3. Data Manipulation in GeoDa

ACE 492 SA - Spatial Analysis
Fall 2003

1 Objectives

This lab deals with basic data manipulations in *GeoDa*, such as reading in data sets, computing the centroids and exporting as a new shape file, creating Thiessen polygons, and some elementary table calculations. Detailed instructions can be found in the *GeoDa User's Guide*, in the chapters on "Manipulating Spatial Data" (pp. 19–34) and "Editing and Manipulating Tables" (pp. 54–64).

The lab consists of two parts. The first is a demonstration of the relevant *GeoDa* features. In the second part, you work through the examples below and try the practice exercises. Remember to refer to the *User's Guide* for detailed instructions, they are *not* repeated below. Nothing must be handed in, these exercises are not graded.

2 Computing Centroids

Load the SIDS sample data set with 100 North Carolina counties, and make sure that FIPSNO is specified as the Key Variable.

2.1 Adding centroids to an existing data set

You will use the SIDS polygon shape file to compute the centroids of the counties (their center of gravity). There are two ways to accomplish this. First, you will simply add the centroids to the data table and then save the table under a new name. Note that this does not create a point shape file, but another polygon shape file for North Carolina counties that now includes the centroids as elements in the data table. Proceed as outlined:

- Right click on the map and select Add Centroids to Table, check both X and Y and keep the default variable names. Click on OK to add the centroids to the data table. Check the table to make sure they are there.
- In the Table menu (on the main menu), select Save to Shape File As and save the shape file under a new name. You can also do this by right-clicking on the table itself to invoke a menu.
- Clear all windows (click on the Close all Windows toolbar button) and load the shape file you just created. Click on the Table icon to check the contents of the table.

2.2 Converting a dBase file to a point shape file

The centroids are contained in a *.dbf file (in dBase format) for a polygon shape file. The *.dbf file only contains data on attributes, not on the shapes, and you can use it as any other data base file. Specifically, you can load the *.dbf file and use the values for the `X` and `Y` coordinates just computed and turn it into a *point* shape file. Proceed as follows:

- Select `Tools > Shape > Points from DBF` and enter the name for the just created file at the prompt for the `Input file`, and a different file name for the point shape file at the prompt for the `Output file`.
- Specify `XCOO` in the drop down list for `X-coord` and `YCOO` in the drop down list for the `Y-coord`. Click on `Create`.
- It seems like nothing happened, but the new file has been created (the message will be there in one of the next *GeoDa* versions). To add it to the current project, click on the `Open New Map Window` toolbar button and select the file from the dialog, or close all windows and open a new project with the point shape file.

In general, any dBase file that contains the values for `X` and `Y` coordinates can be converted to a *point shape file* using these commands. Ascii files containing the `X` and `Y` information can be converted in a similar way (see the *User's Guide* for details).

2.3 Creating a point shape file with centroids

The most direct way to create a point shape file from the centroids of a polygon shape file is to use the `Tools` menu. This can be done without opening a project. Proceed as follows:

- Select `Tools > Shape > Polygons to Points`, and enter the name of the SIDS file as the `Input file`, and a new name for the point shape file as the `Output file`.
- Note how a thumbnail sketch of the polygons appears on the left hand side panel.
- Click on `Create` and notice the progress bar and a new thumbnail on the right hand side panel. The new file will be written to disk. Click `Done` to close the dialog.
- Load the new shape file and check the table. Note how the coordinates have been added.

2.4 Practice

Use the STL_HOM file from the sample data sets on the SAL server [1] to create a new point shape file with the county centroids. Try both approaches and compare the resulting table.

[1] http:\\sal.agecon.uiuc.edu\stuff\data.html

3 Thiessen Polygons

Use the BALTIM point shape file from the *GeoDa* sample data to create a polygon shape file with the Thiessen polygons. This can be done directly from the `Tools` menu, without opening a project. Proceed as follows:

- Select `Tools > Shape > Points to Polygons` and enter the Baltimore point file as the `Input file` and a new file name for the polygon shape file as the `Output file`.
- Note how a thumbnail sketch of the Baltimore points appears on the left hand side panel.
- Click on `Create` and notice a thumbnail of the new polygons. `Done` will close the dialog.
- Close all current windows and load the Thiessen polygon file, check the contents of the table.

3.1 Practice

Create a Thiessen polygon file from the point shapes you constructed for the St. Louis county data set. Compare the layout of the new polygons to the original county boundaries.

4 Table Calculations

GeoDa contains limited functionality to carry out variable transformations and compute new variables in the table. Load the SIDS data set and open the table by clicking on the `Open Table` toolbar button. Note how this data set (as opposed to the SIDS2 one) does not contain any rates for the Sids deaths. You can compute these rates and add them to the table, as follows:

- Select `Table > Field Calculation` from the main menu or right click on the table to do the same.
- Click on the `Rate Operations` tab and leave it to the default methods (`Raw Rate`).
- Overwrite the variable name for the `Result` and enter the name for the rate, say `SIDR74`.
- Select `SID74` as the `Event Variable` and `BIR74` as the `Base Variable`.
- Click `OK` to compute the new variable. Note how it has been added as a new column to the table.

- The rate is not in the usual units. Express the rate as "per thousand" (or any other unit of your choice) by means of the `Field Calculation > Binary Operations` function. Select `SIDR74` as the both the `Result` and `Variable 1`, choose `Multiply` from the drop down list and type in 1000 as `Variable 2`. Note how the value changes in the table.
- Save the table as a new shape file, using the same procedure as in 2.1.

Experiment with some of the other transformations and operations. You can always return to the original table by selecting `Refresh Data` from the menu. This will wipe out all your changes though, so it's all or nothing. Only after the table has been saved does the change become permanent (in the new shape file, not in the current one).

4.1 Practice

Use the St Louis data set to compute a new homicide rate by combining the events from the 79–84 and 84–88 periods (you need to calculate the new events as sums and compute a raw rate). Compare the new rate to the rates in each of the periods.