

Cornell University
Albert R. Mann Library

Mann Library GIS Tutorial

Geocoding Address Data Using ArcGIS

Introduction

Address data can be applied to a variety of research questions using GIS. Once imported into a GIS, you can spatially display the address locations and analyze the data for patterns. You can also filter/display your address information based on certain criteria.

Before address data can be used in a GIS, they must be “geocoded”, assigning a specific point location (e.g. latitude and longitude) to each address by comparing the descriptive elements in the address to those in a reference data source. In some cases where you are working with a limited number of addresses, or don’t require absolute control over methods or accuracy, using an online “batch geocoder” like batchgeo.com will suffice. This tutorial will introduce some of the tools and methods available within ArcGIS for using address data in GIS. It assumes some familiarity with GIS in general and ArcGIS specifically.

In order to geocode addresses in ArcGIS, the following elements are needed:

Address Data: A table containing your properly formatted address information.

Abbreviations should be standardized and grammatical errors corrected. When geocoding street addresses, data must include an address field with street number and name, and can include one or more of the following: street prefix direction, prefix type, street type, and suffix direction. Street intersections may also be included in the address field.

Reference Data: Primary reference data containing descriptive (address) and corresponding spatial information for the geographic area referenced in your address data is needed. Address data can then be matched with reference data and assigned spatial attributes (coordinates). Secondary reference data can also be employed for geocoding “alias” place names (e.g. hospitals, schools, parks) and alternate street names (e.g. highway numbers). Reference data can be file based (shapefile or coverage) or within a geodatabase (GDB).

Address Locator: An address locator uses explicit methods for matching (geocoding) address data to specified reference data based on guidelines unique to the locator style selected and any customizations. Once an address locator is created, files containing its configured properties and pathways to reference data are created. A “geocoding index” is created and stored in a new table. This index, rather than the entire reference table, is used to quickly match addresses when geocoding. If file based reference data is used, the index is stored in the same folder. If the reference data is within a geodatabase, it is stored within the same geodatabase

Geocoding Engine: The geocoding engine consists of several files within the ArcGIS program files that read the directions dictated by the address locator file and process the address being searched, resulting in a location match for the address.

Exercise 1: Creating and Formatting Address Data

In this exercise, we will work with a spreadsheet to create an address data table using Microsoft Excel. If you were using your own data, and it wasn't already in spreadsheet format it could be copied and pasted into Excel from a text file.

The data files you'll need for this exercise are located in the directory `\GIS\GIS_II\Geocode\` in the **Classroom Files** folder (use the desktop shortcut to get there). Check to see if there is a folder named "Temp" at the top level of the `C:\` drive on your computer. If no such directory exists, create one. It's important that there be no spaces or special characters anywhere in the path directories of your data files, so be sure to create this directory at the top level of the `C:\` drive.

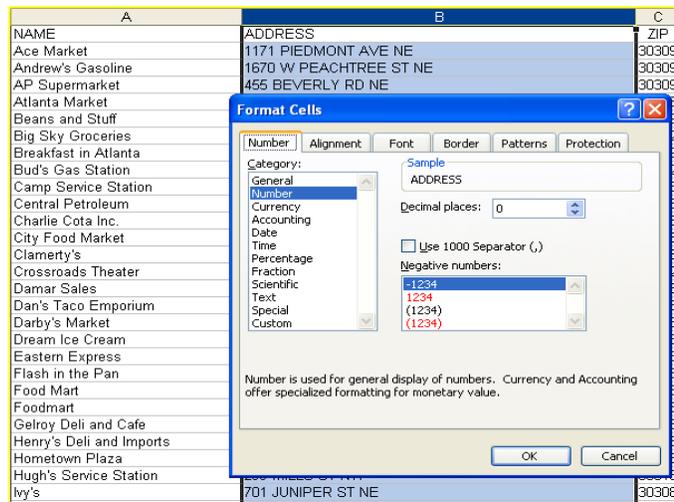
- Copy the entire **Geocode** folder to `C:\Temp` on your computer
- From your `C:\Temp\Geocode` folder open the **customers.xls** file in Excel.

This spreadsheet is a list of businesses in Atlanta, with name, address, sales and business type information. If you were working with your own data, at this point you would want to go through and correct any spelling errors and standardize abbreviations. You should add and edit column headings as necessary, always beginning with letters, shortening to 10 or less characters, removing spaces and special characters and using intuitive self explanatory names. For example, for street address information, column headings should be *Street* or *Address*, and for zipcode, *Zone*, *Zip* or *Zipcode*. This will allow the address matching process to proceed more smoothly.

When creating tables for import into ArcGIS, pay particular attention to cell formatting. When possible, each field/column should contain only one type/format of data. Many times Excel will make assumptions based on entered values. For example, cells containing numbers may automatically be defined as numeric and assigned default properties. You can **open the Format Cells menu** for a cell or column by **right-clicking** on it.

➤ **Set Formatting for customers.xls columns/cells:**

- Notice that the **address column** appears to be formatted as numeric. In order to avoid potential problems importing to ArcGIS, **change to General format**. This is usually advised for mixed format cells.
- The same applies for **zipcodes** – change to **General format**.
- Set the **Sales column** to **Number format**, with **zero decimal places**.
- Once all edits are made, **save the file and close Excel**.



Exercise 2: Setting up Your Address Locator

In order to understand how an Address Locator is constructed and functions, we will briefly review the process for creating your own. For those who desire more in depth coverage of this topic and geocoding in general, please refer later to the ArcGIS publication *Geocoding in ArcGIS* (located in the *C:\Program Files\ArcGIS\Documentation* folder).

Preparing an address locator is a two stage process, involving

1. **Formatting and standardization of reference data**, and
2. **Creation of the address locator** itself which will refer to the reference data.

Reference Data

In order to assign spatial attributes to address data, a primary source of reference data is needed which contains both spatial (point, line or polygon) and matching address elements. Ancillary reference data, including tables containing alternate street names or place aliases can also be used. For the address locator to work properly, the address elements must include the constituent elements and formatting appropriate to the address locator style being used. Reference data can be file based (e.g. shapefile) or from a geodatabase.

For street address geocoding, common address elements include

- prefix direction (“W. Redlands Blvd.”) and type (“Avenue B”),
- Streetname
- Street type
- Suffix Direction
- Zone

In ArcGIS, US Street address locators let you use reference data that includes an address range for each side of a street segment, offering a means for approximating the location of a particular address within the range. It is important to note that this method of address geocoding is indeed that, an approximation of where that address would be expected to fall within a given street segment, based on an even distribution of addresses across the range. US Census spatial data for streets can be used in this way. If more precise geocoding is desired, a more accurate source of reference data like parcel boundaries may be needed.

To illustrate what a reference file looks like, we will use a geodatabase source file.

- **Open ArcCatalog** from the ArcGIS program menu.
- From the left window, find the **C:\Temp\Geocode** folder and
- open the **Atlanta.mdb** (personal geodatabase) file by single clicking on the plus sign to the left.
- Now single click on the **Streets** data layer.

- In the right window, select the **preview tab** at the top, and look at both the **Geography** and **Table preview** options by selecting from the drop down menu at the bottom. Notice that this file contains both spatial and detailed attribute data.

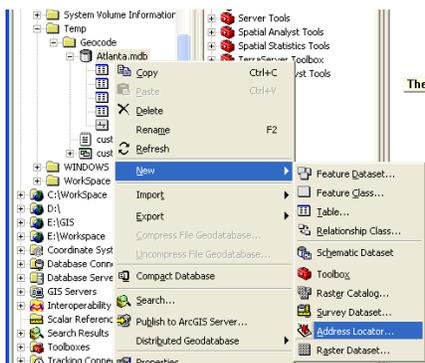
The attribute data found within the table includes columns containing left from address (L_F_ADD), left to address (L_T_ADD), right from address (R_F_ADD), and right to address (R_T_ADD) values, indicating address number ranges for the left and right side of each line segment. Street address reference data containing address ranges for the street as a whole, and corresponding address locator styles can also be used, but plotting location on a particular side of a street is not an option in that case.

OBJECTID*	Shape*	L_F_ADD	L_T_ADD	R_F_ADD	R_T_ADD	PREFIX	PRE_TYPE	NAME	TYPE	SUFFIX	ZIPL	ZIPR	CITYL	CITYR	STATE_ABBR
1	Polyline	800	898	801	899			Drummond	St	SW	30314	30314	Atlanta	Atlanta	GA
2	Polyline			173	199			Hunter	St	SW	30303	30303	Atlanta	Atlanta	GA
3	Polyline	100	128	101	129			Jones	Ave	NW	30313	30313	Atlanta	Atlanta	GA
4	Polyline	130	158	131	159			Jones	Ave	NW	30313	30313	Atlanta	Atlanta	GA
5	Polyline	160	166	161	167			Jones	Ave	NW	30313	30313	Atlanta	Atlanta	GA
6	Polyline	168	214	169	215			Jones	Ave	NW	30314	30314	Atlanta	Atlanta	GA
7	Polyline	216	262	217	263			Jones	Ave	NW	30314	30314	Atlanta	Atlanta	GA

The remaining columns of the streets table contain other address elements including prefix, street name, type, suffix, zipcode (for each side of each street segment), city and state.

When you create an address locator, the address fields in the reference data that you use for the locator should be standardized according to the geocoding rules used by the address locator. ArcToolbox contains a Standardize Address tool for standardizing and parsing elements to match a particular address locator style. Doing so will increase the geocoding match rate that you can achieve with the address locator. If your reference data and output table (geocoded address data) are part of the same geodatabase, you can create a dynamic output. Subsequent edits that you make to the address fields in the input address data will then automatically be reflected in the output table. For this tutorial we will be using preformatted/standardized reference data, so you can skip this step.

Creating an Address Locator



From the left window of ArcCatalog,

- From the **C:\temp\Geocode** folder, right click the **Atlanta.mdb** file
- Select **New > Address Locator**
- Find and select the **US Streets with Zone and AltName** (GDB). Click **OK**. We will be using data from the geodatabase previewed, including tables with alternate street names and alias place names.

- Give the Address Locator a **name** indicating it as an Atlanta street address locator, then
- browse for the **Primary Reference Data** file by clicking the folder icon.
- From the Atlanta geodatabase (**C:\Temp\Geocode\Atlanta.mdb**), find and select the *streets* shapefile.

Notice that all the locator fields have successfully matched to their respective columns in the reference data. In some cases you may need to change or complete this process manually.

There are a variety of settings, including thresholds for acceptable matches, intersection connectors and address table field names. You can leave most of these at their default values, but

- Confirm **Side offset** is set to **20 feet** so your address points are clearly drawn away from the street centerline. In some cases having X and Y coordinates (latitude and longitude) in your output table is helpful as well (checkbox in lower right Output Fields section).

If you ever need to adjust settings later, simply right click the locator within ArcCatalog and select Properties. You can also adjust when geocoding later.

- Now click on the **Alternate Name table** tab, and find the *altname* table in the same Atlanta geodatabase. This table will allow the address locator to match addresses that use alternate street names (e.g. highway numbers).
- A Join ID field is required in the altname table and the streets feature class. This field is used to relate the records in the altname table to the records in the streets feature class. **Select ALTNNAME_ID as the Join ID field.**

The last option for your address locator is a Place Name Alias table. This will allow the locator to match addresses that use place names (e.g. parks, hospitals, schools) rather than street addresses.

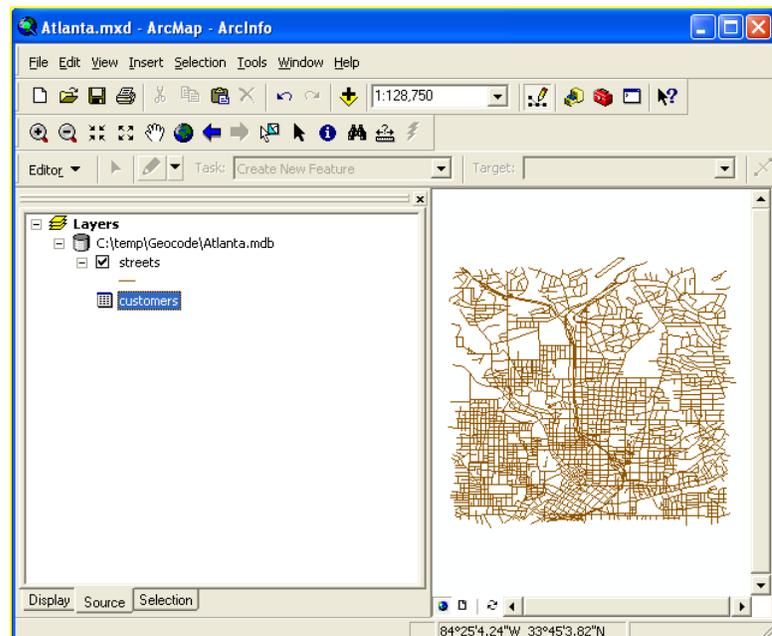
- Click on the **Place Name Alias Table** button (in Matching Options section); find (click folder) the **place_aliases** table in the Atlanta geodatabase.
- Select the **Name** field as the **Alias field**, and make sure **Street or Intersection** is set to **ADDRESS**, and **Zone** to **ZIP**.
- Click **OK**, then again for the main Address Locator. You should see the locator creation process proceed until complete.



Exercise 3: Geocoding and Mapping Your Address Data

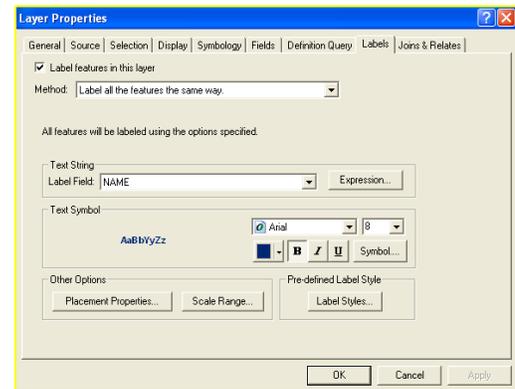
Now that you have your address data and address locator prepared, you are ready to geocode and map addresses.

- Open **ArcMap** from the globe icon on the top toolbar of ArcCatalog;
- Select a **new empty map**;
- Click the yellow **Add Data** “plus sign” button at the top, and select the **streets** feature that we used earlier from the Atlanta.mdb geodatabase (from the C:\Temp\Geocode\ folder).
- Click Add Data button again and select the **customers.xls** file edited earlier, choosing the **customers\$** table.



In order to assist us later in reviewing/rematching addresses,

- turn on the **streets layer** labeling by right clicking in the left window and selecting **Properties**. Click the **Labels tab** and make sure **Name** is selected as the Label field. Check the **Label Features in this Layer** box in the upper left, and click **OK**.

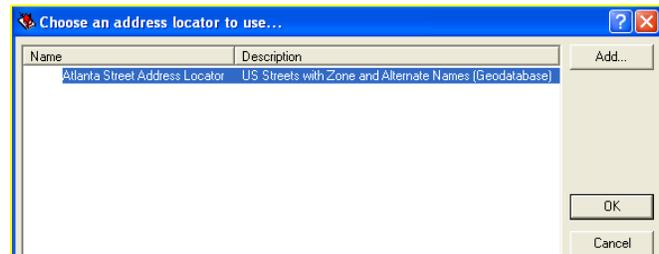


To start the geocoding process,

- from the **Tools** menu select **Geocoding > Geocode Addresses**.

In the Address Locator Manager window,

- click the **Add button**.
- From the *Look In* drop down menu, find and open the the C:\Temp\Geocode\ folder.

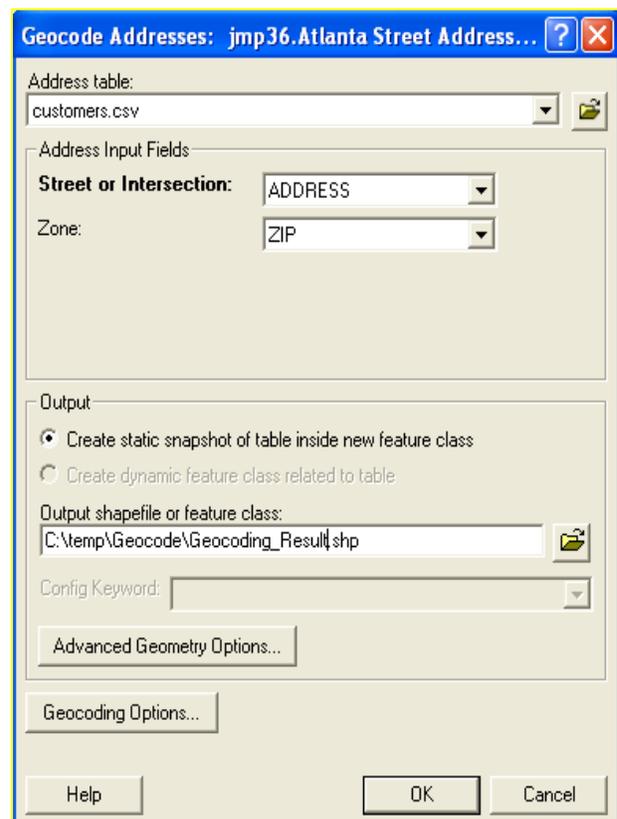


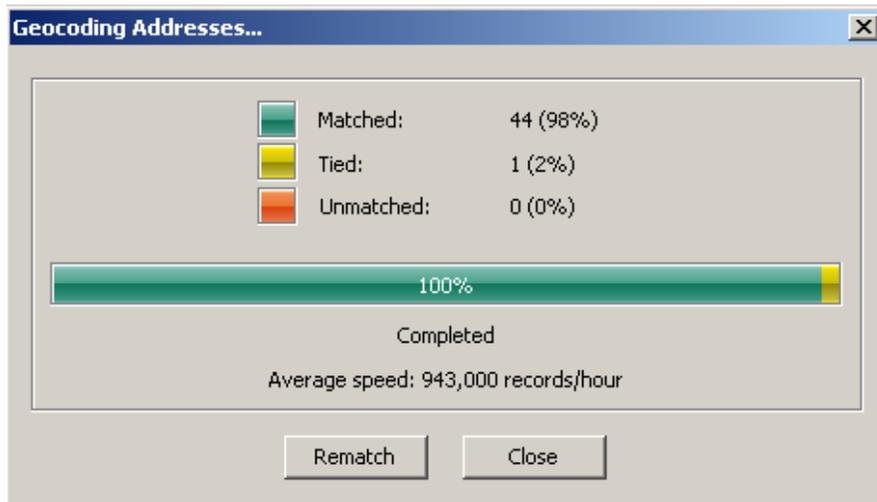
- Select Atlanta.mdb then add the **Atlanta Street Address Locator** you created.
- Now click **OK** in the Choose an address locator window.
- Select the **customers\$** table as the **Address table**, and make sure the Street and Zone fields match up correctly.

If we were using an address and reference table that were both within the same geodatabase, we could select the *Create a dynamic feature class related to table* as an option. This would allow any future updates to our address data to automatically be represented in an updated output data file (in this case a shapefile).

The geocoding options were already configured when we created the address locator so we don't need to change them. The Advanced Geometry options can be used to edit the Spatial Reference properties of the output data. Leave them at the default values as well.

- Give your output shapefile a name and save it in your C:\Temp\Geocode folder.
- Click **OK** to start the geocoding process.



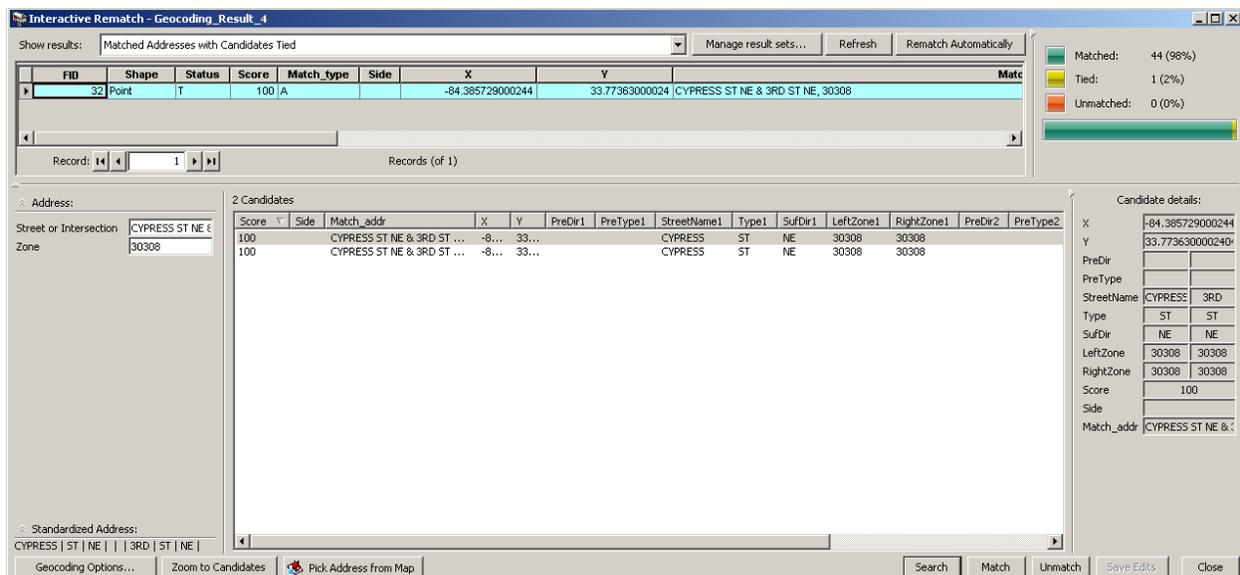


Once the automated geocoding step is complete, a window should appear indicating how many have been matched, and remain unmatched.

You should see that the 45 addresses matched successfully, however one is a match with candidates tied.

- Click the **Rematch** button to open the Interactive Rematch Addresses Dialogue box.

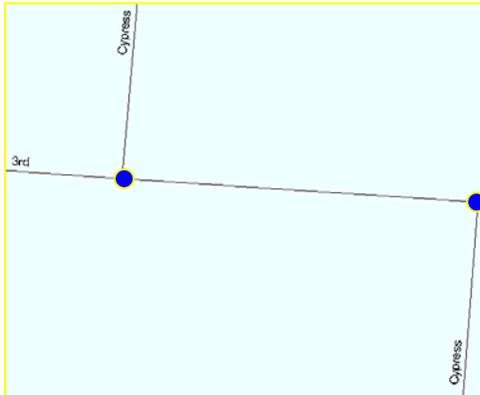
- From the top **Show Results** menu, select **Matched Addresses with Candidates Tied**



In this case, the address locator has successfully interpreted the intersection address, but has found two likely candidates.

- Click the **Zoom to candidates** button to take a closer look.

You may need to move the Interactive Review box around to get a clear view.



Notice that Cypress Street does indeed appear to intersect twice with 3rd. In cases like this you would need to verify by some other means which intersection was the correct one.

- For our purposes **select either** from the Interactive Rematch window and click **Match**. All addresses should now be matched.
- **Click Close** to close the Rematch window.

If you ever need to go back and review or rematch addresses, you can launch the Interactive Review window again through the **Tools > Geocoding > Review/Rematch Addresses** menu.

To create a new spatial file from the geocoding results, right click the Geocoding Result: Geocoding_Result layer, select Data>Export Data, and give the shapefile a name and location.

Exercise 4: Mapping Single Addresses Using Online Tele Atlas Address Finder

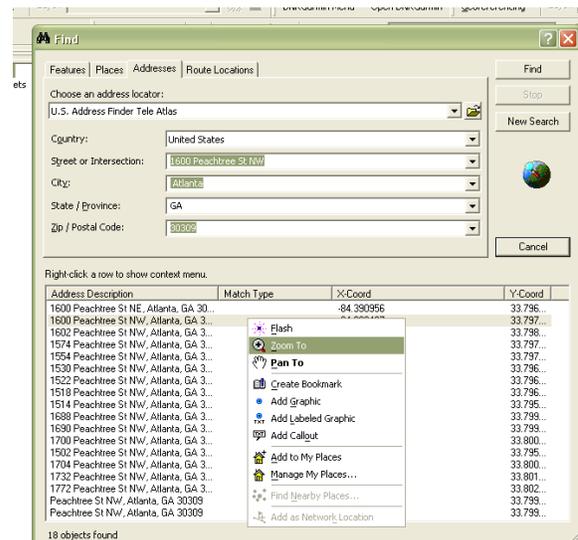
For this exercise, we will be using a licensed online locator provided by ESRI, the company that makes ArcGIS. It uses regularly updated commercial TeleAtlas data so should be fairly accurate.

To use the **Tele Atlas Address Finder** through the **Find Address** feature,

- From the top menu bar, click **Edit > Find**.
- Click the **Addresses** tab.
- Choose **U.S. Address Finder Tele Atlas**.

Now you can enter and find a single address.

- Enter **1600 Peachtree St NW, Atlanta GA, 30309**.
- Click **Find** after the address elements have been entered.
- A number of options will be presented, one with a 100% match.
- **Right click** the top match and select **Zoom to**. You may need to adjust the location and size of the Find window to see the location on the map. You can click the matched address to see on map again.
- Once you are satisfied with the match, **Right click** the candidate and select **Add Graphic** to the map. Other options include a Labeled Graphic or Callout.
- Close the Find window



It is important to remember that this single address locator method will not update or create a spatial file for this address. You can right click the candidate from the Find window and select “bookmark” to save a record of location and the data frame view/zoom when the bookmark was made. You can access and zoom to (as well as create) bookmarked views though the **Bookmarks** menu.

Cornell patrons also have the option for using a “pre-built” **Street_Addresses_US** Address Locator, from the **Classroom Files>GIS>streetmap_na>data** folder (accessible from Mann Library classroom computers only). It can be used for batch as well as individual address matching for US Street addresses. For batch processing, it is accessed though the same **Tools > Geocoding** menu used in the previous exercise.

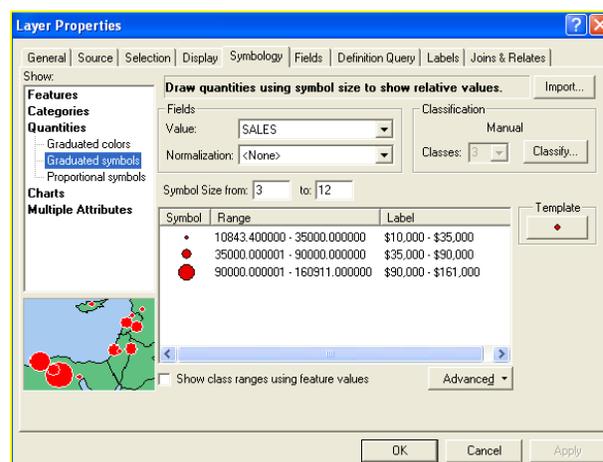
Exercise 5: Mapping Attributes (optional)

Now that you have geocoded and mapped your address points, attributes associated with them can be queried, analyzed and labeled. If you wish to add attributes not found within your existing address shapefile, other attributes can be joined (see Mann Library GIS tutorial *Adding Attribute Data to Spatial Datasets*).

- **Right click** the **Geocoded Result** layer created in exercise 3 and select **Zoom to Layer**
- **Right Click** again and select **Properties**.
- Click the **Symbology** tab to select the type and classification desired.

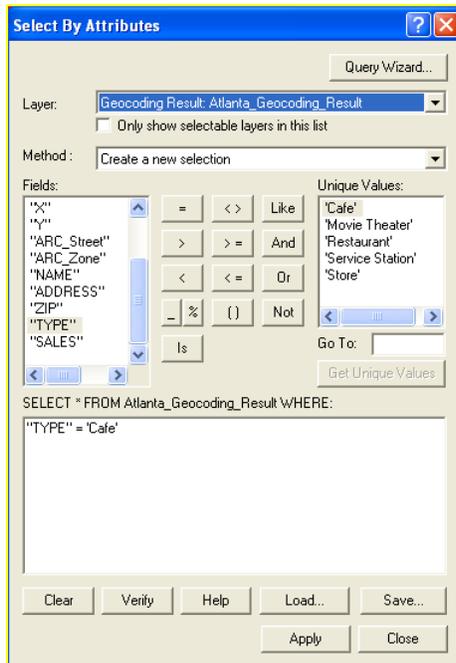
In the example below, a **Quantities- Graduated Symbols** type of symbology has been selected for Sales attributes (designated in the Value Fields box). Symbols, classes, ranges and labels can be modified as needed (labels will be represented in the Layout view on the legend if added).

- Once you are happy with your results click **OK**.



You can also choose to distinguish between business types in your map. Many options exist for representing this. If you wanted each of the business types to have a different color or symbol within the same data layer, you could use the Symbology Properties options above and select Categories –Unique Values.

Another option is to create a new map layer by querying, or selecting attributes from the geocoded data layer.



- From the **Selection > Select by Attributes** menu, this window opens.
- In this case, a “**create new selection**” query has been constructed by selecting (double clicking) the **type** field, then =, and finally **Café** as a unique value (click **Get Unique Values** to display all choices).
- Click **Apply**, then close to make your selection.
- Now right click the Geocoded **shapefile** data layer and
- Choose **Selection > Create Layer** from Selected Features. A new data layer is drawn that shows only Cafes. You can edit the layer and its properties as needed.

You can also create a separate new shapefile that only includes Cafes by right clicking, selecting **Data>Export Data>Selected Features**.