Online Supplement to *GIS and Public Health, 2nd Edition*
Ellen K. Cromley and Sara L. McLafferty

GIS AND PUBLIC HEALTH EXERCISE 9 – ALLOCATION ANALYSIS (ArcGIS 9.3.1)

PREPARATION

Download the **exer9** 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 **exer9**. 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:\exer9** 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.

ASSEMBLING THE DATABASES

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 Allocation. Click OK. The name of the Data Frame in the Table of Contents window should now appear as Allocation.

In this exercise, you will be solving an assignment problem to assign women 40 to 74 years of age residing in New London County, Connecticut, to mammography facilities in New London County certified by the U.S. Food and Drug Administration as shown on the FDA’s web site. In order to perform the analysis, you will model the women’s origins based on the centroids of the 21 towns in New London County, the destinations based on the locations of the 9 mammography facilities in the county, and the distance to mammography facilities based on an origin-destination cost matrix created from a street network database for the county.

You will be using the GIS software to develop the origin-destination cost matrix for the assignment problem, known as the Transportation Problem. Data on travel distance in miles will be used in an Excel spreadsheet with the Solver Add-in to solve the problem.

This exercises uses the same data used in Exercise 8 to model potential accessibility to mammography services from women in different towns but it analyze the data in a different way to answer a different question: what is the assignment of women from towns to mammography facilities that minimize the total travel distance of women while ensuring that the demand from each town is served and the capacity of each facility is not exceeded?

Add Databases of Towns, Town Centroids Origins, and Mammography Facilities

To begin, add a database of towns in New London County, Connecticut. The database was developed from data downloaded from the U.S. Census Bureau TIGER® web site from the 2010 TIGER database.

The data were projected using NAD_1983_StatePlane_Connecticut_FIPS_0600_Feet coordinates. Map units are feet.

Find the button labeled **Add Data** and click the button. You should find the **c:\exer9** folder in your catalog. If not, please connect to the folder using the **Connect to Folder** button. Navigate to **c:\exer9\data\agency\shapes** and add the **towns.shp** shapefile.

Right-click the towns data layer and select **Open Attribute Table** from the menu. There are 21 towns in New London County. Each town has a numeric ID from 1 to 21. The NAME field provides the name of the town. The TOWNNO field is the town number used by the State of Connecticut to identify the 169 towns in the state.

Close the table.

Use the **Save** button or go to **File** ⇒ **Save** to save your map document. Navigate to **c:\exer9\mapdocs** and save the file as **exer9.mxd**.

Next, use the **Add Data** button to add a database of town centroids used to model the origins of women needing mammography services. Navigate to **c:\exer9\data\agency\shapes** and add the **origin.shp** shapefile.

Right-click the origin data layer and select **Open Attribute Table** from the menu. There are 21 town centroids in the database, one for each town in New London County. Each town centroid has a numeric ID from 1 to 21. The NAME field provides the name of the town. The FEM4074 field shows the 2010 population of women 40 to 74 years of age in a town based on the 2010 Redistricting Data Summary File (P.L. 94-171) downloaded from the Census Bureau's American FactFinder web site. These women need an annual screening mammogram.

Next, add a shapefile of mammography facilities in New London County certified by the U.S. Food and Drug Administration, geocoded from address data from the FDA's web site. Navigate to **c:\exer9\data\agency\shapes** and add the **facility.shp** shapefile.

Right-click the facility data layer and select **Open Attribute Table** from the menu. There are 9 facilities in the database. Each facility has a numeric ID from 1 to 9. The NAME field provides the name of the facility. The ADDRESS, TOWN, STATE, and ZIP fields provide the facility's address. The ESTCAPAC field gives the estimated capacity of the facility to provided screenings on an annual basis. ESTCAPAC is not information distributed by the FDA; this field was added by the authors for the purposes of the exercise.

Close the table and **Save** the map document.

Symbolize the Origin and Facility Data Layers

Right click the origin layer and select **Properties** from the pull-down menu. Then click the **Symbology** tab. Under “Show:” select **Quantities** ⇒ **Graduated symbols**. Under “Fields:” in the “Value:” section, select FEM4074 from the pull-down menu as the field containing the values to be mapped. Under “Classification” use the Natural Breaks (Jenks) classification and set the number of classes to 3.

Click the **Template** button to open the “Symbol Selector” window and choose the Circle 1 symbol. Then click OK. Set the Symbol Size from 6 to 18 and click Apply and OK.

Next, right click the facility layer and select **Properties** from the pull-down menu. Then click the **Symbology** tab. Under “Show:” select **Quantities** ⇒ **Graduated symbols**. Under “Fields:” in the “Value:” section, select ESTCAPAC from the pull-down menu as the field containing the values to be mapped. Under “Classification” use the Natural Breaks (Jenks) classification and set the number of classes to 3.

Click the **Template** button to open the “Symbol Selector” window and choose the Square 1 symbol. Then click OK. Set the Symbol Size from 6 to 18 and click Apply and OK.

Save the map document.

Look at the data view to compare the distribution of facilities of different capacity with the distribution of women who need service across the towns.

Add a Network Database

Finally, add a network database of the street network in New London County. This database was created in Network Analyst from data on all lines in New London County in the 2010 TIGER databases downloaded from the Census Bureau’s web site. The data were projected based on the Connecticut State Plane Coordinate system and the road segments were selected from all lines. Map units are feet. A field, LENGTHFT, was added to each record giving the length of the street segment in feet.

Use the **Add Data** button and navigate to **c:\exer9\data\agency\shapes** and add the **tl_2010_09011_streets_Projected_ND.nd** network dataset. When you are asked whether you want to add all feature classes that participate in tl_2010_09011_streets_Projected_ND.nd to the map, click No.

Save the map document.

CREATING THE ORIGIN-DESTINATION COST MATRIX

Enable the Network Analyst Extension and Choose Origin-Destination Cost Matrix

If you do not have the Network Analyst extension, read through this section, then complete the analysis by following the steps outlined in the **Explore the OD Cost Matrix** section below.

If you have the Network Analyst extension, use the **Tools** ⇒ **Extensions** menu to make sure that the “Network Analyst” extension is checked. Then use the **View** ⇒ **Toolbars** menu and check Network Analyst to make sure that the Network Analyst extension toolbar is visible.

Click the **Show/Hide Network Analyst Window** button in the toolbar to open the Network Analyst window. Make sure that **tl_2010_09011_streets_Projected_ND** is shown as the “Network Dataset:” in the pull-down.

From the Network Analyst pull-down menu, select **New OD Cost Matrix**. You should see that the Network Analyst window is updated to show “OD Cost Matrix” and to list Origins (0), Destinations (0), Lines (0), and Barriers (0).

In addition, the Table of Contents has been updated with OD Cost Matrix added as a layer in the Data Frame.

Load the Origins and the Destinations

Right click “Origins (0)” in the Network Analyst window and select **Load Locations** from the menu. This will open the “Load Locations” window.

Next to “Load from:”, select the **origin** database from the pull-down list. Then click OK.

The Origins entry in the Network Analyst window should be updated to show that there are 21 origins or “Origins (21)”. All origins should appear as located in the Data View, based on the symbol for Located Origins in the Table of Contents.

Next, right click “Destinations (0)” in the Network Analyst window and select **Load Locations** from the menu. Next to “Load from:”, select the **facility** database from the pull-down list. Then click OK.

The Destinations entry in the Network Analyst window should be updated to show that there are 9 destinations or “Destinations (9)”. All destinations should appear as located in the Data View, based on the symbol for Located Destinations in the Table of Contents.

Right click the Origins layer under OD Cost Matrix in the Table of Contents and select **Open Attribute Table** from the menu. Look at the attribute table. Then close the table. Do the same with the Destinations layer and then close the table.

Save the map document.

Create the OD Cost Matrix

To create the OD Cost Matrix, click the **Solve** button in the Network Analyst toolbar. You will see that the “Lines (0)” entry in the Network Analyst window is updated to “Lines (189)” and there are 189 line segments in the Data View such that every origin is connected to every destination.

Explore the OD Cost Matrix

Right click on the Lines layer under OD Cost Matrix and select **Open Attribute Table** from the menu. The Name Field gives the name of the town followed by the name of a facility. The OriginID field gives the ID of the town centroid of origin. The Destination ID field gives the ID of the facility destination. The DestinationRank field gives the rank from closest to farthest for each destination from a given origin. The Total_LENGTHFT field gives the travel distance (cost) measured from the origin to the destination along the street network. This is represented as a straight line in the Data View but the distance is the network distance.

Right click on the Lines layer under OD Cost Matrix and select **Data ⇒ Export Data** from the menu. In the “Export Data” window, export All features. Under “Use the same coordinate system as:”, check the radio button for “this layer’s source data”. Set the path and name for the “Output shapefile or feature class” to:

c:\exer9\data\agency\shapes\Lines.shp

and click OK.

If you do not have the Network Analyst extension, use the **Add Data** button and navigate to **c:\exer9\data\agency\lines** and add the **Lines.shp** shapefile. This database was exported from the origin-destination cost matrix created using Network Analyst.

Right click on the Lines.shp shapefile and select **Open Attribute Table** from the menu. The Name Field gives the name of the town followed by the name of a facility. The OriginID field gives the ID of the town centroid of origin. The Destination ID field gives the ID of the facility destination. **Note that this field name is truncated to Destinatio in the Lines.shp shapefile.** The DestinationRank field gives the rank from closest to farthest for each destination from a given origin. **Note that this field name is truncated to Destinat_1 in the Lines.shp shapefile.** The Total_LENGTHFT field gives the travel distance (cost) measured from the origin to the destination along the street network. This is represented as a straight line in the Data View but the distance is the network distance. **Note that this field name is truncated to Total_LENG in the Lines.shp shapefile.**

Save the map document.

Calculate Distance in Miles

Right click on the Lines.shp shapefile and select **Open Attribute Table** from the menu.

Use the **Options** button and select **Add Field** from the menu. Enter LENGTHMI as the field name in the “Name:” section and select **Double** as the field type from the pull-down next to “Type:”. Then, double click in the Precision line and set the precision to 5 and double click in the Scale line and set the scale to 1. Then click OK.

Right click the LENGTHMI field header and select **Field Calculator** from the menu. In the window under LENGTHMI= add the formula:

[Total-LENG]/5280

to convert the length in feet to the length in miles. Then click OK. You should see that the LENGTHMI field is populated with values describing the distance from town centroid origin to facility destination in miles.

Save the map document and use the **Minimize** button in the upper right hand corner of the window to minimize the ArcMap map document.

SOLVING THE TRANSPORTATION PROBLEM

Explore the Spreadsheet and Problem Set-up

Navigate to **c:\exer9\data\agency\excel** and open the Excel spreadsheet named **transport.xls**.

In the worksheet, explore the how the transportation problem is set up in Excel.

In cells B2:V2, the demand values from each town (the number of women 40 to 74 years of age) have been entered with a 3-letter abbreviation of the town name in the cells below.

In the cells A4:A12, the IDs of the mammography facilities have been listed.

Use the cursor to highlight the range of cells B4:V12. These cells contain the travel cost (distance in miles) from each demand site to each supply site. Look at the Name Box in the upper left corner of the interface. You will see that this array has been given the name “costs”.

In the cells W4:W12, the estimated capacities of the 9 facilities are given.

Maximize the ArcGIS exer9 map document window and right click the Lines.shp shapefile and select **Open Attribute Table** from the menu. Right click the Destinatio field header and select **Sort Ascending** from the menu to sort the field based on the destination or facility ID. Compare the length of each segment in miles to the values in the transport.xls spreadsheet to make sure they are the same. Then **Close** the table and **Minimize** the map document.

Cells in the lower section of the spreadsheet offer an initial feasible solution to the problem, an initial feasible assignment of women to facilities. The solution is feasible in the sense that it satisfies all of the constraints.

Use the cursor to highlight the range of cells B15:W23. These cells contain a set of assignments showing the number of women from each origin assigned to each facility. Look at the Name Box in the upper left corner of the interface. You will see that this array has been given the name “flows”.

In the cells B24:V24, a formula has been entered to sum each column to total the number of women sent from each town across all facilities. Use the cursor to highlight one of the cells in this range and look at the formula entered for the cell in the interface above.

In the cells W15:W23, a formula has been entered to sum each row to total the number of women received by each facility across all towns. Use the cursor to highlight one of the cells in this range and look at the formula entered for the cell in the interface above.

The values in cells B26:V26 are the numbers of women who must be served from each town. To satisfy the constraints of the problem, the number of women in the corresponding cell from row 24 must be greater than or equal to (>=) the number in row 26. This means that at least the number of women to be served from the town must be served in a solution.

The values in cells Y15:Y23 are the numbers of women who are served by each facility. To satisfy the constraints of the problem, the number of women in the corresponding cell from column W must be less than or equal to (<=) the number in column Y. This means that no more than the number of women who can be served at a facility will be assigned to it in a solution.

Note that the total capacity of the facilities is 61,200 and the total number of women to be served is 61,158. This means that total capacity exceeds total demand. This is a requirement of the problem. If this were not the case, we could not meet the constraint of serving all women from every town.

The flow values in this table representing an initial feasible solution were determined using the Northwest Corner Rule. Starting with the first town and the first facility, we assigned as many women from the town to the facility as we could, without assigning more women than exist in the town or more women than the facility can serve. We move from left to right and top to bottom through the matrix, assigning women to facilities. Note that all of the excess capacity (the difference between 61,200 and 61,158 (42) is associated with the last facility in the list (the facility with ID 9), which could serve 5,200 women but is only serving 5,158.

Use the cursor to highlight the cell B29. This is the cell storing the result of the analysis. It should contain the total transportation distance based on the costs and flows. Note that the formula for this cell uses the SUMPRODUCT function to multiple corresponding elements in the costs and flows arrays. For this initial feasible solution, the total transport cost of the assignment is 673,576.5 miles. This may not be the minimum that could be achieved. To find the minimum, you will use the Solver function.

Add-in Solver

After you have the spreadsheet set up properly, you must add in the Solver extension.

If you are using an older version of Microsoft Excel (for example, Microsoft Office 2003), click on the **Tools** ⇒ **Add-Ins** menu to open the “Add-Ins” window and check **Solver Add-In**, then click OK to close the window.

If you are using a newer version of Microsoft Excel (for example, Microsoft Office 2007), click the **Microsoft Office** button in the upper left of the Excel interface, then click the **Excel Options** button.

Click **Add-Ins** in the list on the left of the window that opens and then, in the “Manage” pull-down at the bottom of the window, select Excel Add-ins from the pull-down. Then click Go.

In the Add-Ins available box, check the Solver Add-in and click OK. You may need to browse and install the Solver Add-in.

If these instructions do not work, use Help in your version of Excel to find out how to add the Solver.

Solve the Problem

Once you have added the Solver function, click on the **Tools** ⇒ **Solver** to open the “Solver Parameters” window if you are using an older version of Excel.

In newer version of Excel, the Solver function is available on the **Data** tab in the **Analysi**s group. Click on **Solver** to open the “Solver Parameters” window.

Next to “Set Target Cell:” enter $B$29 because this is the cell where the answer will be stored.

Click the radio button next to “Min” to set the value of the target cell to the minimum total transport distance that can be achieved.

Enter flows in the space under “By Changing Cells:” because the array named flows contains the flow values that will be changed or updated when a different, optimal solution is found.

Next to the window under “Subject to Constraints:” click the **Add** button.

Set the Cell Reference to:

$B$24:$V$24

and the relationship to >= using the pull-down list

and the Constraint: to

$B$26:$V$26

and then click **Add**.

Next, enter the second set of constraints.

Set the Cell Reference to:

$W$15:$W$23

and the relationship to <= using the pull-down list

and the Constraint: to

$Y$26:$Y$26

and then click **Add**.

Click cancel and make sure that the constraints appear as follows in the window:

$B$24:$V$24 >= $B$26:$V$26
$W$15:$W$23 <= $Y$15:$Y$23

Next, click the **Options** button and set the Max Time: to 100 seconds and the Iterations to 500. Check the box in front of Assume Linear Model and the box in front of Assume Non-Negative.

Then click OK.

Click the **Solve** button to solve the problem.

You will see that the cells in the flows section of the spreadsheet are updated and that the total distance cost of the new assignment is substantially reduced to 301,614.2 miles. Click on Answer, Sensitivity, and Limits to highlight them under the “Reports” section of the “Solver Results” window. Then click OK.

You will see that additional tabs with the requested reports have been added to the spreadsheet.

Print the sheet with the updated flows associated with the minimum total travel distance solution.

Then **Close** Excel.

If you choose, you can map the optimal flows by selecting the lines associated with flows that entered into the solution, exporting the selected lines as a shapefile, adding the shapefile to the Data View, adding a field to store the flow value, and then updating the field to record the number of women associated with each line. The lines can be symbolized based on graduated symbols (lines of different width).

The Excel Solver is limited in terms of the size of the problem it can solve. You cannot solve a problem with more than 200 flow cells and there may be limits on solving problems of that size based on the number of constraints. If you need to solve larger problems, you can formulate the problem using PROC LP in SAS-OR, CPLEX, or other software.

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