

Geocoding vs. Add XY Data using *Reference USA* data in ArcMap



Written by Barbara Parmenter. Revised by Carolyn Talmadge for ArcMap 10.7.1

GETTING BUSINESS DATA FROM REFERENCE USA BY NAICS AND CITY	2
MODIFY YOUR EXCEL FILE.....	6
STARTING ARCMAP	7
ADDING REFERENCE USA DATA TO ARCMAP USING ITS LATITUDE AND LONGITUDE INFORMATION ERROR! BOOKMARK NOT DEFINED.	
EXPORT YOUR POINTS TO A SHAPE FILE:	10
DO A DATA QUALITY ASSESSMENT!	11
USING GOOGLE MAPS TO GEOCODE YOUR BUSINESS DATA	ERROR! BOOKMARK NOT DEFINED.
GEOCODE BUSINESS DATA USING ADDRESS INFORMATION	11
GETTING STREET CENTERLINES WITH ADDRESS RANGES FROM THE US CENSUS BUREAU	12
PREPARING YOUR STREET CENTERLINE FILE BY BUILDING AN ADDRESS LOCATOR	13
USING ADDRESS INFORMATION TO GEOCODE	15
GEOCODING AGAINST PARCEL POLYGONS OR ADDRESS POINTS	19
WRAPPING UP	23

In this exercise, you will map businesses or services from a database called **Reference USA**. The exercise demonstrates three different methods for geocoding business (or other address-based) information for you to compare.

1. First, you will use **latitude** and **longitude (XY) coordinates** that come with the business database – you can use the **Add XY Data** method for any community in the US.
2. Next, you’ll use the **address information** to address-match (or **geocode**) using Census TIGER street centerlines that the US Census Bureau has formatted for this purpose.
3. Finally, you’ll try **geocoding** to address points for the city of Cambridge – this method could also work if you were using parcel polygons.

This tutorial will use **Reference USA**, an online business database for which Tufts Library has a subscription. You’ll search for businesses on *Reference USA* using the Census *NAICS* code (North American Industry Classification System) and a town name. Once you have a list of businesses, you’ll download an Excel file, modify it as needed, then map it using three different methods.

Using Census.gov to find NAICS Codes

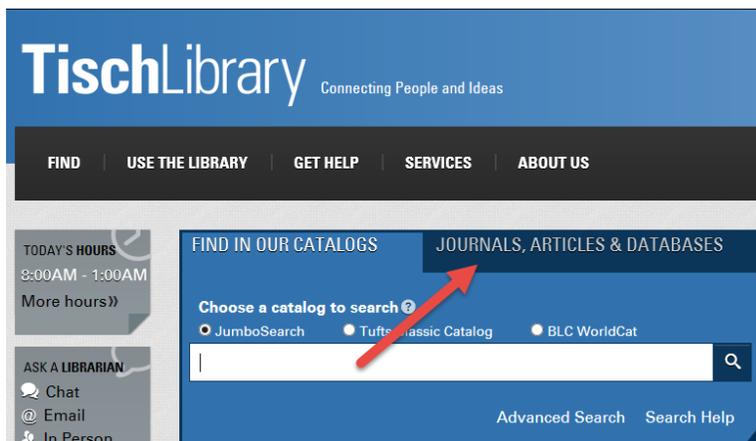
1. Determine what type of business or service you want to search for to geocode. In this case, we'll be using grocery store data.
2. Go to the Census NAICS code web site <http://www.census.gov/eos/www/naics/>
3. Above the 2017 NAICS search box on the left-hand side of the webpage, type in *Grocery Store* and press Search.
4. This will turn up a series of codes you can select from and use to find the list of stores. For grocery stores, we'll use **code 445110**.

NAICS Search:

Enter keyword or 2-6 digit code

Getting Business Data from Reference USA by NAICS and City

1. Go to the Tufts Tisch Library site - <http://www.library.tufts.edu/tisch/>
2. Click on **Journals, Articles & Databases**.



3. Navigate to "R" in the alphabetical list of databases. Then scroll down and click on *ReferenceUSA*.
4. Once in *Reference USA*, click on **U.S. Businesses**.

Available Databases

Select a Database to Get Started

U.S. Businesses 58 Million Businesses 3.3 Million Closed Businesses SEARCH MORE INFORMATION	U.S. Jobs / Internships 2.5 Million Job Postings
---	--

5. Then click on the **advanced search** tab.

U.S. Businesses Database

Quick Search **Advanced Search**

Fill out one or more of the following criteria boxes, then click "View Results" button.

Company Name <input type="text"/>	Executive First Name <input type="text"/>	Executive Last Name <input type="text"/>
City <input type="text"/>	State All ▼	Phone <input type="text"/>

[Additional Filters](#) [Clear Search](#) **VIEW RESULTS**

6. On the left, click on the checkboxes for the buttons **Keyword/SIC/NAICS** and **City/State**.
7. Carefully follow the **5 steps below** – in this tutorial we are searching for grocery stores (445110) in Cambridge. But this is a useful resource when searching for other business locations in other US cities as well.

U.S. Businesses Database

The screenshot shows the 'Advanced Search' interface of the U.S. Businesses Database. It features a left-hand sidebar with various search criteria, a main search area with filters and a search box, and a right-hand sidebar with result summary and action buttons. Red callout boxes with arrows point to specific elements:

- 1. Select Criteria:** Points to the 'City / State' checkbox in the Geography section of the sidebar.
- 2. Click here to search with NAICS:** Points to the 'SHOW 2 - 6 DIGIT CODES' button.
- 3. Select both verified and unverified businesses:** Points to the 'Verified Businesses' and 'Include Unverified Businesses' checkboxes in the Record Type section.
- 4. Type in desired NAICS codes here. You can search for other business and codes above as well.** Points to the input field for NAICS codes.
- 5. Navigate to Cambridge, Ma and double click to move right into the selected box.** Points to 'Cambridge, MA' in the 'Results' list of the 'City / State' search.
- 6. Click UPDATE COUNT to see how many stores are in Cambridge.** Points to the 'UPDATE COUNT' button in the right sidebar.
- 7. Click VIEW RESULTS to see the list of stores.** Points to the 'VIEW RESULTS' button in the right sidebar.

Record Type

- Verified Businesses (Phone verified and quality checked)
- Include Unverified Businesses (Not yet fully verified, may not be accurate)
- Include Closed / Out of Business Records (Suspected to be out of business)

Keyword/SIC/NAICS

- Search All SICs
- Search Primary SIC Only
- Search All NAICS
- Search Primary NAICS Only

SHOW 2 - 6 DIGIT CODES

Results:

Selected:

445110 - Supermarkets And Other Grocery (Except Convenience) Stores

Enter 2 - 6 digit NAICS Codes in the boxes below.

445110					
--------	--	--	--	--	--

City / State

Search

Search for City or State

Results

Results	Selected
Brookline Village, MA City	Cambridge, MA City
Bryantville, MA City	
Buckland, MA City	
Burlington, MA City	
Buzzards Bay, MA City	
Byfield, MA City	
Cambridge, MA City	
Canton, MA City	
Carlisle, MA City	

Note: You can enter more than one NAICS code and more than one city. For example, you could search for both grocery stores (445110) and convenience stores (445120).

- On the results screen, you need to select the businesses of interest (we have 101 returns for Cambridge). We want all of them, so check the box at the top of the first column as shown here:

Page 1 of 4

Review Details Heat Map Summary

<input checked="" type="checkbox"/>	Company Name	Executive Name	Street Address
<input checked="" type="checkbox"/>	7-Eleven		2245 Massachusett...
<input checked="" type="checkbox"/>	7-Eleven	Tesfaye Gebremichael	275 Prospect St
<input checked="" type="checkbox"/>	7-Eleven	Habte M Ayalew	40 Jfk St
<input checked="" type="checkbox"/>	7-Eleven	Tommy Chen	589 Massachusetts...
<input checked="" type="checkbox"/>	7-Eleven	Paul Marino	600 Technology Sq...

- Each page has 25 results. If you have a second page of results, go to page 2 and click on **the same box again**. Do this until all 101 records are selected. The maximum download at a single time is 250 results.
- Click the **download** button – this will download the results from the page(s) you have clicked to checkmark.
- Fill out the form as you see here: In **Step Two**, select **Custom**. Search for *Latitude* and *Longitude* in **Find Fields**. Then add the **Primary NAICS Code** and **Primary NAICS Description** as well:

Step One: Select your file format

Comma Delimited (Most Popular)
 Excel
 Excel (2007/2010)
 Tab Delimited

Step Two: Select your level of detail

Summary - includes name, address, phone number and 17 additional data fields
 Detailed - includes all data
 Custom - select fields to export

Directions: Add the fields you wish to export to the box on the right. Reorder the fields by dragging fields within the box. The fields will export in the order displayed in the list.

Find Fields: CLEAR

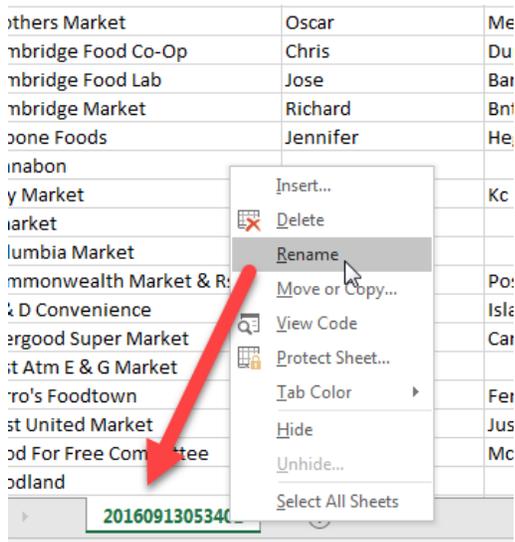
Selected Fields: RESET DEFAULTS

All	Company	Management	Corporate	Total Fields Selected: 12
Latitude				Company Name <small>Remove</small>
				Executive First Name <small>Remove</small>
				Executive Last Name <small>Remove</small>
				Address <small>Remove</small>
				City <small>Remove</small>
				State <small>Remove</small>
				ZIP Code <small>Remove</small>
				Record Type <small>Remove</small>
				Latitude <small>Remove</small>
				Longitude <small>Remove</small>
				Primary NAICS <small>Remove</small>
				Primary NAICS Description <small>Remove</small>

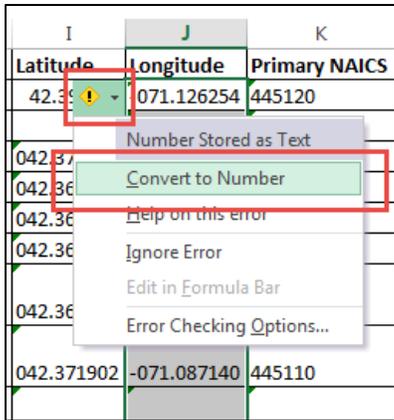
12. When finished, click **Download Records** and choose to open it with Excel – if you get a warning message about formats, choose Yes to open the file.
13. Before proceeding, choose **Save As** and navigate to your H drive. Create a new folder called “GeocodingReferenceUSA”. Then, save the excel file into the folder and name it “CambridgeGroceryStores” – do not use hyphens or spaces in your file name or folder name! **Save it as an Excel 97-2003 Workbook (.xls)** file.
14. If you had more than 250 results, go back to your results table, deselect the first 10 pages of results (click on NONE at the top of the left column), and select the next 10 pages. Run through the download process again. This will create a second Excel file. You can copy and paste the rows of data from the second sheet to the end of the first sheet to make one big Excel table of all your data.
15. You can close out of Reference USA.

Modify and Clean your Excel File

1. Rename the *worksheet* to something more comprehensible like GroceryStores (no hyphens or spaces!)



2. There is also a few additional problems we need to correct in the Excel file.
 - 1) Adjust each column name so there are no spaces in the heading. I.E Company Name → CompanyName.
 - 2) The latitude and longitude columns are text (aka a *string* in ArcMap) and they need to be formatted as numbers.
 - i. Highlight all the *Latitude* and *Longitude* **data cells** (**not** the column names).
 - ii. Click on the little message diamond (⚠) and choose **Convert to Number**.

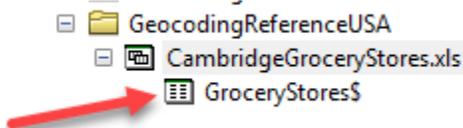


3. **Save your file and Exit** out of Excel. This is important – you *cannot* have Excel file open when you work with it in ArcMap.

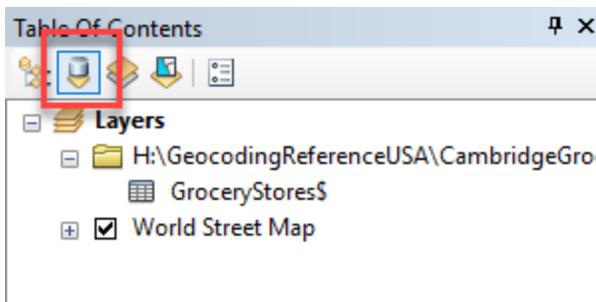
Starting ArcMap & Adding Reference USA data using Latitude and Longitude

Because you have the latitude and longitude coordinates for your Reference USA grocery store data, you can add the business records as points to a map in ArcGIS using the **Add XY Data** Method.

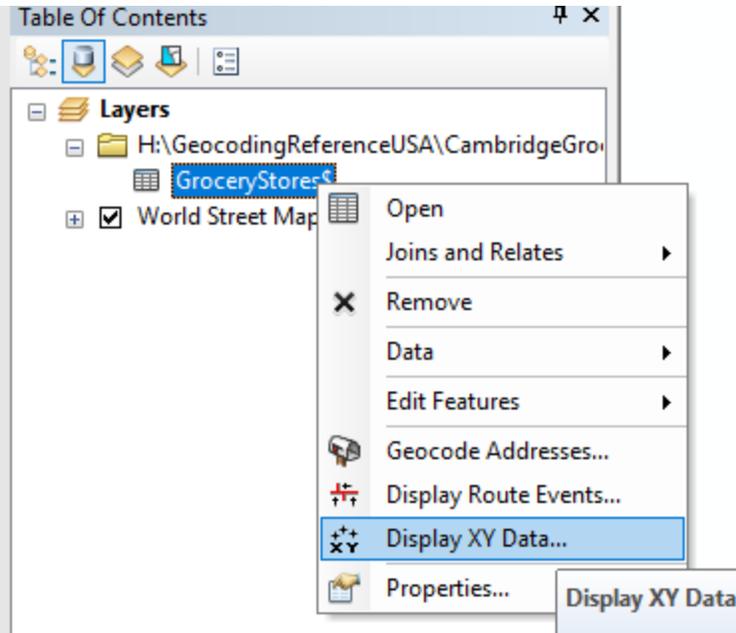
1. Start a blank ArcMap.
2. Add some kind of basemap in ArcMap – this could be a Base Map from ESRI Online (**File → Add Data → Add Basemap**) or a GIS street file from your local or state clearinghouse (if you are at Tufts, try the *dtl_cnty.sdc* and/or *cities_dtl.sdc* data set from M:\Country\USA\ESRIDataMap10\usa\census – this is a detailed county and city polygon layers for the entire country).
3. Note the **data frame’s** coordinate system by clicking on Layers → Properties → Coordinate System tab.
4. Zoom to the area for which you have Reference USA data (in this case Cambridge, Ma).
5. In Catalog, drag in your Excel **sheet** into ArcMap (you have to navigate one step beyond the .xls file to choose the individual sheet, e.g., ‘Grocer Stores\$’).



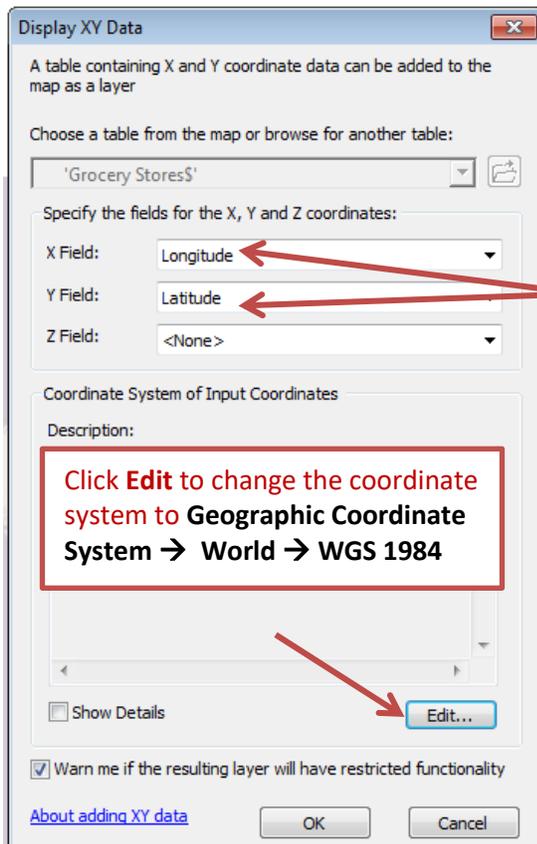
Notice how the Table of Contents view now changes to “list by data source” so we can see the excel table. This also shows us where each file is saved, but we are not able to move layers in this view.



6. In the **Table of Contents**, right-click on your Excel file (e.g., *grocerystores*) and choose **Display XY Data**.

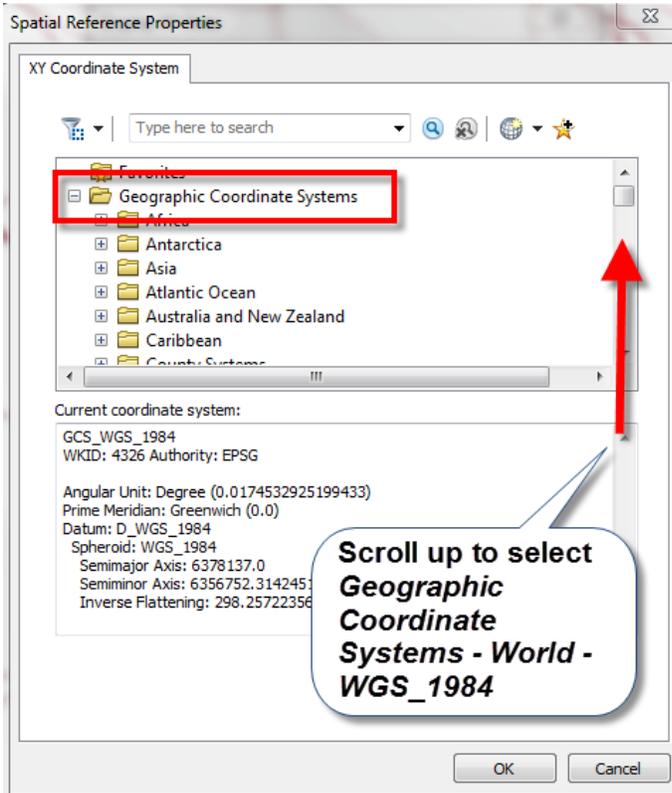


7. Fill the dialog box out as follows below. Make sure you choose *Edit* to select the **coordinate system** that's appropriate. Choose *Geographic Coordinate System* → *World* → *WGS 1984*. Make sure that it is not a projected coordinate system (projected coordinate systems use Meters or Feet, geographic coordinate systems use **decimal degrees**).

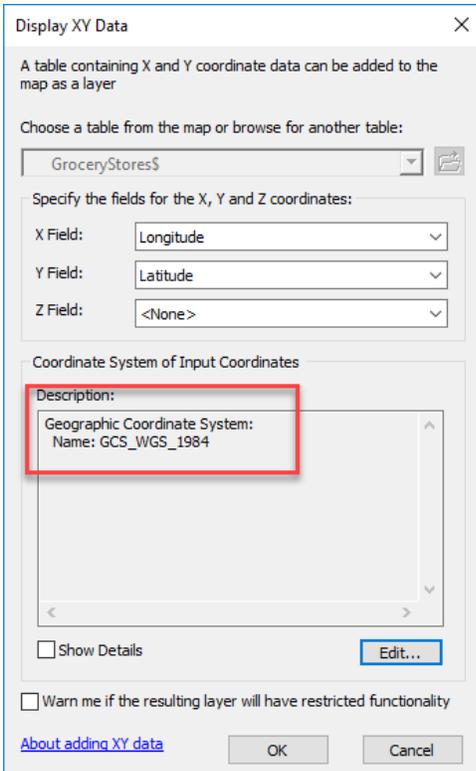


Check that these are correct.
 - X is Longitude!
 - Y is Latitude!
 (Many people switch these by accident!)

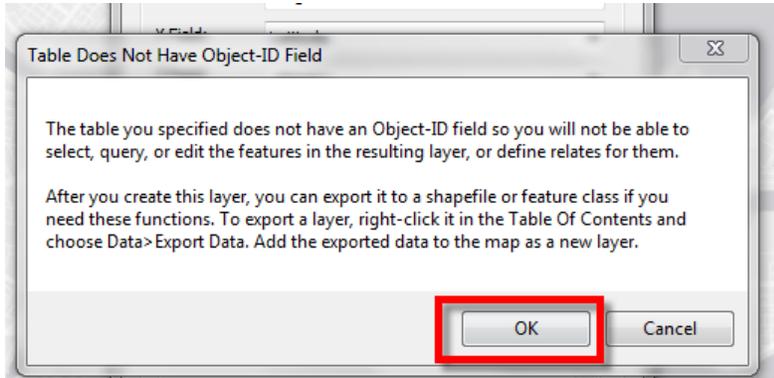
Click **Edit** to change the coordinate system to **Geographic Coordinate System** → **World** → **WGS 1984**



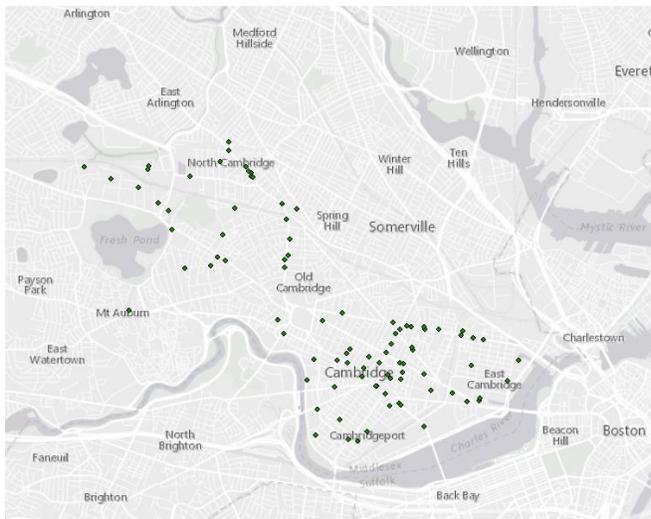
8. Make sure your Display XY Data box has updated with the correct Coordinate System info! That is VERY important, or else the points will be in the very wrong spot.



9. Press **OK twice**. Read the warning if it pop us but then press **OK** again.



10. The points should appear on your map like so. If necessary, right click on the layer and “zoom to layer”. **This new layer is NOT a shapefile.** This is just a visualization of your excel data in ArcMap (that’s what the \$ Events represents). To save it as a permanent point shapefile, follow the directions in the next section.



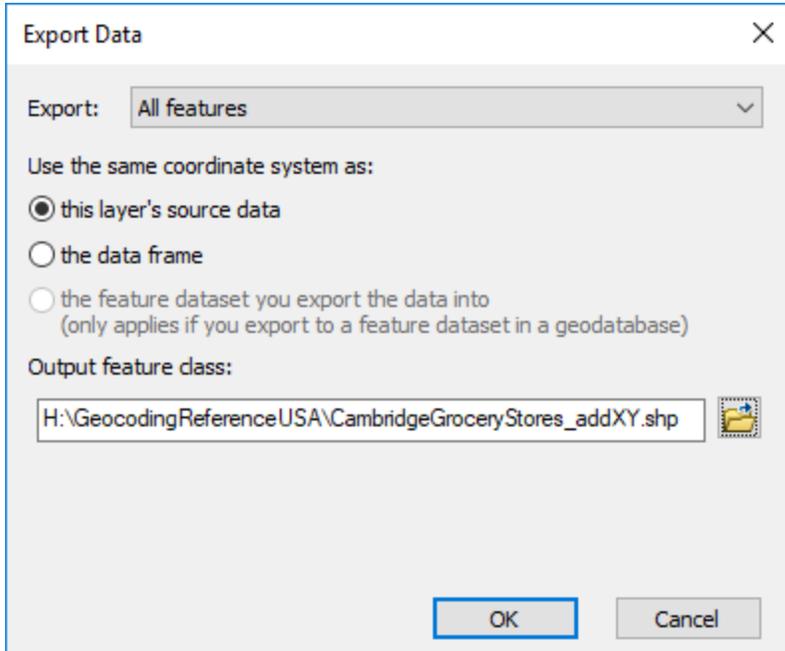
Export Points as a Shapefile

When the data initially comes up as points in a map, ArcMap refers to it as an “\$events” layer – this is a temporary, virtual view of your tabular data. That’s what the warning was about.

To make it into a permanent shapefile, which you can use in analysis:

1. Export the “events” layer to a shapefile by **right-clicking on the *GroceryStore\$ Events* layer.**
2. Choose **Data → Export Data.**
3. Press the folder button and navigate to your **H://GeocodingReferenceUSA** folder to save the shapefile there.
4. Name the shapefile, “**CambridgeGroceryStores_addXY**” so we know these are the points we created using *the Add XY data method*. Make sure you save it as a shapefile in the **Save as type** dropdown.

5. Press **Save** and then **ok**. When asked if you what to add the exported data to the map as a layer, press **Yes**.



6. Remove the **Grocery Store \$Events** point layer from your table of contents (don't remove the excel sheet though). You'll work directly with the shapefile.
7. Change the symbology of this point layer to red triangles, so we can easily tell them apart later.
8. Save your ArcMap document into H://GeocodingReferenceUSA folder. Just incase arcmap crashes.

Do a Data Quality Assessment!

You should now have points on your map. But are they in the right place? Explore the placement of your data points to see if they are accurate enough for your purposes. Some ways to do this:

- Add the **Imagery** from ArcGIS Online for reference (File → Add Data).
- Check specific addresses against an online mapping service like Google Maps or Yahoo Maps.
- Use Google Streetview (in Google Maps or Google Earth) to see if you see a particular business on that street or find it's more exact location.

Geocode Business Data Using Address Information

Often you will have a list of addresses you want to map, but the list does not have latitude and longitude, only street addresses. This process is called **geocoding** or **address-matching**. In the next two sections, we'll see two ways to use address information to put points on a map. Neither one is perfect, so you have to be very careful in checking the results! Using the *Reference USA* data, you'll have an opportunity to compare your geocoding results based on your own address-matching with what Reference USA provided for latitude and longitude.

You will use your Reference USA data again for this part of the exercise, but you will use the address information instead (address and zip code). But first you have to download a GIS data set from the US Census Bureau that has street centerlines with address ranges!

Getting Street Centerlines with Address Ranges from the US Census Bureau

The Census has street centerline files for the entire US as part of its TIGER geography, and for most metropolitan areas they have a data set that has address ranges for each side of street segments – you will use this information to geocode your Reference USA file based on address and zip code.

1. Using a web browser, go to <https://www.census.gov/>
2. Click on the **Browse by Topic** tab and then go to **Geography**.



3. Scroll down to Tiger/Line Shapefiles.

Mapping Files

MAPPING FILE

TIGER/Line Shapefiles

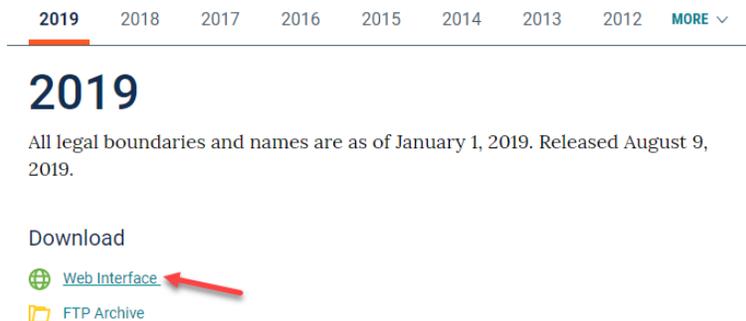
View all the available mapping files from the Geography program.

MAPPING FILE

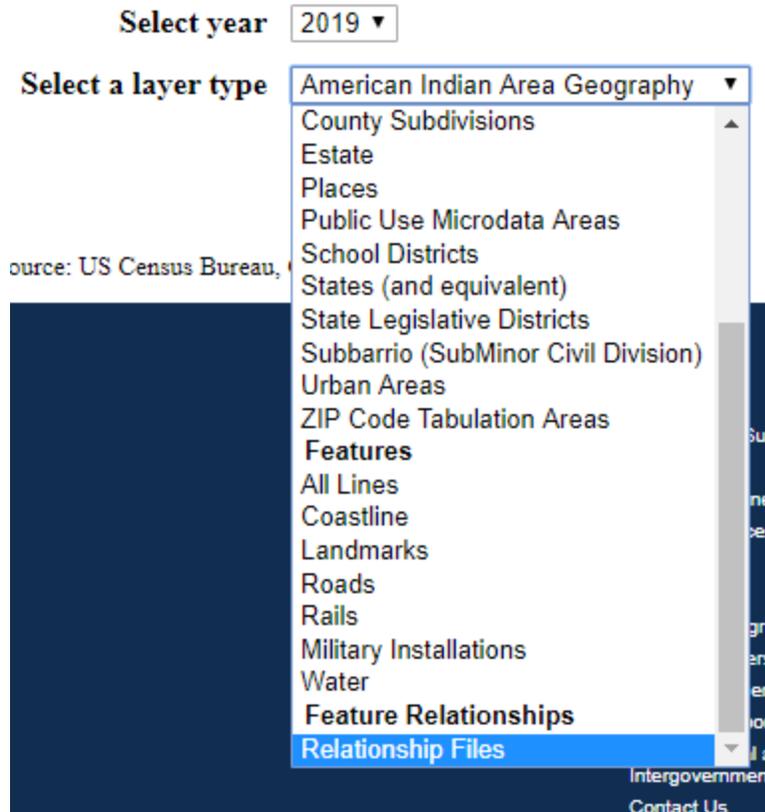
Cartographic Boundary Files - Shapefile

View the available shapefile cartographic boundary files from the Geography program.

4. Click on the **2019** tab, then select **Web Interface** under Download.



- Under **Select a Layer Type**, scroll down to **Feature Relationships** and click on **Relationship Files**. Click Submit.



- Go to the choice **Address Range – Feature Shapefile** (be very careful to get the right one – there are several with similar names!) and select your state (MA) and county (Middlesex) and click **download**.

Address Ranges Relationship File
 Select a State:

Address Range-Feature Shapefile
 Select a State:

Select a County:



- A zipped file will download. Copy it to your H://GeocodingReferenceUSA folder. Right click on it and press **Extract Here**.

Preparing your Street Centerline file by Building an Address Locator

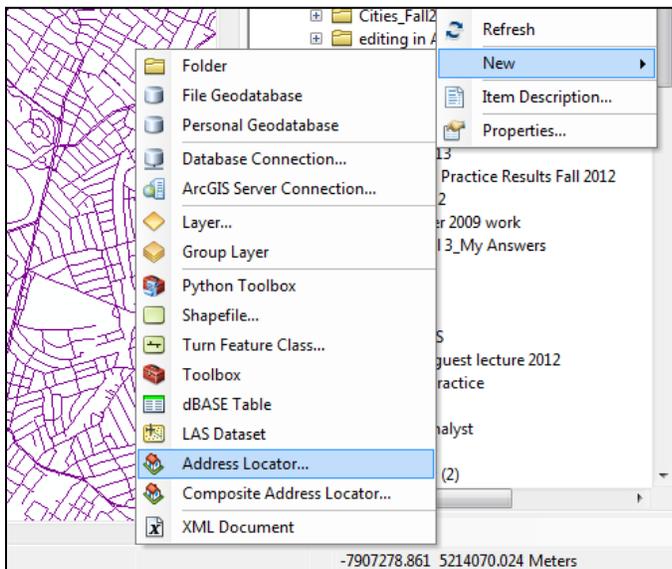
Before you can geocode, you need to prepare your **geographic reference file** (our TIGER roads in this example) so that you can match your business addresses against it. This involves creating an *Address Locator* for that reference file.

- In Catalog, right click on you H drive and **Refresh**. It is important to do that whenever we extract a new shapefile and ArcMap is already up and running. Or else catalog won't see the new file.
- Add your Census Tiger street centerlines to ArcMap – it will have a name like **tl_2019_25017_addrfeat**. This is the **Tiger Line** file for **2019** for Mass (**25**) Middlesex County (**017**) – Another county FIPS code would be different.

3. If you get an error about a coordinate system, it is telling you that the streets are in a different coordinate system than the data frame. Ignore it for now, but don't forget about it.
4. Open the street file's attribute table to see how it codes address ranges – you see the *street name column* (FULLNAME), the *Left from House Number* (LFROMHNN), *Left To House Number*, etc. Leave this table open for reference.

TLID	LINEARID	FULLNAME	LFROMHNN	LTOHN	RFROMHNN	RTOHN	ZIPL	ZIPR
86909071	110340818174	Greenleaf St	100	198	101	199	01821	01821
87083636	110340822815	Technology Dr	2	598	1	599	02453	02453
86862731	110340810110	Wykeham Rd	99	1	98	2	02465	02465
87126310	110340821809	Half Crown Cir	82	98	61	99	01721	01721
86906017	110340812983	Hathaway Rd	24	26	25	27	01887	01887
86935978	110340810977	Caldwell Dr	2	98	1	99	01886	01886
86865413	110340800830	Spruce St	66	198	71	199	02472	02472
86857085	110340807979	Perkins St	2	8	1	5	02180	02180
86882366	110340809621	Waltham St	372	384	369	373	02421	02421
87097797	110340778335	Lakin St	21	5			01463	
86854308	1103747079686	Pleasant St S	26	30	29	33	01760	01760
86865778	110340813201	Hillside Rd	42	68	39	67	02461	02461
86935886	110340802268	Andover St	1816	1836	1817	1837	01876	01876
86935886	110340769537	State Rte 133	1816	1836	1817	1837	01876	01876

5. In Catalog, Right click on your *GeocodingReferenceUSA* Folder and choose **New → Address Locator**.



6. Fill out the dialog box for the Address Locator as you see on the next page VERY CAREFULLY. If any part of this is not filled out correctly, it will not work. Refer to the TIGER street attribute table as needed. Be sure to give the output address locator a name like **TigerAddressLocator**.
 - a. **Note:** If you try this address locator tool several times and it keeps failing, pull in the tiger line file saved in this folder: S:\Tutorials & Tip Sheets\Tufts\Tutorial Data\Geocoding_Cambridge Example\tl_2019_25017_addrfeat
 - b. **Navigate to the folder in Catalog and pull in the line file.** Try the tool again with this line file.

Address Locator Style
US Address - Dual Ranges

Reference Data

Reference Data	Role
tl_2019_25017_addrfeat	Primary Table

Field Map

Field Name	Alias Name
Feature ID	FID
*From Left	LFROMMHN
*To Left	LTOHN
*From Right	RFROMMHN
*To Right	RTOHN
Left Parity	ParityL
Right Parity	ParityR
Full Street Name	FULLNAME
Prefix Direction	<None>
Prefix Type	<None>
*Street Name	FULLNAME

Output Address Locator
H:\GeocodingReferenceUSA\AddressLocator_TL2019

Configuration Keyword (optional)

Enable suggestions (optional)

Click here to select Dual Ranges style

Select your street

Make sure these drop downs match exactly. Scroll down to make sure zip codes are filled out too.

Save it in your folder with this name.

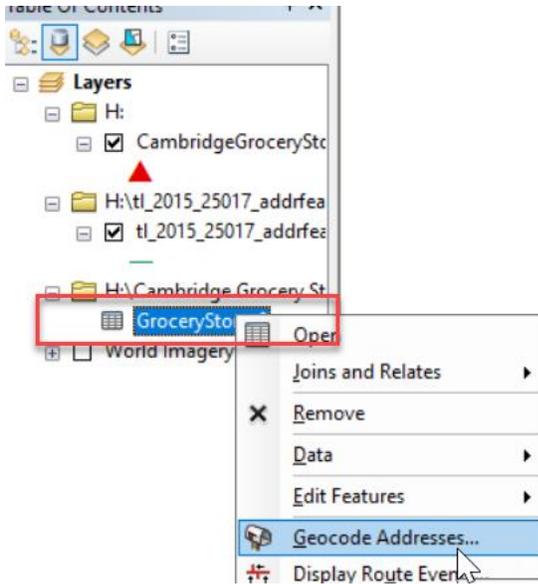
7. Click **OK** when you are done filling out the form. This **process may take 5 minutes**. Relax, stretch!
8. If you continue to have issues, you can use the address locator that has already been created and is saved in this folder. S:\Tutorials & Tip Sheets\Tufts\Tutorial Data\Geocoding_Cambridge Example\AddressLocator

Using Address Information to Geocode

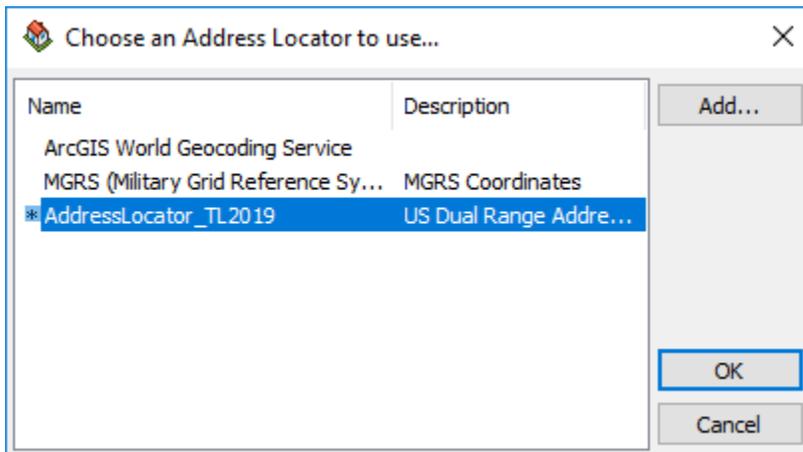
Now you're ready to geocode against the **Tiger Road centerlines** file using the address locator you created.

You should have your Excel file with grocery store data in your ArcMap session.

1. In the Table of Contents, right-click on your *Excel GroceryStores\$* file and choose **Geocode Addresses**.

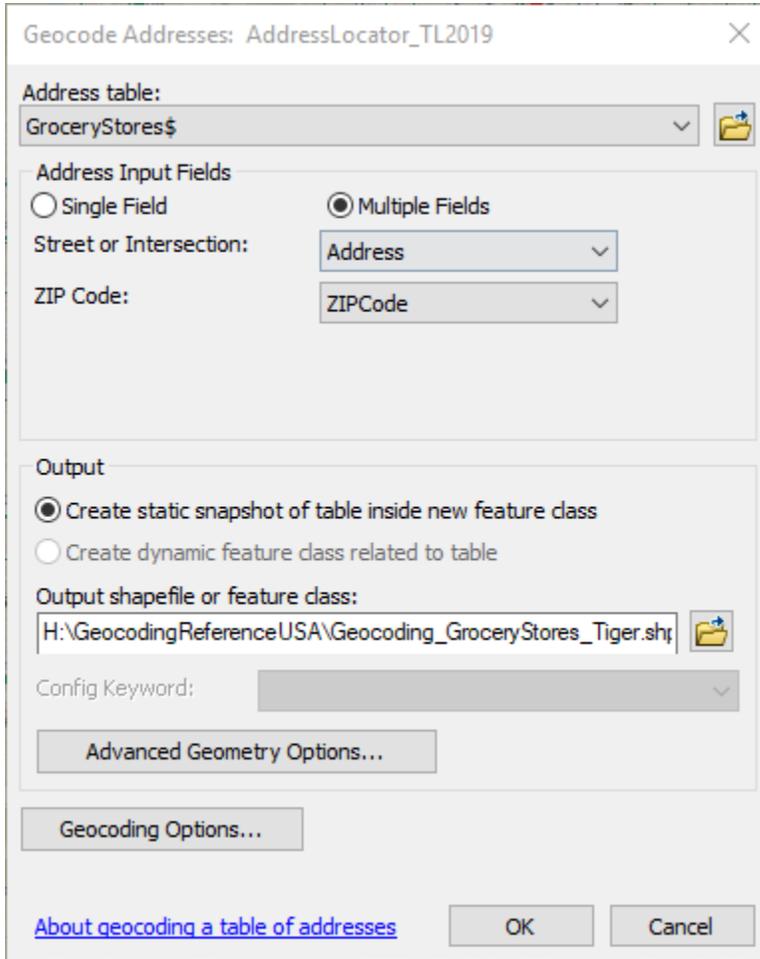


2. For your Address Locator, choose your **Tiger address locator** and click OK:

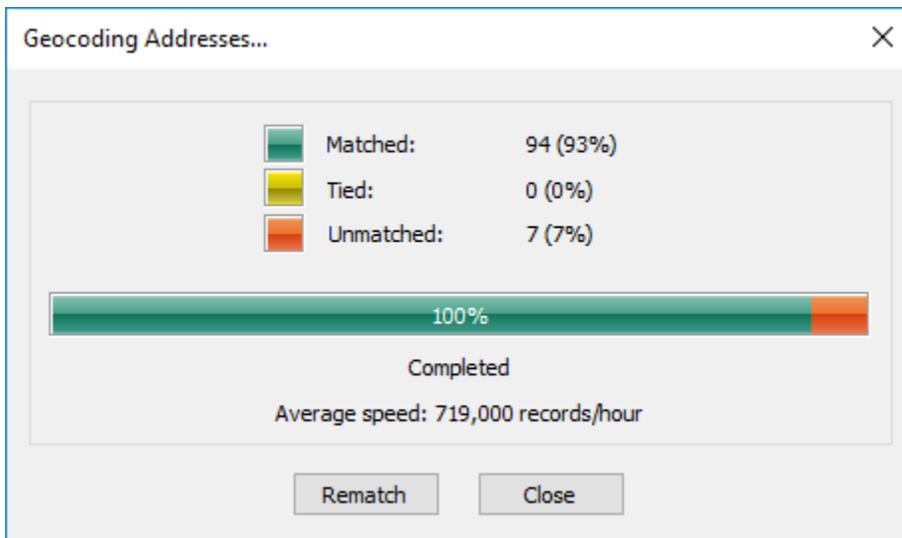


*If you are using the address locator that has already been created. Click on Add and navigate to **S:\Tutorials & Tip Sheets\Tufts\Tutorial Data\Geocoding_Cambridge Example\AddressLocator** and select the address locator in this folder.

3. Fill out the dialog box as you see below (choose your Excel worksheet as the address table). Make sure to save the file as **Geocoding_CambridgeGrocery_Tiger** so that we know this was the shapefile created using tiger roads and the geocoding method. Then click OK.



4. You will see a screen that tells you your progress and how many matches you got. Click **Close** when the process is finished.



5. Change the symbology of these points to yellow circles.
6. Explore your results and compare them against what happened when you used Latitude and Longitude from Reference USA to add the points. Zoom to this **Red Square** area near Fresh Pond on the Western Side of

Cambridge:





What kinds of differences do you see? Why do you think the dots aren't in the same place? Which one is more accurate? How do they compare with other sources like Google Maps or StreetView?

Notice how the Lat/Long (Add XY Data) method puts the points on the actual building! Geocoding puts the points right on the street (essentially where the mailbox might be located along the street).

Geocoding against Parcel Polygons or Address Points

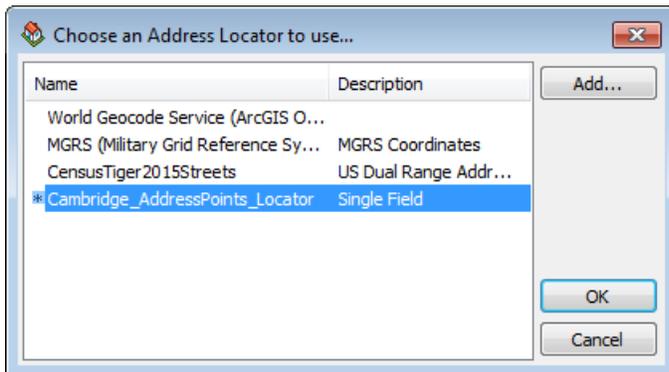
Some localities have **parcel polygons** or address points GIS layers – you can address match against these if they have address information in the attribute tables. In this example, we'll use Address Points GIS data set from the *City of Cambridge* and we'll geocode the Cambridge Grocery Store Excel file from Reference USA against it as a test.

You will need to examine the attribute fields of your Parcel or Address Point GIS data set before you create the address locator to see how to map the fields.

1. The following example uses the data sets listed below, both found in *S:\classes\UEP_ENV\Geocoding Practice2015\City of Cambridge Address Points – Add Address points from the City of Cambridge, MA, called ADDRESS_AddressPoints.shp*
2. Examine the attribute table of the *ADDRESS_AddressPoints* data set – in the case of Cambridge we see that there is a single field for the address in the attribute table – it is called **Full_Addr**:

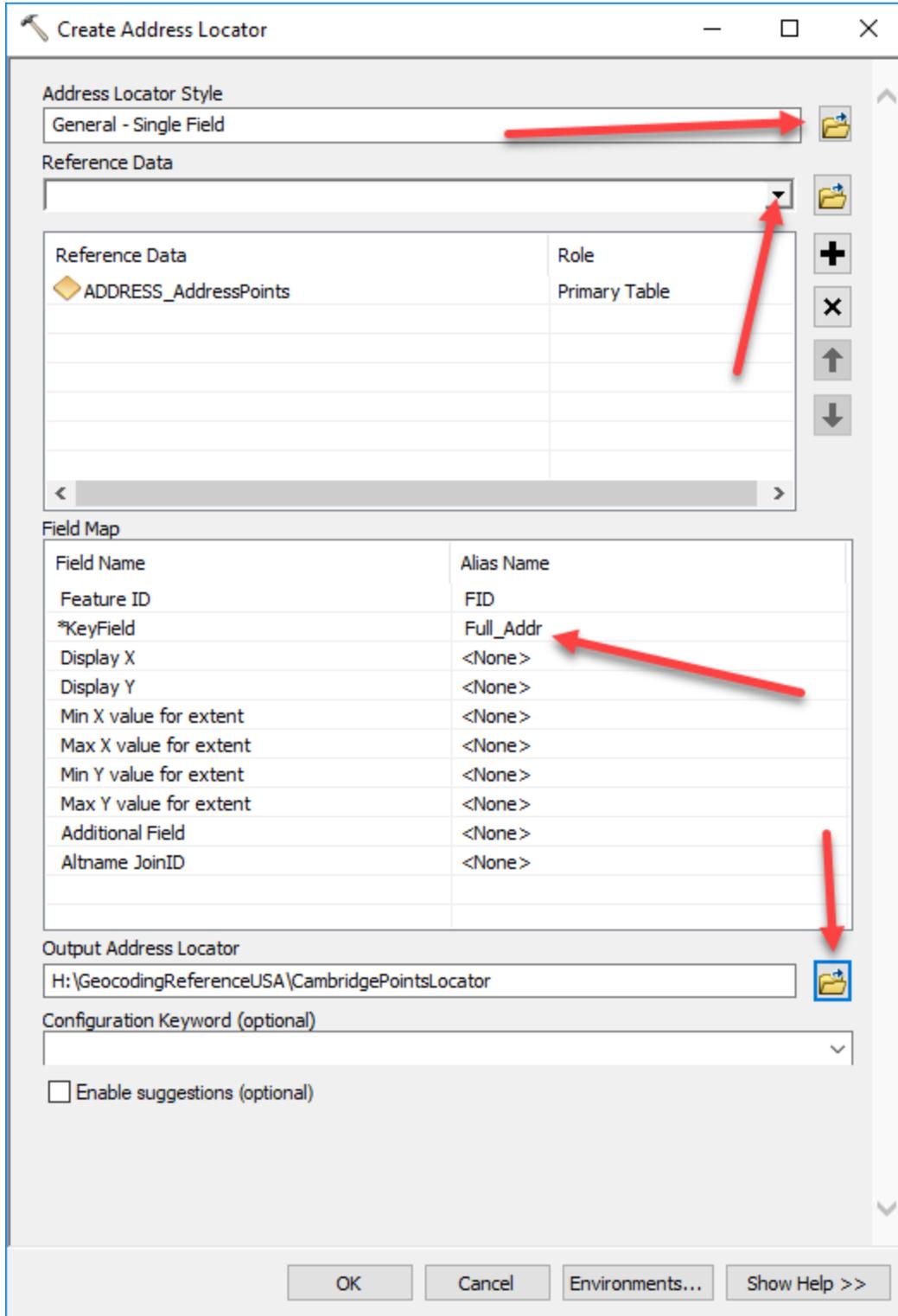
FID	Shape*	address_id	BldgID	ml	StlNm	Full_Addr	TYPE	EditDate
0	Point	866	370-4	84-105	1300	1300 Cambridge St		2010
1	Point	869	370-5	84-101	1280	1280 Cambridge St		
2	Point	3767	339-23	83-69	18	18 Clary St		
3	Point	3315	343-5	83-5	16	16 Springfield St		
4	Point	3862	343-4	83-4	18	18 Springfield St		
5	Point	2816	522-20	26-140	157	157 Fifth St		
6	Point	515	623-26	125-131	113	113 River St		
7	Point	516	615-3	122-31	47	47 Howard St		
8	Point	518	578-47	123-90	205	205 Western Ave		

3. Right click on your **GeocodingReferenceUSA** folder and choose **New → Address Locator**.
4. Fill out the **Create Address Locator Dialog Box** as shown below – remember that the *Cambridge AddressPoints* GIS data set had a single field in its attribute table for the address – the name of this field was **Full_Addr**.
 - a. We will set the Address Locator Style to General – Single Field.
 - b. Tell ArcGIS that we are using the **ADDRESS_AddressPoints** GIS data set as our *Reference Data layer* and that the *KeyField* is **Full_Addr**.
5. Click OK when finished – the process of creating the *Address Locator* will take a few minutes to complete depending on the size of the file.
6. When the *Cambridge Points Address Locator* is complete, you can geocode addresses using it. Right click on the *excel data table* that has your address data (e.g., “grocery stores”) and choose **Geocode Addresses**.
7. In the first dialog box, scroll to find your new address locator as shown below:

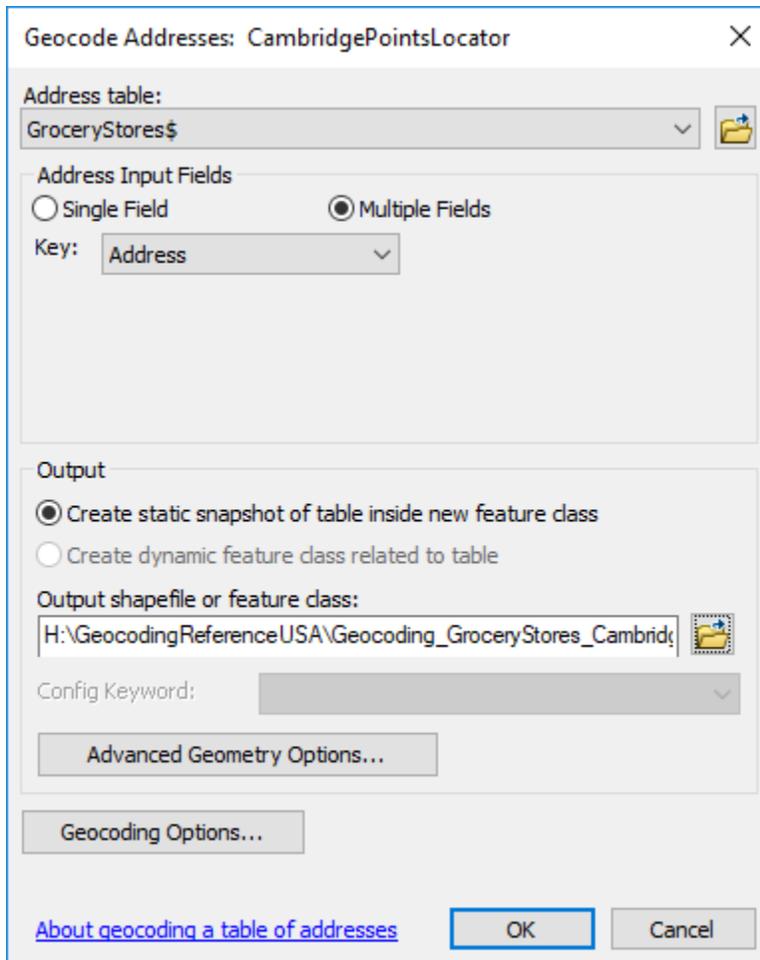


8. Click OK.

9. Fill out the dialog box as follows and press ok.



10. Then, Geocode the excel file again using this new Address locator.



11. Click OK. When the geocoding results come up, click **Close**. Notice how many match versus those that don't match.
12. Change the symbology of this layer to blue squares and turn off your **Address_AddressPoints** layer.
13. Inspect the new points added to your map.

Here's the Fresh Pond example again with the points using **Add XY Data Method**, the points using **Geocoding with Tiger Road layer**, and the points using **Geocoding with Cambridge's Address Points layer**. Which is the best? Which is the worst? Why? Which reference layer would you use for your project?



Wrapping Up

You've seen multiple ways to map **point data** in this exercise.

- 1.) Add XY Data using Lat/Long coordinate information in decimal degrees that exist within the grocery store database.
- 2.) Geocoding with different "reference" layers
 - a.) US Census Tiger Line road files – a national dataset.
 - b.) AND Town of Cambridge Address Points, created by the local government.

Whichever method you use, you will need to carefully inspect the results to see if the **accuracy** is appropriate for your application.