Introducing Spatial Analysis Using GeoDa
Workshop presented at "Teaching Economics: Instruction and Classroom Based Research"
Robert Morris University
February 2005

Jo Beth Mertens

GeoDa™ Exercise

Outline

1. Introduction

2. Background

3. The data set

4. Data description

5. GeoDa introduced

In this exercise, all "commands" are printed in **Times New Roman (Bold).**

This lab and data sets were created to demonstrate the power of analyzing data spatially using GeoDa software. The software is available free from the GeoDa web site http://sal.agecon.uiuc.edu/geoda_main.php.

1. Introduction

Present students with a spreadsheet of data and instruct them to "play with it. See what you can discover. What sorts of questions come up as you look at the relationships between the numbers? What can you tease out?" Imagine the looks—rolled eyes, incredulous stares; the body language—slumped in their seats, eyelids closing; and the sounds—heavy sighs, what is she, nuts?

Convincing students that data contains fascinating information and that examining it, or playing with it, is the way to get to know it and have fun with it is an almost impossible task. This is especially true with students today, who were raised with videogames and GameBoy. If they can't "see" it, they really are not interested. And "seeing" information in a table of numbers is difficult. But motivating students to examine data with statistical and econometric techniques is hard if they can not think of any interesting questions to ask that the data might answer. Somehow, their curiosity needs to be piqued. Once that happens, they develop questions and become interested in learning the techniques they must use to answer those questions. One way of piquing that interest is to use spatial analysis.

Luc Anselin and colleagues have developed a free software program, *GeoDa*™, which is an excellent tool for spatial analysis. With *GeoDa*, users can perform simple mapping, data analysis, and even spatial regressions. A major feature is the interactive nature of the program, because maps are linked to statistical graphics, such as histograms, box plots, and scatter plots. This allows users to actually "see" relationships between variables and their location, and this naturally leads to all kinds of questions.

In this workshop, I will introduce GeoDa™ using economic and population data for the 48 contiguous United States. GeoDa is an intuitive, easy to use program; therefore, the goal of this workshop is to familiarize you with some of its capabilities and provide you with a comfort level using the program so that you can use it with students to enhance their learning.

2. Background[1]

GeoDa™ is a trademark of Luc Anselin.

> GeoDa is the latest incarnation in a long line of software tools developed by Anselin and co-workers designed to implement techniques for exploratory spatial data analysis (ESDA) on lattice data (points and polygons). It is intended to provide a user friendly and graphical interface to methods of descriptive spatial data analysis, such as spatial autocorrelation statistics and indicators of spatial outliers.
>
> The design of GeoDa consists of an interactive environment that combines maps with statistical graphics, using the technology of dynamically linked windows. Its origins trace back to early efforts to develop a bridge between ESRI's ArcInfo GIS and statistical software. Its immediate precursor was the DynESDA extension for

[1] This lab exercise is based on "Santa Barbara Housing Data Exercise" developed by Stuart Sweeney and Rob Farrell, University of California—Santa Barbara for the July 2004 Santa Barbara SPACE workshop.

> ArcView 3.x, which introduced linked windows and brushing in a GIS environment. In contrast to the extension, the current software is freestanding and does not require a specific GIS system. GeoDa runs under any of the Microsoft Windows flavored operating systems. It also runs under the Virtual PC windows emulator on Mac operating systems (MacOS 9 and MacOS X). Its installation routine contains all required files and libraries. (http://sal.agecon.uiuc.edu/geoda_main.php.)

Luc Anselin suggests http://www.csiss.org/learning_resources/content/syllabi#gis for extended course notes and examples dealing with an introduction to spatial 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).

3. The data set

In this exercise, we are using BEA and Census data on gross state product, state tax revenue, state population, and state education. A complete listing of the data is in the Data Description section.

Please note that this exercise is meant to illustrate some of the capabilities of GeoDa for use in the classroom. It is not intended to cover the full functionality of GeoDa.

4. Data Description

VARIABLE NAME	All data is for 2000
GSP	Total gross state product, in millions of current dollars
PRIV_IND	Private industries as a percent of GSP
GOVT	Government as a percent of GSP
AGRI	Agriculture, forestry, fishing, and hunting as a percent of GSP
MINING	Mining as a percent of GSP
UTILITIES	Utilities as a percent of GSP
CONSTR	Construction as a percent of GSP
MANUF	Manufacturing as a percent of GSP
WHSL_TRADE	Wholesale trade as a percent of GSP
RETAIL	Retail trade as a percent of GSP
TRANSPORT	Transportation and warehousing, excluding Postal Service, as a percent of GSP
INFORMATION	Information as a percent of GSP
FIN_INS	Finance and insurance as a percent of GSP
REAL_ESTATE	Real estate, rental, and leasing as a percent of GSP
PROF_SERV	Professional and technical services as a percent of GSP
MGMT	Management of companies and enterprises as a percent of GSP
ADM_WASTE	Administrative and waste services as a percent of GSP
EDUC_SERV	Educational services as a percent of GSP
HEALTH	Health care and social assistance as a percent of GSP
ARTS	Arts, entertainment, and recreation as a percent of GSP
ACC_FOOD	Accommodation and food services as a percent of GSP
POP2000	2000 Population
PER_CENT_1	Percentage of state population over 25 with at least a bachelor's degree
TOTAL_TAXES	Total State Tax revenues in millions
TAXES_GSP	State tax revenues as a percent of GSP

5. GeoDa introduced

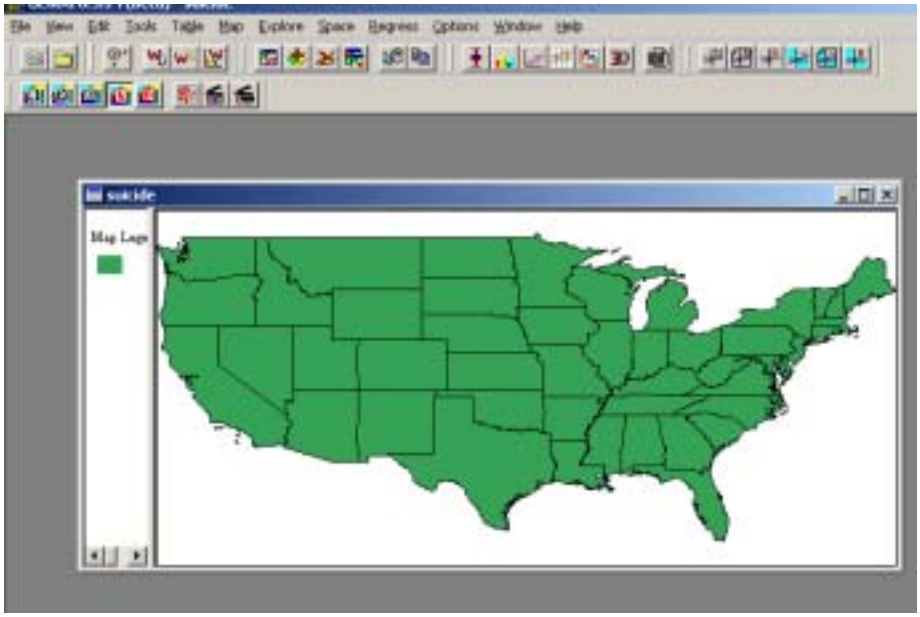
Start GeoDa

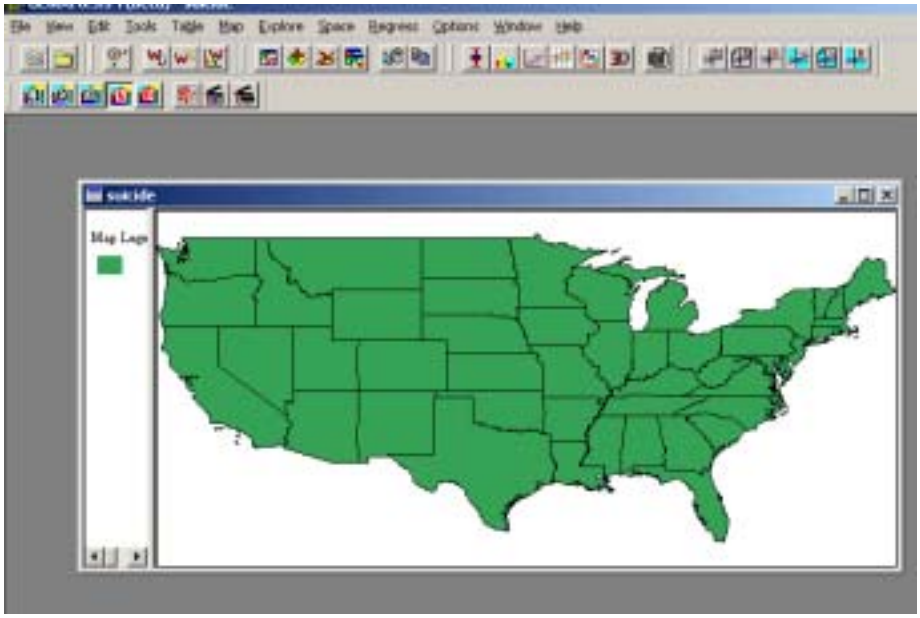
a) **File** > **Open Project** (This opens the GeoDa Project Setting dialogue box.)

b) **Browse** through the folder to find and select the shape (.shp) file, c:\work\final.shp.

c) Select **STATE_NAME** as the “key Variable” (by scrolling and clicking on the variable name.) The key variable must have a unique value for each observation (in this case, state name.) The unique value is used to implement the link between maps and statistical graphs.

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





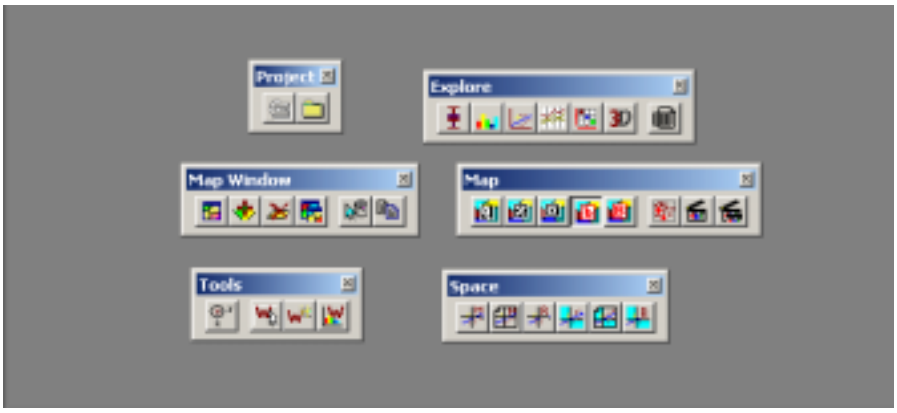
(you will have to **click** on the left side of the map window and then drag to the center of the map to view the legend area.)

e) To maximize the GeoDa window, **Click on the full screen button for GeoDa**. All windows can be re-sized and positioned anywhere within the main program window.

The GeoDa menu bar has 12 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

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



The **File** Menu (**Project Toolbar**) contains the standard project management commands (e.g., open a new project, close project, etc.)

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.

Basic Data Exploration

a) First, change the background color to improve clarity. Right-Click on the map, and choose **Color > Background**, and choose a light gray.

b) Create histograms. For example, we could explore the distribution of gross state product.

c) To select a variable, choose **Edit > Select Variable**. This opens a variable settings dialog box. (Note: for univariate operations, the choice of a second variable is ignored.)

d) In the dialog box select the variable of interest **(GSP)** by scrolling down the 1st variable Y listing.

e) Notice that the box "select variable as default" is not 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 box is not checked, the variable selection dialog box will open for each mapping or statistical operation.

f) Click **OK.**

Note: Tables of data may appear following certain operations. The user may close, move, or minimize the table. These instructions will assume the table is closed.

g) To produce a basic (non-spatial) histogram of GSP, select **Explore** > **Histogram** (if **GSP** is not checked as the default, then select **GSP**.) Note the lowest value is 17,798 million (Wyoming) and the highest is 1,296,637 million (California). Values proceed left to right.

h) The spatial component of this distribution can be viewed, however, using the Map Movie functionality. Choose **Map** > **Map Movie** > **Cumulative**. Make sure the new window and the histogram are visible and set the speed of play (using the continuous toggle) to about 700 (the higher the number the slower the speed). Choose **Play.** Note the geographic relationship of the poorest and wealthiest states (with respect to GSP). Also note how the histogram display is linked to the map movie.

i) Now repeat steps g-h for the variable **TAXES_GSP**

j) When finished, close the map movie and histogram windows. Maximize the map and click anywhere on it (this de-selects those points selected by the map movie function.)

k) Now let's map individually the variables for population, agriculture as a percent of GSP, and tax revenue as a percentage of GSP.

l) Choose **Edit** > **Select Variable**, and choose **POP2000** (1st variable—don't worry about choosing a second variable. That only applies in multivariate analysis, which we will be doing later.)

m) Choose **Map** > **Quantile** and select 9 categories. How much clustering is present in this variable? Select **Window** > **Tile Horizontal** and simultaneously examine the quintile map and the histogram.

n) Repeat steps k and l for the other two variables, **AGRI**, and **TAXES_GSP.**

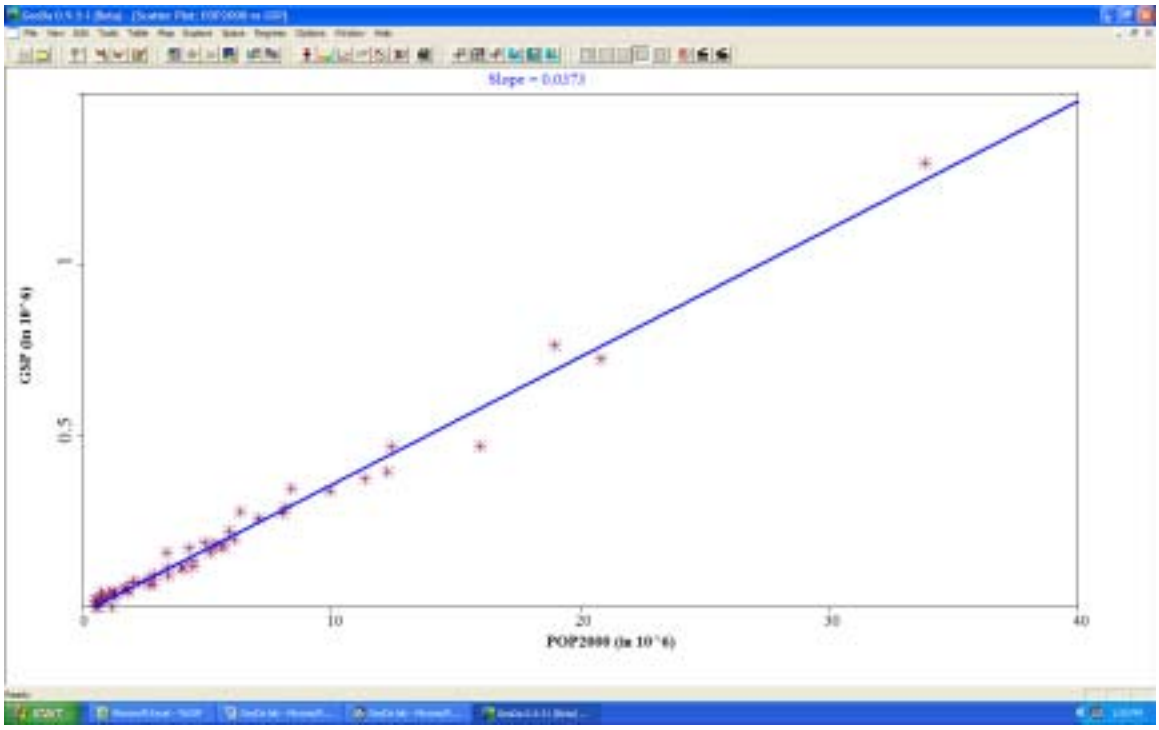
Bivariate and multivariate exploratory analysis

Scatterplots

a) Create scatterplots for GSP versus population, and taxes as a percent of GSP against population. To do this, you must set the variables using the Edit menu.

b) Choose **Edit > Select Variables**, and select **GSP** and **POP2000** as the variables.

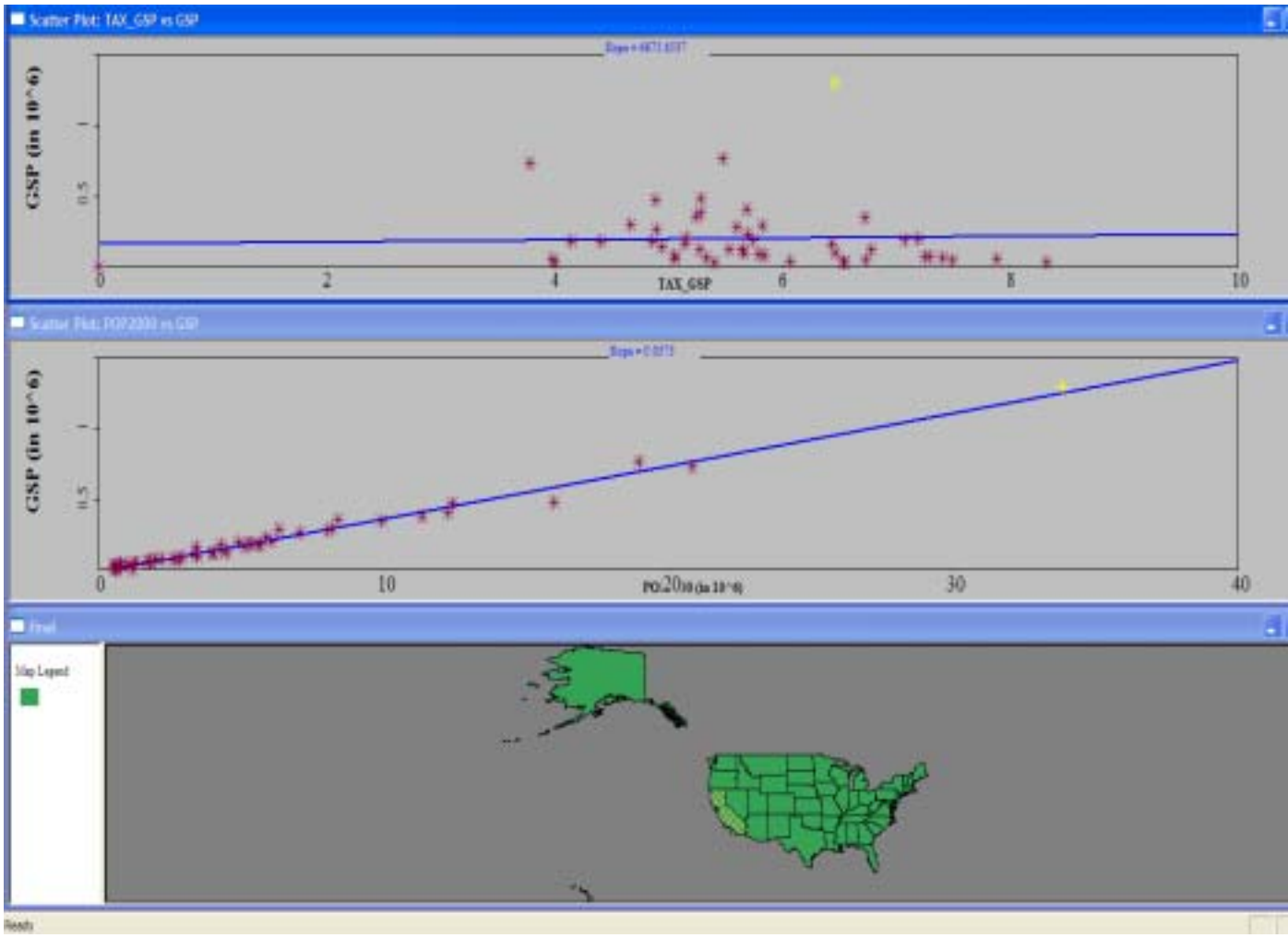
c) Choose **Explore > Scatterplot.** The scatter plot should look something like this…..



What does this mean?

d) Repeat b & c for taxes as a percent of GSP.

Note how GeoDa links the windows. If you click on the highest point in the Tax_GSP to GSP scatter plot, the corresponding point in the other scatter plot as well as the state on the map are highlighted (California.) It should look like this…..



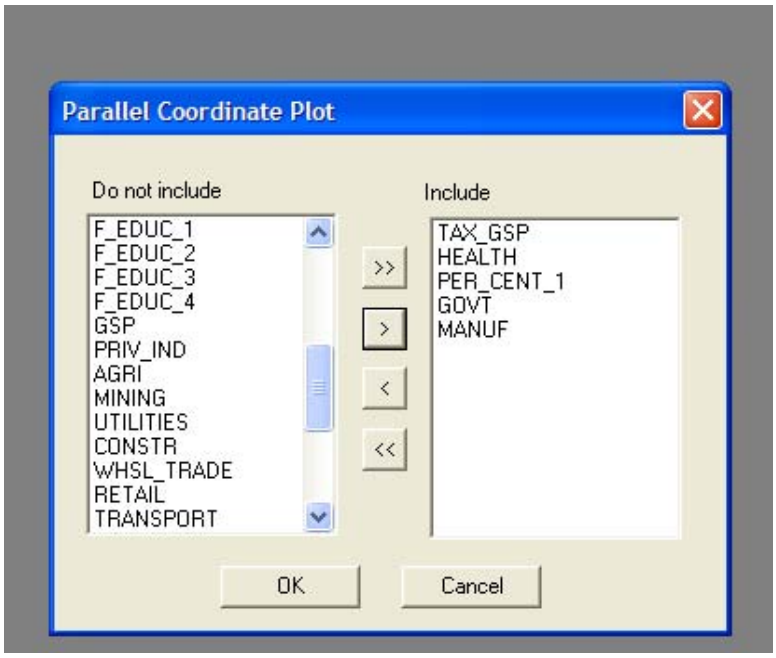
e) Note that GeoDa has the ability to link selections across multiple windows. To examine some of the points in the scatter plot simultaneously, tile your map and the scatterplot window so that you can see them both. Click and drag to select a group of points in the scatter plot and notice that they become highlighted in the map window.

f) Now with the **Ctrl** button depressed, click, drag, and release to create a small box in the scatter plot window. It will flash for a coupe of seconds, and then become continuously active. You can move it around the scatter plot and dynamically highlight portions of then plot ("brushing"), all the while viewing the active selections in the map ("linking"). Click the mouse to end the brushing.

g) Close the scatterplot window.

Parallel Coordinate Plots

a) Choose **Explore > Parallel Coordinate Plot.** A window will appear allowing you to choose the variables you wish to examine….



b) Add these variables in the following order: TAXES_GSP, HEALTH, PER_CENT_1, GOVT, MANUF and then click **OK.**

c) Maximize the plot window and make the background color gray (using **right-click > background color**)

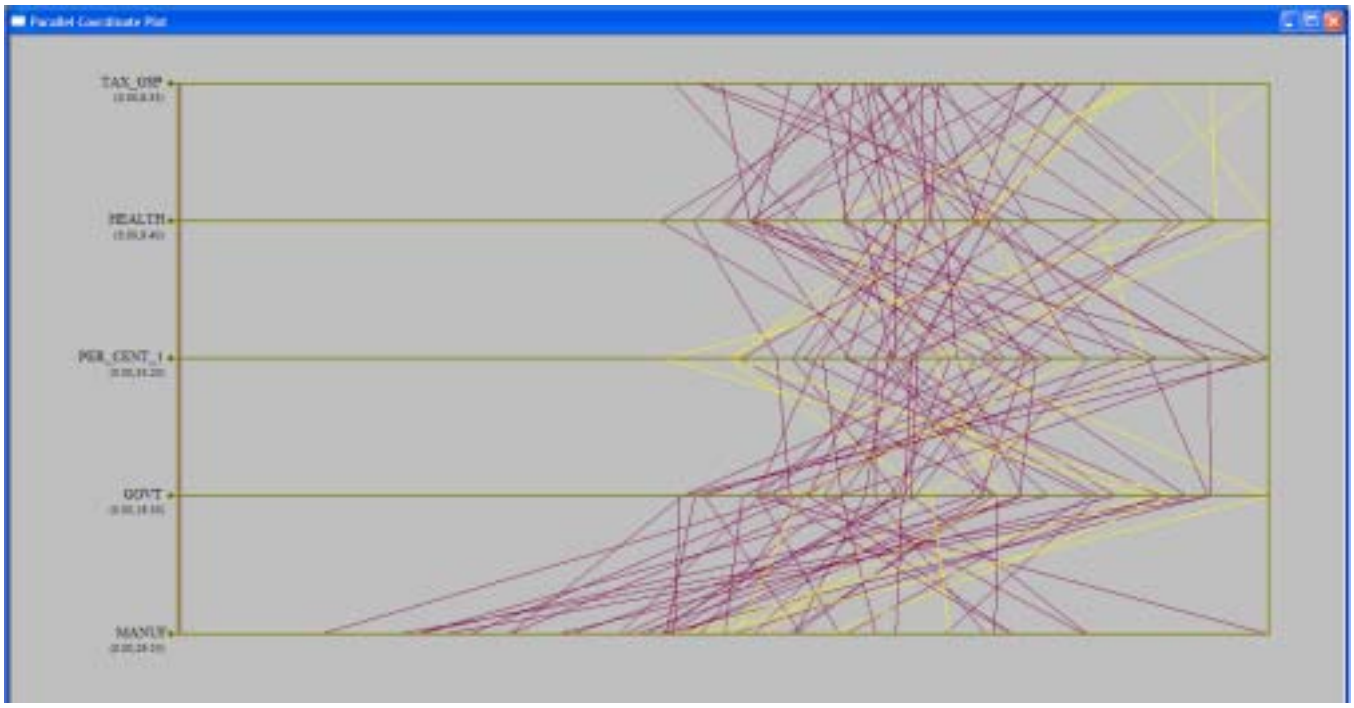
d) Note the vertical axis has each variable name, and in parentheses below the variable are the low and high values for that variable.

e) By right-clicking, the user can toggle back and forth between original data and standardized scores.

f) Examine the data by selecting groups of data points based on some variable and assessing their simultaneous values on all the others. For example, note the relationship between manufacturing and education.

g) Using the histogram on **TAX_GSP**, click on high and low Tax revenue to GSP states and note the relationships between the other variables.

The parallel coordinate plot should look something like this….



h) Just as in previous exercises, the selected items are linked to the map window, so that the states currently selected are highlighted in the map window and any other window open.

I now encourage you to sit and play with the program and look at spatial relationships that are surprising to you. Note that GeoDa can perform multiple regression analysis, as well as spatial regression analysis. (But that is saved for another lab session.)