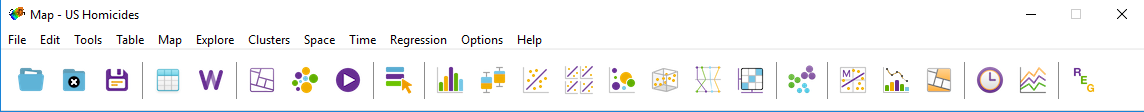
**Session 2 Exercise: Spatial weights, spatial clustering analysis**

1. **Spatial Weights**

In a computational environment, data management and the theoretical background should be relevant all the time, because they ensure the quality and robustness of the analysis (and save you from trouble afterwards, actually). Beyond those requirements, however, the first practical issue is the generation of spatial weights. Spatial weights formalize neighborhood relationships in the observations. The entity to-be-generated is a text file (but it does not have familiar extensions, such as \*.txt or \*.csv). This file is a requirement to work with spatial statistics.

In our class we are using a free program called GeoDa. The program looks like the following image, with additional floating windows opening depending on the operations:



GeoDa is able to generate contiguity and distance-based weights:

• Contiguity spatial weights files have the extension .gal

• Distance-based spatial weights files have the extension .gwt

Keep in mind that:

• These are connectivity files, later converted to weights, but we will name them weights,

• Contiguity weights use common polygon edges,

• Distance-based weights use Euclidean or other measures.

After spatial weights are computed, you have everything in place to conduct statistical tests for

spatial autocorrelation and compute relevant metrics, and then move on to estimating spatial

regression specifications.

0 – Locate software and data, change decimals notation if needed

Software: You will find GeoDa at the lab’s computer.

Data and Working folder: You’ll find the data for each exercise in the corresponding section in

Moodle. Please use your workstation’s hard drive or USB drive to create a working folder for the

exercise and download and unzip the data in that working folder.

Decimals mark: Go to Control Panel > Clock, Language, and Region > Change

the date, time, or number format. Or type region and language in the start menu’s

search box and press enter. In the window that opens, click the button Additional Settings

and in the new window make sure that the decimal symbol is the “point” (.) and not “comma” (,).

1 – Inspect the data, Generate contiguity weights

Contiguity-based weights work with polygons only, because they are computed based on

common edges, or edges and vertices. Two common types are:

• “Rook” contiguity: polygons with common edges are considered neighbors

• “Queen” contiguity: polygons with common edges and vertices are neighbors

Regardless of rook or queen contiguity, you can further distinguish:

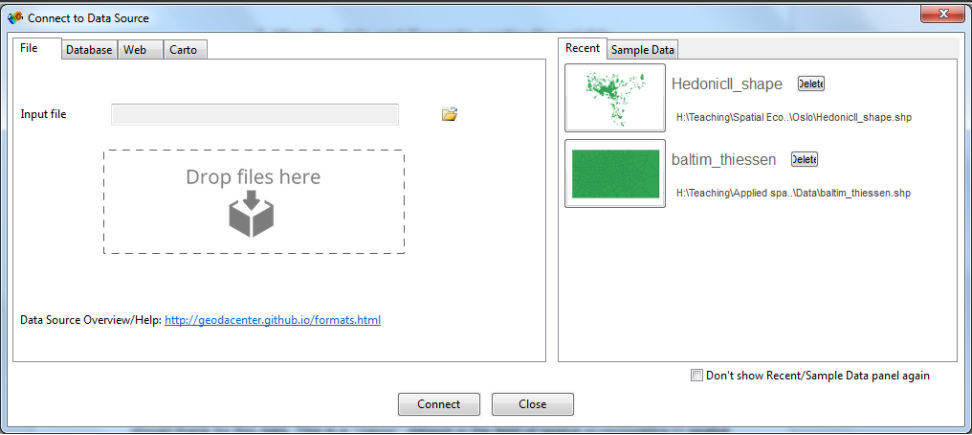
• First-order contiguity: the immediate “ring” of neighbors

• Higher-order contiguities: additional rings of neighbors.

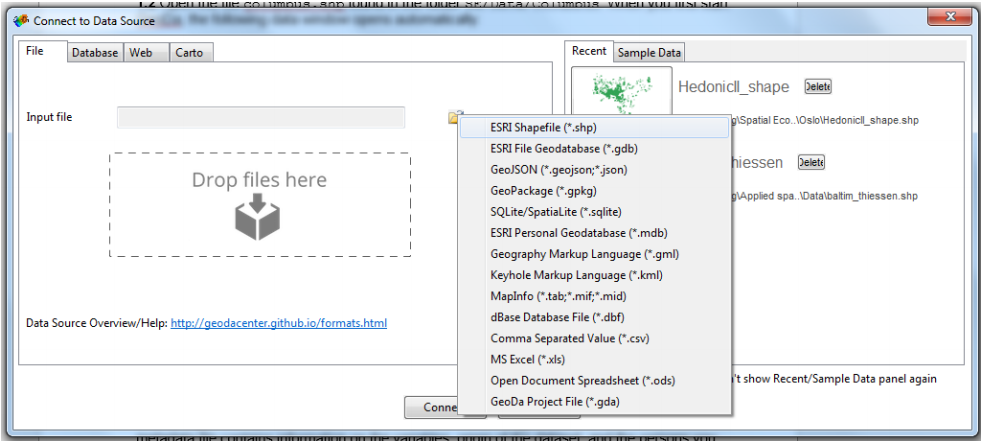
1.1 Start GeoDa.

1.2 Open the file columbus.shp. When you first start GeoDa, the following data window opens

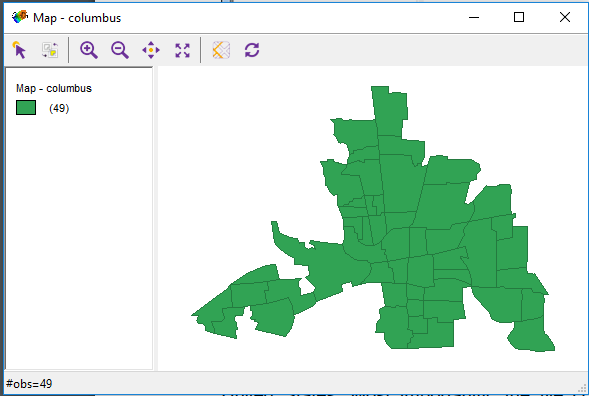
automatically:



Press the folder icon that is next to the Input file field, then click on ESRI Shapefile (\*.shp) and navigate to the file columbus.shp. Alternatively, just drag and drop the file in the area named Drop files here.



A new window will open, showing the polygons of the various districts of the city:



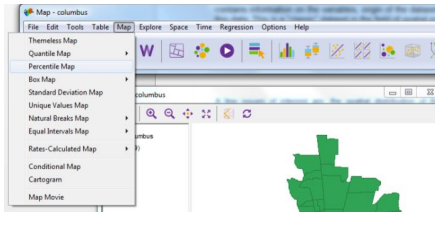
1.3 The shapefile contains the districts of Columbus, which is the capital city of Ohio State in the United States. Most importantly, the file contains a number of socio-economic variables per district. Open the file columbus.html. This metadata file contains information on the variables, origin of the dataset, and the persons you should thank for this data. This is a “classic” dataset in the field of spatial econometrics.

1.4 We will make a few thematic maps of Columbus for the crime variable CRIME, income

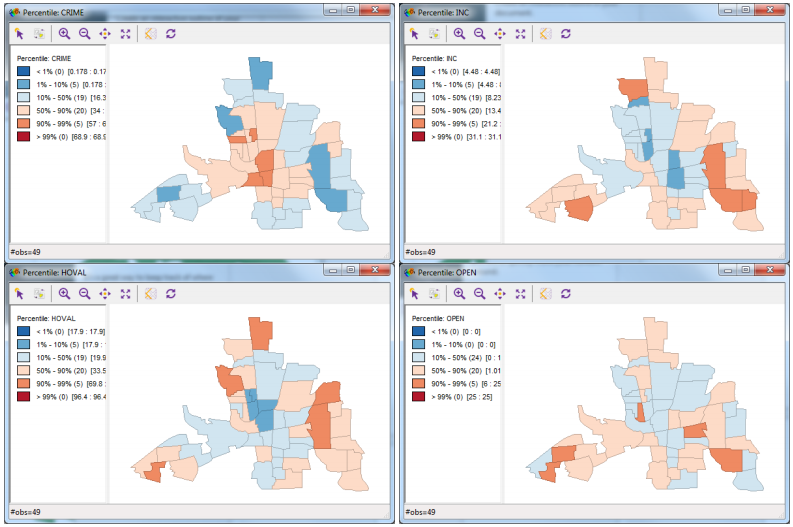
variable INC, and housing value variable HOVAL. This can be done through the Map tab of

GeoDa’s menu bar, selecting, among others, Quantile Map, Percentile Map, or Natural

Breaks Map:



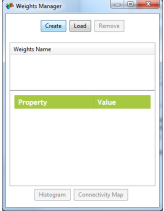
Percentile maps for crime (top left), income (top right), property value (bottom left), and percent open space (bottom right):

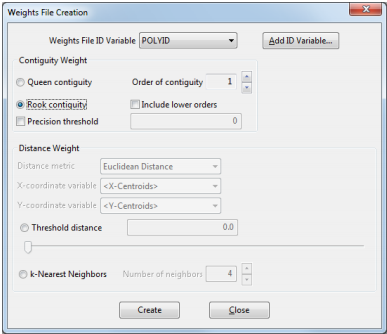


1.5 Now, go to Tools > Weights Manager

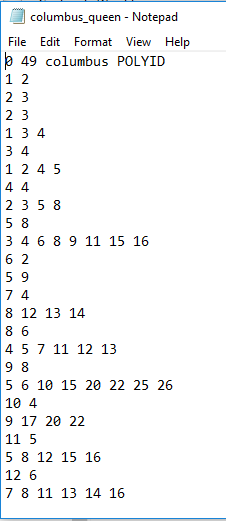


The “Weights Manager” window opens. Press Create.





The “Weights File Creation” window contains several options for weights creation. Most importantly, the shapefile must contain a field with unique numeric (integer) ID’s for each polygon, other than the FID or other internal fields of ArcGIS. In the case of the Columbus data POLYID is such a field so you do not have to create a new one. When you work with files without such a field, you can create one on the fly with GeoDa by clicking the Add ID Variable button, which will walk you through the creation of such a field. 1.6 Create two weights files, one by Rook contiguity (i.e. von Neumann neighborhood) and one by Queen contiguity (i.e. Moore neighborhood): • From the drop-down list Weights File ID Variable select the POLYID variable as your unique ID. • Select the Rook Contiguity radio button. • Set the Order of Contiguity = 1, as shown in the picture above. • Click the Create button. In the new window that opens, save the file as columbus\_rook. You do not need to enter the extension .gal manually – GeoDa takes care of this. • Do the same for Queen contiguity and save the file as columbus\_queen.

As mentioned earlier, the weights files can be viewed with any text viewer such as Notepad. For example, the columbus\_queen.gal file looks like the image on the left. Notice that the first row is a header, that is, it displays information on the file. The rest of the rows display groups of polygons that are considered as neighbors. The polygons are marked by the unique identifier (POLYID) that you set earlier while creating the weights. Technically speaking, this is not a weights matrix, but a sort of connectivity index. Nevertheless, this file is used by the software to calculate one.

2 – Generate k-nearest-neighbors weights

Distance-based weights, such as the commonly used k nearest neighbors (knn, with k = 1, 2 … n) is another way to model the neighborhood structure of your data. Contrary to rook and queen contiguities (which look at common edges/vertices), distance-based weights search either for a specified number of nearest entities, or for any entity within a specified distance. Thus: • A knn weights procedure identifies the k closest entities of each item and assigns them to its neighbor list – usually by nearest centroid; • A pure distance weight finds anything within a given distance and assigns it as a neighbor. Distance can be Euclidean distance or more complex distances such as cost distance, socioeconomic distance, Manhattan distance, and so on. While very complex neighbor rules can be created, it is preferable to not use complex expressions because you might end up imposing too much pre-conceived structure on your data. Of course, there might be theoretical reasons to do that, in which case more complex modeling will be understandable. 2.1 Use the Weights Manager (as in steps 1.5-6 above to) create a 4 nearest neighbors weights file. Use the options as shown in the image below, and save the file as columbus\_4nn. Again, there is no need to manually type the extension .gwt (denoting this type of weights, while .gal denoted contiguity weights in the previous cases):

3 – Visualize weights with GeoDa

GeoDa has useful capabilities for exploring and visualizing weights. Visualizing and comparing the various weights can help you understanding whether alternative neighborhood conceptualizations alter significantly the imposed connectivity structure of your observation (remember: with spatial statistics we make explicit assumptions about space). For instance, it helps understand why spatial regressions may yield different coefficients when using different weights. 3.1 In the Weights Manager window (Tools > Weights Manager), you can click the Load button to navigate through the different weights files you might have created. 3.2 Now, make sure you have a map window open displaying the Columbus shapefile. If not, you can open a new one by clicking Map > Themeless Map. Inside the map, select one district of the city by clicking on it. The selected polygon will remain vivid, while the rest of the (unselected) polygons will be dimmed:

