Lab Exercise #5: Geocoding and Onscreen digitizing and editing **Due Oct 3/2024**

CRP 4080: Introduction to Geographic Information Systems
Fall 2024

Instructor: Wenzheng Li (wl563)
Lab TAs: Gauri Nagpal (gn247), Anika Sinthy (ats243), Shubham Singh (ss3736)
Location: Sibley 305, Barclay Gibbs Jones Computer Lab
Points Possible: 40

Overview of Weekly Exercise:
1. Geocoding an address file
2. Create and edit shapefiles in ArcMap

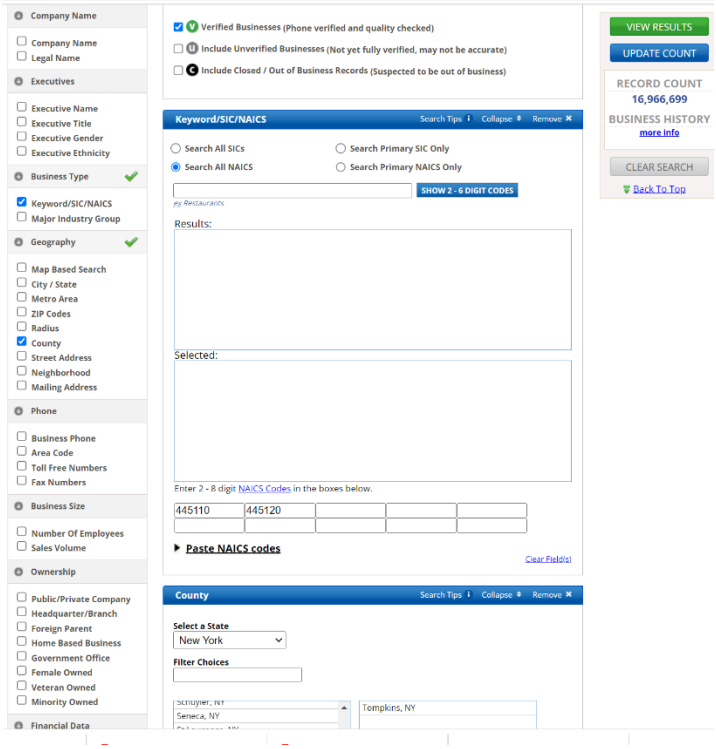
Part 1: The Geocoding process

Locating and plotting addresses and point locations is a fundamental function of GIS software. The process of spatially locating tabular address data is called **geocoding**. In the lab today, we are going to geocode a tabular dataset containing information on grocery stores in Tompkins County. We will download the data, detail a web-based geocoding workflow, then look in-depth at how to clean, parse, and ultimately geocode addresses using a Desktop GIS.

OBJECTIVE 1: OBTAIN DATA

Let's get started! The first step is to acquire a tabular address dataset. This dataset can be any tabular data that contains addresses for each feature. In this exercise, we are going to download a table of grocery and convenience store addresses from a business database called Data Axle (formerly Reference USA), which contains information on business locations and characteristics, and is a subscription service. Cornell has a subscription, and so do any major public libraries across the country. It is a valuable source of business-related information. Note if you are off campus, you may need to log in with your Cornell username.

Now click the subscription link, and click Connect to Data Axle (formerly Reference USA). Click on U.S. Businesses. Then select Advanced Search. The dialog below will open. In the Custom Search dialog, we want to specify the data we are looking for.
Fill out the form with the following:



- **Record Type:** Include only Verified Businesses
- **Business Type:** Keyword/SIC/NAICS - Search All NAICS. Enter the digit codes: 445110 (Supermarkets and Grocery Stores) and 445120 (Convenience Stores). See the entire listing of NAICS here: http://www.naics.com/search Note that we could also have searched using general terms (ie restaurants, etc.) if we did not know specific NAICS codes
- In the **geography** tab, click County and find Tompkins County, NY

Side Note: What is NAICS? From the Census website: "The North American Industry Classification System (NAICS, pronounced Nakes) was developed under the direction and guidance of the Office of Management and Budget (OMB) as the standard for use by Federal statistical agencies in classifying business establishments for the collection, tabulation, presentation, and analysis of statistical data describing the U.S. economy. […] NAICS is based on a production-oriented concept, meaning that it groups establishments into industries according to similarity in the processes used to produce goods or services."

By clicking 'Update Count', we see 35 results, but as this database is regularly updated, you may get slightly more or fewer results. It will contain listings for all the verified grocery stores and convenience stores in Tompkins County. Click 'View Results" - notice that not all of the locations have complete attributes; there are a few stores that do not have specific address information. This could present an issue later when we try geocoding. Select all data points (be careful, you may have some on the second page), and Click Download to download a tabular table of the data and save it to your working directory.

After clicking download, select:

- Comma Delimited
- Summary
- Download Records and open up the table.

OBJECTIVE 2: FORMAT DATA

The first step towards working with tabular data is cleaning and preparing the dataset. To do this, open the file in Microsoft Excel. You may see a warning message reading 'The file format and extension of 'Summary[timestamp].xls don't match. The file could be corrupted or unsafe. Unless you trust its source, don't open it. Do you want to open it anyway?' This is fine – go ahead and open it.

When you open the data, notice that there is an **'Address'** field that contains the street number and name. Also note there are **City, State, and Zip Code** fields. These fields are necessary for our geocoding, as they contain the information we need to create an Address Locator and locate points. Other fields might be useful as well, though we won't necessarily use them this time around.

Delete the Fax Number Combined, IUSA Number, Primary SIC Code, Primary SIC Description, SIC Code 1, and Record Type field.

Finally, we also need to create **a unique identifying column**. Create a new column called 'ID' at the left of your sheet, place a number 1 in the first row. In the second row, use the formula **=A2+1** to increment by 1. Then fill this to the bottom of the table. Click the column, copy it, right click the same column, and select Paste Special > Values.

We also need the address to be in one field, so we need to format it to include Address, City, State, and Zip Code. Combining several fields into one is called concatenation.

- Create a new field called **Address2**.
- Click on the cell and copy the following formula so that the address is in one line.
- =CONCATENATE(F2,", ",G2,", ",H2,", ",I2)

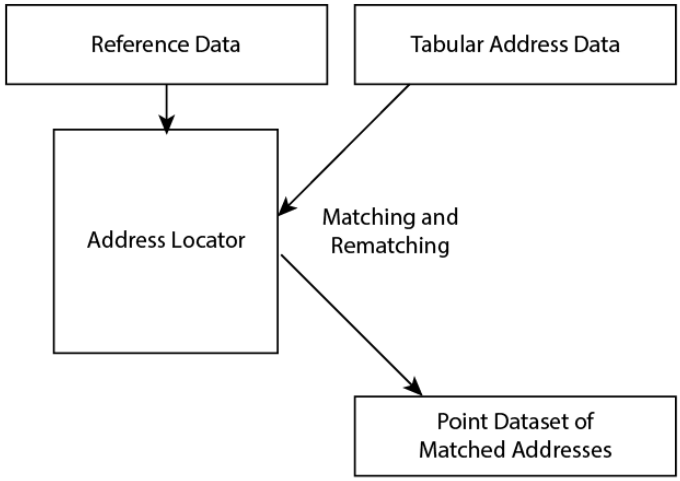
Notice that formula has added spaces and comma, so the program knows the differences between the town, state, and zip. Double-click on the right-hand side of the cell to make the formula go all the way down to the addresses.

- Now we need to make this concatenated address permanent. Same as before, select the column, copy it, right click the same column, and select Paste Special > Values. You will know you did it correctly if you don't see a formula when you click in any of the Address2 cells.

- Save the prepared file as '**Tompkins_stores_cleaned.csv'**. If you get a message box saying that 'some features in your workbook might be lost, click 'yes', and save the file. This will be our starting point for additional cleaning work!

PART 2: PARSING ADDRESSES AND GEOCODING

In order to geocode, you need to have two sources of information: **a table containing addresses**, and a **reference file or database that contains streets and information about their name, address number ranges, and zip codes**. With these two files, you can create what is called an **address locator**, which is a functional

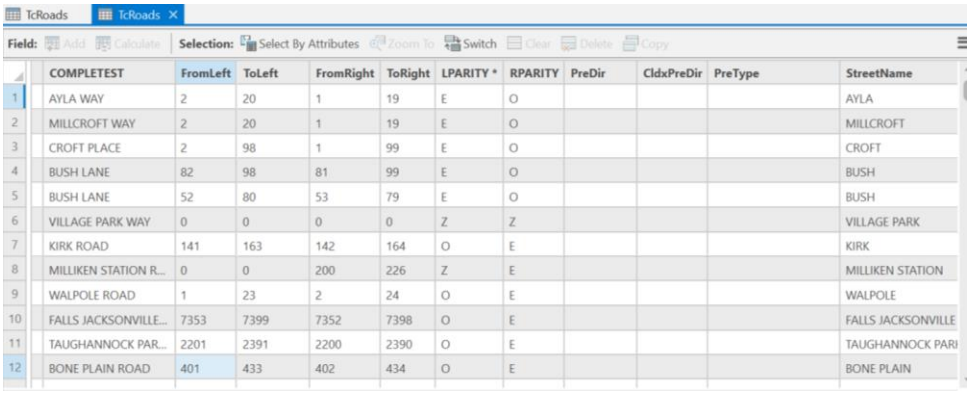


tool in ArcGIS that runs geocoding based on criteria you set. The address locator will match tabular address data to a reference file. The address locator will not match everything, so we must match unmatched addresses manually using a troubleshooting process. Finally, we can export our geocoded points to a shapefile. The workflow for Address Locators is based on the following model.

OBJECTIVE 2: CREATE AN ARCGIS ADDRESS LOCATOR

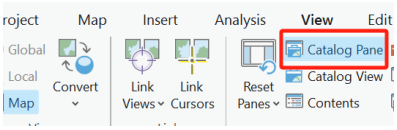
To begin geocoding the data we must now create an Address Locator in ArcGIS. Address Locators tell the software how to convert tabular data into spatial points using a reference file. The reference file contains line work with an address range and street name attributes, and interpolates where our address is located along the line work in the network.

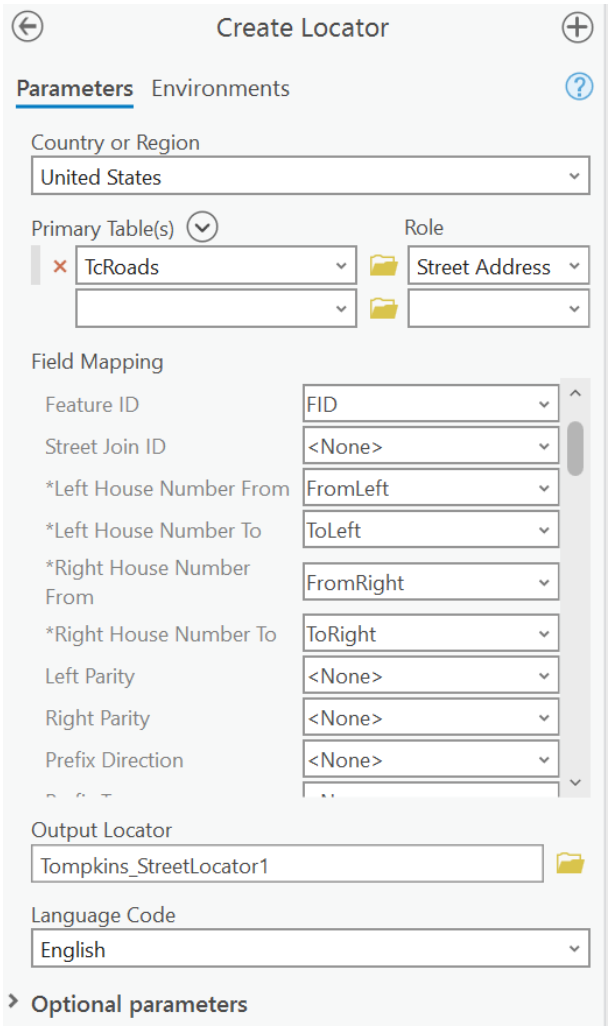
- Open ArcMap and create a new project
- Select 'Add Data' and go to the ArcGIS Online portal. DO a search for Tompkins County Road centerlines (the owner is TompkinsGIS) and add the layer (there will be a number of similar layers). TcRoads should appear in your TOC. Open the attribute table. You should see attribute information that indicates the geometry has been geocoded (FromLeft, ToLeft, etc.)



	COMPLETEST	FromLeft	ToLeft	FromRight	ToRight	LPARITY	RPARITY	PreDir	CldxPreDir	PreType	StreetName
1	AYLA WAY	2	20	1	19	E	O				AYLA
2	MILLCROFT WAY	2	20	1	19	E	O				MILLCROFT
3	CROFT PLACE	2	98	1	99	E	O				CROFT
4	BUSH LANE	82	98	81	99	E	O				BUSH
5	BUSH LANE	52	80	53	79	E	O				BUSH
6	VILLAGE PARK WAY	0	0	0	0	Z	Z				VILLAGE PARK
7	KIRK ROAD	141	163	142	164	O	E				KIRK
8	MILLIKEN STATION R...	0	0	200	226	Z	E				MILLIKEN STATION
9	WALPOLE ROAD	1	23	2	24	O	E				WALPOLE
10	FALLS JACKSONVILLE...	7353	7399	7352	7398	O	E				FALLS JACKSONVILLE
11	TAUGHANNOCK PAR...	2201	2391	2200	2390	O	E				TAUGHANNOCK PARI
12	BONE PLAIN ROAD	401	433	402	434	O	E				BONE PLAIN

- Open the ArcCatalog Pane. Under folders, you should see your project folder, right click the folder and select New -> Locator.  The Create Address Locator dialog will appear.





Give your output Locator a name (**Tompkins_StreetLocator**) and save it into your folder.

Under Primary Table, select Tcroads. Set the role for it to 'Street Address'

Under Field Mapping, we match fields in our reference dataset to components of the Address Locator as follows (you can double check the attribute table to be sure):

Field Name	**Alias Name**
Left House Number From	From Left
Left House Number to	ToLeft
Right House Number From	FromRight
Left House Number to	ToRight
Street Name	Full Street
Left ZIP	ZipLeft
Right ZIP	ZipRight (since there is only one zip code, its not necessary to complete both)

Click Run. The Locator should be created with warnings.

OBJECTIVE 4: GEOCODE ADDRESSES

In ArcMap, add the tabular address data you just created (**Tompkins_stores_cleaned.csv'**). In the Table of Contents, right click on the tabular grocery address data, select **Geocode Table.** An outline of the process is visible. Click Start.

Step One: What locator are you using? A dialog will appear allowing you to choose a geocoding service. A list will appear as well. If you don't see the locator you just created - Click Add - and browse to the Address Locator we just created. Add it, highlight it in the dialog, and click OK to select it as our Address Locator service for the geocode. Click Next.

Step Two: About your Table. The input Table should be selected. Remember that we concatenated the Addresses, so select 'One Field' in the drop down below (note that you can also open the table to check). Click Next.

Step Three: Mapping the fields in your table. Select 'Address2' as the Data Field. Select Next.

Step Four: Output. Select a name and location for the output feature class. Leave the Preferred Location type as ‘Address’ and the output Fields as ‘All’. Make sure ‘Add output to map after completion’ is checked.

Skip the Optional Step and select ‘Finish’. Review the workflow and select ‘Run’

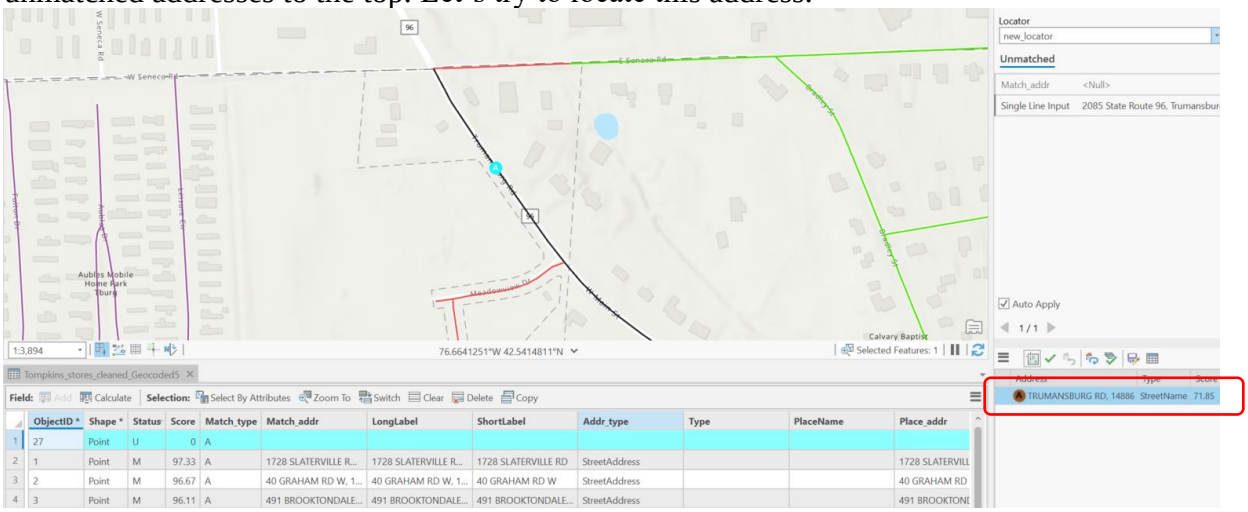
Click OK. This will return you to the previous, main geocoding dialog. Click OK again and the geocoding process will start. This may take some time to run. The Address Locator is locating your addresses.

When complete, addresses should be tied to points. A dialog will appear that will show how many addresses were matched (34), how many were tied (0), and how many were not matched (1) (again, yours may be somewhat different). In the Geocoding table, these will be marked either with a ‘U’ or an ‘M’ accordingly. These are very good results! We are given the option to begin the rematch process.

Start the Rematch process to match all unmatched addresses.

The **Interactive Rematch** window will open. Rematching is an iterative process of troubleshooting. We’ll be using it to manually rematch records that didn’t have an exact match during the automatic geocoding process.

Select **Unmatched Addresses** to only see the addresses with no match.
In the attribute table, right click on Status and select ‘Sort descending’. This will bring the unmatched addresses to the top. Let’s try to locate this address.



There is not much information, but scrolling through the table, we find that this is the Trumansburg Shur-Save, located on route 96 (a quick google search reveals that this is the correct address). Selecting the location brings up some possible candidates in the lower right. In this case, there is only one on Trumansburg Road.

This could be the source of the issue: the road has 2 official names: State Route 96 and Trumansburg Road (not an uncommon occurrence!). Once we are confident, we can click the

green arrow to match the candidate. Once selected with the green check, the attribute information will update, and the Status will change from ‘U’ to ‘M’.

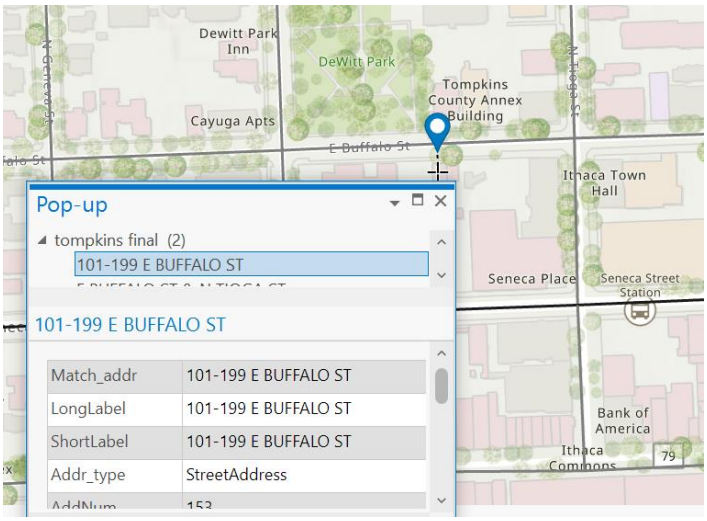
Note: So far, you might still have an unmatched record with the address listed as PO Box 1010, Ithaca, NY, 14851. This is a very vague and difficult-to-locate address, so you can simply delete it in the attribute table.

Export your result to a new point shapefile by right-clicking on your ‘Geocoding Result’ file. Data -> Export Features. Make sure to save the file as a Shapefile into your workspace.

Notes on Geocoding

1. While the geocoding worked quite well (no doubt due to the quality of the street data), there can be instances where the street data is poor, and you have many more unmatched or tied candidates. In this case, you may have to adjust some of the settings on your address locator to allow for more errors. To do so, in the ArcCatalog tab, right click on your Address Locator and select Properties. Select Geocoding options. Note that the default minimum match score is currently set to 75; you could adjust it downward to better account for spelling and syntax differences. You could also lower the minimum candidate score in order to not miss any possible matches.

2. If you feel there are mistakes with the attribute table address, you can locate businesses interactively. In the map extent, right click on the lot or building, and go to ‘Whats here?’ A pop-up will then provide the rough address.



Visual discrepancies between a street address and the actual location of a building is a relatively common error to find while geocoding.

*If at any point during the exercise you closed out of the **Interactive Rematch** dialog. You can get back to it by opening up the Geocoding Toolbar. You can find it by right clicking the layer you are working on, Data-> Rematch Address. You can then click on the ‘Review/ Rematch’ icon to reopen the Interactive Match Dialog.*

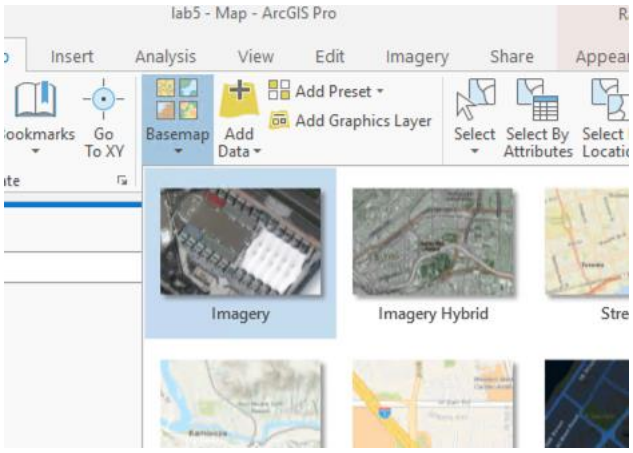
Map 1: Create a map layout of all geocoded points for Tompkins County. Label with ‘Company Name’. You can turn off the roads layer to make the final layout more readable.

Part 2: On-screen digitizing and creating new data

In this section, we will utilize onscreen screen digitizing to edit and create shapefiles in ArcGIS Pro.

1) Insert a new map.
2) Add the following data files:
   a) Tompkins County roads (*roadcl.shp*) and
   b) building footprints (*buildings.shp*)
   c) Image basemap (below)

3) Zoom to the Route 13 corridor (just south of Walmart).
4) Notice that there are several new buildings constructed since the original 'buildings' layer was created.

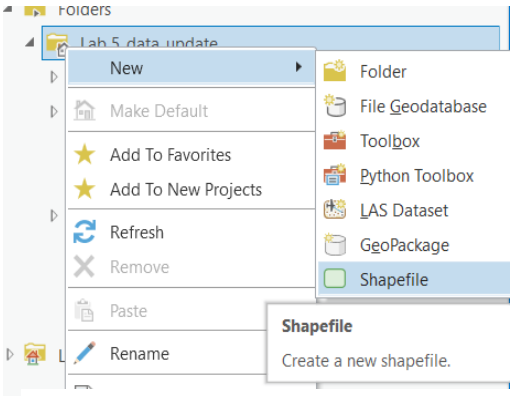


Zoom over to the Route 13 corridor (just south of Walmart). Notice that there are several new buildings constructed since the original 'buildings' layer was created.

Creating a new shapefile

Instead of editing the 'Buildings' layer, let's instead create a new shapefile of some of the building footprints not yet digitized.

1) Go to Catalog Pane (View->Catalog Pane).
2) Right click on the folder you are using for this lab.
3) Go to New and then Shapefile.
4) For name use 'New Buildings'.
5) Change the Feature Type to "polygon."

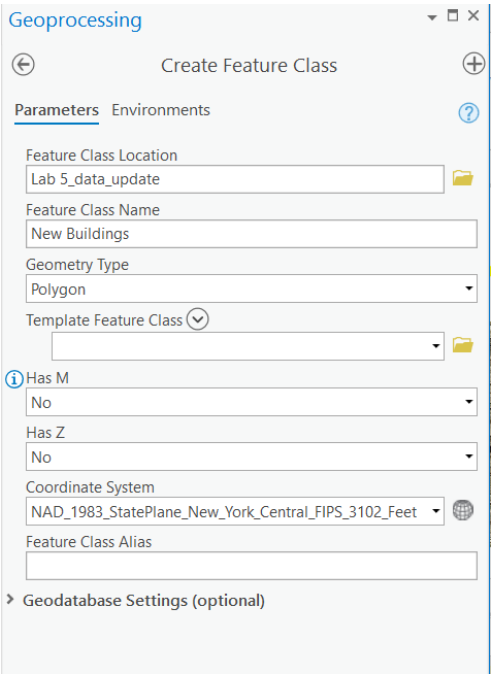


Creating a new shapefile for drawing new buildings

6) Set the coordinate system to buildings (so that they are of the same coordinate: New York State Plane central.
7) Click Run.

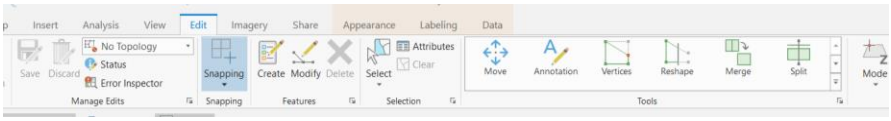
8) Back in the Map View, the "New Buildings" should be added. It's empty, so nothing comes on the screen, but if you check the properties of the layer, you will see that the projection information has been set. So has the scale and map extent.

9) Click on the Edit tab at the topic, and a number of tools will appear.

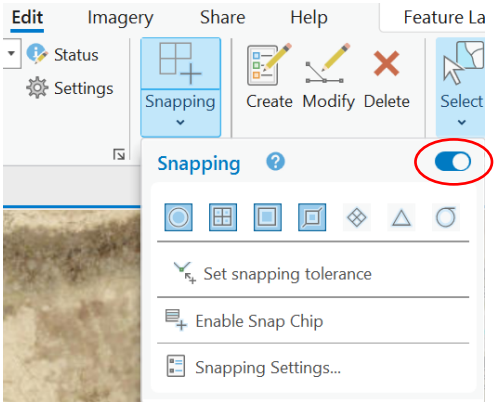


Parameter setting for the new shapefile

Setting the snapping options: Open the snapping options (blue box indicates active)



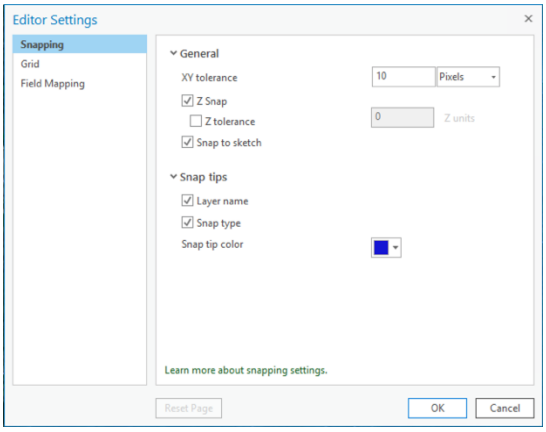
Make sure your snapping tool is activated. If not, click a snapping agent to enable it.



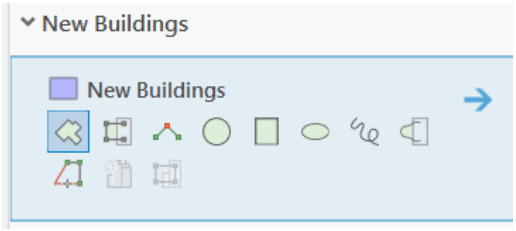
5) Hover over each icon to see which option each represents.
6) Click on the "Snapping Settings…" at the bottom of the snapping drop-down. This should display a window that allows you to set the XY tolerance or snapping tolerance.
   Note that we can set either pixels or Map units. It is typically best to set the projection, which also specifies the map units, and set a snap distance in map units.
7) Set your XY tolerance to some number between 2 and 4 map units and
8) click on OK. The end use specifies the data accuracy requirements, but if you set the snapping too small you end up with inefficient digitizing.

10) Click the ‘Create’ tool 

and the Create Features pane should appear.
11) Select ‘New Buildings’, as this is the shapefile we are interested in editing.
12) This opens up some additional tools for digitizing.
13) Select the Polygon tool.

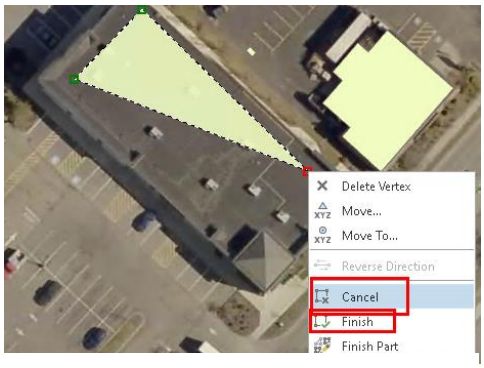


14) If you pan over the Map, cross hairs should become active.
15) Outline the new building. As you do so, notice a floating tool bar. If you make a mistake, click cancel  or right click on Cancel icon in the pop-up box and this will allow you to start over.



16) If you just want to erase the last step, you can undo the last action by clicking the undo icon  at the top or simply using Ctrl+Z.
17) To finish editing a polygon, right click on the finish icon in the pop-up box  or simply clicking F2.

Try adding several other buildings. Note that we have not actually created any attribute information – we would need to separately edit the attribute table.



Finish or cancel editing a feature

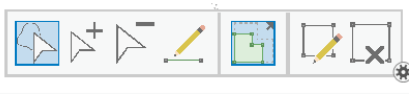
Adding Vertices to an Arc

Sometimes you will want to add vertices into a straight segment of an already existing polygon, in order to improve the accuracy.

1) Activate the Modify tools with the icon found along the main Edit tools ribbon: 

This will highlight a number of alignment and reshaping tools.

2) If you click the Edit vertices tool  Another floating toolbar will allow you to Add and delete vertices.



3) Experiment with adding and deleting vertices. Once added, you can also move the vertex.
4) To add or delete vertices, you need firstly select a feature  ,
5) click vertices tool, and then add or delete a vertice  .

Moving a Vertex

Sometimes you may want to move an existing vertex of a polygon.

1) Select the normal tool  put the mouse on a vertex you want to move,
2) wait for the cursor to change, and then
3) drag the vertex to its new location. Make sure not to move vertices in such a way that the line ends up crossing itself. Remember you can always undo the change by clicking the undo icon  or Ctrl+Z.
4) Once you've edited this boundary to your satisfaction commit your changes.

Commit the Changes

The changes you make are not committed until you save so.

1) Go ahead and commit to the changes now (finish). 
2) Save your edits often. 

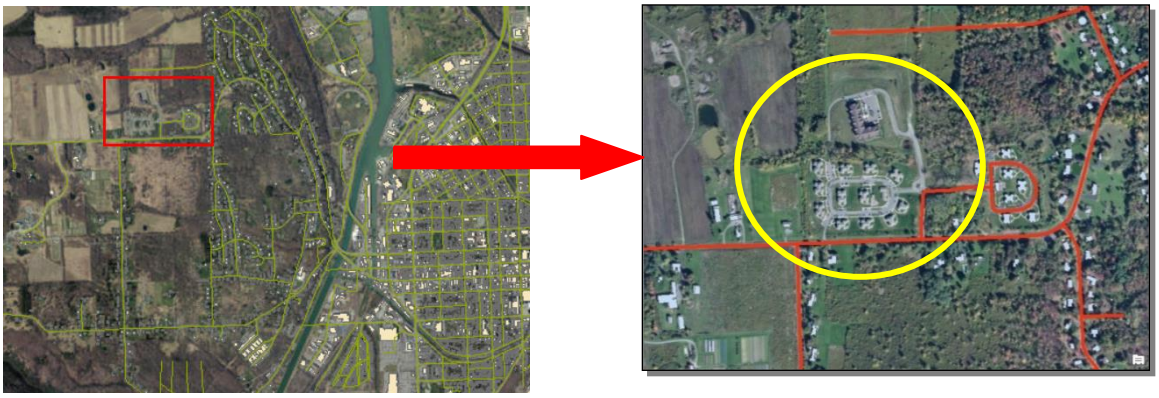
.
3) When finish adding new buildings, close the modify panel and clear any selected features.

Digitizing Lines

If you have not yet done so,
1) add *roadcl.shp* and
2) change the symbology so that it is visible against the aerial.
3) Pan over West Hill and note that there is new housing and roads that have been constructed off of Mecklenburg Road (Route 79).

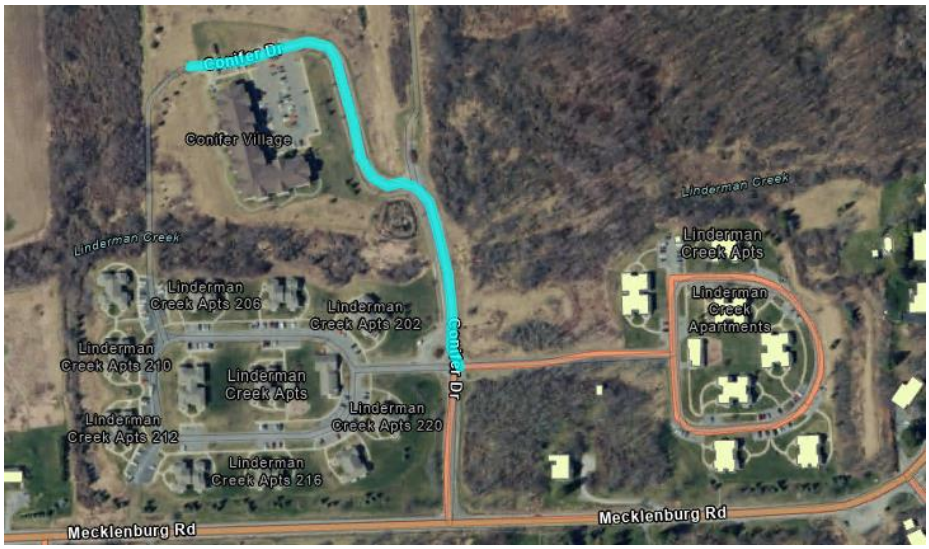
4) We will edit the existing roadcl layer to reflect these changes.
   a) In the Create Features box 

, highlight the 'roadcl' layer.
   b) Select 'Line' 
   c) Note that as you pan your cursor over the map, a little circle or square picks up the vertices of the road layer.
   d) Notice the little red circle around your cursor? That's a snapping indicator.



The new housing that has been constructed appears within the yellow circle. These features have not been digitized.

5) Now begin digitizing.
6) Snap to the end of the existing road and start adding line segment to add road. It's important to make sure that ends of roads snap into other roads, partly just for visual appeal, but also because any potential network analysis you do won't work right if they don't.



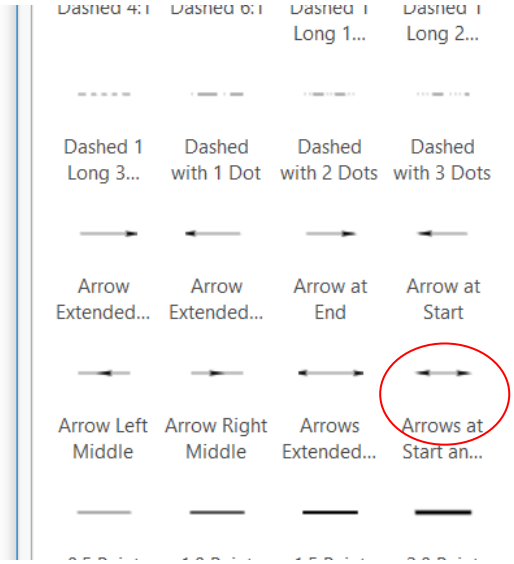
Editing a new road. Basemap: Imagery Hybrid

7) Whenever you come to a place where the road would change its name (presumably at the corners), you want to finish one arc and start another.
8) Also, as much as possible, do each road as one line segment.
9) To easily identify the name of each road, you can add "Imagery Hybrid" or "Imagery with labels" as Basemap.
10) When finished, commit the changes. Your finished road should now be highlighted.

Merging Lines

Often when digitizing, we may want to merge several road segments together. To be able to see where one segment ends and another begins,

1) change the line symbol for *roadscl* to the one called "arrows at start and end."



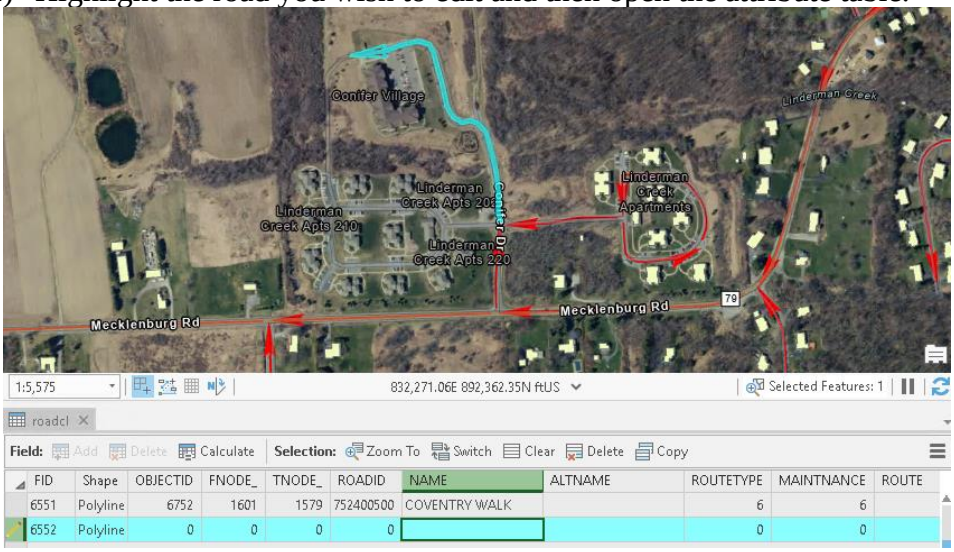
2) If you want to merge two-line segments, or a road that was done in multiple pieces, select all the pieces with the select tool (hold shift to select multiple features), and

3) use the ***Merge*** function 

in the Edit toolbar to combine the parts into one arc (order doesn't matter, as none of the roads have attributes yet. This way you can ensure that an entire road consists of only 1 segment.

Enter Arc Attributes

1) Once the digitizing is done, you are ready to enter the attribute information for the roads.
2) Highlight the road you wish to edit and then open the attribute table.



Editing Road Name

3) Since the editing session is still open, you should be able to edit the appropriate text columns directly (Name, etc.), assuming we know this information (you may have to refer to outside sources). We could then use this to label the streets.
4) Finally, remember to save your edits 

Map 2: Create a map of the development (zoom in!) depicting your newly digitized roads (include the aerial in the layout). Note: completing the circular road presents some difficulties, as you are not able to snap to a road segment you are currently editing. You may have to use multiple segments and then merge them together. Find out what these streets are called, edit the attribute information, and then include **labels** in your map

LAB 5 DELIVERABLES

1. **Map 1:** Map of the geocoded points 10 points
2. **Map 2:** Map of the completed, up to date roads (include the aerial image) 15 points
3. **Map 3:** (15 points): download an additional address table of your own choosing from Data Axle and mapping the geocoded points