field name BORONAME.

Through research you've decided School District 3 should be excluded from the areas of consideration for your site selection.

9. Using the same process, restrict the school districts layer (layer named SCHOOLDIST) to anything but (!=) School District 3. The expression should look like "SCHOOLDIST" != 3 . The final result should look something like this.
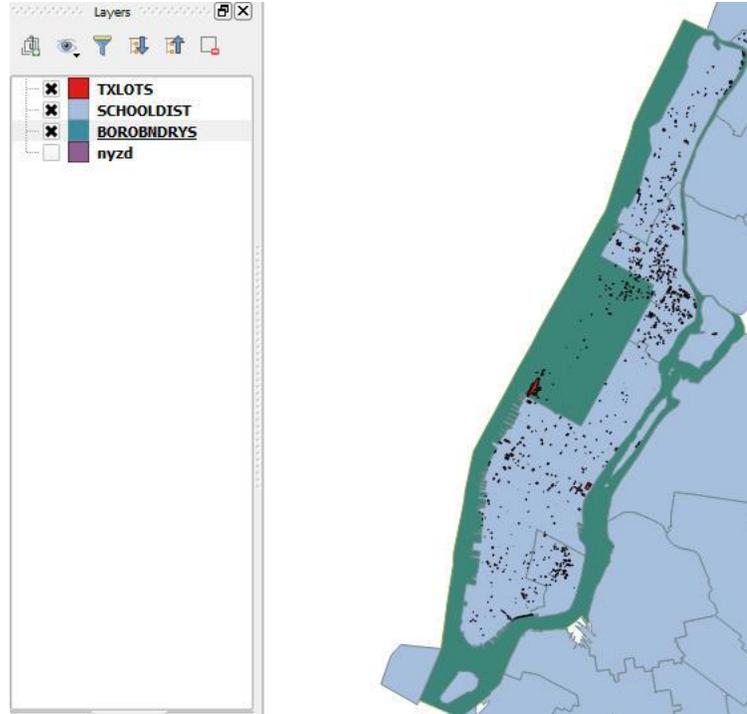
Create a new field in the attribute table and use the Field Calculator to create an expression

We also want to restrict to areas that have been zoned for residential use. If you open the attribute table (right click data layer nyzd), you'll notice that ZONEDIST gives information on zoning that is too specific for our purposes. We just want to know the first letter (R = residential).

1. At the top left of the attribute table box, click the pencil icon 🖉 (toggle editing).
   a. Click the Field Calculator icon 📠.
   b. Check the box next to create a new field.
   c. Make the output field name ZONECAT and output field type text.
   d. Make the output field width 1.
   We want ZONECAT to contain the first letter of ZONEDIST.
2. In the Functions list, click on the + next to String, and double click on left.
3. Type ZONEDIST next to the parenthesis, then type a comma, a space, then a 1, and click on the right parenthesis icon. Your expression should look like this: `left(ZONEDIST, 1)`
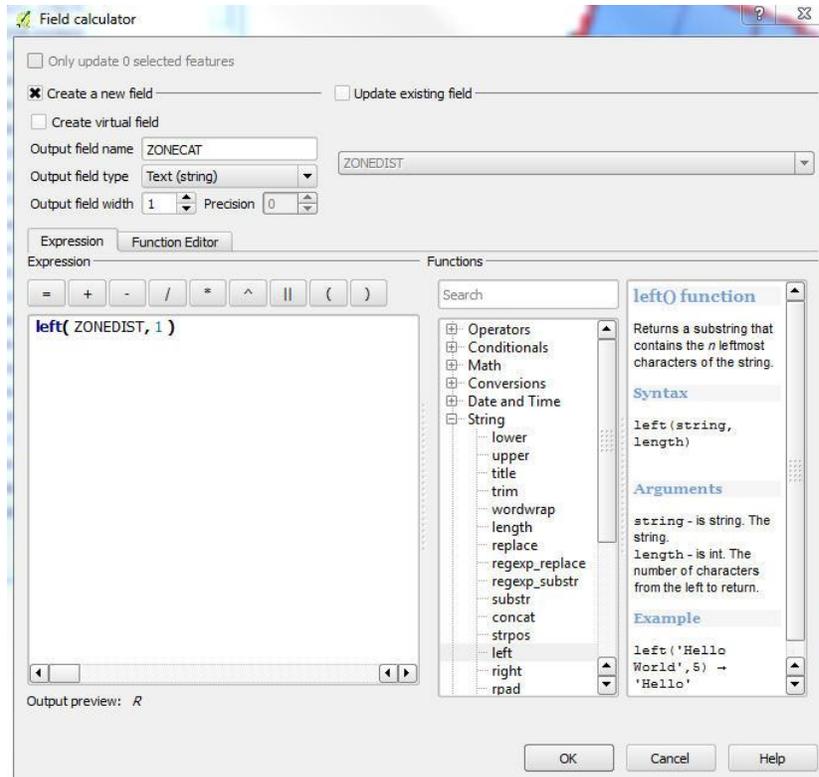
4. Click Ok. Click the Save Edits icon [icon]. Click the toggle editing icon [icon] again to turn off editing, close the table.
5. Use the query builder in the layer's properties to see where the zoning category (ZONECAT) is R (residential). Once you have set up the query builder, close the nyzd properties dialog box.

Add XY Data and Projections

QGIS is also able to display points based on X and Y fields in a table. In our site selection, we want to incorporate the location of subway stations in New York, but we only have a table of where the stations are.

1. Open the StationEntrances.csv in excel.

Scroll all the way to the right and observe the Latitude and Longitude columns. These numbers represent the locations of each subway station entrance. Since our XY data is in latitude and longitude, we know that these coordinates are based on an unprojected, degree-based coordinate system. However, these numbers are each missing a decimal after the second digit, so we'll fix this before continuing. To create usable data, the numbers need to be expressed as decimal degrees. We will create two new columns for our work.

2. Add two new column headings called X and Y in columns AA and AB.
3. In row 2 of the column under X, create this expression: =Z2/1000000
4. Copy this expression. Click in row 3 of the column under X and hold down Ctrl + Shift + End at the same time to extend the selection to the last cell used on the sheet in this column. Paste the copied expression. If you don't have an End key on your keyboard, highlight the cell in

which you just created an expression, hover over the bottom right corner of the cell until you see a solid + sign, and double click.

5. Repeat this for Y, but using the Latitude field instead. Make sure the Xs are in the -70s and the Ys are in the 40s (otherwise your data would map in the wrong hemisphere). Save the table as StationEntrances2.csv. If you are using a Mac, save the file as a .csv windows (near the bottom of the list when you Save As). If you don't save it this way you won't be able to open it in QGIS.

6. Close the file after saving.

7. In the QGIS project, click on Layer (in the top menu bar) and then Add Layer > Add Delimited Text Layer. Browse to StationEntrances2. Choose X for the X field and Y for the Y field. Click OK.
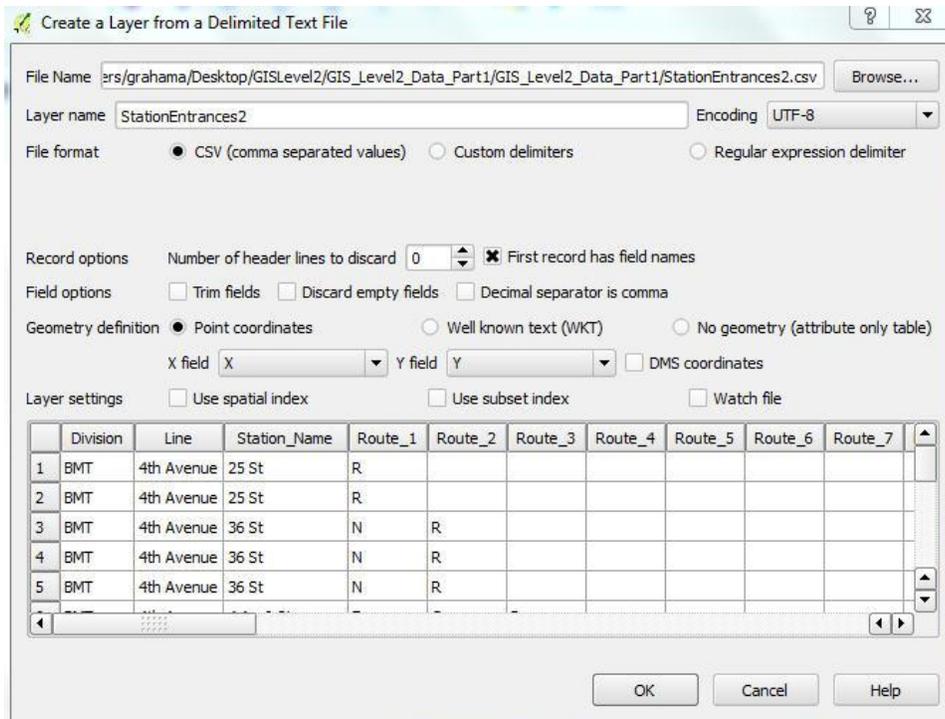
The Coordinate Reference System Selector will open so that you can specify the coordinate system.

We already know that the data uses an unprojected coordinate system. We did not find metadata specifying which datum to use, but since we are in North America we will guess the North American Datum NAD83 was used. (Another common datum, especially for data outside North America is the World Geodetic System WGS84).

The data frame was set to the coordinate system we chose for our project, which was: NAD_1983_StatePlane_New_York_Long_Island_FIPS_3104_Feet – a coordinate system with the world **projected** onto a flat surface. QGIS detects that this layer has no projection specification. This file is using latitude longitude coordinates and is **unprojected.**

8. In the Filter box, type NAD 83, and choose NAD83 for the Coordinate Reference System, EPSG:4269. Click Ok.
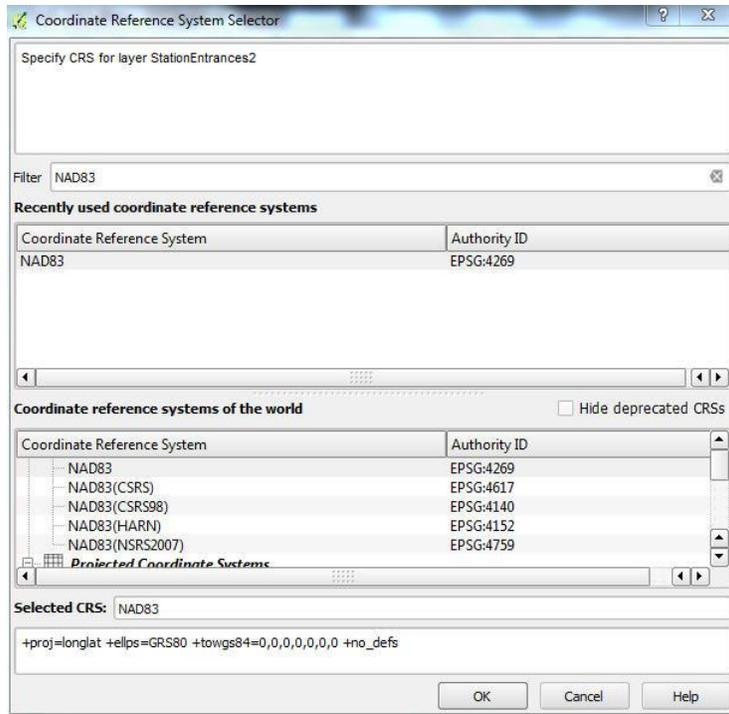
9. You now have an on-the-fly point layer, but this is not yet its own shapefile (It is still a .csv.), and it is still in an unprojected coordinate system. We will save it as a shapefile with the same projection as the other files.

   a. Right click on the new point layer (StationEntrances2) > Save as .
   b. Make the Format ESRI Shapefile.
   c. Browse to your folder and save the file as StationEntrances2.
   d. Choose the CRS that has EPSG:102718 – NAD_1983_StatePlane_ New_York_Long_Island_ FIPS_3014_Feet, which is the same projection you set at the beginning for the Project.
   e. Make sure Add saved file to map is checked.
10. Click OK.
11. Remove the StationEntrances2.csv from the table of contents by right clicking and selecting Remove. Hover over the file name to see the extension if you have trouble identifying the older layer.
12. Save your map document, especially if you haven't done so recently.
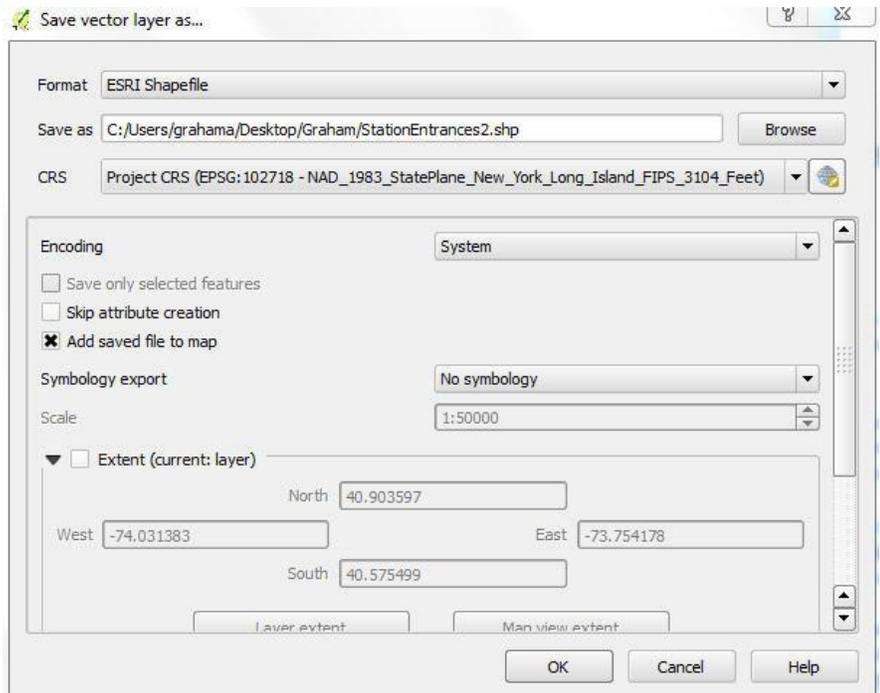
More information on displaying coordinate data and understanding map projections can be found through the GIS Services Website under Learn GIS > Tutorials and Help Resources:
http://libguides.mit.edu/content.php?pid=347508&sid=2844938

Clip

A clip analysis creates a new layer from the features of another layer that are completely inside a polygon feature. We are going to clip the Subway Stations (Station Entrances2), the zoning districts (nyzd) and the School districts (SCHOOLDIST) files to the borough of Manhattan, which we chose as the only borough to show within the BOROBNDRYS file.

1. Open the clip tool (Vector > Geoprocessing Tools > Clip).
2. Select SCHOOLDOST as the input layer and the borough boundaries as the clip layer.
3. Click the Browse button and navigate to your folder on the desktop. Save the layer as clippedschools.
4. Click OK.
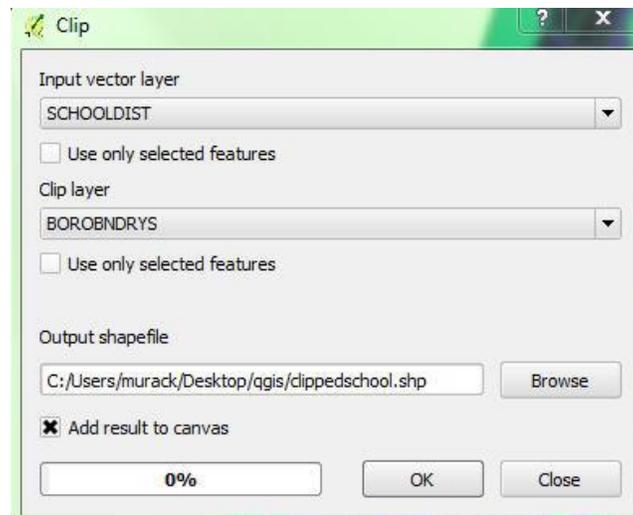
5. The new data layer will be added to your QGIS project. The Clip tool will remain open so **repeat** this process with the **station entrances** and **zoning districts** (nyzd).
6. Close the Clip tool after you are finished.

Processing modeler: Buffer, Intersect, and Dissolve

Create a Model

1. Click on Processing > Toolbox.
   a. In the toolbox, expand the Models menu and then the Tools menu.
   b. Double click on Create new model.
   c. Enter the model name model1 and the group name group.
2. Click on the Save as icon ⊞ and save your model called Model1.
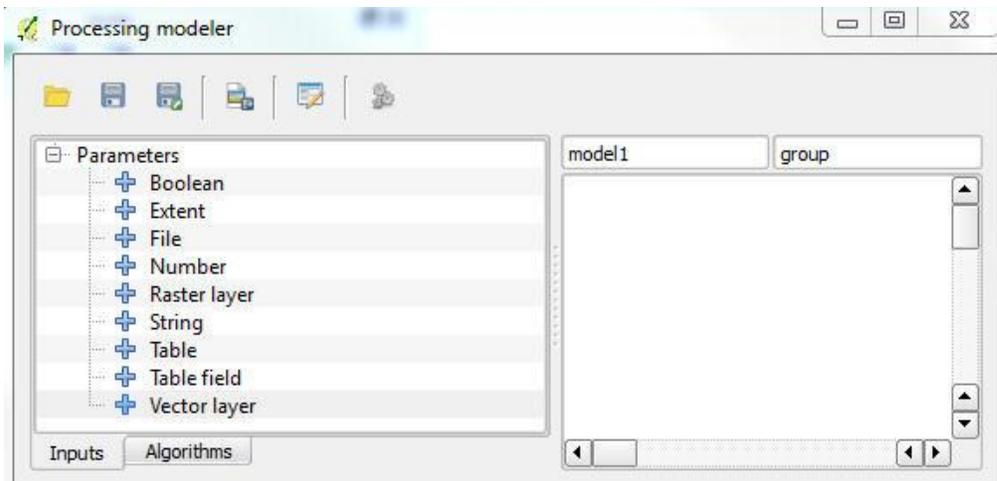
If you ever need to come back to your model after closing it, find it in the toolbox, right click on model1, and click Edit model.
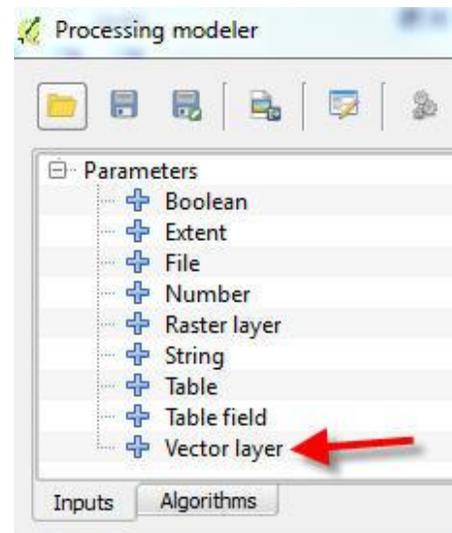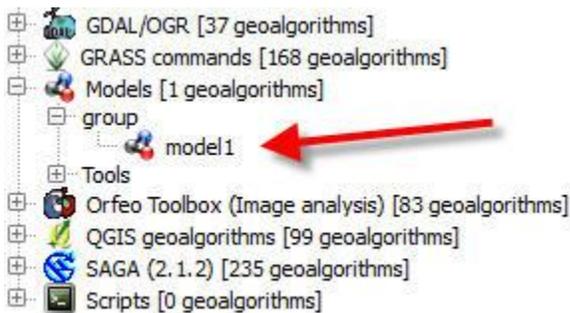
3. You need to tell the modeler which vector layers you will use.
   a. Double click on **+** Vector layer.
   b. Enter a parameter name that is similar to the name of each clipped data layer. The file names do not need to be exact as long as you recognize them.
   c. Do this for **EACH** clipped layer you just created.
   d. The subways should be "Shape type: point", while the zoning districts and school districts are polygons.
4. Save the model by clicking on the Save icon .
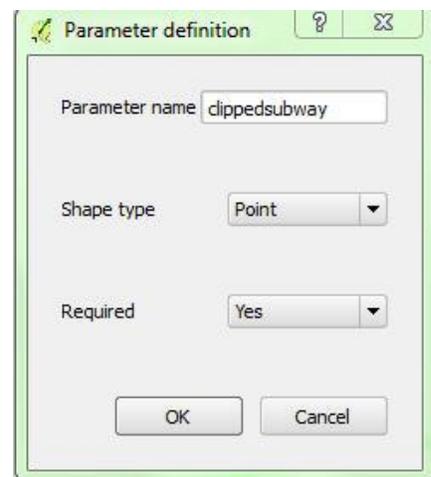
Buffer

A buffer analysis creates a polygon feature based on a specified distance around an input feature. It is therefore useful for finding out what is around something. We want to perform a buffer analysis on the subway stations, as we want a site that is within 500 feet of a station entrance.

1. Add the Fixed distance buffer tool (Algorithms tab > QGIS geoalgorithms>Vector geometry tools > Fixed distance buffer) to the model window.
   a. Your input layer is the clipped subway stations.
   b. Name the file subway_stations_clip_buffer.
   c. The distance unit is 500 feet
   d. Leave Segments at 5.0.
   e. The dissolve result is Yes. This means that we don't care about keeping individual features; we just want a shape to show what space is within 500 feet of a subway station entrance.
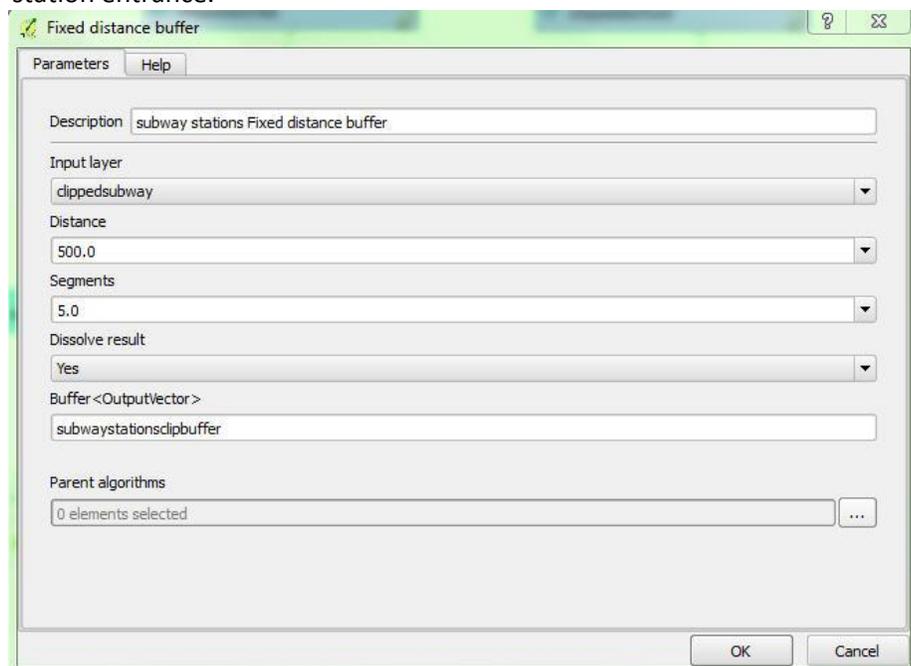
2. Click OK.

The buffer output will be a polygon layer representing all space within 500 feet of a Manhattan subway station entrance.
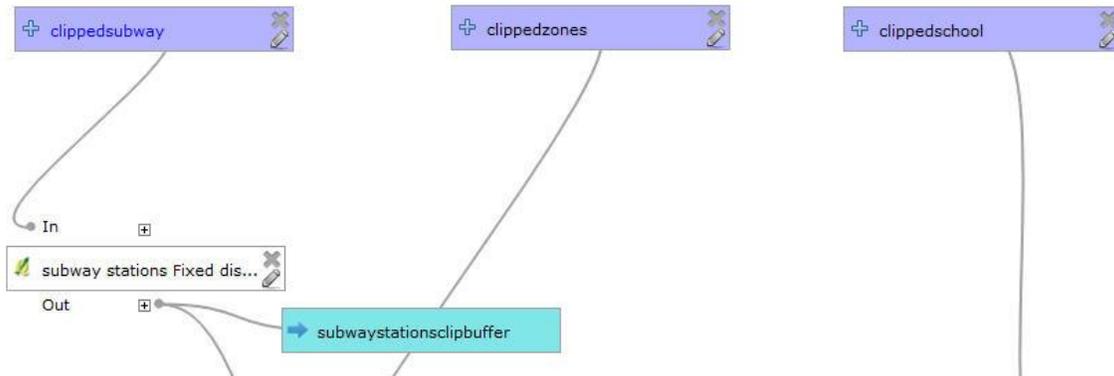
Intersect

An intersect analysis creates a layer based on where two or more features overlap. This will be the final step in the Processing modeler. The intended result is a layer describing the area in Manhattan that satisfies the following criteria: within 500 feet of a subway station entrance, NOT in School District 3, and zoned residential use.

1. Add the intersect tool (Algorithms tab > QGIS geoalgorithms > Vector overlay tools > Intersection) to the model window. The Intersection box will automatically open so that you can to edit the parameters.

QGIS cannot intersect more than one file at a time, so we will set up 2 intersections.

2. For the first intersection, specify the following:
   a. Description: "Intersection Zones"
   b. Input layer: subway station buffers ('Buffer' from algorithm 'subway stations Fixed distance buffer')
   c. Intersect Layer: clipped zoning layer.
   d. Intersection <OutputVector>: "almost_suitable_area".
   e. Click OK.
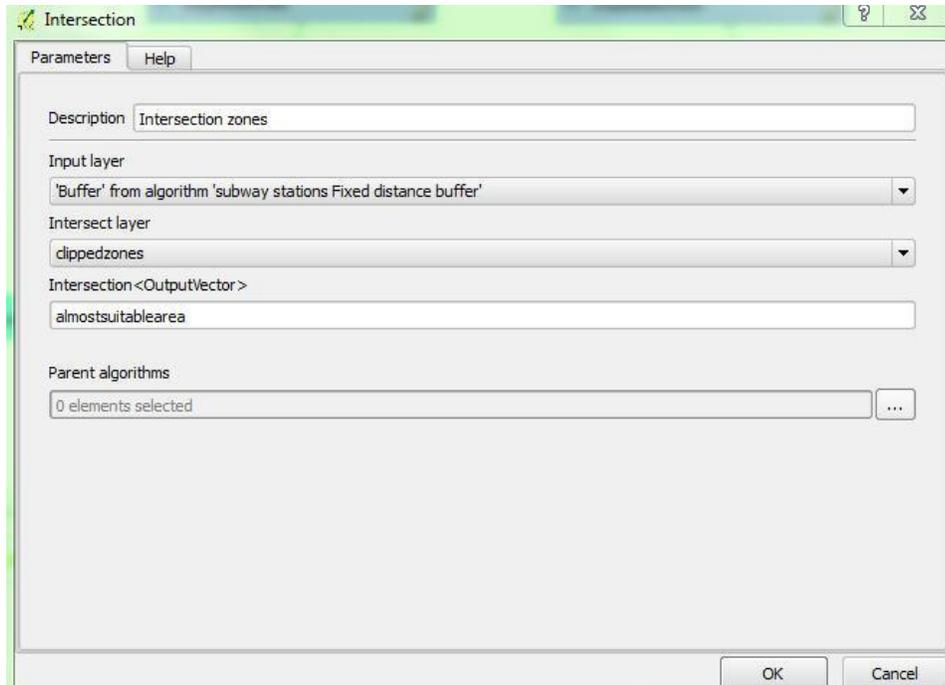
3. The second intersection will be the output of the first intersection and the clipped school district layer. Specify the following:
    a. Description: Intersection School
    b. Input layer: Clipped Schools
    c. Intersection layer: 'Intersection' from algorithm 'Intersection Zones'
    d. Intersection<OutputVector>: "suitable_area" .
    e. Click OK.

The output is the suitable area shapefile.
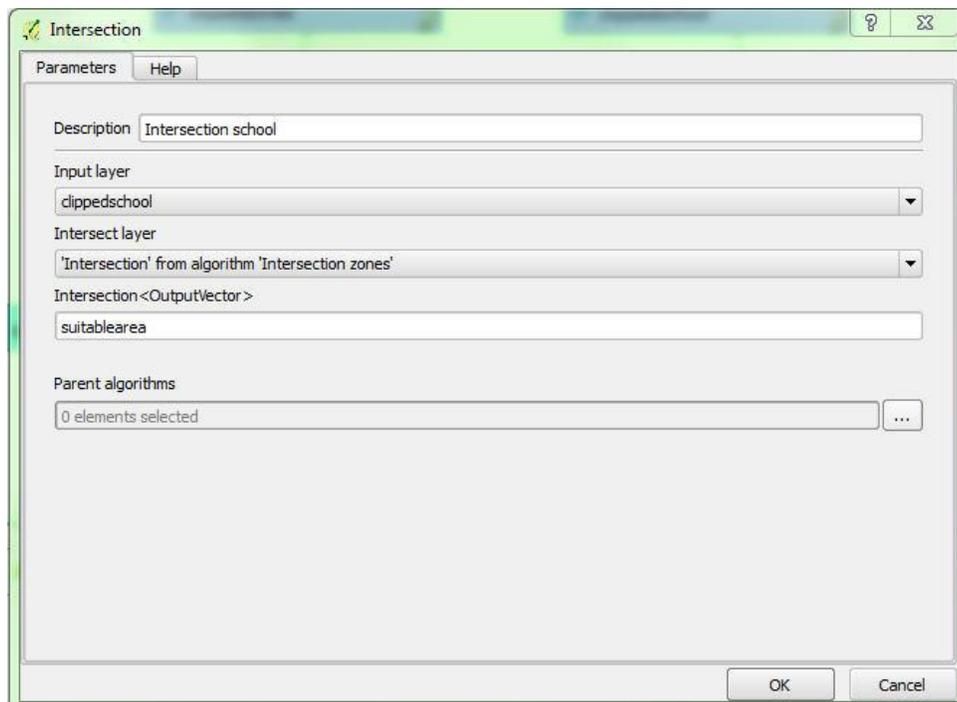
Dissolve

A dissolve allows a user to collapse features based on an attribute or set of attributes. In this case, a new layer will be created, where extra borders in suitable_area are removed and an aggregated boundary file based on the school district boundaries is created.

1. Add the dissolve tool (Algorithms tab > QGIS geoalgorithms > Vector geometry tools > Dissolve) to the model window. The Intersection box will automatically open so that you can to edit the parameters.
2. Specify the following:
    a. Input layer: The result of your second intersection ('Intersection" from algorithm 'Intersection school')
    b. Dissolve all (do not use field): No
    c. Unique ID field: SCHOOLDIST (type this exactly the same as in the attribute table)
    Choosing a dissolve field creates a separate feature for each distinct entry, which in this case is each unique school district. Our output file retains a distinction between suitable areas that fall in different school districts.
    d. Dissolved<OutputVector>: suitable_area_dissolve
3. Click Ok.
4. Save your model.
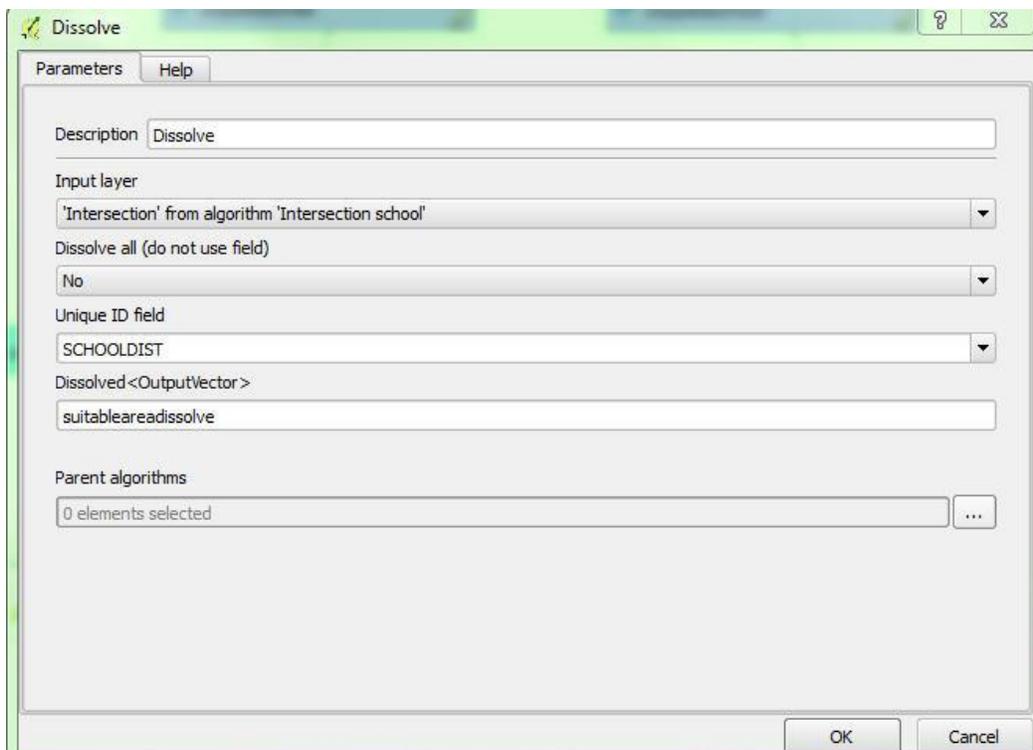
Run your model

Your model should look something like this:
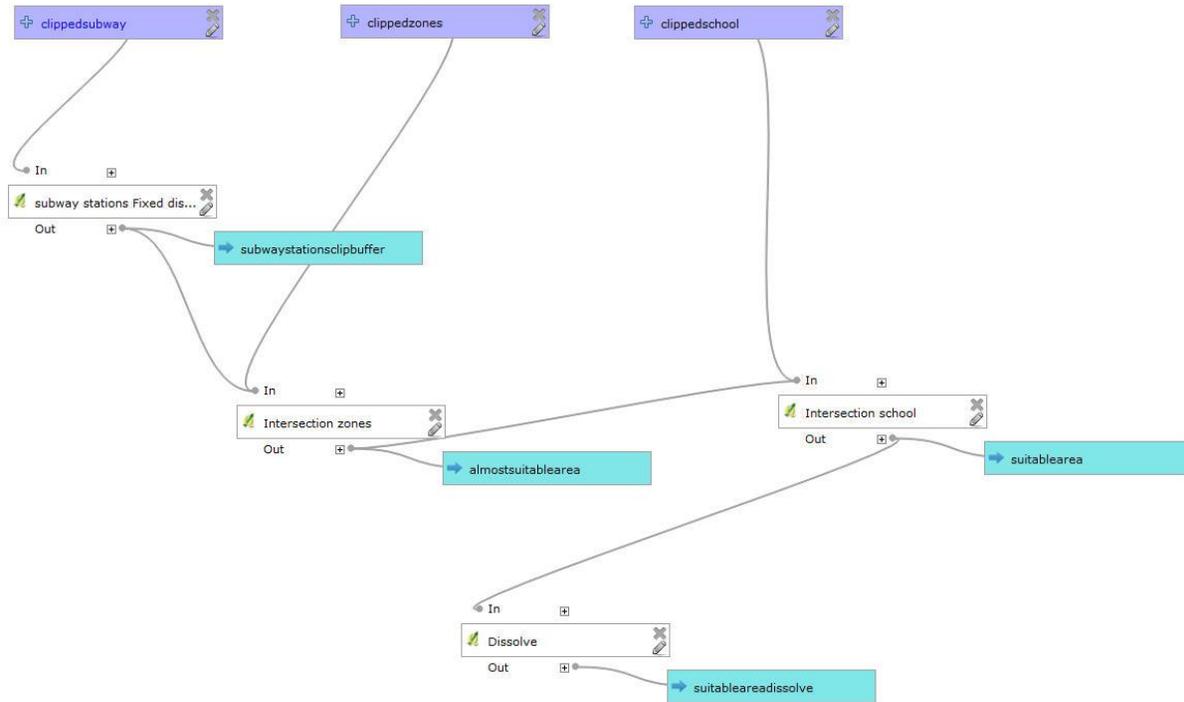
1. This is final model. Run it using the model run button ⚙️ .
2. A box will open and you will need to choose which of your data layers corresponds to each of your input layers. Select the appropriate data layers using the drop down menu (Choose clippedstations under clippedsubway, clippednyzd under clippedzones, etc.).
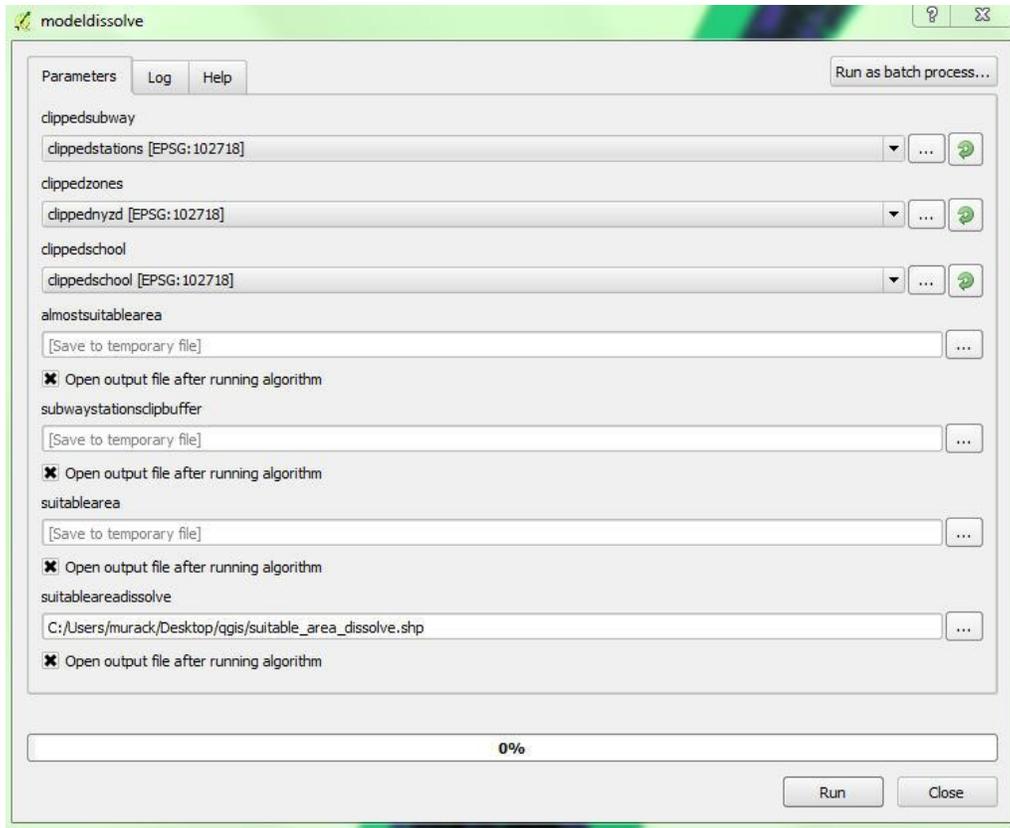
3. Under "suitableareadissolve" click the … button [...] and select Save to file… Since this is the final output we will save it to our working folder, rather than as a temporary file.
4. Click Run. One by one, the steps will be completed.

Back in the Processing modeler window, by clicking on the export as image icon, you can export a PNG so you can include an image of your model in documentation, like in your thesis. We will not export the model right now.

8. Save and close your model.
9. Turn off all layers except suitable_area_dissolve. Your output should look something like this.
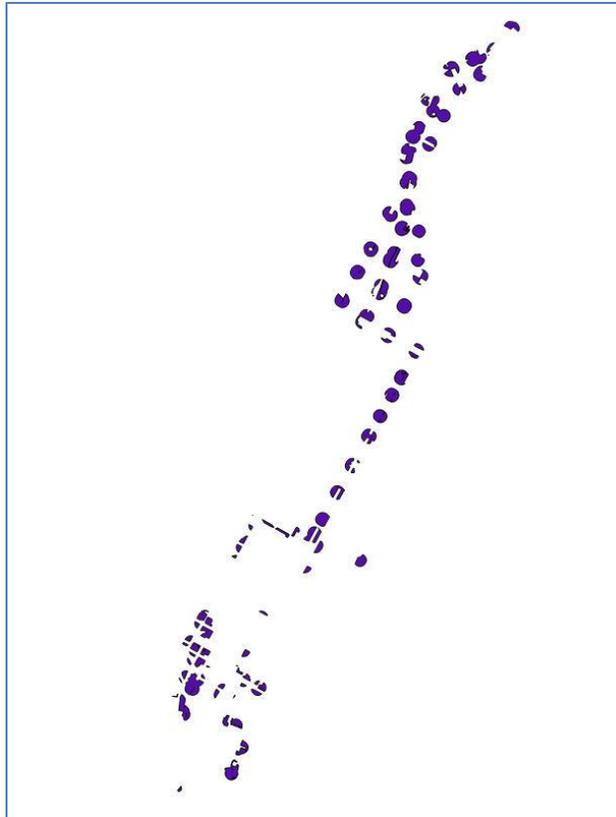
10. Open the attribute table and notice there is one row for each school district. This is because we dissolved the data layer using the school district column. Close the attribute table.

Select by Location and Select by Attributes

QGIS can also select features based on their spatial location relative to another layer. In this way, we can select vacant parcels that are completely within our suitable area.

1. Turn on the tax lot layer.
2. Click on Vector > Spatial Query
   a. Select features from tax lots (TXLOTS)
   b. Where the feature: Within

c. Reference features of: suitable_area_dissolve
d. And use the result to: Create new selection
3. Click Apply. The blank white box at the bottom will give you a report of the processing progress. When the selection process is finished, the number of selected features will be identified and listed.
4. Click on the button that makes a new layer of just the selected features.
5. Click Close.
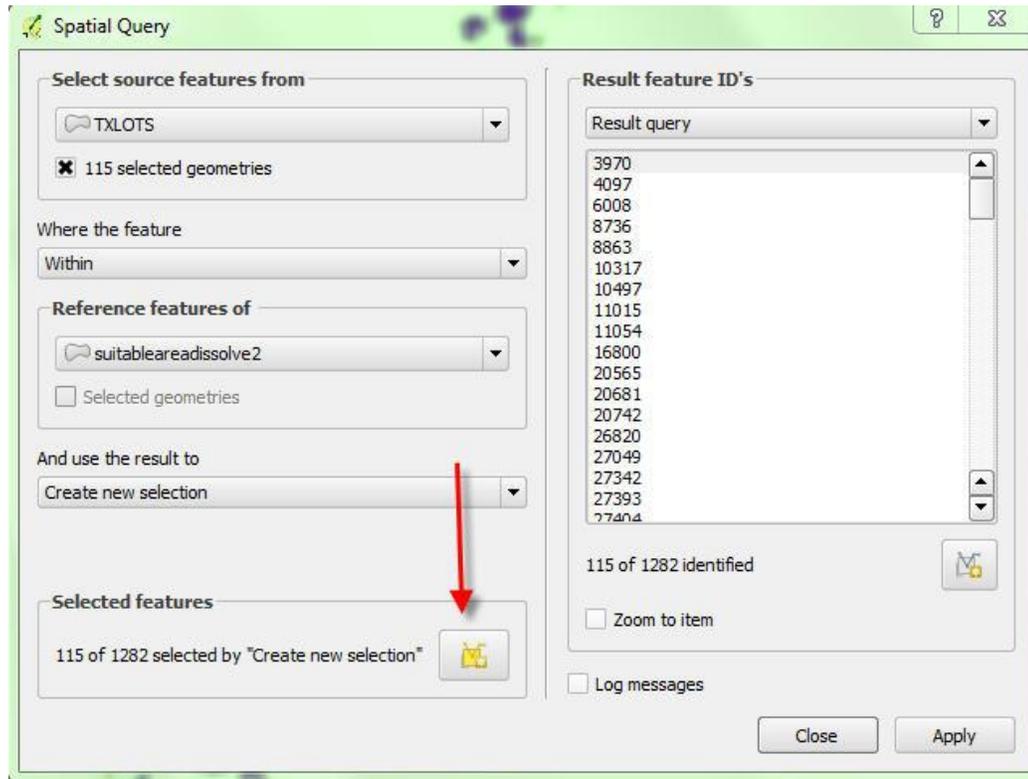
6. Open the attribute table of the new layer. At the top of the table box, notice that 115 vacant tax lots are suitable for our apartment building so far.

We have one more constraint to apply, and that is lot area. Our apartment building is going to take up a lot of space, so we need to select a parcel that is at least 10,000 square feet. Our tax lot layer already has an attribute called LOTAREA, expressed in square feet.

7. At the top of the attribute table, click the Select Features Using Expression button (   ).
8. The expression should be: `LOTAREA >= 10000`. Click Select. Now there are only 4 parcels to choose from which satisfy our criteria. Click Close.

9. Use the Zoom to selected features button   to see the yellow, selected lots on your map.
10. Right click on the TXLOTS selected layer. Select Save As. Name the new file TaxLots4GoodOnes.shp.
11. Place an X next to Save only selected features and Add saved file to map. Click OK.

Join attributes by location

We'd like to join the table information from the suitable areas with the tax lots, where the two files spatially intersect with each other.
1. Click on Vector > Data Management Tools > Join attributes by location.
2. Enter the following criteria in the Join attributes by location dialog box:
    a. Target vector layer: TaxLots4GoodOnes
    b. Join vector layer: suitable_area_dissolve
    c. Output Shapefile: save to your folder on the desktop and name it taxlots_suitablearea_join.shp
    d. Output table: Only keep matching records

3. Click OK, click OK if you get a CRS warning, and Yes to add the new layer to the Table of Contents. Close the Join attributes by location box.
4. Open the attribute table of taxlots_suitablearea_join.shp.

Notice this file only contains the 4 records that were selected in from the taxlots data layer. If we had not done the selection process first our output file would include all records from the taxlots file.

In addition to the many details that were in the tax lot file, we can now quickly also find out the school district where the tax lots are located. If we had instead run the spatial join with the undissolved suitable area shapefile as the join feature, we would see a row of information for each parcel in the school district.
5. Close the attribute table.

Create and Edit Features

QGIS also allows the creation and editing of shapefiles. In this section, we're going to create a new layer for the building footprint of our potential apartment complex.
1. Choose one of the suitable lots that you want to use for your apartment building and zoom in close to it.
2. Add an imagery base layer by clicking on Web > OpenLayers plugin.
3. Choose Bing Maps > Bing Aerial. OR Google Maps > Google Satellite.
    Important: If you do not see the OpenLayers plugin, click Plugins at the top of the screen and select Manage and Install Plugins. Select OpenLayers from the list of plugins. Click Install Plugin and close the window. Then you will be able to add imagery to the map as described above.
4. Move the imagery layer below the taxlots_suitablearea_join layer.

    If the imagery does not load or line up properly, proceed without it.

5. Make your taxlots_suitablearea_join layer translucent. This can be done by double clicking on the layer and choosing Style from the left hand menu and sliding the Transparency to something around 50%. Click Ok.
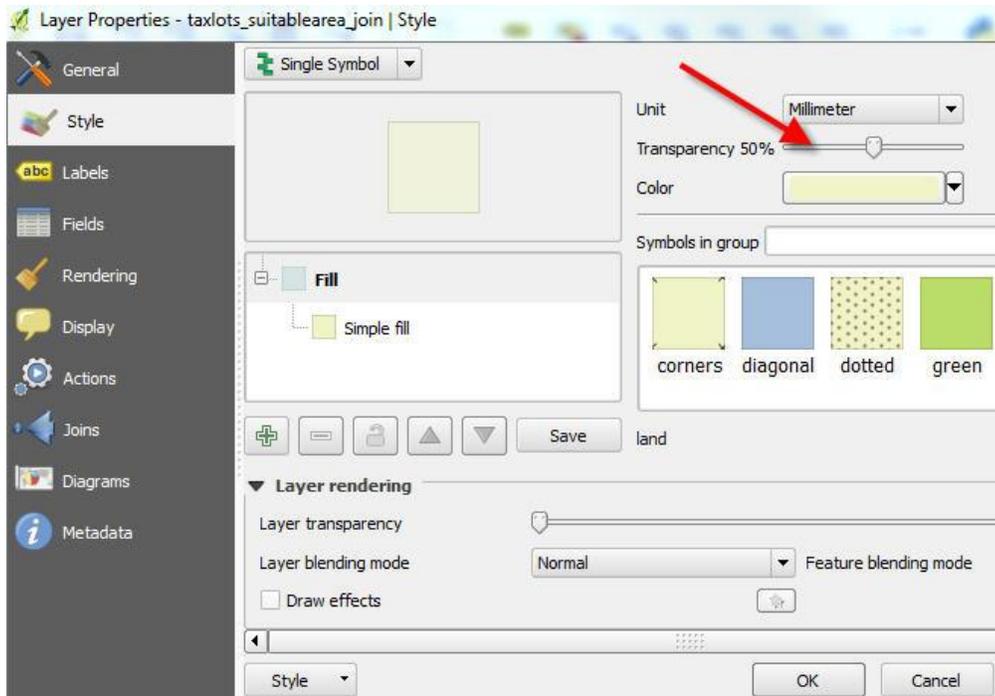
If the location you have chosen is unsuitable (one of them is over a sports field), go back (right click taxlots_suitablearea_join and click zoom to layer) and choose another. **If you cannot see the imagery properly, choose any lot.**

6. Create a new layer: Click Layer > Create Layer > New Shapefile Layer
7. Type is polygon, as we want to create a building footprint.
8. The Coordinate System (CRS) is the same as the other layers in our map. Click on the Coordinate

   Reference System Selector icon. Select
   NAD_1983_StatePlane_New_York_Long_Island_FIPS_3104_Feet
9. Click OK.
10. Name the attribute "building". Choose the type: Text data. Keep the width 80. Leave precision blank.
11. Click Add to attributes list. Click OK
12. Choose a name for your new shapefile and put it in the folder with your other data.

The layer will be automatically added to your map. Open the attribute table. There are no Features in the table. There is an attribute column for an id field that QGIS adds by default and the attribute column you created called outline. We're going to create features for this layer using the Editor.

1. Close the attribute table.
2. Right click on the buildings file name. Toggle editing.
3. Click on the add feature icon. The cursor icon will change to a circle.
4. One by one, click on the map where you want the corners of your building.
5. After you've finished the last corner, right click to finish the sketch.
6. Enter an ID number and a name for your building.
7. Unclick the toggle editing button and save your edits.

Calculate the area of your building

1. Open the attribute table of your buildings file. Click on the Field calculator icon.
2. Make sure Create a new field is checked.
3. In the 'Output field name' box, type "area"
4. Choose field type 'Decimal number (real)', field width 10 and precision 2.
5. From the Functions menu, open Geometry and double click $area to add it to the Expression box on the left. Click OK.
6. The area field will appear as a new column on the right. The area of your building will be calculated. Because the map projection is in feet, the area is in square feet.

Optional bonus exercises: Geocoding with QGIS

We have a list of addresses of other job sites that we can geocode to see their proximity to our proposed apartment building. In ArcMap you can geocode using a file from ESRI Data & Maps (available in the GIS lab and online on the same IS&T web site from which you can download ArcMap). Instructions for geocoding using ArcMap are available here:
http://libguides.mit.edu/ld.php?content_id=8586639

Geocoding in QGIS is even simpler, although slower, because of access to online geocoders.

Here is a brief rundown of the main differences between using ArcGIS and QGIS for geocoding.

- **QGIS Geocoding**
  - QGIS is free and has the advantage of being able to tap directly into Google or OpenStreetMap geocoding APIs, and does not require an association with MIT or ESRI.
  - Note that Google limits the number of addresses that can be geocoded for free (2,500 per day). There is no limit for OpenStreetMap.
  - Google and OpenStreetMap geocoding in QGIS can be used for addresses around the world, though accuracy may vary.
  - Easy to use, and relatively fast for small numbers of addresses. It can be slow for medium to large (>1,000) numbers of addresses.
- **ESRI Geocoding**
  - ESRI's geocoding service has many more customization options, and allows you the opportunity to interactively review and rematch address locations.
  - ESRI's geocoding service is much faster for large numbers of addresses.
  - ESRI geocoding is only available if you have access to a license of ArcGIS (i.e. active students, faculty, and staff at MIT)
  - At MIT, ESRI's geocoding service is only available for the United States and Canada.

1. If you don't see MMQGIS in the top menu, install and activate it by navigating to Plugins→ Manage and Install Plugins, and selecting and installing mmqgis. Then close the Plugins window.
2. Navigate to MMQGIS→ Geocode→Geocode CSV with Google / OpenStreetMap.
   a. For your input CSV file, browse and select current_projects.csv from your data folder. Note that any table you want to geocode must be in .csv format.
   b. Select an address field, city field, state field and country field using the drop down menu. QGIS may do this automatically, but make sure it has selected the correct fields.
   c. Select Google Maps as your Web Service.

d. Choose your working folder as the location for your output shapefile and not found output list.
e. Press OK.
3. Once complete, QGIS tells you at the bottom of the screen how many addresses were matched, and the geocoded addresses should have been automatically added to your map. Save your QGIS project.

Resources

- MIT GIS Services: http://libguides.mit.edu/gis
- NYC Department of City Planning: nyc.gov/html/dcp/home.html
    - BYTES of the BIG APPLE: nyc.gov/html/dcp/html/bytes/applbyte.shtml
- NYC Open Data: https://nycopendata.socrata.com/
- MTA Developer Resources: mta.info/developers/download.html

MIT OpenCourseWare
http://ocw.mit.edu

RES.STR-001 Geographic Information System (GIS)Tutorial
January IAP 2016