Lab Exercise 8: Geocoding

In this lab, you will learn how to convert x,y coordinates or street addresses into point theme in ArcView.

A. Data for This Lab

In this lab we will use the following themes and tables:

<table>
<thead>
<tr>
<th>Description</th>
<th>Directory</th>
<th>File</th>
<th>Type</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston TIGER file</td>
<td>$MASSGIS_1997_TIGER\b</td>
<td>boston</td>
<td>arc</td>
<td>Click on the folder icon to the left of the word &quot;boston&quot; to display the list of layers, then select the &quot;arc&quot; layer</td>
</tr>
<tr>
<td>Cambridge TIGER file</td>
<td>$MASSGIS_1997_TIGER\c</td>
<td>cambridge</td>
<td>arc</td>
<td>Click on the folder icon to the left of the word &quot;cambridge&quot; to display the list of layers, then select the &quot;arc&quot; layer</td>
</tr>
<tr>
<td>Massachusetts public airports</td>
<td>/mit/11.520/labdata</td>
<td>maPuAp.dbf</td>
<td>DBF</td>
<td>Extracted from Department of Transportation on Lecture #21.</td>
</tr>
<tr>
<td>Cambridge Public School</td>
<td>/mit/11.520/labdata</td>
<td>school.dbf</td>
<td>DBF</td>
<td>Extracted from Massachusetts Department of Education Website on Lecture #21.</td>
</tr>
<tr>
<td>Mass. major roads</td>
<td>$MASSGIS_2000_STATE</td>
<td>majrdmhd</td>
<td>arc</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------------------------</td>
<td>----------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Mass. water boundaries</td>
<td>$BOSTON_METRO</td>
<td>msa_water.shp</td>
<td>polygon</td>
<td></td>
</tr>
<tr>
<td>Mass. town boundaries</td>
<td>$MASSGIS_2000_STATE</td>
<td>towns</td>
<td>polygon</td>
<td></td>
</tr>
<tr>
<td>Boston-area 1990 Census tract boundaries</td>
<td>$C11_520_DATA</td>
<td>msa_tr90</td>
<td>polygon</td>
<td></td>
</tr>
<tr>
<td>Boston-area 1990 Census data by tract</td>
<td>$C11_520_DATA\info</td>
<td>msa_tr90.dat</td>
<td>INFO table</td>
<td></td>
</tr>
<tr>
<td>Eastern Massachusetts road network</td>
<td>$C11_520_DATA</td>
<td>east-mass-roads.shp</td>
<td>arc</td>
<td>Extracted from StreetMap USA 1.1.</td>
</tr>
</tbody>
</table>

**B. XY Geocoding in ArcView**

Suppose you're involved in a airport security program designed to prepare for emergency. You want to know exactly where the public airports are located.

- Launch ArcView
- Add ma_pu_ap.dbf table
- Go to "Add Event Theme..." under View menu
- Select ma_pu_ap.dbf for the table. Also specify X field and Y field as shown below.
After click OK, you will get the event theme.

Notice that your event theme is in Lat/Lon coordinate system.

You need to convert it to Mass State Plain coordinate system in order to overlay on top of other available MassGIS layers. First, go to View properties and set projection to State Plain - 1983 Massachusetts Mainland. Then, go to "Convert to Shapefile.." under Theme menu. Click OK at the warning of coordinate conversion due to projection.

Bring back the converted shapefile into a new View and elaborate your map with other layers.

C. Address Geocoding in ArcView

In this part, you'll use address geocoding to convert an external set of a textual data into mappable points. You'll create the map points by matching the addresses from a table of addresses. To geocode addresses, you need a reference theme (usually a street theme) that is prepared for geocoding (this is called making the reference theme "matchable").

Reference Theme

Launch ArcView and add the arc theme from the boston TIGER coverage into a new view. Note that this is not the default layer that is added to your project if you double click on the name 'boston'; you must click on the folder icon to the left first, then select 'arc'. This coverage includes all the streets in Boston.

* Refer back to the Lab section.
• Select the Properties from the Project menu and set the work directory to your 11.520 user directory (/mit/11.520/users/<your_username>).

• Make the View window active. Select the Properties from the View menu and set the map units to meters and the distance units to miles.

• Make the boston theme active.

• Go to Theme > Properties and click on the Geocoding icon (you will need to scroll down the list of icons to see it).

• By clicking on this icon, ArcView automatically checks the items from your boston theme that will be used to do the geocoding. ArcView has selected US Streets with Zone in the Address Style field (if not, select US Streets with Zone from the Address Style drop-down list). Address components required by this style have check marks beside them. ArcView pairs the field names in the reference theme with the address components they correspond to. Make sure that each of the address components uses the right field as shown in the table below. If the correct features are not filled in, you probably didn't add the arc feature class from the boston coverage. If so, carefully add the theme again, this time clicking once on the icon to the left of the coverage name, then selecting the arc feature class of the coverage.

<table>
<thead>
<tr>
<th>Address Style</th>
<th>US Streets with Zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>LeftFrom</td>
<td>Leftadd1</td>
</tr>
<tr>
<td>LeftTo</td>
<td>Leftadd2</td>
</tr>
<tr>
<td>RightFrom</td>
<td>Rgtadd1</td>
</tr>
<tr>
<td>RightTo</td>
<td>Rgtadd2</td>
</tr>
<tr>
<td>PreDir</td>
<td>Fdpree</td>
</tr>
<tr>
<td>PreType</td>
<td>none</td>
</tr>
<tr>
<td>StreetName</td>
<td>Fname</td>
</tr>
<tr>
<td>StreetType</td>
<td>Ftype</td>
</tr>
<tr>
<td>SufDir</td>
<td>Fdsuf</td>
</tr>
<tr>
<td>LeftZone</td>
<td>Zipcolef</td>
</tr>
<tr>
<td>RightZone</td>
<td>Zipcorgt</td>
</tr>
</tbody>
</table>

• Click OK to set the geocoding theme properties.

• The Build Geocoding Index dialog box may appear (if the theme has been already set for geocoding, this dialog box does not display). Click Yes. ArcView builds geocoding indexes on the address fields in the attribute table to make address geocoding run faster.

• Can you guess what each of these fields means and how it is used as part of the geocoding? Take a look at the attribute table for the boston arcs. Be sure you understand the meaning of the address ranges in the 'leftfrom', 'leftto', etc. columns. (If you like, you
can refer to the ArcView online help under Help Topics > Index > Geocoding and look at 'address components', 'address style', etc.)

Interactive Address Geocoding

- When the index is built and the View window is once again active, you will see the normal ArcView toolbars. Click on the Locate Address button (It's the 7th icon from the left in the upper row of icons, between the binoculars and the hammer!) By moving the pointer over the icon, ArcView displays the function of each icon at the left lower corner of the ArcView window. This icon is used to **locate an address in the active, matchable theme.** Click on it.
- A Locate window pops up when you click on this icon. Enter the address below and then click on OK:

Enter Address: 4 Yawkey Way Zone: 02215

A marker is displayed showing the location of the address. You just geocoded the address of Fenway Park in Boston! Try this again typing in the word **Street** instead of the abbreviation **St.** (Clear the old marker away first using Edit > Select All Graphics, then Edit > Delete Graphics. Address markers are treated as graphics objects in ArcView.) Does the address matching work?

Note that if you click on the Preferences button in the Locate Address dialog box, you can adjust the matching score the way you like.

Matching a List of Addresses: Overview

Next, you will perform the address matching and geocoding using an address table that contains the addresses you want to locate as points on the map. For this exercise, you will use the table school.dbf, which contains the addresses of Public Schools in Cambridge.

The geocoding process consists of three steps:

1. You prepare a reference theme for geocoding.
2. You match a 'batch' of addresses that are stored in a table (the dbf table you just added). Each matched address is added as a point in a new geocoded theme.
3. You re-match the addresses that couldn't be batch matched. ArcView adds these points to the geocoded theme.

Matching a List of Addresses: Step-By-Step Procedure

After preparing reference map, the next step in address geocoding is to "batch match" addresses in a table to the reference theme. During batch matching, ArcView completes two processes at once. The first is address matching, where addresses in the address table are compared to address
ranges in the reference theme. The second is geocoding, where each matched address in the address table is assigned geographic coordinates and made a point feature in a new theme.

- Add the cambridge tiger file
- Make the cambridge tiger file "matchable" like above
- Add the school.dbf table to your project if you haven't already done so
- Make the View window active
- Choose Geocode Addresses from the View menu. The Geocode Addresses dialog box displays
- Make sure that Cambridge is selected as a Reference Theme.
- From the Address Table drop-down list, choose school.dbf
- Select:
  - Address1 as the Address Field
  - Zip as the Zone Field
  - Sch_name as the Display Field
- Set the filename for saving the Geocoded Theme to be camb_pub_schools.shp.
- Leave other values at their defaults
- Click Batch Match to start address matching. When it is finished, the results are summarized in a new window that pops up:

![Re-match Addresses dialog box]

The Re-match Addresses dialog box informs you that 12 of the 18 records in the school.dbf table had good matches, with scores of 75 or above, but that 6 records weren't matched (A score of 75 is a good match. Sixty is a partial match. Both good and partial matches are geocoded).

- We'll accept the 12 out of 18 matches for now, so click Done in the Re-match Addresses dialog box. ArcView creates a shapefile that contains a point feature for each of the 12 matched addresses. The shapefile is added to the view as a theme, camb_pub_schools.shp.
- Click the check box of camb_pub_schools.shp theme to display the locations.
- Open the attribute table of camb_pub_schools.shp. Widen the table and scroll to the right to see all of the fields in the table. ArcView copied all the records from the original school.dbf table into the attribute table for the camb_pub_schools.shp theme. Some new fields were also created and all 18 sites are listed even though 6 sites aren't yet mappable since their Av_score value wasn't high enough for the automatic geocoding to
assign a location. Note, also, that your original table, school.dbf, is unchanged. Do you understand what information was used to match the addresses and generate the points that you see on the map?

Handling Unmatched Addresses

The last step in address geocoding is to re-match the addresses that ArcView couldn't batch match. Addresses may fail to match because of misspellings or other errors. Interactive re-matching lets you look at candidate addresses from the reference theme and decide if any of them are close enough to accept (according to your judgement).

- In the Re-match Addresses dialog box, click the Interactive Re-match button. The Geocoding Editor dialog box displays. (If you already made this dialog box go away, you can get it back by selecting the camb_pub_schools.shp theme in the view, then choosing Theme > Re-match Addresses.)
- You will see that the second unmatched address is "GRANITE STREET" for "MORSE".
- Given that there are 3 candidates, and you have no street number to guide you, you probably should pass on matching this one until you have more information. Click on the Next button to move on.
- The third address to consider will be "LINNAEAN ST" for "Peabody". Although again you are missing the street number, here, however, the sixth one is a strong candidate. Click on the Match button to accept the sixth one.
- Go through all 6 unmatched items and see how many you can provide satisfactory matches for. You'll find that the addresses fail to match for a number of reasons:
• The address number doesn't fall within the address ranges for the link (as above)
• There's no address number at all (e.g., "MEGUNKO RD")
• The address is given as an intersection ArcView can't find (e.g., "COMMERCE & ATLANTIC")
• ArcView parses the address incorrectly (e.g., "63 SOUTH AVE BLDG 6"). Click on the Edit Standardize button to see how ArcView parsed the address and how it went wrong.
• The address is something too informal (e.g., "OFF SALEM TURNPIKE")
• The address is misspelled, or other parts of the address (ZIP code, city, etc.) are wrong.

• Note that you can edit the text that ArcView processes by editing the fields. This can help a lot when ArcView gets a bit confused. Take a look at the example below:

• In addition to editing the address itself, if you're not happy with the way ArcView parsed an address, you can click on the Edit Standardize button and change the parsed result directly.

• When you are done matching as many as you can, click the Done button.

Note: Manually geocoding addresses is an exercise in judgement. However, by closely investigating the data, you can make your decision with your rational reasoning. For example, you may be able to confirm if the address is correct by referring to a secondary source. You can also see if the location of the candidate is the same as the location geocoded by some on-line geocoding engines. Here are some useful links:

• YAHOO Yellow Pages
• YAHOO Maps
• MapQuest
• Superpages

Back to Lecture*

What you turnin

Your final task is to draw two maps for each geocoding.

* Refer back to the Lab section.