# GeoDa
# Exploratory Spatial Data Analysis

## 1. Background

GeoDa is a trademark of Luc Anselin.

GeoDa is a collection of software tools designed for exploratory spatial data analysis (ESDA) based on dynamically linked windows, and this software replaces the DynESDA Extension for Arcview 3.x. GeoDa is freestanding and does not require a specific GIS system. GeoDa has evolved from efforts to couple SpaceStat and DynESDA with ESRI products (e.g., Arcview 3.x) via extensions. GeoDa adheres to ESRI's shapefiles as the standard for storing the information, using MapObjects LT2 technology for spatial data access, mapping and querying.

Luc Anselin suggests http://www.csiss.org/learning_resources/content/syllabi#gis for extended course notes and examples dealing with an introduction to spatial data analysis, and requests that users ***please*** **report "anything that seems like a bug"** to anselin@uiuc.edu (or post to the Openspace mailing list: mailto:openspace@sal.agecon.uiuc.edu).

The software can be downloaded from http://sal.agecon.uiuc.edu/geoda_download.php , and http://sal.agecon.uiuc.edu/geoda_support.php offers GeoDa Tutorials and other support materials, including release notes and User's Guide.

GeoDa is highly interactive and the subcomponents you will explore in this lab are well integrated with one another. In the latter components of this lab you will have the opportunity to dynamically select cases (census tracts) from any one of several representations (histogram, map, scatterplot, box-plot, or parallel coordinate plot) and see those cases highlighted in the other representations that are active.

## 2. Data set

In this exercise we will use US 2000 Census data from New York City. (Note: GeoDa comes with several data files as samples (e.g., Crime in Columbus [tracts], SIDS in North Carolina [counties])).

The zipped data file is on the S4 Training page and can be downloaded from http://www.s4.brown.edu/S4/about.htm click on S4 tutorials and scroll down to GeoDa 1.

The shapefile nyc2000.shp is the map of New York City with Census 2000 data from summary file 3. These are socioeconomic attributes for 2219 Census tracts in five boroughs. It includes the following variables:

nyc2000.shp

| Vraiable name | Label |
|---|---|
| POLYID | Polygon ID |
| STATE | State FIPS |
| COUNTY | County FIPS |
| TRACT | Census Tract ID |
| sctrct00 | FIPSID |
| hvalue | Median housing value |
| t0_pop | Total population |
| t0_nhw_f | Total number of non-Hispanic white persons |
| t0_nhb_f | Total number of non-Hispanic black persons |
| t0_hsp_f | Total number of Hispanic persons |
| t0_asn_f | Total number of Asian persons |
| t0_min | Total number of minority persons |
| pctnhw | Percent non-Hispanic white persons |
| pctnhb | Percent non-Hispanic black persons |
| pcthsp | Percent Hispanic persons |
| pctasn | Percent Asian persons |
| pctmin | Percent minority persons |
| chn00 | 2000 Chinese |
| fil00 | 2000 Filipino |
| jap00 | 2000 Japanese |
| ain00 | 2000 Asian Indian |
| kor00 | 2000 Korean |
| m0_mex | 2000 Mexican(Mumford estimates) |
| m0_prn | 2000 Puerto Rican(Mumford estimates) |
| m0_cbn | 2000 Cuban(Mumford estimates) |
| m0_dom | 2000 Dominican(Mumford estimates) |
| t0_afa | 2000 African American |
| t0_car | 2000 Afro-Carribbean |
| t0p_own | Percent homeowners |
| t0p_vac | Percent vacant housing |
| t0p_coll | Percent college educated |
| t0p_prf | Percent of people employed in professional/managerial occupations |
| t0p_uemp | Percent of people unemployed |
| t0p_nat | Percent persons born in the United States |
| t0p_for | Percent foreign born persons |
| t0p_rec | Percent recent immigrants |
| t0p_old | Percent older immigrants |
| t0p_only | Percent persons who speak only English at home |
| t0p_oth | Percent persons who speak language other than English at home |
| t0_minc | Median household income |
| t0_pcinc | Per capita income |
| t0p_poor | Percent total population below poverty |
| m0p_poor | Percent minority population below poverty |

## 3. 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.`**

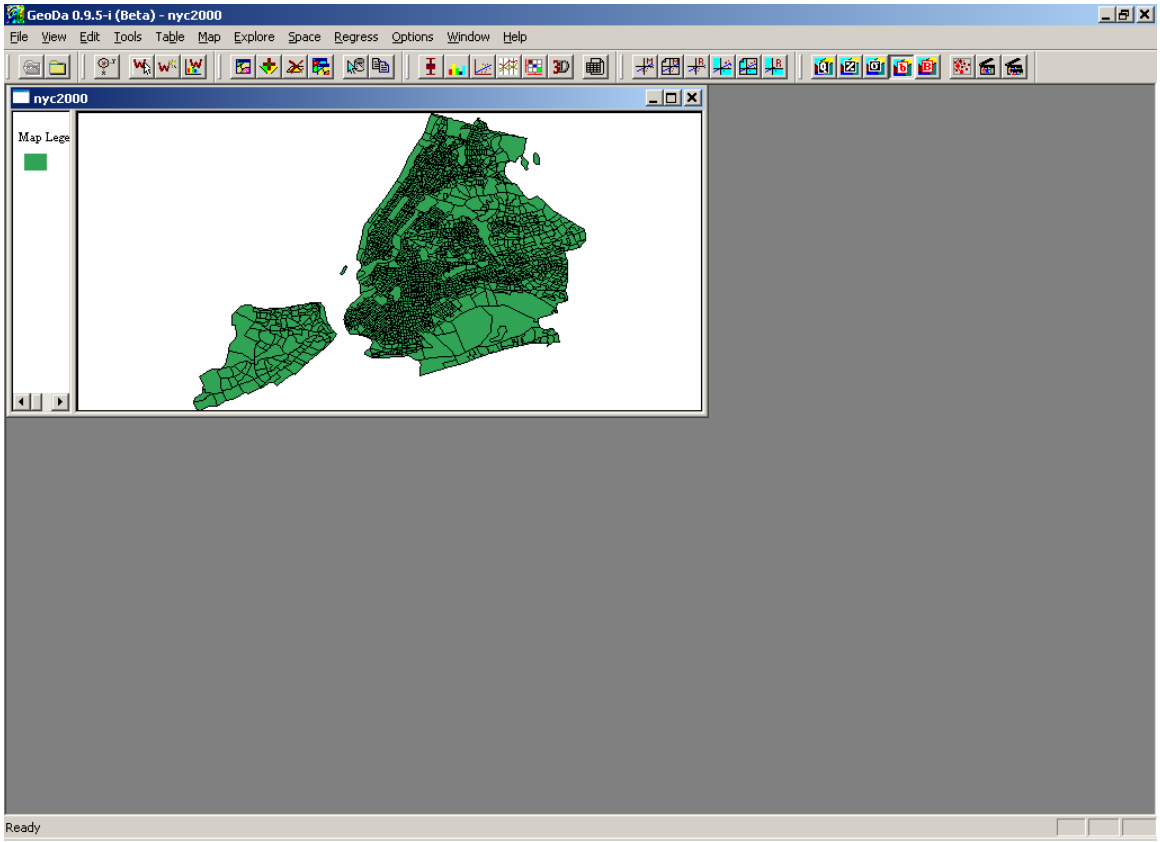
Start **GeoDa** by Click the **`Start > Programs > Instructional Tools > Geoda`**

1) Go to **File > Open Project** (This opens the GeoDa Project Setting dialog box).

2) **Browse** through the folder to find and select the shape file, **nyc2000.shp**.

3) Select **`POLYID`** as the "Key Variable" (by scrolling and clicking on variable name). The Key Variable must have a unique value for each observation (i.e., in this case Census tract). The unique value is used to implement the link between maps and statistical graphs.

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



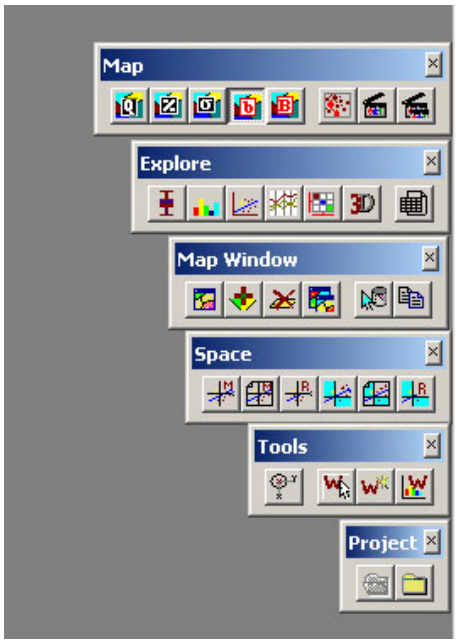

5) It is a good idea to maximize the GeoDa window. **Click on the full screen**

**button for GeoDa.** Note, all windows in GeoDa can be resized and positioned anywhere within the main program window.

6) The GeoDa menu bar contains twelve menu items

1. File (Project Toolbar)
2. View
3. Edit (Edit Toolbar)
4. Tools (Weights Toolbar)
5. Table
6. Map
7. Explore
8. Space
9. Regress
10. Options
11. Window
12. Help

with the important menu items matched by a "button" on the toolbar. Toolbar components [Project, Tools, Edit/Map Window, Explore, Space, and Map] can be moved and docked anywhere within the main program window. For example,

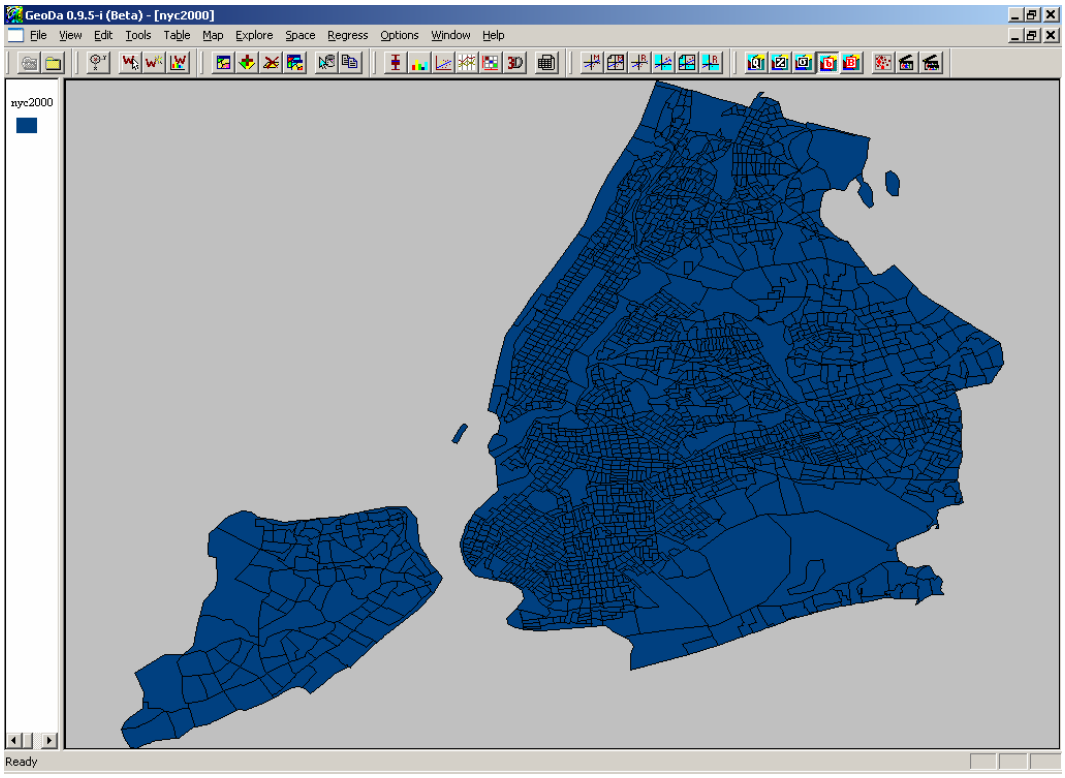



7) Now let's make some adjustments to the view:

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

Go to **`Options`** **> Color > Background**, and choose a light gray.
b) Then, change the map color into a different color.

Go to **`Options`** > **Color > Map**, and choose a deep blue. The map looks like this…




c) zoom in and zoom out.

Go to **`Options`** > **Zoom > Zoom In** to anywhere of your interest,

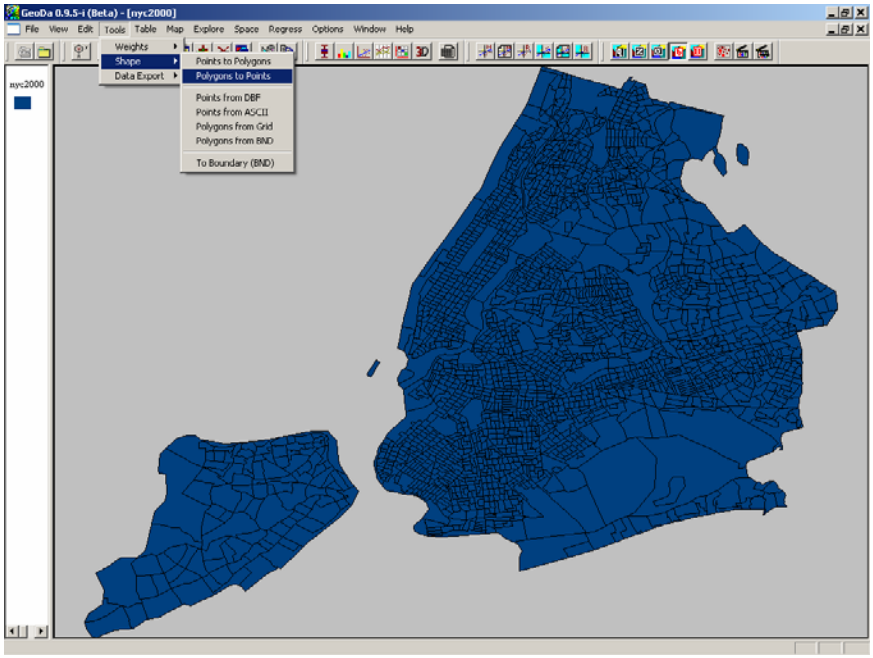
And try **`Zoom Out`**, and zoom out to **Full Extent**.

### 4. Shape conversion and map layers

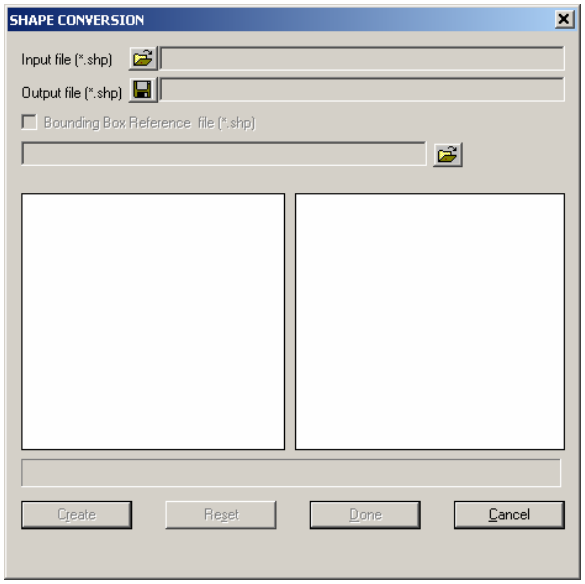
The ***Edit*** Menu ***(Map Window Toolbar)*** allows the user to (a) manipulate maps and layers (e.g., **`new map, duplicate map, add layer, remove layer`**), (b) **`select variables`**, and (c) make use of the **`Windows clipboard`**.

1) **GeoDa** can import shapefiles of several different formats, and enable shapefile conversion between polygons to points with corresponding data. Here, we will convert our original polygon file into a point file.

Go to **`Tools > Shape > polygons to points`**



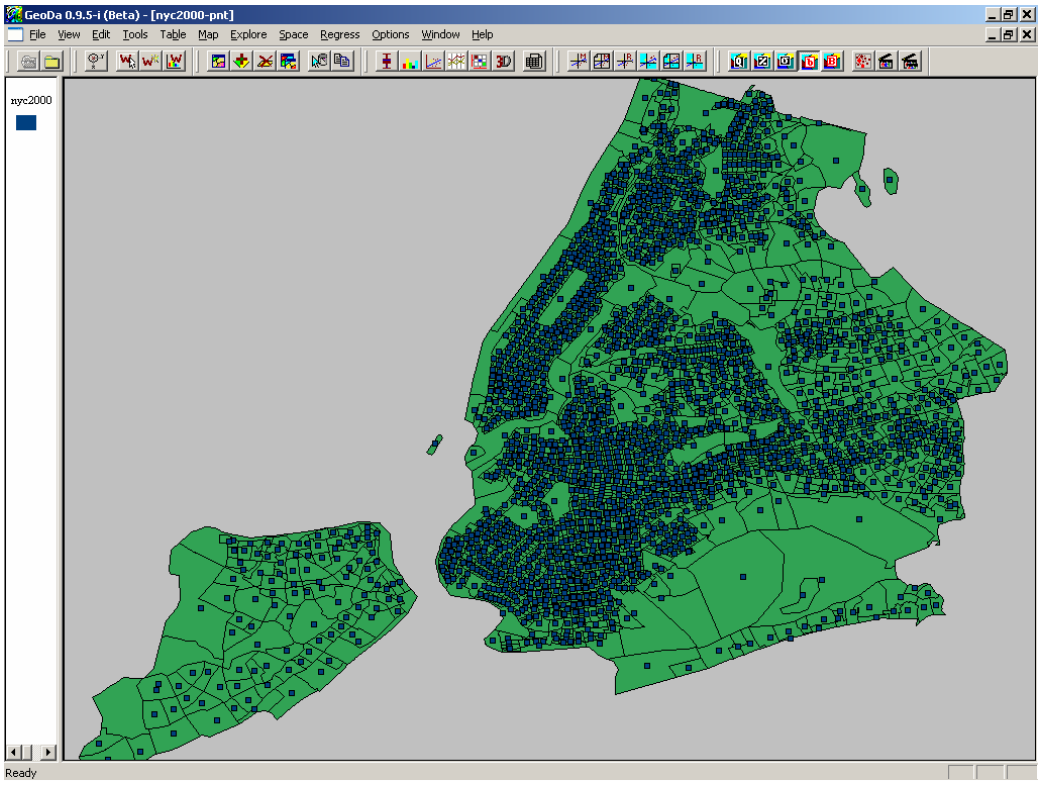

Select input file (polygon): **nyc2000.shp**
Name output file (point): **nyc2000-point.shp**




Click **Create** … Click **Done** when finished.

2) Add the point layer to the current map

Go to **Edit > Add Layer** … select **nyc2000-point.shp**… It should look like this…



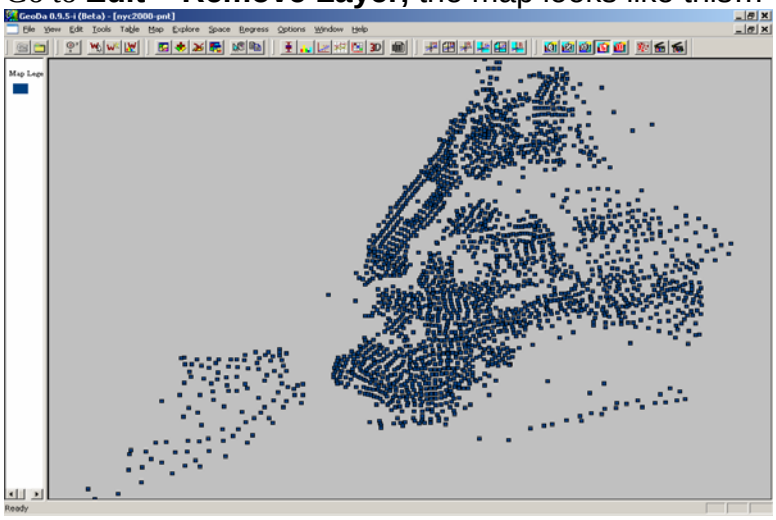

## 5. Data exploration

(Note: Most of the GeoDa exploration functions can be applied to either polygon or point shapefiles, here we will use the point file as an example.)

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

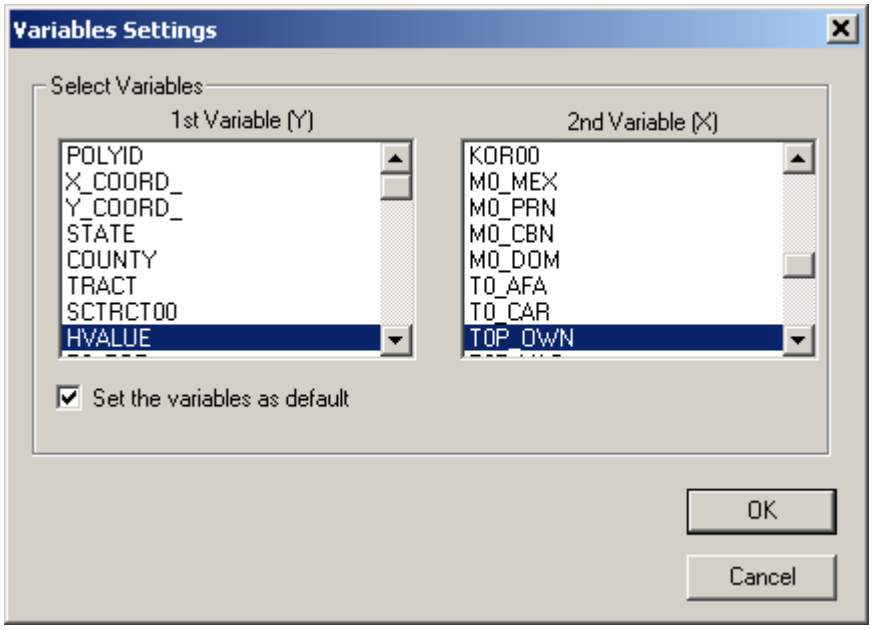
a) For the interest of clarity, let's remove the polygon layer

Go to **Edit > Remove Layer,** the map looks like this…



b) To select the variable,

Go to **Edit > Select Variable**. This opens a variable settings dialog box.




c) In the dialog box **select** the variable of interest (**HVALUE**) by scrolling down the 1st variable Y listing, (Note: for univariate operations, the choice of second variable is ignored). Click **OK**. (Note: If the table nyc2000-point appears, minimize the window.)

Notice that the box "select variables as default" is checked. If the variable is set as a default (i.e., checked) then all mapping or statistical graph options assume this is the focal variable. If this check box is not marked, the variable selection dialog box will open for each mapping or statistical operation.

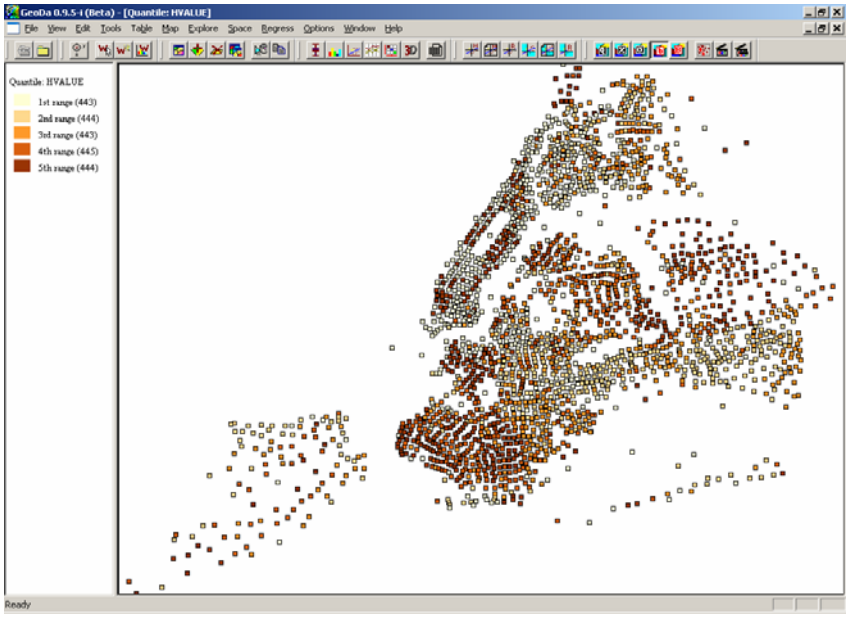
d) Four kinds of maps of housing value distribution

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,** and **type "5"** in the box of **# of classes/groups,** Click **OK.**

Now, we have a map of five-class quantile distribution of housing values.



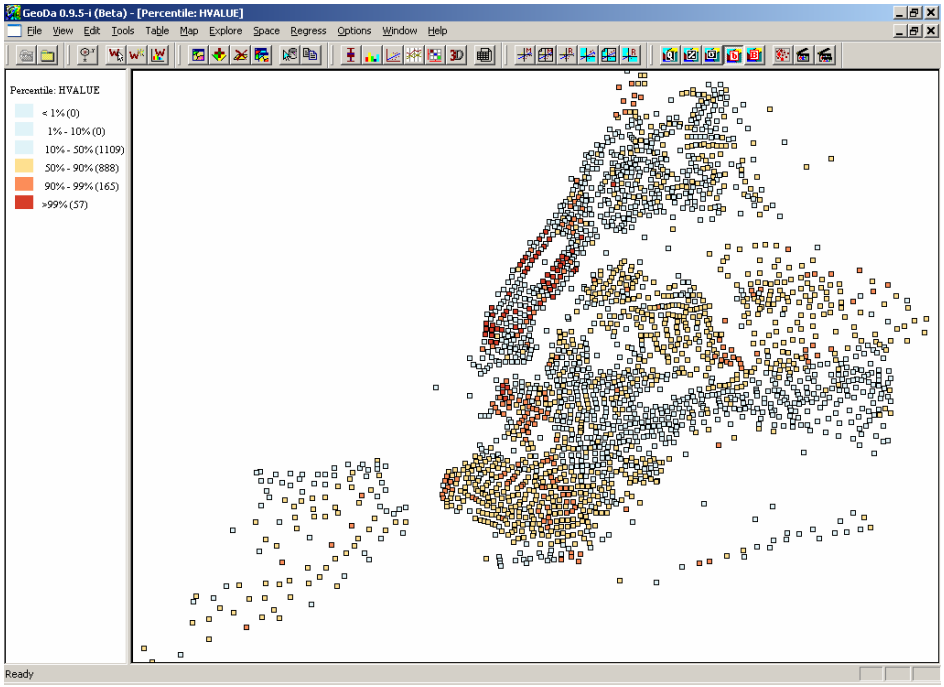

Drag the vertical bar between the TOC and map to the right so that the legend will show properly.

II) Percentile distribution

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

Since we would want to keep the **Quantile** map, let's create an additional window.

Go to **Edit > Duplicate Map,** then, go to **Map > Percentile**

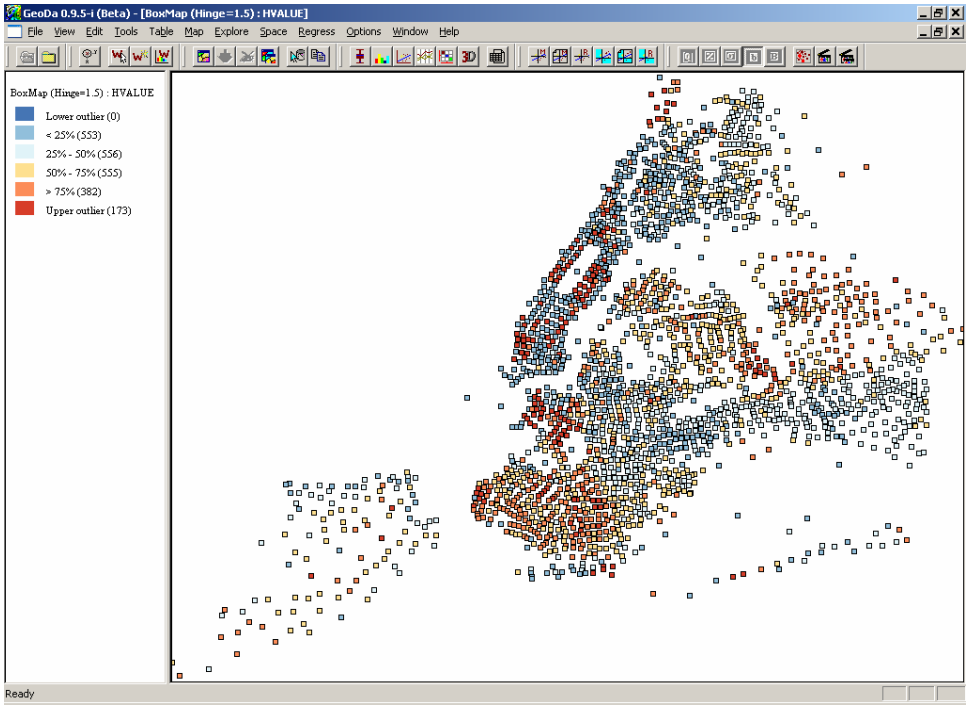


III) Box Map

(Box Map is designed to show quartile distributions with outliers defined by upper and lower hinges.)

Let's create an additional window:

Go to **Edit > Duplicate Map,** then, go to **Map > Box Map >** Select **Hinge = 1.5** (the hinge value, either 1.5 or 3 represents 1.5 or 3.0 times the interquartile range (the difference between the 25% and 75% value, which are also mapped).

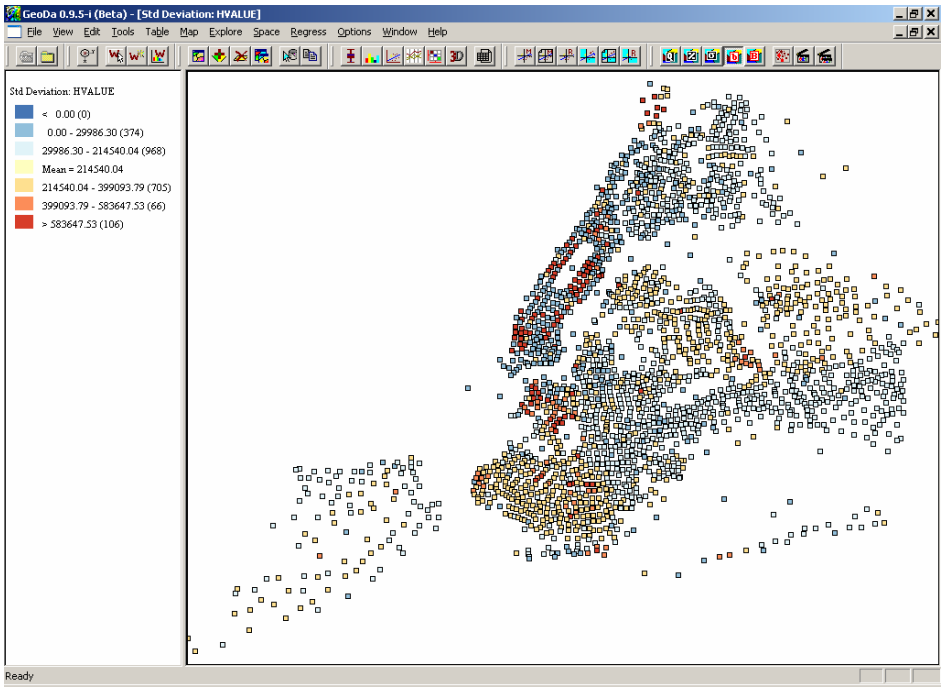



IV) Standard Deviation map

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

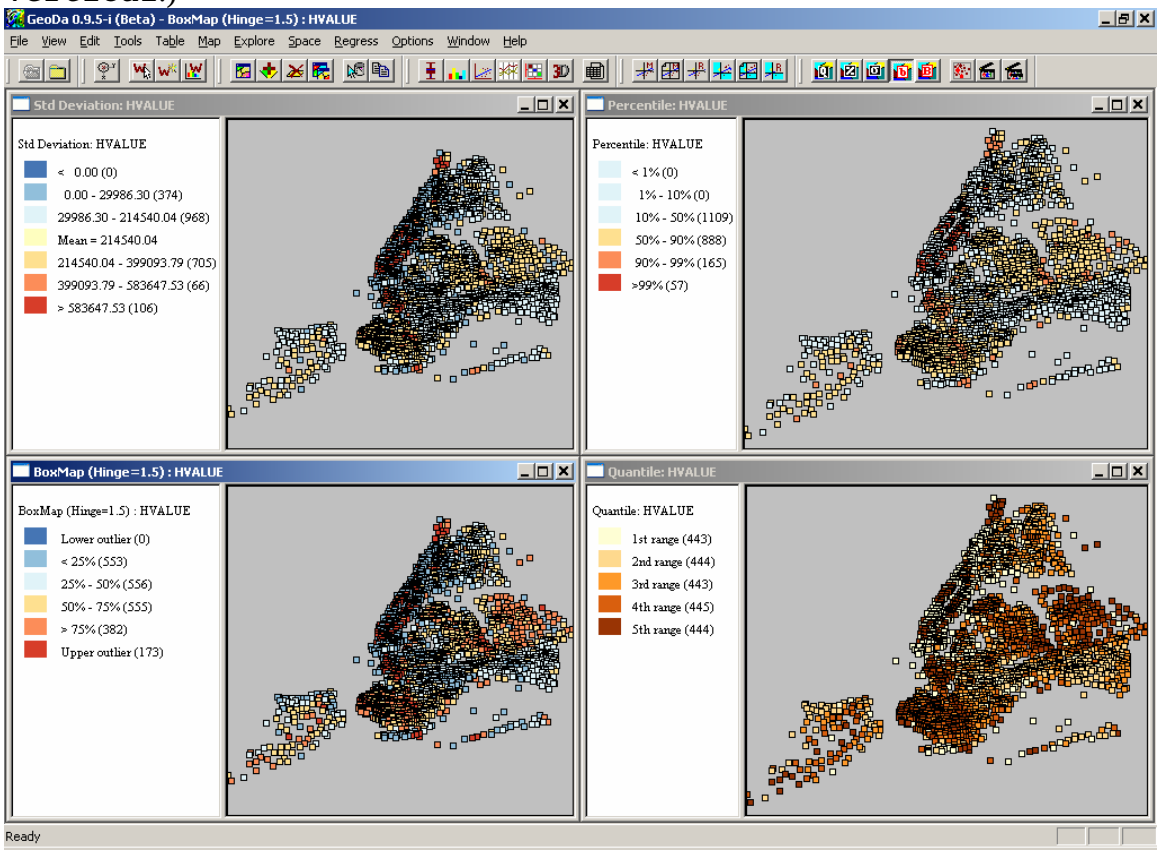
Let's create an additional window:

Go to **`Edit > Duplicate Map`,** then, go to **`Map > Std Dev`.**




V) Arrange and save the created maps

Go to **`Window > Tile Horizontal`** (You can also choose to **`Cascade`** or **`Tile Vertical`**.)



Now, press **`alt + PrintScreen`**

Open a blank **Word** document, and click **`Paste.`**

Save the document.

e) **Cartogram**

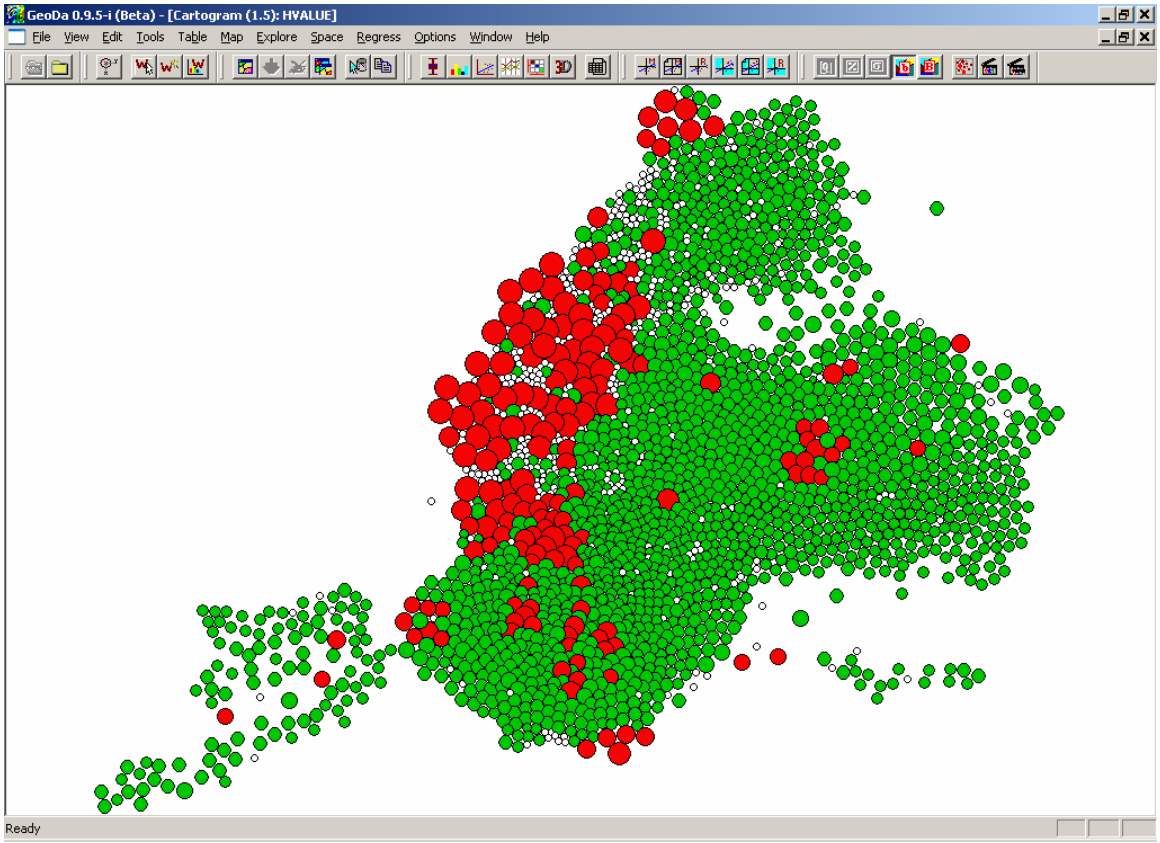
Cartogram is another method to examine variable distribution. It is a technique that allows you to map locations according to the values of a selected variable. (Note: Cartogram can only be made on polygons.)

I) **`Switch back`** to **GeoDa** and **`minimize`** all windows within it.

Let's first open our polygon shapefile.

II) Go to **Edit > New Map** `and select the shape file` **nyc2000.shp,** and **Maximize** the new window**.**

III) Go to **Map > Cartogram**




Note: color code: red- upper outlier; green- default; transparent- zero.

IV) Save the map image- Let's try another way:

Go to **Edit > Copy to Clipboard,** and **switch back** to the **Word** document **Geodalab1.doc** , and click **Paste**, then click **Save.**

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**, go to **Map > Map Movie >** select **Cumulative** (the **Single** option will highlight one location at a time)**,** then, set **Speed Control** to 10, Click **Play.**

Click **Reset** and **Play** again.

The linking to the previous cartogram map can be observed by
Go to **Window > Tile Vertical, then,** Click **Reset** and **Play.** Notice the simultaneous actions in both windows.

**In each of the following sections you will be able to select individual cases (points, lines, or map symbols) and see that same selection highlight in other GeoDa elements (all other displayed graphs or maps). You can also select more than one case by drawing (dragging) a box on the window and selecting neighboring cases (either geographic neighbors or plot neighbors).**

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.

**Minimize** all windows, then, go to **Explore > Histogram**.

**Maximize** the Histogram window, you can change the number of intervals by **Right Clicking** on the Histogram, **selecting intervals** and **typing "50"** in the interval dialogue box**,** then, click **OK.**

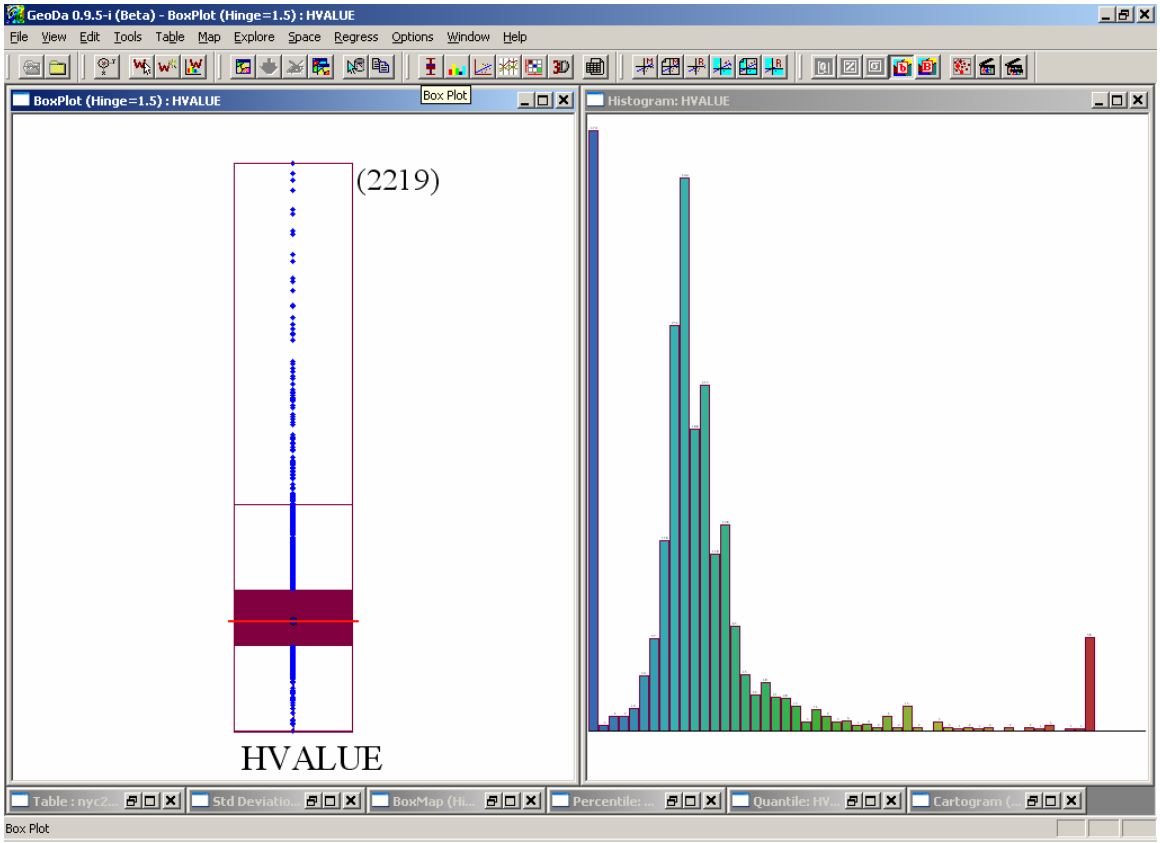
h) Box Plot

(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.**

i) Arrange and Save the results

Go to **Window > Tile Vertical**




Press **`alt + PrintScreen,`** and **switch back** to the Word document **Geodalab1.doc**, and click **`Paste,`** and click **`Save.`**

**`2)` Multivariate exploration**

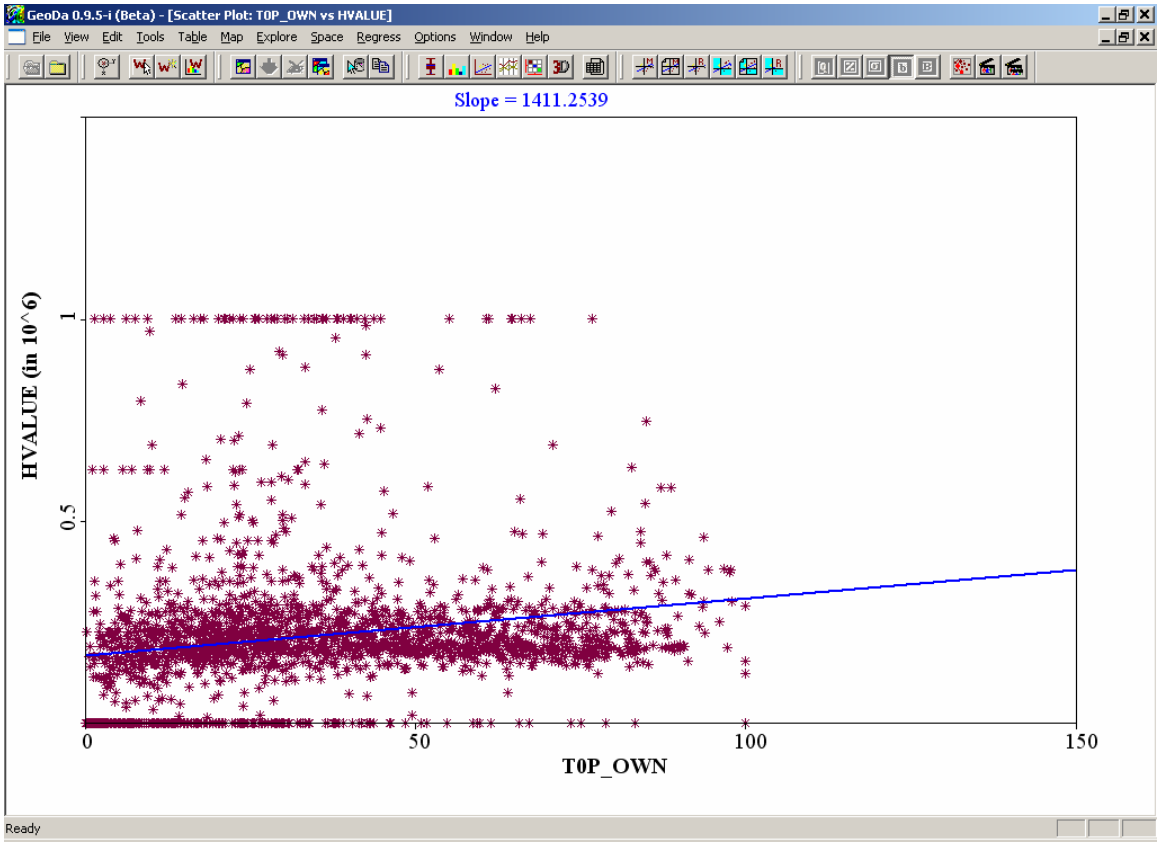
a) Scatter plot

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

I) Go to **Edit > Select Variable,** in the dialog box **select `TOP_OWN`** (percent of homeowners) as the second variable.

II) Go to **`Explore > Scatter Plot.`**

**You can also make a similar selection in the maps you created or in the other plots (scatter, histogram, or box) and see those selections dynamically displayed in the other representations you created.**



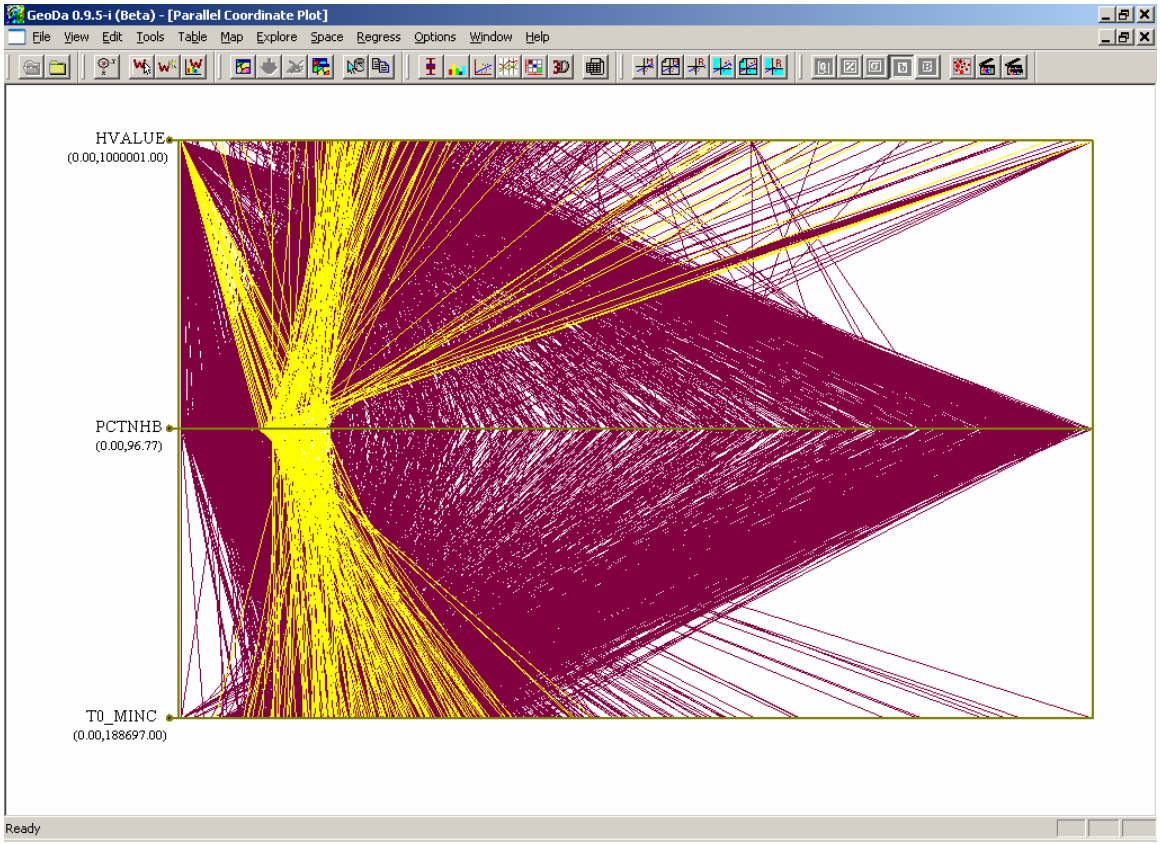

b) Parallel Coordinate Plot (PCP) and brushing

(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"). Simply click the mouse to end the brushing.




## 6. Close project

Go to **`File > Close All`**.

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

## 7. Extension

Use **Percent Homeowners (t0p_own)** or a variable of your own interest and try again. It may take a while to become comfortable with the linking and brushing, particularly in the multivariate setting.

## 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.