

# Geocoding vs. Add XY Data using *Reference USA* data in ArcMap 10.5.1

---



Written by Barbara Parmenter. Revised by Carolyn Talmadge 2/15/2017

|  |    |
|--|----|
| GETTING BUSINESS DATA FROM REFERENCE USA BY NAICS AND CITY .....                       | 2  |
| MODIFY YOUR EXCEL FILE.....  | 5  |
| STARTING ARCMAP .....  | 6  |
| ADDING REFERENCE USA DATA TO ARCMAP USING ITS LATITUDE AND LONGITUDE INFORMATION ..... | 6  |
| EXPORT YOUR POINTS TO A SHAPE FILE: .....  | 8  |
| DO A DATA QUALITY ASSESSMENT! .....  | 9  |
| USING GOOGLE MAPS TO GEOCODE YOUR BUSINESS DATA .....                                  | 9  |
| GEOCODE BUSINESS DATA USING ADDRESS INFORMATION .....                                  | 9  |
| GETTING STREET CENTERLINES WITH ADDRESS RANGES FROM THE US CENSUS BUREAU .....         | 10 |
| PREPARING YOUR STREET CENTERLINE FILE BY BUILDING AN ADDRESS LOCATOR .....             | 12 |
| USING ADDRESS INFORMATION TO GEOCODE .....   | 16 |
| GEOCODING AGAINST PARCEL POLYGONS OR ADDRESS POINTS .....                              | 18 |
| WRAPPING UP.....   | 21 |

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

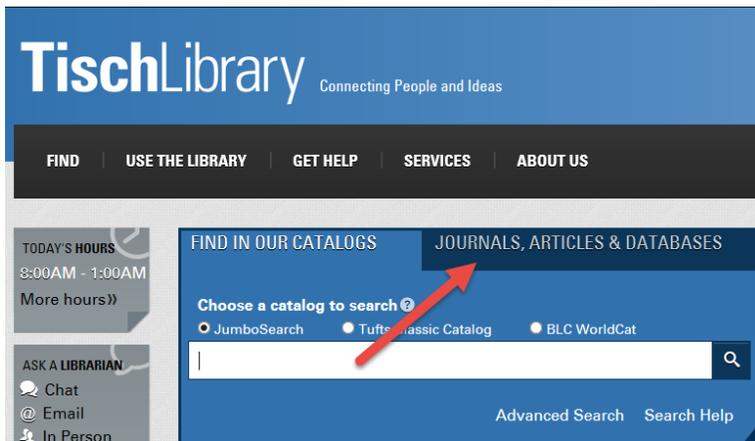
1. Determine what type of business or service you want to search for to geocode. In this case, we'll be using grocery store data.
2. Go to the Census NAICS code web site <http://www.census.gov/eos/www/naics/>
3. Above the 2012 NAICS search box on the left-hand side of the webpage, type in *Grocery Store* 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:**

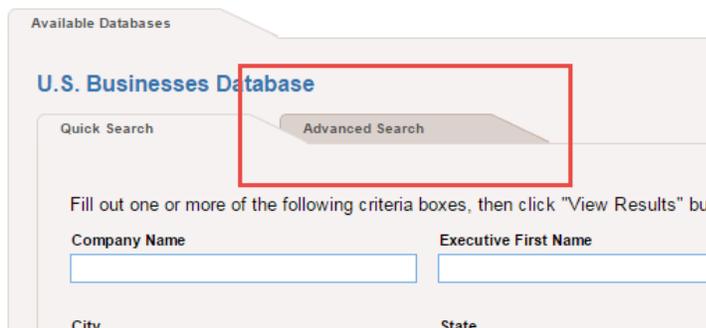
Enter keyword or 2-6 digit code

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



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 example we are searching for grocery stores (445110) in Cambridge, but you can choose other NAICS and other cities:

U.S. Business Database

**1. Select Criteria**

**2. Click here to search with NAICS**

**3. Click on "Include Unverified ..."**

**4. Type in desired NAICS code(s) here**

**5. Select State and City here. Move to right column**

**6. Click to view results**

**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 99 returns for Cambridge). We want all of them, so check the box at the top of the first column as shown here:



- 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 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**. If you downloaded information for more than one NAICS, 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

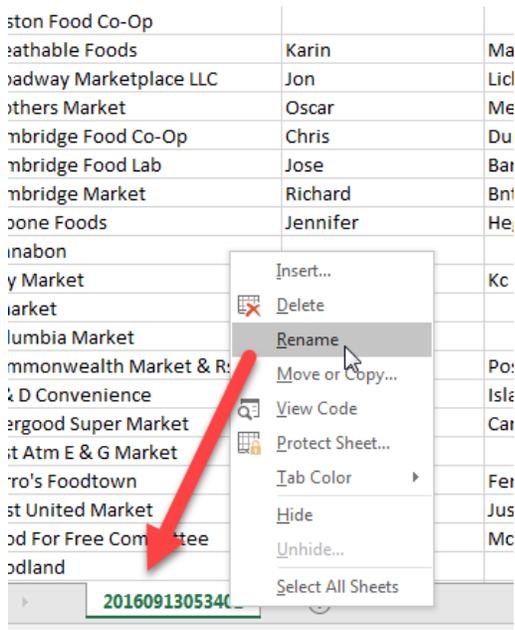
| All | Company | Management | Corporate | Total Fields Selected: 12                       |
|-----|---------|------------|-----------|---|
|     |         |            |           | Company Name <small>Remove</small>              |
|     |         |            |           | Executive First Name <small>Remove</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> |

- 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.

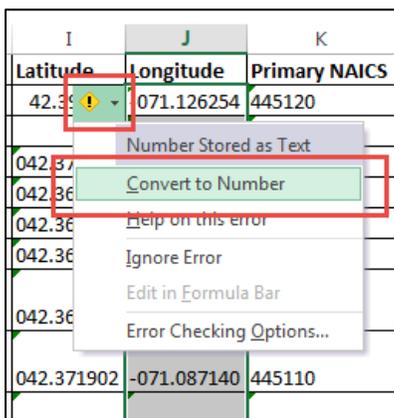
13. Before proceeding, choose **Save As** to save the file to your H: drive with a name like “Cambridge Grocery Stores” – do not use hyphens in your file name! **Important: Save it as an .xlsx file.**
14. If you had more than 250 results, go back to your results table, deselect the first 10 pages of 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 your Excel File

1. Rename the *worksheet* to something more comprehensible like Grocery Stores (no hyphens!)



2. There is one additional problem we need to correct in the Excel file – the latitude and longitude columns are text (aka string in ArcMap) and they need to be numbers.
  - a. Highlight all the *Latitude* and *Longitude* data cells (**not** the column names).
  - b. 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 your Excel file open when you work with it in ArcMap).

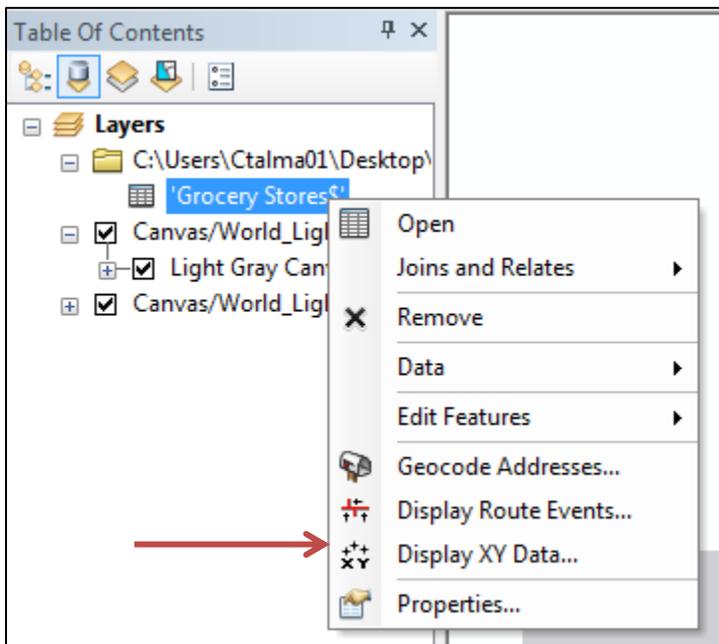
## Starting ArcMap

1. Start 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 got Reference USA data.
5. Add your Excel **sheet** to ArcMap (you have to navigate one step beyond the .xlsx file to choose the individual worksheet, e.g., ‘Grocery Stores\$’).

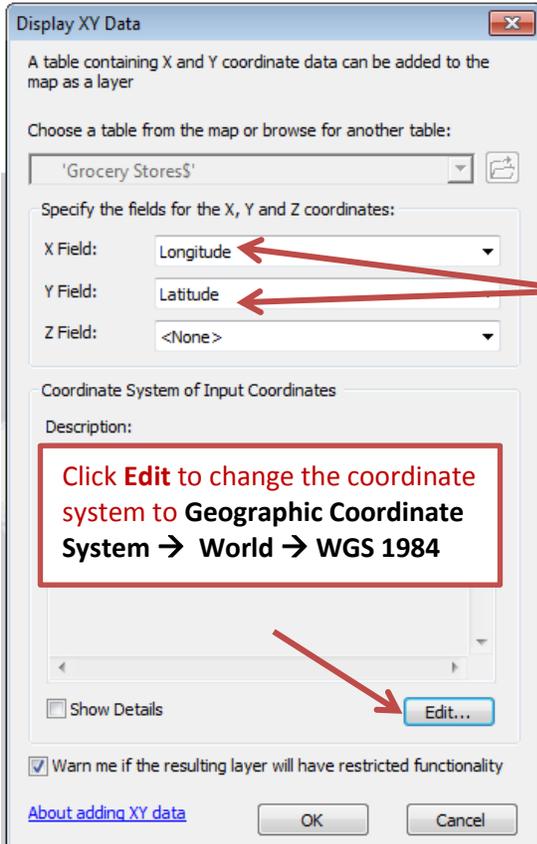
## Adding Reference USA data to ArcMap using its Latitude and Longitude Information

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

1. In ArcMap Table of Contents, *right-click* again on your Excel file (e.g., *Boston grocery stores*) and choose **Display XY Data**.



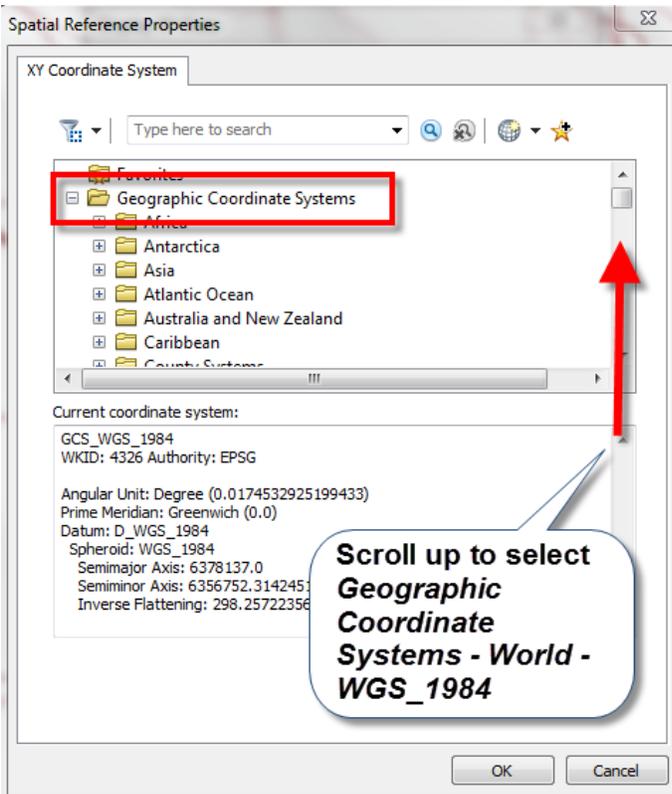
2. Fill the dialog box out as follows. Make sure you choose *Edit* to select the **coordinate system** that’s appropriate. Choose *Geographic Coordinate System* → *World* → *WGS 1984*. Make sure that it is not a projected coordinate system (we need to work with **decimal degrees**).



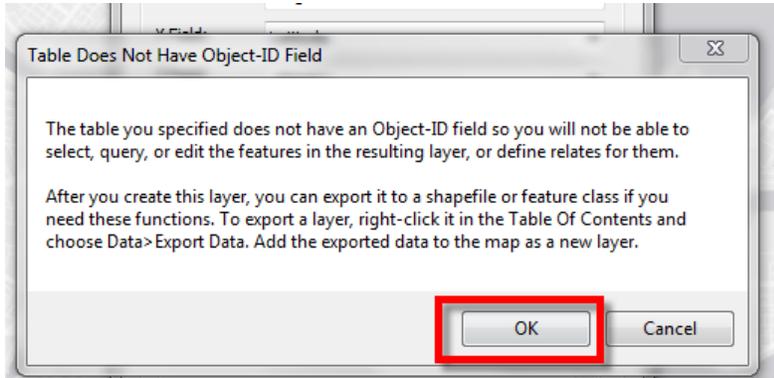
Check that these are correct.

- X is Longitude!
- Y is Latitude!

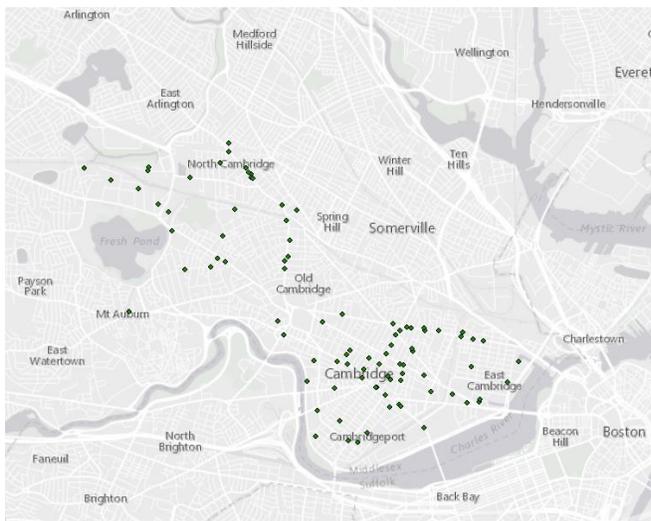
(Many people switch these by accident!)



3. Press **OK** twice.
4. Read the warning but then press **OK** again.



5. The points should appear on your map like so. Note: this is NOT a shapefile. This is just a visualization of your excel data. To save it as a shapefile, follow the directions below.



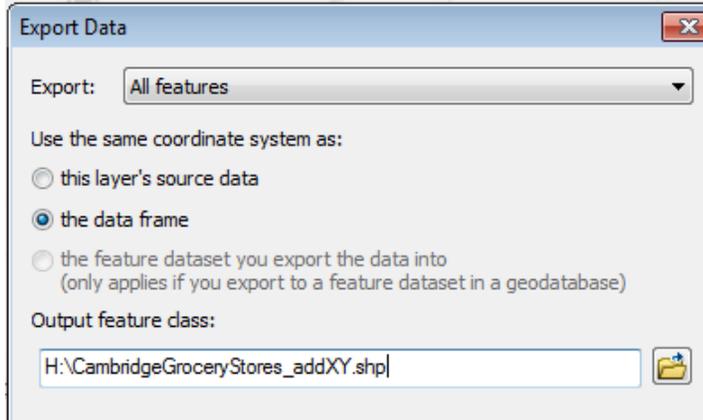
## Export Points to a Shapefile

When the data initially comes up as points in a map, ArcGIS 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 edit and use in analysis:

1. Export the “events” layer to a shapefile by **right-clicking on the *Points* events layer**.
2. Choose **Data → Export Data**.
3. When the export dialog box comes up, you can choose to export the data into the data frame’s coordinate system so that it matches your other data or you can leave it in the GCS\_WGS84 for now.
4. Press the folder button and navigate to your H drive.
5. Name the layer, **“CambridgeGroceryStores\_addXY”** so we know these are the points we created using *the Add XY data method*.
6. Make sure you save it as a shapefile in the **Save as type** dropdown.

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



8. Change the symbology of this point layer to red triangles, so we can easily tell them apart later.

## 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.

## Using Google Maps to Geocode your Business Data

This is optional but potentially useful. If you have a Google account, you can use Google Maps to geocode your Excel sheet. See this [tip sheet for instructions](#) – note in the example we use address information and compare the results to using the latitude and longitude data. **Do NOT use this method to geocode addresses that are subject to privacy restrictions (i.e., IRB restrictions).**

## 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 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 your *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 <http://census.gov>
2. Click on the **Geography** tab and then go to **Maps & Data**.



### Geography

- Main
- About
- Maps & Data
- Reference
- Partnerships
- Education
- Research
- GSS-I
- Contact Us

#### Maps & Data

- Maps & Data Main Page

#### Maps

- Census Data Mapper
- Reference
- Thematic
- Maps Available for Purchase

#### Data

- TIGER Products**
- Census Geocoder
- Partnership Shapefiles
- Relationship Files
- Gazetteer Files
- Block Assignment Files
- Name Lookup Tables
- Tallies
- LandView

#### Maps

[Census Data Mapper](#)

The Census Data Mapper is a web mapping application intended to provide users with a simple interface to view, save and print county-based demographic maps of the United States.

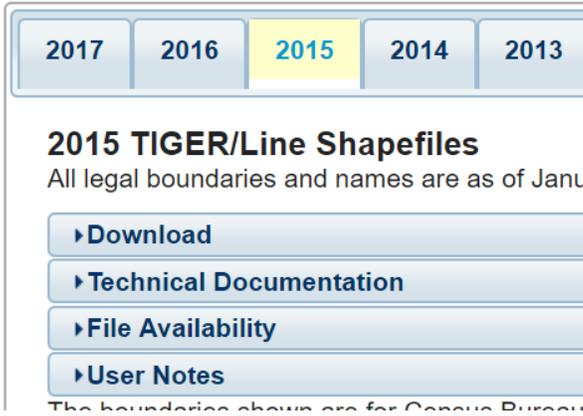
[Reference Maps](#)

3. Click on **Tiger/Line Shapefiles**.

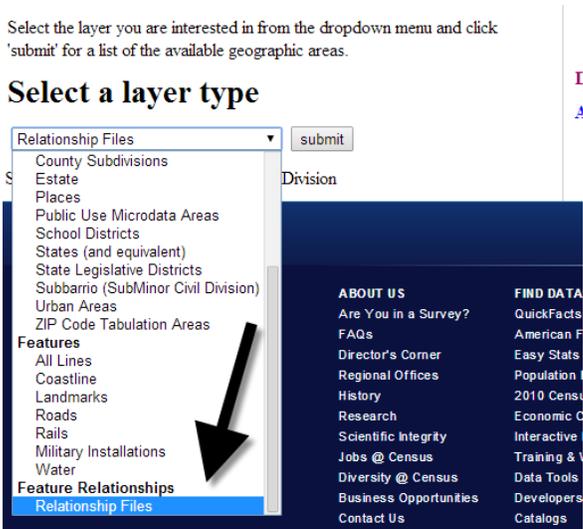
Which product should I use?

| Product  | Best For...   |
|--|---|
| <a href="#">TIGER/Line Shapefiles</a>  | Most mapping projects--this is our <i>most comprehensive dataset</i> . Designed for use with GIS (geographic information systems).  |
| <a href="#">TIGER Geodatabases</a>   | Useful for users needing national datasets or all major boundaries for by state. Designed for use in ArcGIS. Files are extremely large.   |
| <a href="#">TIGER/Line Shapefiles &amp; Geodatabases with Demographic Data</a> | Demographic analysis from selected attributes from the 2010 Census, 2006-2010 ACS 5-year estimates, 2007-2011 ACS 5-year estimates, and 2008-2012 ACS 5-year estimates for selected geographies. Designed for use with GIS. |

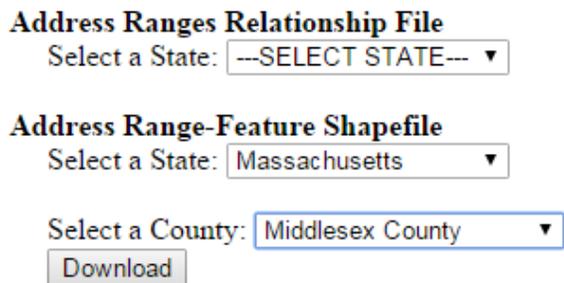
- Click on the **2015** tab, **Download** and then **Web Interface**. Make sure it's 2015 and not 2016 or 2017.



- Again, change the year to 2015 and then 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 and county and click download.



- A zipped file will download. Extract it to your H drive using PowerArchiver or another decompression program.

## 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.

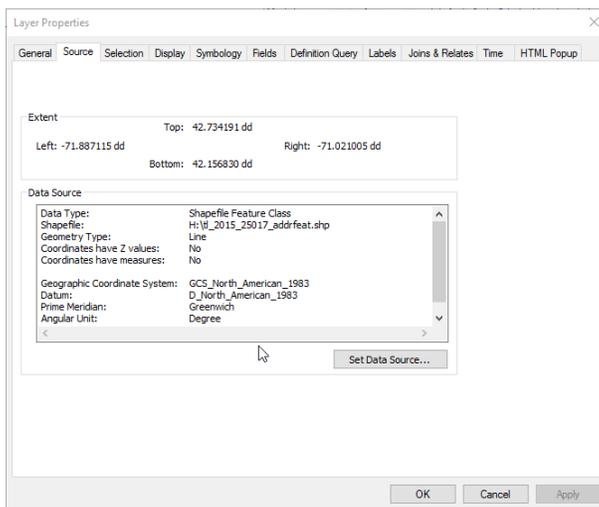
1. Add your Census Tiger street centerlines to ArcMap – it will have a name like **tl\_2015\_25017\_addrfeat**. This is the **Tiger Line** file for **2015** for Mass (**25**) Middlesex County (**017**) – Another county FIPS code would be different.
2. Open it's attribute table to see how it codes address ranges – you see the *street name column* (FULLNAME), the *Left from Address* (LFROMHN) – HN stands for House Number), *Left To Address*, etc. Leave this table open for reference.

Table

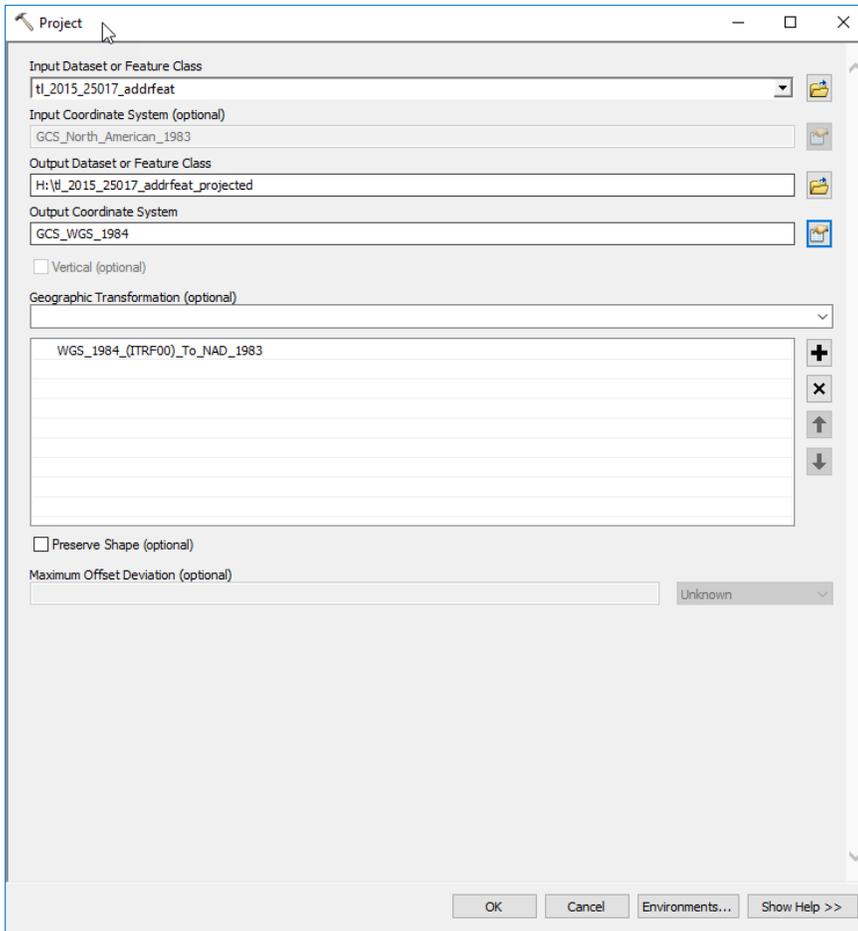
tl\_2013\_25017\_addrfeat

| TLID     | LINEARID      | FULLNAME       | LFROMHN | LTOHN | RFROMHN | 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 |

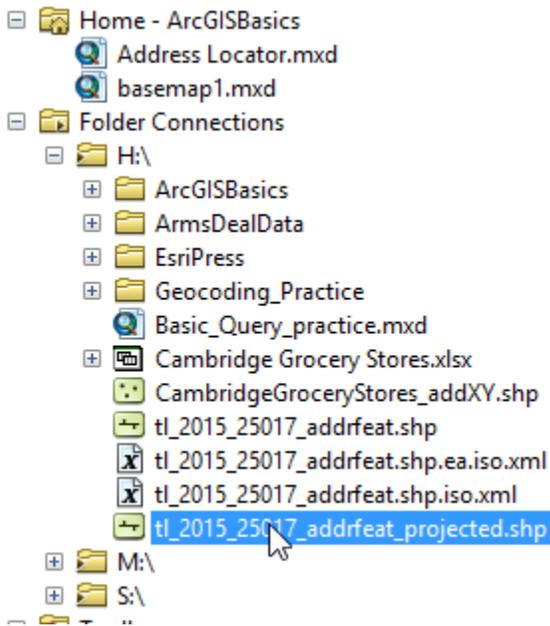
3. If you don't have the Catalog visible in ArcMap already, click on **Windows** → **Catalog**.
4. **Before creating your address locator make sure that your tl\_2015\_25017\_addrfeat file is projected to WGS 1984.** To make sure the file is projected to WGS 1984 right click and open its properties.
  - a. **If you see this:**



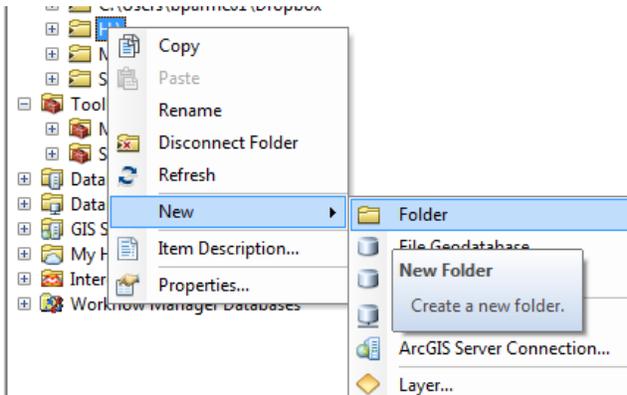
Make sure you use the **Project** tool to change the file to WGS 1984 like so.



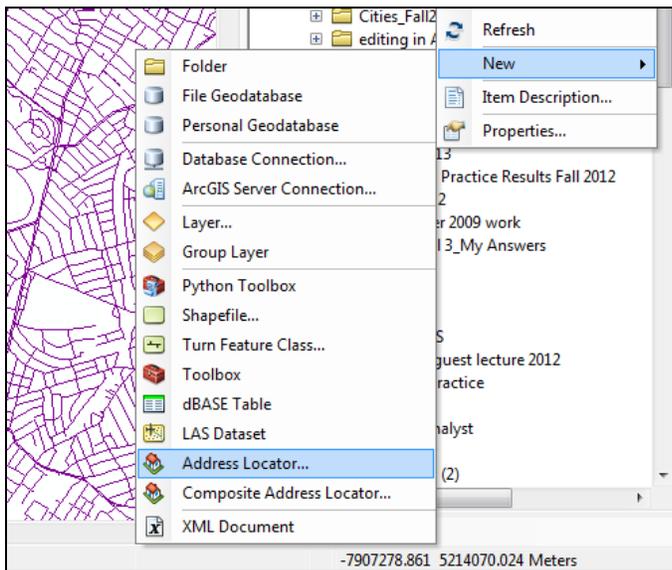
Click Ok. Go to your H drive and add the new projected street file. Remove your old street file.



- Practice good data management and create a folder structure to support geocoding. We need to create an Address Locator for our business data, so let's make a new folder in your H: drive called **Geocoding\_Practice**, then a subfolder called **Address Locators**.



- Right click on your new *Address Locators* folder and choose **New → Address Locator**.



- Fill out the dialog box for the Address Locator as you see on the next page. Refer to the TIGER street attribute table as needed. Be sure to give the OUTPUT ADDRESS LOCATOR a name like **CensusTIGER2015Streets**.

**Click Here to select this locator style**

**Click here to find your tiger data set**

| Field Name       | Alias Name |
|------------------|------------|
| Feature ID       | FID        |
| *From Left       | LFROMHN    |
| *To Left         | LTOHN      |
| *From Right      | RFROMHN    |
| *To Right        | RTOHN      |
| Left Parity      | PARITYL    |
| Right Parity     | PARITYR    |
| Full Street Name | <None>     |
| Prefix Direction | <None>     |
| Prefix Type      | <None>     |
| *Street Name     | FULLNAME   |

**Make sure each of these rows mirrors this exactly. Do not change anything below \*Street Name**

**Save it in your H drive - Address Locator folder and give it a good name like CensusTiger2015Streets**

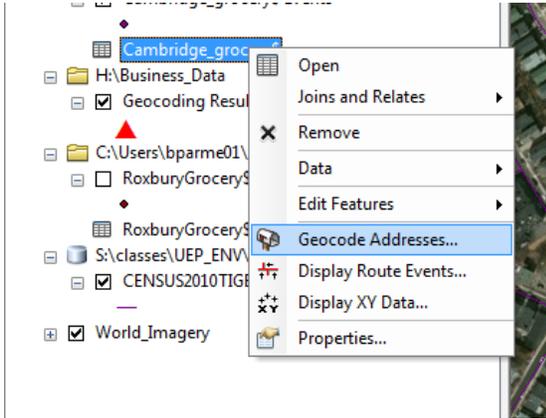
8. **Click OK** when you are done filling out the form
9. Click OK. This **process may take 5 minutes**. Relax, stretch!

## Using Address Information to Geocode

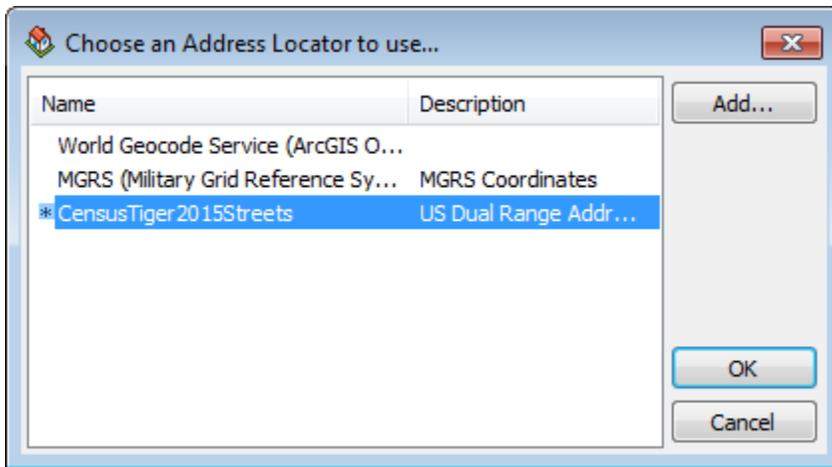
Now you're ready to geocode against the **Tiger Road centerlines** file using the address locator you created.

You should have your Excel file with grocery store data in your ArcMap session. In the example below, we are using the Cambridge grocery stores excel file.

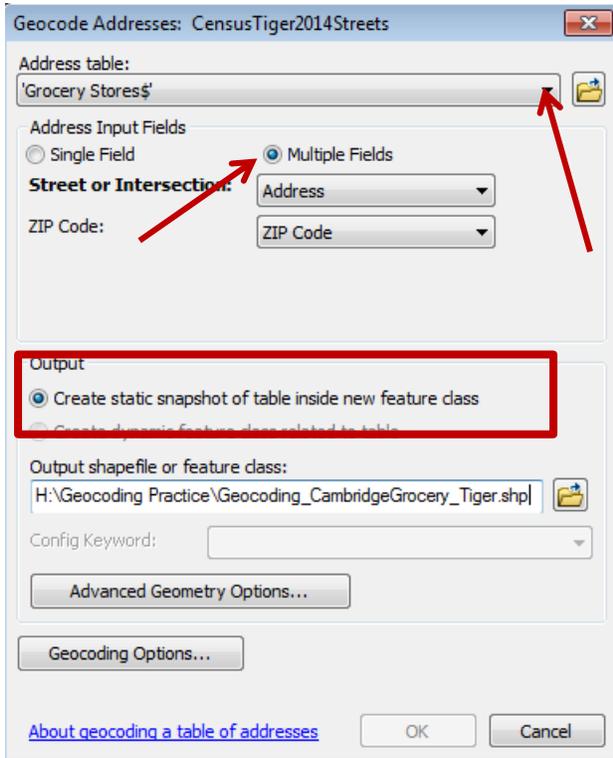
1. In the Table of Contents, right-click on your *Excel worksheet file* and choose **Geocode Addresses**.



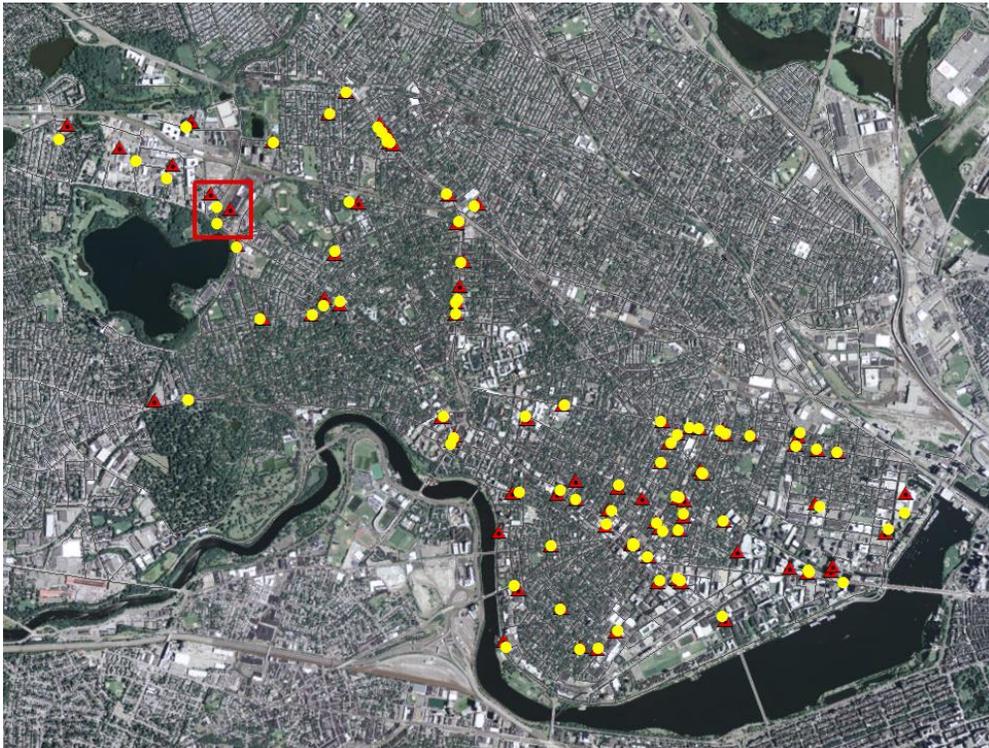
2. For your Address Locator, choose your **Tiger address locator** and click OK:



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. Change the symbology of these points to yellow circles.
5. Explore your results and compare them against what happened when you used Latitude and Longitude from Reference USA. 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?

## 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.

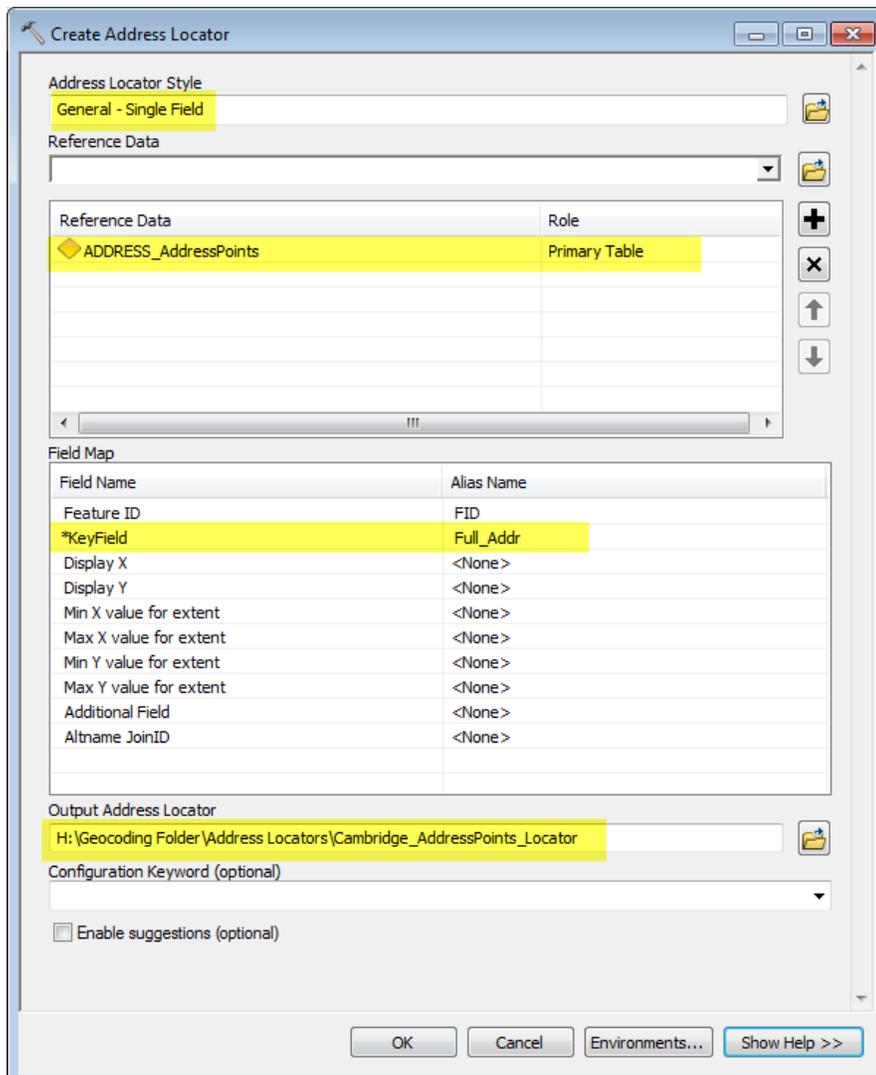
You will need to examine the attribute fields of your Parcel or Address Point GIS data set before you create the address locator to see how to map the fields.

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*

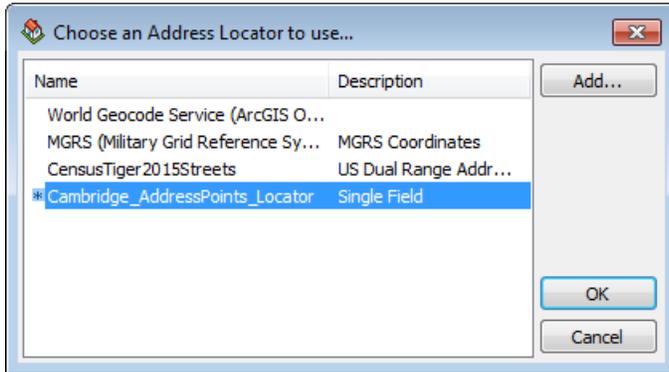
- Examine the attribute table of the *ADDRESS\_AddressPoints* data set – in the case of Cambridge we see that there is a single field for the address in the attribute table – it is called **Full\_Addr**:

| FID | Shape * | address_id | BldgID | ml      | StNm | 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   |      |          |

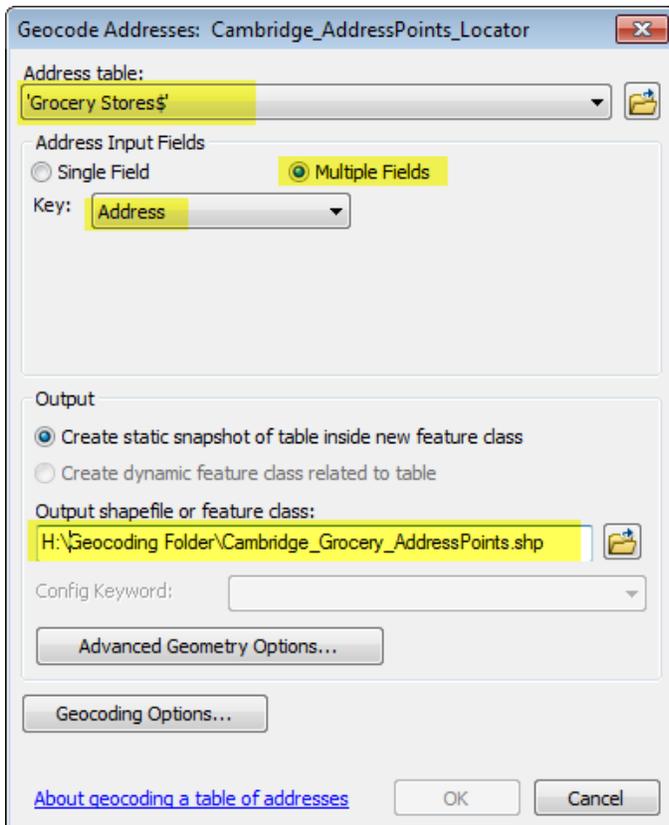
- Right click on your **Address Locators** folder and choose **New → Address Locator**.
- Fill out the **Create Address Locator Dialog Box** as shown below – remember that the *Cambridge AddressPoints* GIS data set had a single field in its attribute table for the address – the name of this field was **Full\_Addr**.
  - We will set the Address Locator Style to General – Single Field.
  - 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:



10. Click OK.
11. When the geocoding results come up, click **Close**.
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.