

GIS AND PUBLIC HEALTH EXERCISE 4 - MAKING DATA MAPPABLE (ArcGIS 10)

PREPARATION

Download the **exer4** folder you will need for this exercise from the online supplement.

All of the databases and files used in the exercise will be stored in various subfolders within the folder called **exer4**. The following instructions are written for this folder to be located on the **c:** drive. If the folder is located on another drive, the path names shown below should be modified accordingly. Some of the folders are empty. They have been included because you may need to save the results of an operation to one of these folders.

The map documents created using ArcGIS10 reference the spatial databases and tables in the application based on the directories and paths where the data are stored. Changing the locations of databases in the system can prevent a GIS application from working properly.

Connecting to the Exercise Folder

Go to **Start ⇒ Programs ⇒ ArcGIS ⇒ ArcCatalog 10** to start ArcCatalog.

Find the button labeled **Connect to Folder** and click the button. Navigate to **c:\exer4** then click OK and look at the Catalog Tree in the left window to see that the folder has been added.

Within the data folder, data can be organized in folders identifying the agency that produced the data and then by the format of the data. For these exercises, you will consider yourself to be working for the organization called “agency” that is creating the GIS.

As you work through the exercises, you will be retrieving data from and saving data to specific folders. Please make sure you understand the System Design for the exercises.

Use the **File ⇒ Exit** menu to close ArcCatalog.

PREPARING TO ADDRESS MATCH GEOCODE

Go to **Start ⇒ Programs ⇒ ArcGIS ⇒ ArcMap 10** to start ArcMap.

In the “ArcMap – Getting Started” window, close the window you would use to open an existing map document or make a new map using a template.

Rename the Layers data frame by right clicking the word Layers and selecting the **Properties** item in the menu. Then select the **General** tab and enter the name Address Match. Click OK. The name of the Data Frame in the Table of Contents window should now appear as Address Match.

Address match geocoding involves matching a database table containing records with streets addresses to a street database with address ranges for each street segment based on a set of matching rules. In this exercise, you will first add and explore the table, then add and explore the street network database, and then create an Address Locator which describes the rules for matching the table and the street database. Then, you will perform the geocoding analysis.

Add a Table of Physician Office Locations

To begin, add a table of selected office locations of selected internal medicine and family medicine practitioners in Hartford County, Connecticut. This table of data was prepared from public sources of data including online directories for the purposes of this exercise. Find the button labeled **Add Data** and click the button. You should find the **c:\lexer4** folder in your catalog. If not, please connect to the folder using the **Connect to Folder** button.

Navigate to **c:\lexer4\data\agency\dbase** and add the **physician.dbf** data table.

Use the **Save** button or go to **File ⇒ Save** to save your map document. Navigate to **c:\lexer4\mapdocs** and save the file as **exer4.mxd**.

Right click on the physician table, click **Open** to open the table, and explore the fields. There is a physician ID number, specialty, address, town, state, and 5-digit ZIP code field for each physician. Note that the ZIP field is a character field. ZIP codes in Connecticut begin with "0".

Add an Address-Ranged Street Database

Next, add a street database with address-range fields. The database has street segments for Hartford County, Connecticut. It was extracted from a database of all TIGER[®] lines for Hartford County downloaded from the U.S. Census Bureau web site. Using the ROADFLG field, street segments were extracted. Then, the data were projected using the Connecticut State Plane Coordinate System NAD-83 Feet. Find the button labeled **Add Data** and click the button. You should find the **c:\lexer4** folder in your catalog. If not, please connect to the folder using the **Connect to Folder** button.

Navigate to **c:\lexer4\data\agency\shapes** and add the **tl_2009_09003_roads_projected.shp** shapefile.

Then, click the **Add Data** button and add the **tl_2009_09003_cousub_projected.shp** shapefile stored in the same folder. This database shows boundaries of the 29 towns in Hartford County, Connecticut. Towns are county subdivisions. A database of county subdivision boundaries in Hartford County was downloaded from the Census Bureau web site and then projected.

These databases are stored in the agency subfolder because they have been modified by the user and are no longer the same data downloaded from the Census Bureau site.

Click the rectangle in the legend of the **tl_2009_09003_cousub_projected.shp** to open the "Symbol Selector" window. Change the "Fill Color:" to No Color and increase the "Outline Width:" to 2.00 to better visualize the locations of the town boundaries.

Use the **Save** button or go to **File ⇒ Save** to save your map document. Navigate to **c:\lexer4\mapdocs** and save the file as **exer4.mxd**.

Right click the **tl_2009_09003_roads_projected.shp** and **tl_2009_09003_cousub_projected.shp** shapefiles in turn and select **Open Attribute Table** from the menu to explore the fields in each database. Then close each table.

Create an Address Locator

In addition to preparing a table of addresses and a street database with address ranges, you will need to specify rules for matching the table of addresses to the street database. In ArcGIS 10, this is accomplished by creating an Address Locator.

To create the Address Locator, click the **Catalog window** button to open ArcCatalog.

Navigate to the **c:\lexer4\locator** subfolder. Right click the subfolder name in the Catalog Tree and select **New ⇒ Address Locator** from the menu. This will open the “Create Address Locator” window.

Under “Address Locator Style”, click the Browse button and scroll through the pull-down list to highlight “US Address – Dual Ranges” then click OK. There are a wide range of styles you can select. US Address – Dual Ranges is a simple style requiring for street data with address ranges on both sides of the street.

Under “Reference Data”, click the Browse button and navigate to **c:\lexer4\data\agency\shapes** and select the **tl_2009_09003_roads_projected.shp** shapefile. This file name should be added to the box below as the Reference Data and the Role of the table should be Primary Table. If it is not, click under Role next to the shapefile name and use the pull-down to change the Role to Primary Table.

The Field Map below will be populated with values. Next to the *Street Name field, click in the box under Alias and select FULLNAME from the pull-down menu of fields. This is the name of the field containing the Street Name in the tl_2009_09003_roads-projected.shp shapefile.

Under Output Address Locator, click on the Browse button and navigate to **c:\lexer4\data\locator** and name the locator **dual**. Then, click OK.

Before you used the Address Locator, you can adjust its properties. In ArcCatalog, navigate to the **dual** Address Locator and double click it to open the “Address Locator Properties” window.

The Name and Primary table properties can be viewed but not modified.

Now set the parameters for the “Input Address Fields”. The field containing the Street name will be recognized by the software if the field is named Address, Street, or others in the table of addresses you wish to geocode. If the field storing the address in the table of addresses has some other name, you can add that field name using the Add button. Next, in the window on the left, click ZIP to highlight it. You should see that the software will recognize fields named ZIP, ZIPCode, and others as the ZIP field in the table of addresses. If the field storing the field you wish to use as a ZIP in the table of addresses has some other name, you can add that field name using the Add button.

Once you are certain that the software will recognize the street and ZIP fields from the table of addresses, you can complete the Address Locator by specifying the “Matching Options”.

Set the “Spelling sensitivity:”, “Minimum candidate score:”, and “Minimum match score:” values to 80 for each. This will ensure that the initial matches are accurate and allow you to explore unmatched addresses.

Because the building associated with the addresses are set back from the street centerline, use a distance of 50 in Reference data units as the “Side offset:”. The reference data units for the street database are feet. This will place the geocoded location 50 feet to the right or left of the street centerline. Leave the end offset at 3%.

Check the box to show matches for candidates with tie scores.

Check the boxes under “Output Fields” to create fields storing X and Y coordinates, the reference data ID, and percent along the street segment as fields describing the geocoded points.

Then, click OK to modify the Address Locator based on the parameters you specified.

When the dual locator has been added to the Locator folder, use the **Autohide** button to close ArcCatalog.

ADDRESS MATCH GEOCODING

Batch Geocoding

Click the **ArcToolbox window** button to open the ArcToolbox window. Click **Geocoding Tools** ⇒ **Geocode Addresses** to open the “Geocode Addresses” window.

Use the Browse button to add the **c:\lexer4\data\agency\dbase\PHYSICIAN.dbf** table as the Input Table of addresses.

Click the Browse button and navigate to:

c:\lexer4\locator

and add the **dual** Address Locator to the window.

Once you have chosen the dual Address Locator to use for the analysis, the table below will be populated with field names for the input address fields and the Aliases used in the PHYSICIAN.dbf table. For example, the Street field is named ADDRESS in the PHYSICIAN.dbf table.

For the Output Feature Class, used the Browse button and navigate to **c:\lexer4\data\agency\shapes** and save the snapshot as:

c:\lexer4\data\agency\shapes\Geocoding_Result.shp

Then click Save and then click OK to initiate the geocoding process.

When the data have been geocoded, close the Geocoding Addresses box.

Exploring Unmatched Records

When the processing is completed, right click the Geocoding_Result layer in the Table of Contents and select **Open Attribute Table**. Right click the Status field and select **Sort Descending** from the menu. You should find that 15 (5%) of the 310 addresses were left unmatched. There were 275 records (89%) which matched with scores of at least 80 out of 100. There were 20 addresses (6%) with tie scores among candidate matches.

Also, In addition to the Status, Score, Match_type, and Side fields, you should see that X and Y coordinate fields have been added showing the geocoded point locations in Connecticut State Plane Coordinates NAD 1983 (Feet), and that the percent along (Pct_along) field have also been added as called for in the dual Address Locator.

Look at the records in the table to identify reasons why some records may not have matched. Some of these errors can be corrected to match the addresses. Then close the table.

Click on the symbol for the Geocoding_Result layer in the Table of Contents to open the “Symbol Selector” window. Increase the size and change the color of the symbol to enhance your ability to see the geocoded physician office locations. Then **Save** the map document.

Review/Rematch Addresses

To review and rematch addresses in ArcMap, go to **Customize** ⇒ **Toolbars** and check the **Geocoding** toolbar to add it to the interface. **Save** the map document.

Click on the **Review/Rematch Addresses** button in the toolbar to open the “Interactive Rematch – Geocoding_Result” window.

You should see a table of all records with the first record highlighted. Use the pull-down menu next to “Show results:” and select Unmatched Addresses from the list to display only the unmatched addresses. Scroll down the records until you find the first unmatched record (the Status field has a U in it). Scroll across to see that this record is for 6 Northwester Dr. in Bloomfield, CT 06002. Click on the small box at the left of the record to highlight the record.

In the area below the table under “Address:” where the Street or Intersection and ZIP Code are displayed, change the “Street or Intersection” to 6 Northwestern Dr. and then use the Enter key to update the Street or Intersection.

You should see 1 candidate match listed in the window to the right. Click the candidate entry to highlight it, then click the Match button below.

You should see that the record in the table at the top of the “Interactive Rematch – Geocoding_Result” window is now updated to show a match and the overall match results are also updated.

Continue to scroll down through the table of results and highlight each unmatched record.

265 Ellington Rd., East Hartford, CT, 06108 is a complete address but address ranges for segments along Ellington Rd. do not include 265. This situation has to be investigated to see whether the street number is wrong for the two records with this number or the address ranges are wrong. Errors would have to be corrected to match these records.

300 Western Blvd., Glastonbury, CT 06033 has an incorrect street name. Change the Street or Intersection to 300 National Dr. and leave the ZIP Code as 06033 and match the record for each of the 2 records with this address.

320 Western Blvd. Glastonbury, CT 06033 is a complete address. There address ranges are missing for this street segment in the street database. This error would have to be corrected in the street database to obtain a match for this one record.

100 Asylum Ave., Hartford, CT 06105 has a typographical error in the street number. Change 100 Asylum Ave. to 1000 Asylum Ave. and match the record.

1 Liberty Sq., New Britain, Connecticut, 06051 is a complete address but the street network database does not have a street segment corresponding to this roadway which connects to a commercial development. Four records have this address.

2275 Silas Dean Hwy., Rocky Hill, CT 06067 has an incorrect street name. Change the Street or Intersection to 2275 Silas Deane Hwy. and leave the ZIP Code as 06067

2800 Tamarack Rd., South Windsor, CT 06074 is a complete address but the street network database does not have a street segment corresponding to this roadway which connects to a commercial development. Two records have this address.

1753 Meriden Waterbury Rd., Southington, CT 06467 has an incorrect street and ZIP. Change the Street or Intersection to 1753 Meriden-Waterbury Tpke. and the ZIP Code to 06479 then match the record.

You have now corrected all of the Unmatched records you can correct. Now use the pull-down list next to "Show results:" and select "Matched Addresses with Candidates Tied" to choose the address to match from among the ties.

Find the record for 8 Canal St. in Avon. You should see 2 candidates with the same match score. For the purposes of this exercise, select the 8 Canal Ct. record and match the record.

923 Farmington Ave., Bristol, CT 06010 is a complete address but the address ranges for segments along Farmington Ave. in Bristol do not include 923. This situation has to be investigated to see whether the street number is wrong or the address ranges are wrong. Errors would have to be corrected to match this one record. Do not match the record.

893 Main St., East Hartford, CT, 06108 is a complete address but the address ranges for segments along Main St. do not include 893. This situation has to be investigated to see whether the street number is wrong for the four records with this number or the address ranges are wrong. Errors would have to be corrected to match these records. Do not match these records.

745 Main St., East Hartford, CT, 06108 is a complete address but the address ranges for segments along Main St. do not include 745. This situation has to be investigated to see whether the street number is wrong or the address ranges are wrong. Errors would have to be corrected to match this one record. Do not match these records.

1027 Farmington Ave., Farmington, CT 06032 is a complete address but the address ranges for segments along Farmington Ave. do not include 1027. This situation has to be investigated to see whether the street number is wrong or the address ranges are wrong. Errors would have to be corrected to match this one record. Do not match this record.

27 Sycamore St., Glastonbury, CT 06003 has an incorrect ZIP. Change the ZIP Code to 06033 using the method you used to correct the ZIP Code and match the record.

80 Seymour St., Hartford, CT 06102 has an incorrect ZIP. Change the ZIP Code to 06106 using the method you have learned to correct the ZIP Code and match the 6 records with this address.

85 Seymour St., Hartford, CT 00610 has an incorrect ZIP. Change the ZIP Code to 06106 using the method you have learned to correct the ZIP and match the record with this address.

151 Farmington Ave., Hartford, CT 06156 has an incorrect ZIP. Change the ZIP Code to 06105 and match the record.

269 Church St., Hebron, CT 06231 is a complete address. The town of Hebron, however, is not in Hartford County so this address falls outside the study area. Do not match this record.

10 N. Main St., West Hartford, CT 06117 has an incorrect ZIP. Change the ZIP Code to 06107 and modify the record. You should see 2 candidates for matching in the table below. Click the candidate offering the 100% match and then click the Match button to match the record to the candidate with the higher score.

Some of the other records that matched have errors in them. You may wish to correct these matches.

15 Palomba Dr., Enfield, CT, 06105 has an incorrect ZIP. Change the ZIP to 06082 and match the record.

574 Middle Turnpike E., Manchester, CT 06042 has an incorrect ZIP. Change the ZIP Code to 06040 and modify the record. You should see 2 candidates for matching in the table below. Click the candidate offering the 100% match and then click the Match button to match the record to the candidate with the higher score.

100 Grant St., New Britain, CT 06052 has a misspelled street name. Change the Street or Intersection to 100 Grand St. and match the record.

100 Grand St., New Britain, CT 06050 has an incorrect ZIP. Change the ZIP Code to 06052 for each of the two records with this error and match them.

162 Moutain Rd., Suffield, CT 06078 has a misspelled street name. Change the Street or Intersection to 162 Mountain Rd. and match the record.

Once you have made these changes, you should see that you have now successfully matched 293 of the 310 addresses for a match rate of 94%.

The two records for address 469 East Main St., Meriden, CT 06450 have a complete address and these records geocoded. The town of Meriden, however, is not in Hartford County so these two addresses fall outside the study area and should not be included.

Because 3 of the addresses were in towns located outside of Hartford County, the number of physician offices in the study area is 307. You successfully geocoded 291 of the 307 records for a match rate of 95%. With additional research, it might be possible to raise the match rate higher.

You can close the “Interactive Rematch – Geocoding_Result” window at any time and return later to complete the review and rematch process.

You can also match by highlighting a record and using the Pick Address from Map button to click a location in the Data View and match the record to the location you clicked. Records matched by address-match geocoding will have a Match_type value of A in the table. Records matched by picking a point on the map will have a Match_type value of PP in the table.

You can also unmatch a record by highlighting it and clicking the Unmatch button.

Close the “Interactive Rematch – Geocoding_Result” window.

Then **Save** the map document.

Create a Shapefile of Matched Addresses

Use the **Selection ⇒ Select by Attributes** menu to open the “Select by Attributes” window. Select Geocoding_Result as the Layer:. Then double-click “Status” from the list of fields to begin creating a selection query. Click the “=” button and then click the Get Unique Values button and double-click ‘M’. Then click the And button, double-click “TOWN” from the list of fields, click the “<>” button and click the Get Unique Values button and double-click ‘Meriden’. The query string should appear as below:

"Status" = 'M' AND "TOWN" <> 'Meriden'

This will select all of the records that matched that are not in the town of Meriden. Click Apply and OK to perform the query. You should have 291 records selected. Close the window.

Right click on the Geocoding_Result layer in the Table of Contents and select **Data ⇒ Export Data** from the menu.

Export Selected features checking the radio button for Using the same coordinate system as “this layer’s data source”. Export the selected records to:

c:\exer4\data\agency\shapes\offices.shp

Add the data layer to the map document.

Use the **Save** button to save the map document and then use the **File ⇒ Exit** menu to close ArcMap.