

## Geographical Information Systems

### Social Science Research Lab

#### GeoDa: Exploratory Spatial Data Analysis

---

##### A. Background

From [geodacenter.asu.edu](http://geodacenter.asu.edu):

“[GeoDa](#) is a free software program that serves as an introduction to spatial data analysis. OpenGeoDa is the cross-platform, open source version of Legacy GeoDa. While Legacy GeoDa only runs on Windows XP, OpenGeoDa runs on different versions of Windows (including XP, Vista and 7), Mac OS, and Linux. It is written in C++ and no longer relies on ESRI's MapObjects library (it uses wxwidgets instead). We are working towards eventually releasing OpenGeoDa as an open source program.

GeoDa is the flagship program of the GeoDa Center, following a long line of software tools developed by Dr. Luc Anselin. It is designed to implement techniques for exploratory spatial data analysis (ESDA) on lattice data (points and polygons). The free program provides a user friendly and graphical interface to methods of descriptive spatial data analysis, such as spatial autocorrelation statistics, as well as basic spatial regression functionality. The latest version contains several new features such as a cartogram, a refined map movie, parallel coordinate plot, 3D visualization, conditional plots (and maps) and spatial regression.

The development of GeoDa and related materials has been primarily supported by the [U.S. National Science Foundation](#)/ the [Center for Spatially Integrated Social Science \(CSISS\)](#) (Grant BCS-9978058).

Reference: Anselin, L., I. Syabri and Y Kho. (2005). [GeoDa : An Introduction to Spatial Data Analysis](#). Geographical Analysis 38(1), 5-22.”

OpenGeoDa can be downloaded at: <http://geodacenter.asu.edu/software>

##### B. Data set

In this exercise we will use American Community Survey (ACS) 3 year data, ending in 2016. There is also a dataset for the same geography (New York City) from the 2000 decennial census, if you are interested. (Note: GeoDa comes with several data files as samples (e.g., Crime in Columbus [tracts], SIDS in North Carolina [counties])).

The data is in “GeoDa2018” in the GIS Data folder.

The shapefile NYC\_Data.shp is the map of New York City with ACS 2016 data. These are socioeconomic attributes for census tracts in the five boroughs. It includes the following variables:

NYC\_Data.shp

PCT_WHT	Percent of Population that is white
STATEFP	State FIPS* code
COUNTYFP	County FIPS* code
TRACTCE	Extended tract id
AFFGEOID	GeoID for joining
GEOID	
ALAND	area of tract
GEOID1	Copy of GeoID
GEOD2	Copy of GeoID
CT	Census Tract #
TOT-POP	Estimate of total population
TOT_WHT_POP	Estimate of total White population
TOT_BLK_POP	Estimate of total Black population
TOT_ASN_POP	Estimate of total Asian population
HSHLDS_TOT	Estimate of total households
FAM-TOT	Estimate of total families
Highschool	Estimate count of population with a high school diploma
Postsec	Estimate count of population with some post-secondary education
Bachelo_o	Estimate count of population with a Bachelor's degree or higher
MED_House	Median House Value
MED_income	Median Household income (derived by compiling three other median income variables)

\*Federal Information Processing Standard

## C. GeoDa introduction

(Note: The following handout is an introduction of selected features of GeoDa, and not intended as a comprehensive review.)

**GeoDa** employs **ESRI** shapefiles as its primary data format, making a convenient program to use in conjunction with ArcGIS.

In this handout, all commands are printed in **Courier - Bold**.

Start **GeoDa** by Click the **Start > All Programs > GeoDa Software > GeoDa**

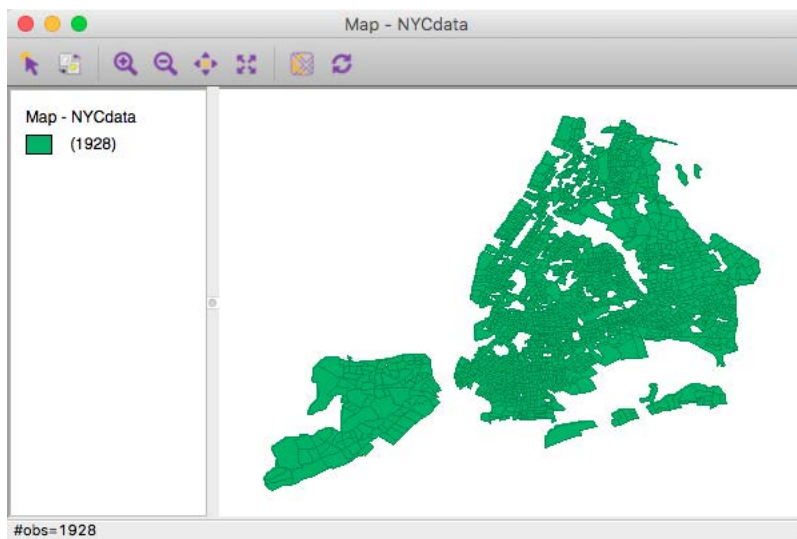
You should see something like this:



1) Go to **File > New**, under **file** click the folder icon, choose **ESRI shapefile**, and navigate to our data location.

2) **Browse** to **NYC\_Data.shp**, click **Connect**.

3) Click **OK**. Your screen should now look something like ...



The holes in the spatial data are places (census tracts) with no data for several of the variables we are using. While the entire country is “traced” (or covered by census tracts) not all of them have consistent data. The removed tracts include parks, open space, industrial areas, or are otherwise differently or less populated than other places.

4) The GeoDa menu bar contains nine menu items

1. File (Project Toolbar)
2. Edit
3. Tools (Weights Toolbar)
4. Table
5. Map
6. Explore
7. Space
8. Time
9. Regression
10. Options

5) Now let's make some adjustments to the view:

a) We will improve the background color for more clarity.

**Right click** on the background and choose **color**, and choose **background color** and select a light gray.

b) Then, change the map color to a different color (right-click on the rectangle on the table of content, choose **color for category**).

c) zoom in and zoom out.

Zooming in and out is a little crude in GeoDa, you can use the **Zooming** mode to zoom in and **Fit to Window** will zoom to the map's full extent.

## D. Data exploration

(Note: Most of the GeoDa exploration functions can be applied to either polygon or point shapefiles, you can use either.)

### 1) Univariate exploration with the variable of housing value (HVALUE)

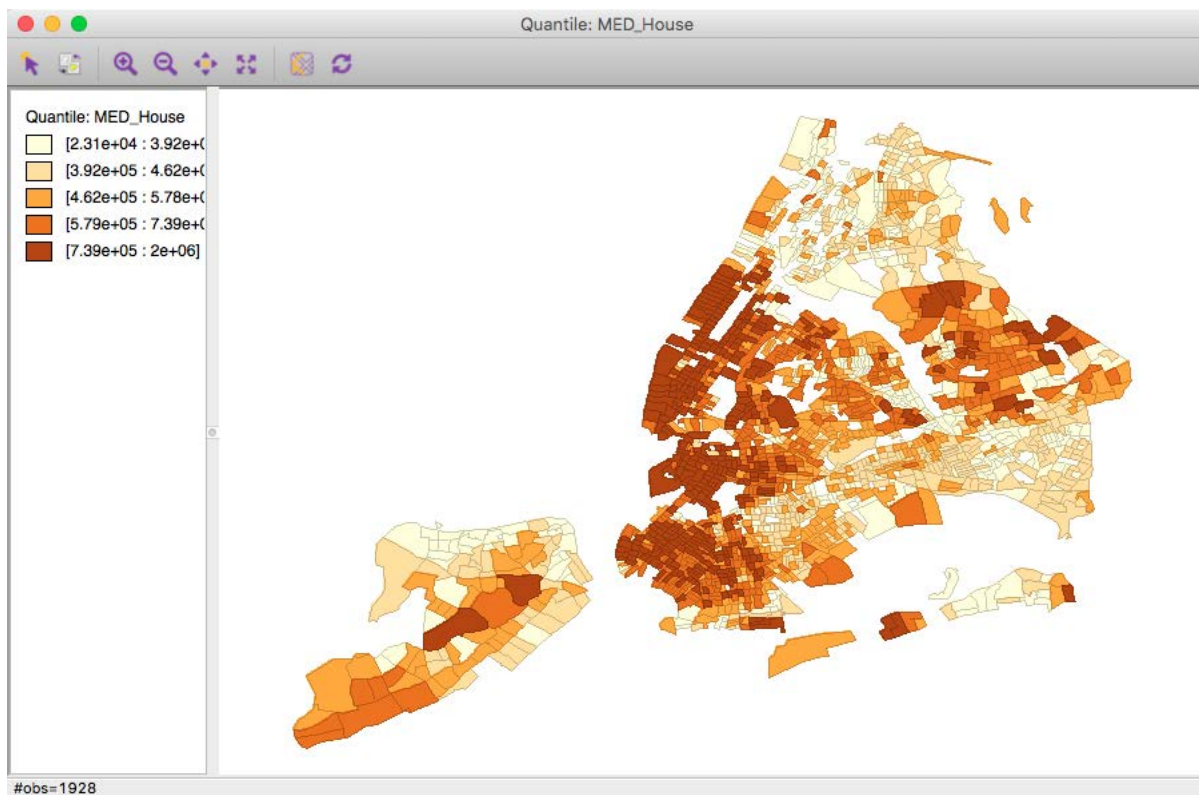
I) Quantile distribution

The **Quantile** map function in **GeoDa** allows you to specify up to 9 categories, the default is 4. Let's choose 5 categories.

Go to **Map > Quantile**,

(you can also access and change map types by right-clicking and rolling over the **Change Current Map Type** option)

c) In the dialog box **select variables** select the variable of interest (**HVALUE**) by scrolling down the variable list. Click **OK**. You will now be asked how many classes, type 5.



Drag the vertical bar so that the legend will show properly.

## II) Percentile distribution

**GeoDa** has six preset categories for percentile maps.

As we make additional maps through the main menu bar, a new map window will be created for each one.

## III) Box Map

A Box Map is designed to show quartile distributions with outliers defined by upper and lower hinges. The “hinge” values allow us to identify outliers based on the values for the interquartile ranges (IQR). A hinge value of 1.5 will identify high and/or low outliers as those observations that are greater or less than the 75<sup>th</sup> or 25<sup>th</sup> percentile (respectively) by more than 1.5 times than the IQR.

Let's create an additional window:

Go to **Map > Box Map > Select Hinge = 1.5**

#### IV) Standard Deviation Map

GeoDa Standard Deviation map maps mean, single and double standard deviations, and beyond.

Let's create an additional window:

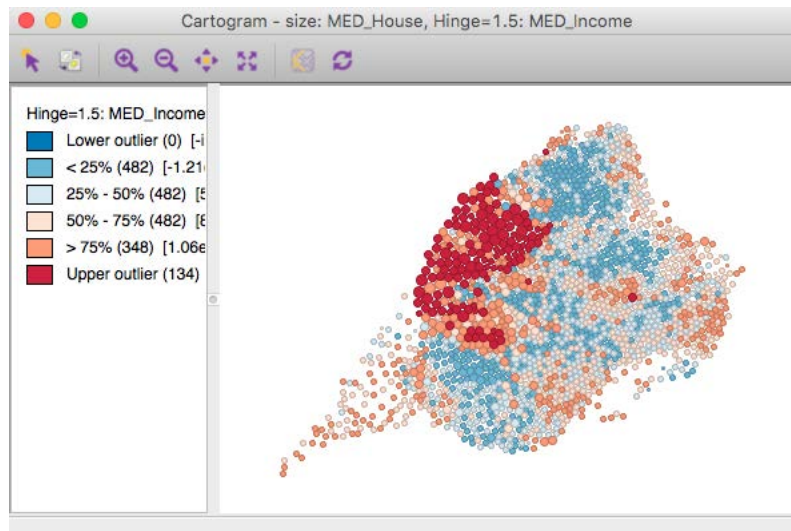
Go to **Map > Standard Deviation Map**

#### e) Cartogram

Cartogram is another method to examine variable distribution. It is a technique that allows you to map locations according to the values of a selected variable using size as the symbol variable that changes with the corresponding non-spatial variable. (Note: Cartogram can only be made on polygons.)

III) Go to **Map > Cartogram**

Cartogram has been buggy this week in the lab; if it freezes your computer, restart GeoDa and skip. When making a cartogram we can symbolize two variables, one with symbol size (MED\_House) and one with color (MED\_Income).



#### f) Show map movie

Map movie is a tool that allows you to see the variable distribution in an animated fashion. Specifically, it highlights the locations of selected variable in ascending order.

**Switch back to GeoDa**, from the main menu select **Table > select animation >** then, set **Speed Control** selector towards the right end, Click **Play**. The animation displays median housing values from lowest to highest across the mapped area.

### g) Histogram

A histogram is a bar graph that shows frequency data. The horizontal axis should be the independent variable and the vertical axis should be the dependent variable.

Click **Explore > Histogram**. Select **HValue** (or another variable), click OK.

**Increase** the Histogram window size. You can change the number of intervals, **Right Click** on the Histogram, select **Interval** and **type “50”** in the interval dialogue box, then, click **OK**. You can also unclick **Display Statistics** to clean up the window.

### h) Box Plot

(Box Plot is designed to show several critical distributional measures in a single graph, we will be able to see median, upper and lower quartiles, and outliers defined by upper and lower hinges.)

Go to **Explore > Box Plot**.

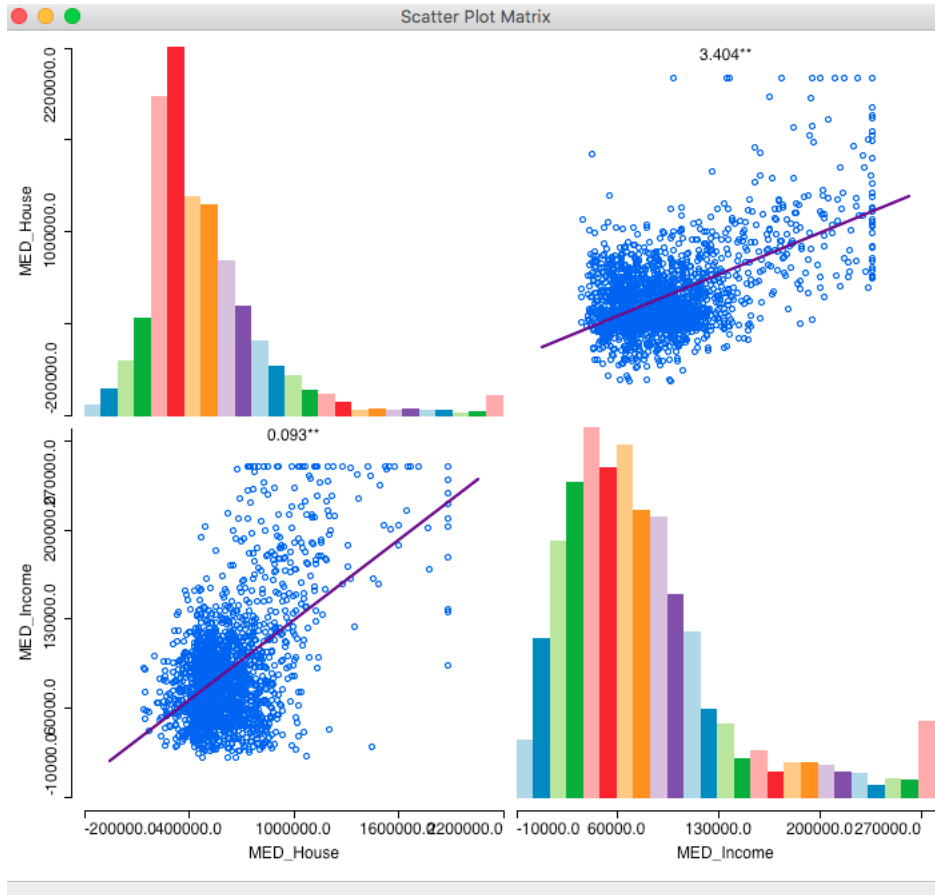
## 2) Multivariate exploration

### a) Scatter plot

Scatter plot explores bivariate relationship. Let's first select one more variable.

I) The Scatterplot tool doesn't immediately allow us to set a second variable (a bug in the program), first we have to **close all** from the **file** menu then open the nyc2000 shapefile again.

II) Go to **Explore > Scatter Plot matrix** and select **MED\_House as X axis** and **MED\_income as Y**.



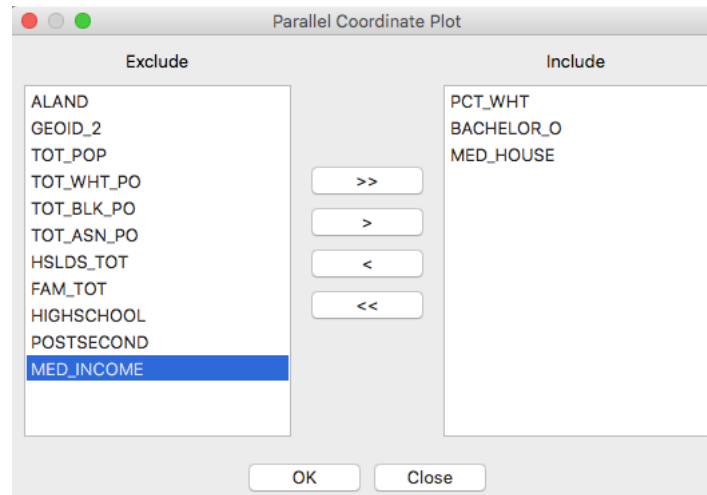
## b) Parallel Coordinate Plot (PCP) and brushing

(PCP) allows you to observe the relationship between multiple variables

I) Go to **Explore > Parallel Coordinate Plot >**

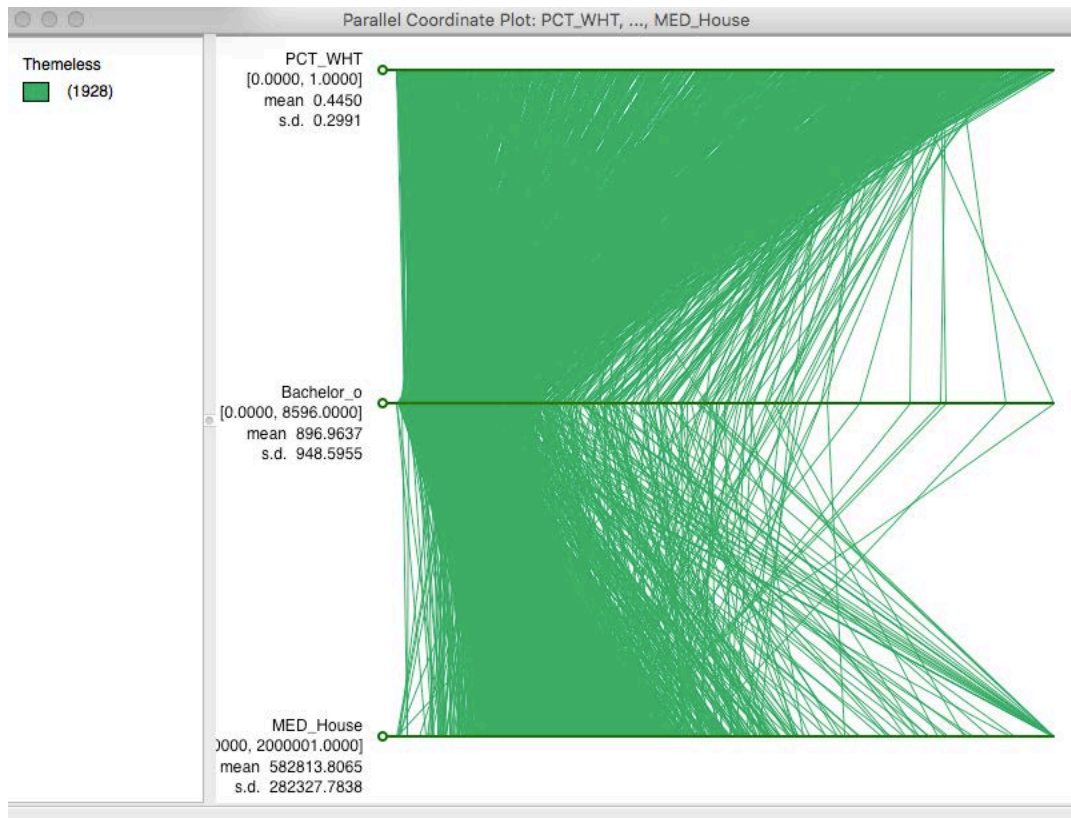
Select three variables: **HVALUE**, **PCTNHB** (percent of non-Hispanic black), and **T0\_MINC** (median household income), and click **OK**.





II) Brushing- You can move the mouse around the PCP and highlight portions of the plot (“brushing”) to observe the multivariate relationship in a dynamic fashion.

**Maximize** the PCP window. Then, with the **Ctrl** button depressed, click, drag, and release to create a small box in the scatter plot window. It will flash for a couple of seconds, and then become continuously active. You can move it around the PCP and dynamically highlight portions of the plot (“brushing”), all the while viewing the active selections in the map (“linking”). Brushing makes most sense when you move the selection “box” along one of the three parallel variable axes (top, middle, or bottom). Doing this allows you to visualize a set of observations (in this case each observation is a census tract) with similar values for one of the three variables and how they vary across the range of the remaining two variables. Simply click the mouse to end the brushing.



## F. Close project

Go to **File > Close Project**.

To exit **GeoDa**, go to **File > Exit**.

## Summary

Now you have learned the main mapping functions and options of GeoDa. You should know your way around GeoDa. You should be able to carry out basic exploratory data analysis and understand the ideas behind those techniques.