

GIS AND PUBLIC HEALTH EXERCISE 4 - MAKING DATA MAPPABLE (ArcGIS 9.3.1)

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 ArcGIS 9.3.1 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** 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** to start ArcMap.

In the “ArcMap Start using ArcMap with” window, click the radio button labeled “A new empty map” and then click OK.

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 9.3.1, this is accomplished by creating an Address Locator.

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

Navigate to the **c:\lexer4\locator** subfolder. With the subfolder open, right click in the window for the **Contents** tab to the right of the Catalog tree window and select **New ⇒ Address Locator** from the menu. This will open the “Create New Address Locator” window.

To “Choose an Address Locator Style:”, scroll through the pull-down list to highlight “US Streets with Zone” then click OK. There are a wide range of styles you can select. US Streets with Zone is a simple style requiring an address field but allowing you to choose a ZIP code, town, or other unit as a zone within which to search for a match. This is useful because many communities may, for example, have an address match for “10 Main St.”

In the “New US Streets with Zone Address Locator” window, specify the rules for matching the table of addresses to the street database.

In the “Name:” window enter the name zipzone so that you will know this locator uses ZIP codes as zones.

To specify the street database which will be used as the reference data, click on the button and navigate to **c:\lexer4\data\agency\shapes** and select the **tl_2009_09003_roads_projected.shp** shapefile. Then, complete the description of this as the “Primary Table”.

Select LFROMADD as the “House from Left:” field from the pull-down list.

Select LTOADD as the “House to Left:” field from the pull-down list.

Select RFROMADD as the “House from Right:” field from the pull-down list.

Select RTOADD as the “House to Right:” field from the pull-down list.

Skip “Prefix Direction:” and “Prefix Type:”.

Select FULLNAME as the “Street Name:” from the pull-down list.

Skip “Street Type:” and “Suffix Direction:”.

You should see that the ZIPL field has been selected as the “Left Zone:” and the ZIPR field has been selected as the “Right Zone:”. If this is not the case, select these fields from the pull-down lists.

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, Addr, or Street 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 Zone to highlight it. You should see that the software will recognize fields named Zip, Zipcode, City, or Zone as the zone field in the table of addresses. If the field storing the field you wish to use as a zone 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 zone 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, standardized address, and percent along the street segment as fields describing the geocoded points.

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

Wait while the software creates an index. When the zipzone locator has been added to the Locator folder, use the **File** ⇒ **Exit** menu to close ArcCatalog.

ADDRESS MATCH GEOCODING

Batch Geocoding

Go to **Tools** ⇒ **Geocoding** ⇒ **Geocode Addresses** to begin geocoding.

The window "Choose an Address Locator to use" will open. If you do not see the zipzone Address Locator listed, click the Add button and navigate to:

c:\lexer4\locator

and add the zipzone Address Locator to the window. Highlight it in the window and click OK.

It is possible to create different address locators to work with different data or to set different parameters for the analysis.

Once you have chosen the Address Locator to use for the analysis, the "Geocode Addresses: zipzone" window will open.

Use the pull-down arrow to select the field from the input table of addresses containing the address. In the case of the physician.dbf table, this is the ADDRESS field. Then, select ZIP as the zone field.

For the Output, click the "Create static snapshot of table inside new feature class" radio button. Navigate to **c:\lexer4\data\agency\shapes** and save the snapshot as:

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

Then click OK to initiate the geocoding process.

Exploring Unmatched Records

When the processing is completed, the "Geocoding Addresses" window will open reporting the outcome of the analysis.

You should find that 268 of the 310 addresses (86%) matched with scores of at least 80 out of 100. There were 41 addresses (13%) left unmatched and there was 1 address with a tie score among candidate matches. For now, click the Close button to close the window.

Click on the symbol for the Geocoding Result: 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.

Right click the Geocoding.Result: Geocoding Result layer name in the Table of Contents and select **Open Attribute Table** from the menu to investigate unmatched records. 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 standardized address field (Stan_addr) and percent along (Pct_along) field have also been added as called for in the zipzone Address Locator.

Look at the records in the table to identify reasons why some records may not have matched. Because the addresses in this table were complete with street numbers, street names, street types, and ZIP codes for every record, most of the errors will be due to errors in the ZIP codes. For example, 211 New Britain Rd., Berlin, CT matches when the ZIP code is 06037 but not when the ZIP code is 06027. These errors can be corrected to match the addresses. Then close the table.

Review/Rematch Addresses

To review and rematch addresses, go to **Tools ⇒ Geocoding ⇒ Review/Rematch Addresses** and select the Geocoding Result:Geocoding_Result file to review.

In the "Interactive Rematch – Geocoding_Result" window, you should see a table of all records with the first record highlighted. 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 211 New Britain Rd., Berlin, CT 06027. 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 Zone are displayed, change the "Zone" to 06037 and then use the Enter key to update the zone.

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.

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

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.

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.

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.

15 Palomba Dr., Enfield, CT, 06105 has an incorrect ZIP. Change the ZIP to 06082 using the method you used to correct the zone for 211 New Britain Rd. and match the record.

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.

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

27 Sycamore St., Glastonbury, CT 06003 has an incorrect ZIP. Change the Zone to 06033 using the method you used to correct the zone for 211 New Britain Rd. and match the record.

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.

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

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

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.

151 Farmington Ave., Hartford, CT 06156 has an incorrect ZIP. Change the Zone 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.

574 Middle Turnpike E., Manchester, CT 06042 has an incorrect ZIP. Change the Zone 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.

469 East Main St., Meriden, CT 06450 is a complete address. The town of Meriden, however, is not in Hartford County so these two addresses fall outside the study area.

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 Zone to 06052 for each of the two records with this error and match them.

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.

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 Zone to 06479 then match the record.

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

10 N. Main St., West Hartford, CT 06117 has an incorrect ZIP. Change the Zone 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.

Once you have made these changes, you should see that you have now successfully matched 290 of the 310 addresses for a match rate of 94%.

Now, scroll through the addresses to find the record with a Status of T (for tie). This is 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. Then click Close. You should see that you have now successfully matched 291 records.

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.

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.

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: 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’ to complete the query:

“Status” = ‘M’

Click Apply and OK to perform the query. You should have 291 records selected.

Right click on the Geocoding Result: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.