Center for Spatially Integrated Social Science

Exploratory Spatial Data Analysis and GeoDa

Luc Anselin
Spatial Analysis Laboratory
Dept. Agricultural and Consumer Economics
University of Illinois, Urbana-Champaign
http://sal.agecon.uiuc.edu

Outline

- CSISS Tools Project
- Overview
- GeoDa Demonstration

CSISS Tools Program

CSISS - http://www.csiss.org

- Center for Spatially Integrated Social Science
  - NSF Infrastructure Project 1999-2004
  - headquartered at UC Santa Barbara
  - Software Tools Development at UIUC
- Mission
  - promote spatial thinking and spatial analysis in social sciences

Center for Spatially Integrated Social Science (CSISS) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites History Links

Address http://www.csiss.org/ Go

The CSISS Mission recognizes the growing significance of space, spatiality, location, and place in social science research. It seeks to develop unrestricted access to tools and perspectives that will advance the spatial analytic capabilities of researchers throughout the social sciences. CSISS is funded by the National Science Foundation under its program of support for infrastructure in the social and behavioral sciences.

CSISS News Apply now for the 2002 CSISS Summer Workshops!

Core Programs	Learning Resources	Spatial Resources	Spatial Tools
These six infrastructure programs form the core of the Center's activities.	These introductory materials include **CSISS Classics** and **select video clips** from the CSISS summer workshops.	CSISS has compiled e-journals, bibliographies, and other spatial resources for the social sciences.	**Spatial Tools Search Engine** **Select Tools** **Links to Portals**

Search Engines	CSISS Events	Community Center	About CSISS
Try CSISS's custom search engine to find spatial analysis resources on the Internet.	Here's where you'll find information and registration for workshops, conferences and specialist meetings.	Join one of the forums on topics such as spatial equity, spatial externalities, and spatial econometrics.	CSISS people, programs and the original NSF proposal are described here.

Core Programs | Learning Resources | Spatial Resources | Spatial Tools | Search Engines | CSISS Events | Community Center | About CSISS
Site Map | Site Search | Contact CSISS | Plug-ins | Privacy Policy | Site Credits | Home

Internet

CSISS Tools Project Mission

- Goals:
  - facilitate dissemination of spatial analysis software tools to social scientists
  - develop a library/libraries of spatial data analysis modules
  - develop prototypes implementing state of the art methods
  - initiate and nurture a community of open source developers
- Industrial Partner: ESRI

http://www.csiss.org/clearinghouse/index.php3

Center for Spatially Integrated Social Science

CSISS Tools Clearinghouse

The ***CSISS Tools Clearinghouse*** is intended to grow into a robust collection of spatial analysis software, software links, and links to information about tools for spatial analysis. The development of these tools is a lively research area and the goal of this clearinghouse is to provide up-to-date information on available tools. The clearinghouse is comprised of:

Search Engine

Search a continuously updated, comprehensive index of the CSISS Select Tools and Links to Portals.

Select Tools

Browse through tools particularly suited to the analysis of spatial phenomena.

Portal Links

A listing of useful collections of software tools for anyone interested in Spatial Analysis, or those looking for specific tools.

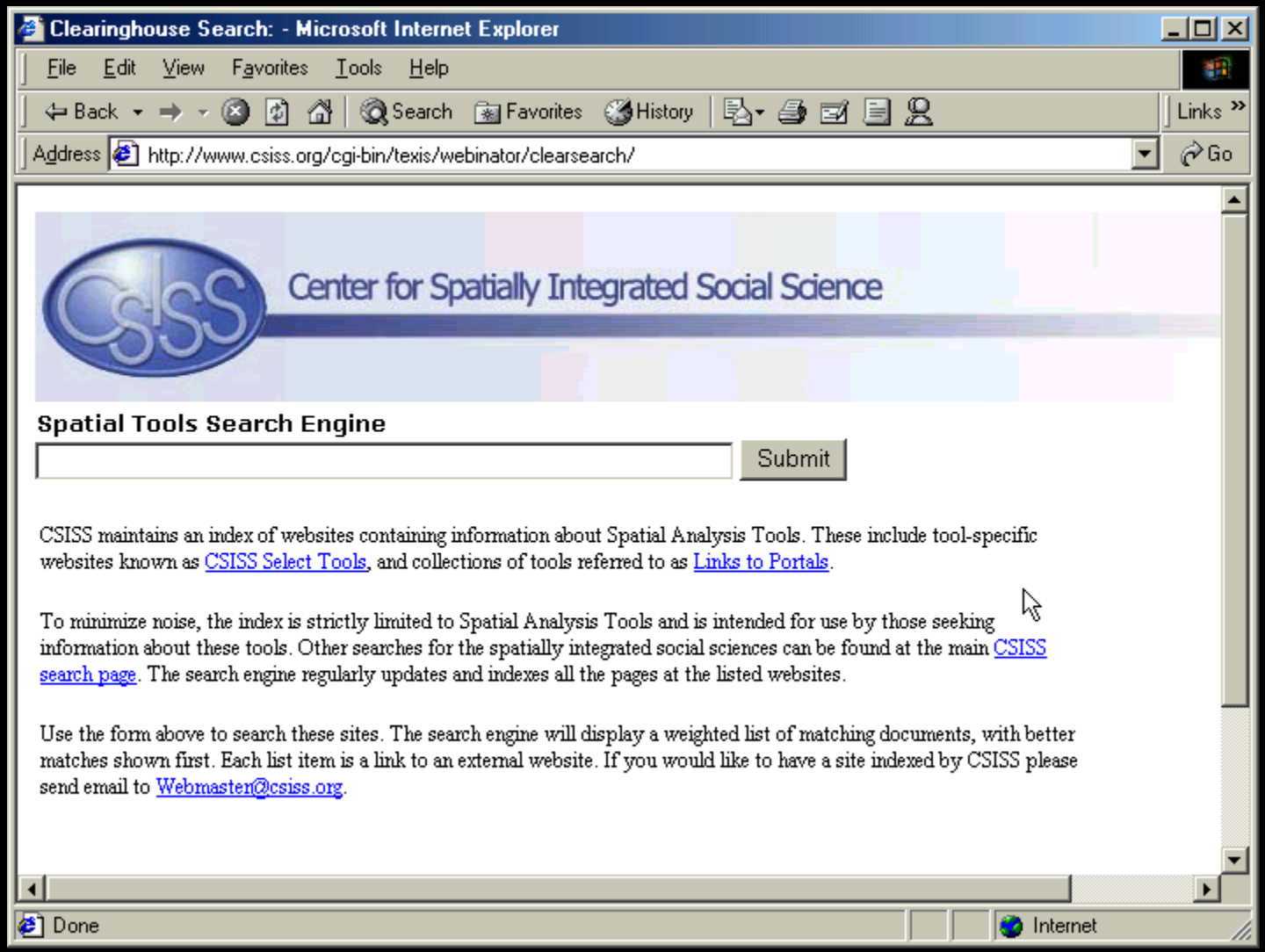
CSISS Tools (offsite)

The home of the software tools development efforts under CSISS, carried out in the Spatial Analysis Laboratory of the Department of Agricultural and Consumer Economics at the University of Illinois, Urbana-Champaign.

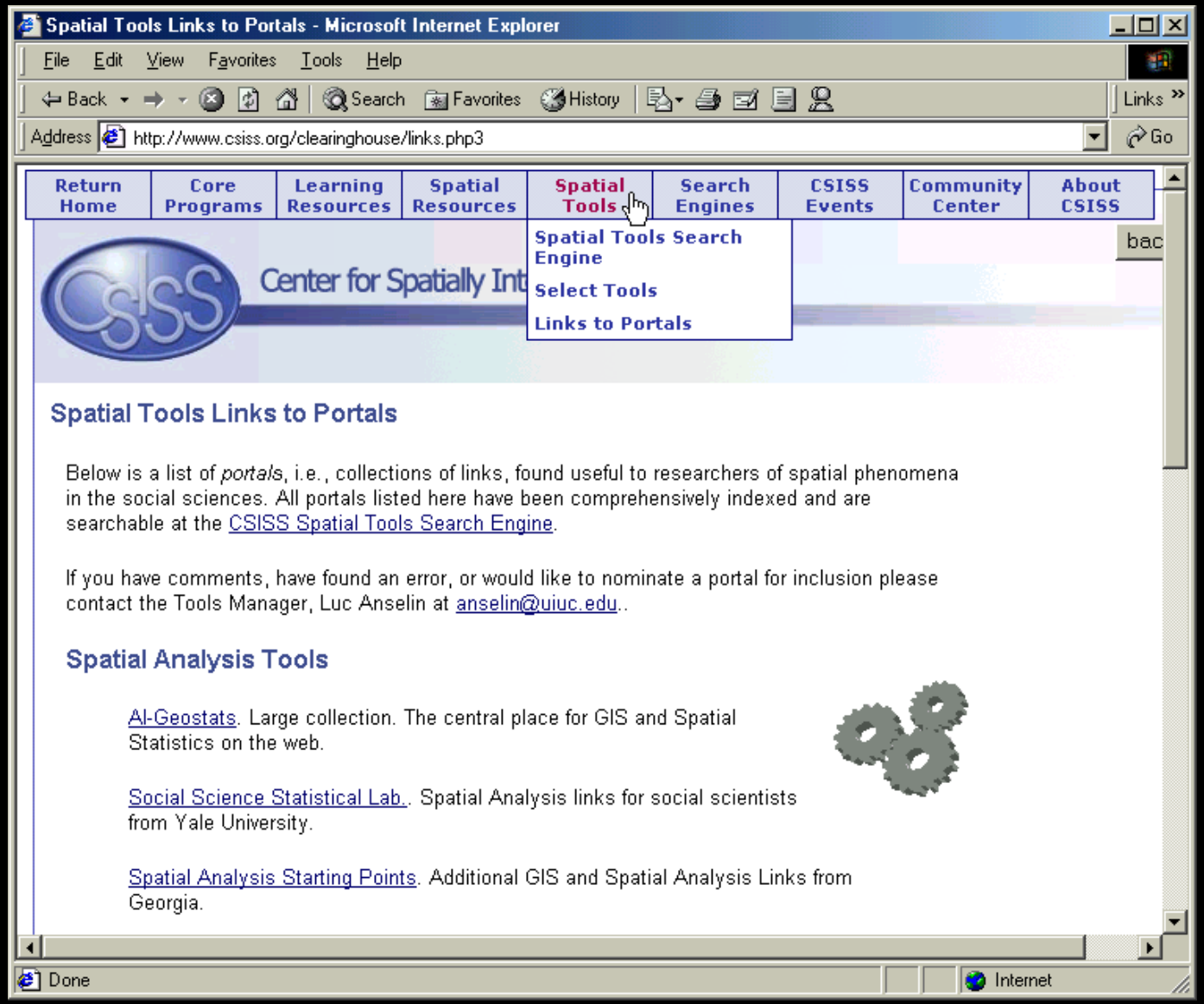
New - GeoDa 0.9, beta release software for ESDA with dynamically linked windows.

Your help is requested in suggesting tools, collections of tools, other portals, and methods that should be represented in this collection - please send these to the Tools Manager, Luc Anselin at anselin@uiuc.edu.

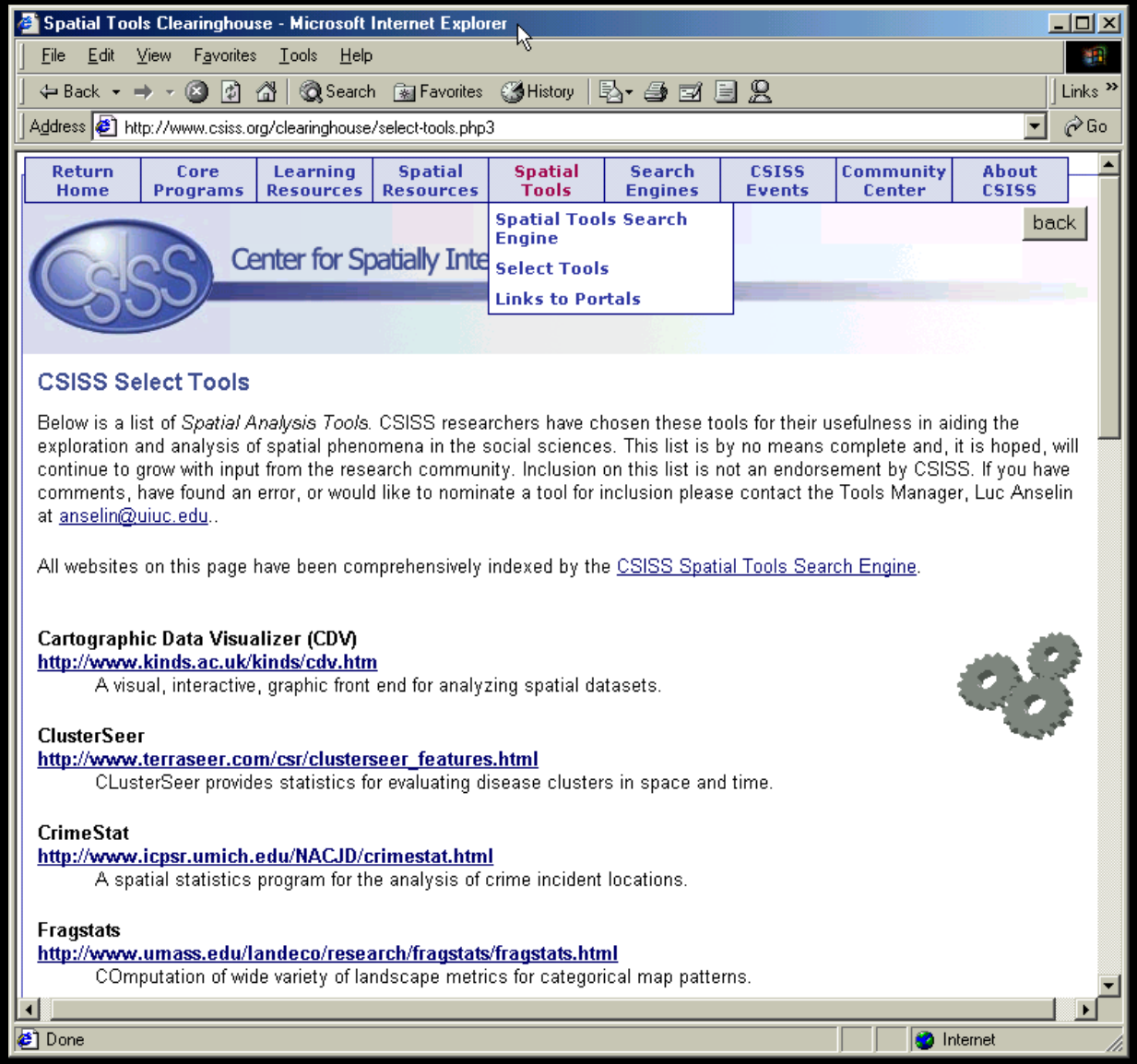
Spatial Tools Search Engine



Links to Spatial Tools Portals



Select Tools



SAL - http://sal.agecon.uiuc.edu

- Spatial Analysis Laboratory
- Research Unit in ACE
  - development, implementation and application of methods of spatial analysis to policy issues in agricultural, consumer and environmental sciences
  - home of CSISS software tools

Address http://agec221.agecon.uiuc.edu/csiss/ Go Links

CSISS Software Tools Project

Mission

The Center for Spatially Integrated Social Science (CSISS) is a five-year project funded by the U.S. National Science Foundation under its program of support for infrastructure in the social and behavioral sciences. CSISS promotes an integrated approach to social science research that recognizes the importance of location, space, spatiality and place.

One of the CSISS programs is devoted to "Spatial Analytic Tools" for the social sciences. It is directed by Luc Anselin and housed in the Spatial Analysis Laboratory of the Department of Agricultural and Consumer Economics at the University of Illinois, Urbana-Champaign. The Tools Project aims to develop and disseminate a powerful and easy to use suite of software for spatial data analysis, to advance methods of statistical analysis to account for spatial effects, and to integrate these developments with GIS capabilities.

Current Activities

- Openspace Mailing List
- Spatial Software Tools Clearinghouse
- GeoDa - Exploratory Spatial Data Analysis with Dynamically Linked Windows
- OpenSpace - Java Applets and Applications for Spatial Data Analysis
- PySpace - Spatial Statistical Analysis in Python
- Large Data Set SAR
- Supporting Materials
- R-Geo - Spatial Data Analysis and the R Project

Overview

Software Development Efforts

- GeoDa
  - Visual and interactive data exploration
- Large Data Spatial Regression
  - Develop and implement new algorithms
- OpenSpace
  - Java tools (beans) for ESDA and spatial regression
- PySpace
  - A collection of object-oriented cross-platform software modules for ESDA and spatial regression

GeoDa

- An Introduction to Spatial Data Analysis
  - ease of use, point and click
  - visual methods
  - natural extension of mapping/GIS analysis
- ESDA with Dynamically Linked Windows
  - freestanding
  - reads/writes ESRI shape files
    - points and polygons
  - MapObjects LT2 technology
  - free
- Download
  - http://sal.agecon.uiuc.edu/csiss/geoda.html

Address http://agec221.agecon.uiuc.edu/csiss/geoda.html Go Links

GeoDa

- About GeoDa
- Tutorials
- Sample Data
- New in GeoDa 0.9.3
- Upgrade to GeoDa 0.9.3
- Download Geoda 0.9.3

About GeoDa

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 an 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 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.

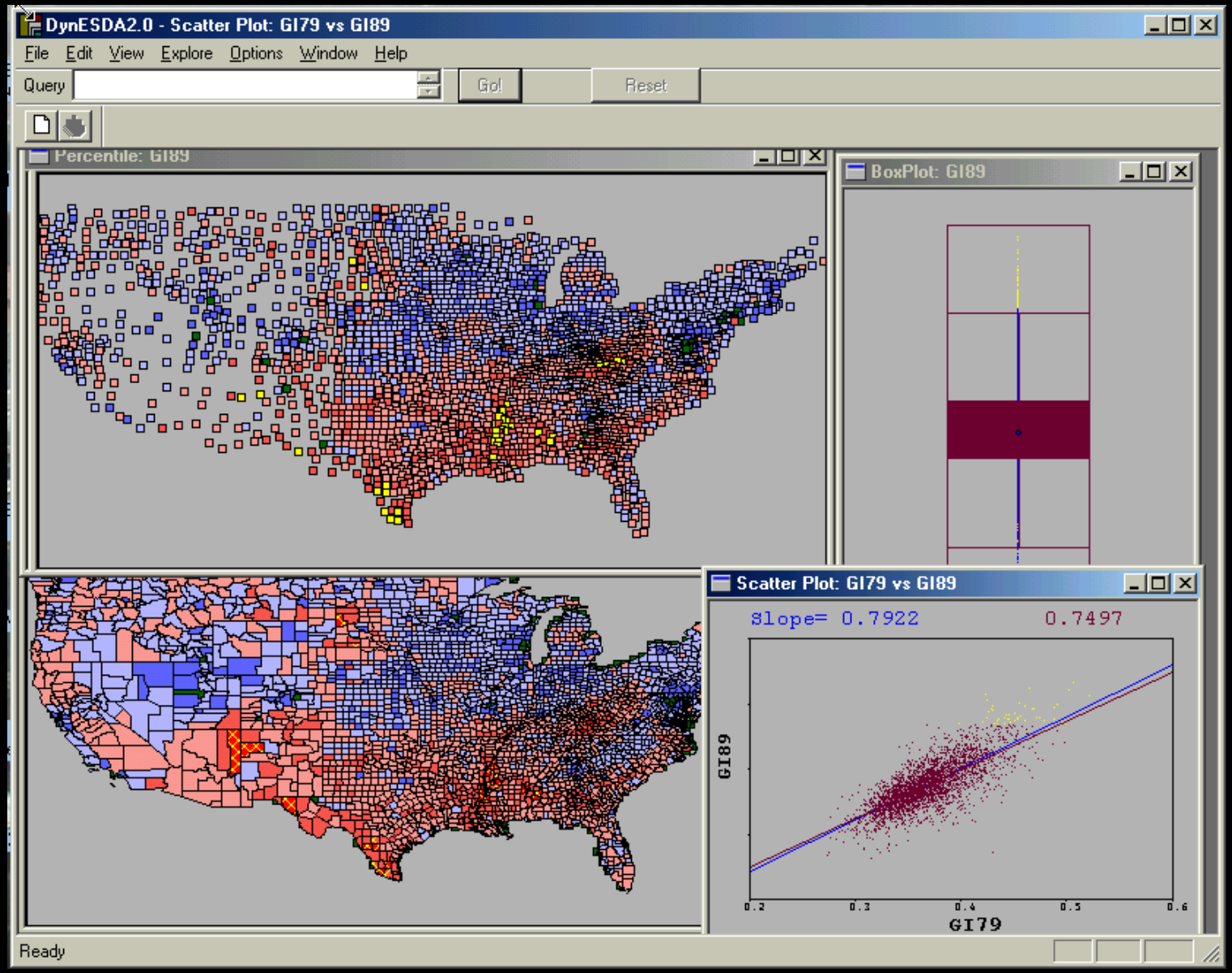
Over the years, the development of GeoDa and its precursors has been supported in part by research projects funded by the U.S. National Science Foundation. Most recent is grant BCS-9978058 to the Center for Spatially Integrated Social Science (CSISS). Earlier funding was provided by grants SBR-9410612, SBR-9513040 (to the National Consortium on Violence Research, NCOVR) and SES-8810917 (to the National Center for Geographic Information and Analysis, NCGIA).

For technical background on an earlier version of the software, see Luc Anselin, Ibnu Syabri and Oleg Smirnov (2002), Visualizing Multivariate Spatial Correlation with Dynamically Linked Windows, In L. Anselin and S. Rey (Eds.), Proceedings, CSISS Workshop on New Tools for Spatial Data Analysis, Santa Barbara, CA, May 10-11, 2002. Center for Spatially Integrated Social Science, CD-ROM (pdf file, 20pp, 517K).

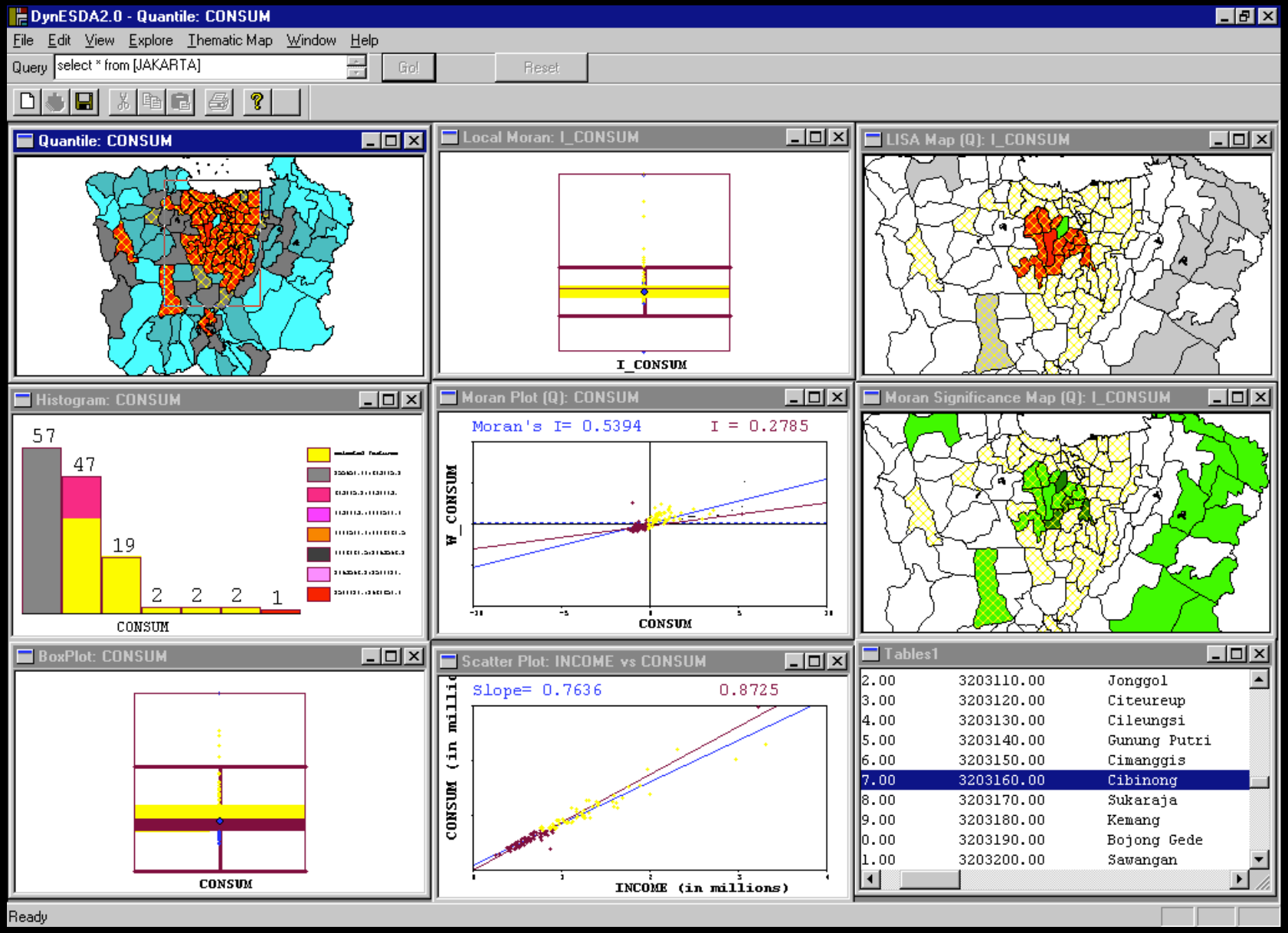
GeoDa Functionality

- Dynamic Linking and Brushing
  - maps and statistical graphs
- Outlier Maps
  - box map, cartogram, map movie
- Rate Smoothing
- Spatial Autocorrelation Analysis
  - global and local
  - univariate and bivariate (space-time)
- Utilities
  - spatial weights construction, spatial lag, point-polygon conversions, data import/export
- Spatial Regression
  - coming soon

Linking Point and Polygon Maps



Dynamically Linked Windows



Large Data Spatial Regression

- Maximum Likelihood Estimation
  - spatial lag and spatial error model
  - specialized algorithms
  - works with large sparse spatial weights
    - 1,000,000 in +/- 7 minutes

Large Data Set SAR Example

```
C:\DEVEL\SRC\large3\mlcp_lag>ml_lag grid1m.gwt sim1m.txt
 spatial weights matrix:                grid1m.gwt
 precision (max length of polynomial): 41
 grid1m.gwt:
 dimension: 1000000    number of dependent variables: 2
matrix has 1000000 rows  and 3996000 nonzero entries.
  computing polynomial - 41 ...
  --- finished computing polynomial
................................................
 iterations: 47  convergence error: 6.37593422503e-10
     SOLUTION:
         coefficient of spatial association (spatial lag model): 0.0612687124729

         log-likelihood: 1361364.24818
         coefficients for spatial regression:
                        0.000812801096478
                        0.212760164716
     ---  end solution ---
 time elapsed: 440653
```

PySpace

- Goal
  - an open source, cross-platform, object-oriented suite of tools for ESDA and spatial econometrics
- Python as the glue
  - collection of scripts to manipulate data bases (mySQL,PostgresQL), compute statistics, drive visualization tools (vtk)

PySpace Prototype

- Spatial Linear Regression
  - spatial weights construction and characteristics
  - diagnostics (LM, etc.)
  - ML estimation (lag, error)
  - IV and GM/GMM estimation
- Spatial Panel Data
  - pooled cross-section/time series
  - diagnostics
  - ML/IV estimation
- Spatial Probit
  - diagnostics
  - prototype simulation estimators

PySpace Example

```
>>> ivlag(crime,xdep0,stdw(colw),['Crime'],['Inc','Hoval'])


************************************************************
****     INSTRUMENTAL VARIABEL ESTIMATION FOR LAG MODEL     ****
************************************************************

The weight matrix IS row-standardized
The instruments are the spatial lag of exogenous variables
Spatial lag of dependent variable created, name starts with W_
Anselin-Kelejian test (Int. Reg. Sci. Review, 1997) computed

Variable      Coefficient     Std. Dev.     z-value     Prob>|z|
--------      -----------    ----------    -------     --------
Constant          45.9474     11.515338       3.990     0.000066
Inc              -1.05273      0.394278      -2.670     0.007584
Hoval           -0.259841      0.092534      -2.808     0.004984
W_Crime          0.410451      0.193866       2.117     0.034243
============================================================

Dep. variable = Crime        Observations = 49
Deg. of free. = 45           Variables =3
Sum sq. resid.= 4706.3391
Sigma-squared = 104.5853 ( 10.2267 )
Anselin-Kelejian test=0.384446 (Prob>|stat|=0.649676)
LM error [df.=1]= 0.288855 (Prob.>stat= 0.590955)
************************************************************
```

R-Geo

- Spatial Data Analysis in R
  - open source and cross-platform
  - clearing house for specialized packages
  - links to resources
  - tutorials and sample data sets
- Spdep
  - spatial autocorrelation and spatial regression

Address http://agec221.agecon.uiuc.edu/csiss/Rgeo/index.html Go Links

Spatial data analysis
R spatial projects

Packages on CRAN
Maps
Point pattern
Geostatistics
Areal
GIS interfaces
Others

Other Packages
Details

Mailing list
R-sig-Geo

Links
Back to SAL

R spatial projects

This collection of web pages is intended to be a guide to some of the resources for the analysis of spatial data using R, and other associated software. Corrections and contributions are very welcome, and may be made through the mailing list R-sig-Geo, or directly to the site maintainer.

Please note that all software and documentation here provided or described is done so "as is" without warranty of any kind. Package maintainers (email addresses on CRAN descriptions, or package homepages) will be grateful for bug reports, or other questions sent after the package documentation, including where necessary the source code, has been read and found unsatisfactory or insufficient. Contributions of code, documentation, good examples, and fruitful questions are alway welcome.

Spatial data analysis with R

A key insight in spatial data analysis is that the "spatial" may add something extra - location may matter in grasping what is driving the data. But it does not have to matter, and good spatial data analysis must also be good data analysis, meeting general requirements for care in handling data and in drawing conclusions. Because R is a very rich environment for general data analysis, it invites spatial analysts to demonstrate clearly that "space" does add insight to analysis, not just assume that this is the case, because the data are spatial.

A further insight is that "spatial" may apply to many fields of data analysis in which the absolute or relative position of observations in relation to each other may have importance, and that methods applied in, for example, sociology or education have direct parallels in "spatial" analysis. Medical imaging is another field of relevance, as indeed are many methods used in examining the relative positions of observations in attribute space. The focus here is on "geographical" spatial data, where observations can be identified with geographical locations, and where additional information about these locations may be retrieved if the location is recorded with care. The date and time of observation are also often of importance. In this connection, geographical information systems (GIS) are very relevant software applications, so that both GIS data formats, and ways to construct maps (thematic or statistical cartography) matter.

Maps and R

Graphical data analysis has always been a strength of S, and thus also of R. S has had a legacy map() function, based on an internal database format of som sophistication, using a topological representation. It is not easy to add to this database, and work in R has been more to permit the import of foreign formats.

Address http://agec221.agecon.uiuc.edu/csiss/Rgeo/index.html Go Links

Areal/lattice data analysis packages on CRAN

Spatial data analysis
R spatial projects

Packages on CRAN
Maps
Point pattern
Geostatistics
Areal
GIS interfaces
Others

Other Packages
Details

Mailing list
R-sig-Geo

Links
Back to SAL

Areal or lattice data is observed on, or often aggregated up to often arbitraty spatial units, like census tracts, counties, or countries. Often, such data is analysed as though the observations were independent, ignoring possible spatial dependence. In practice, dependence may stem from the misfitting of unit boundaries to phenomena being measured (labour markets may not match city limits), from missing variables resolving dependence, or from the actual impact of neighbours on each other (spillover).

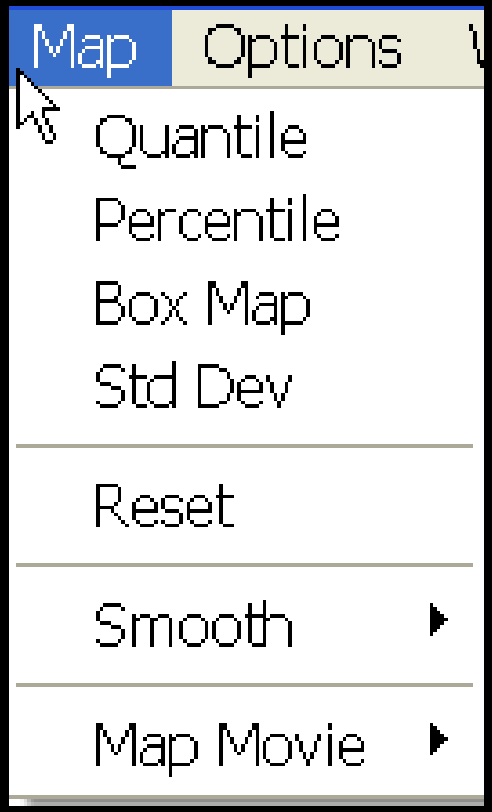
spdep

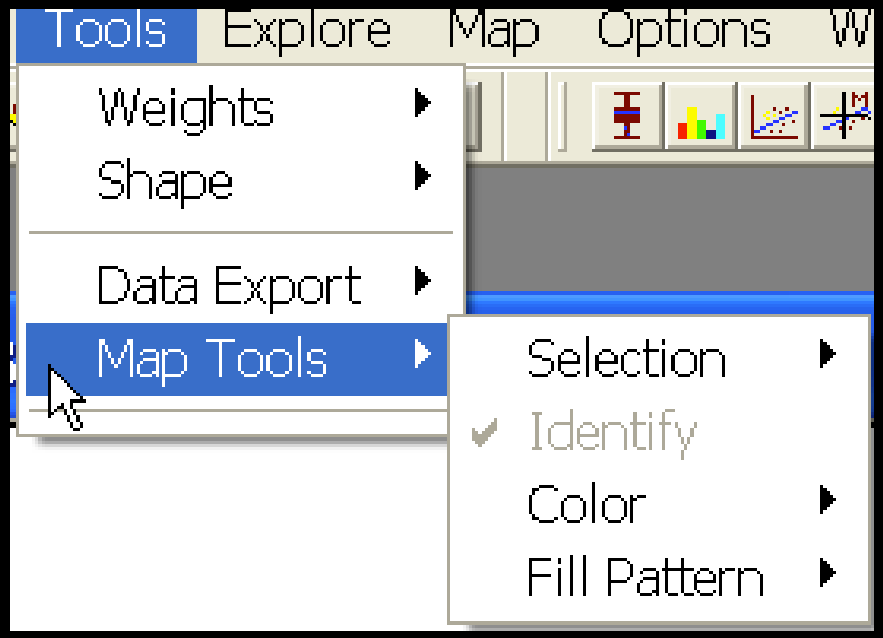
spdep is a collection of functions to create spatial weights matrix objects from polygon contiguities, from point patterns by distance and tesselations, for summarising these objects, and for permitting their use in spatial data analysis; a collection of tests for spatial autocorrelation, including global Moran's I, Geary's C, Hubert/Mantel general cross product statistic, Empirical Bayes estimates and spatial Index, and Getis/Ord G, local Moran's I and Getis/Ord G, saddlepoint approximations for global and local Moran's I; and functions for estimating spatial simultaneous autoregressive (SAR) models. It has no homepage as such, but draft releases may be posted here.

Last modified: June 20, 2003 by Roger Bivand

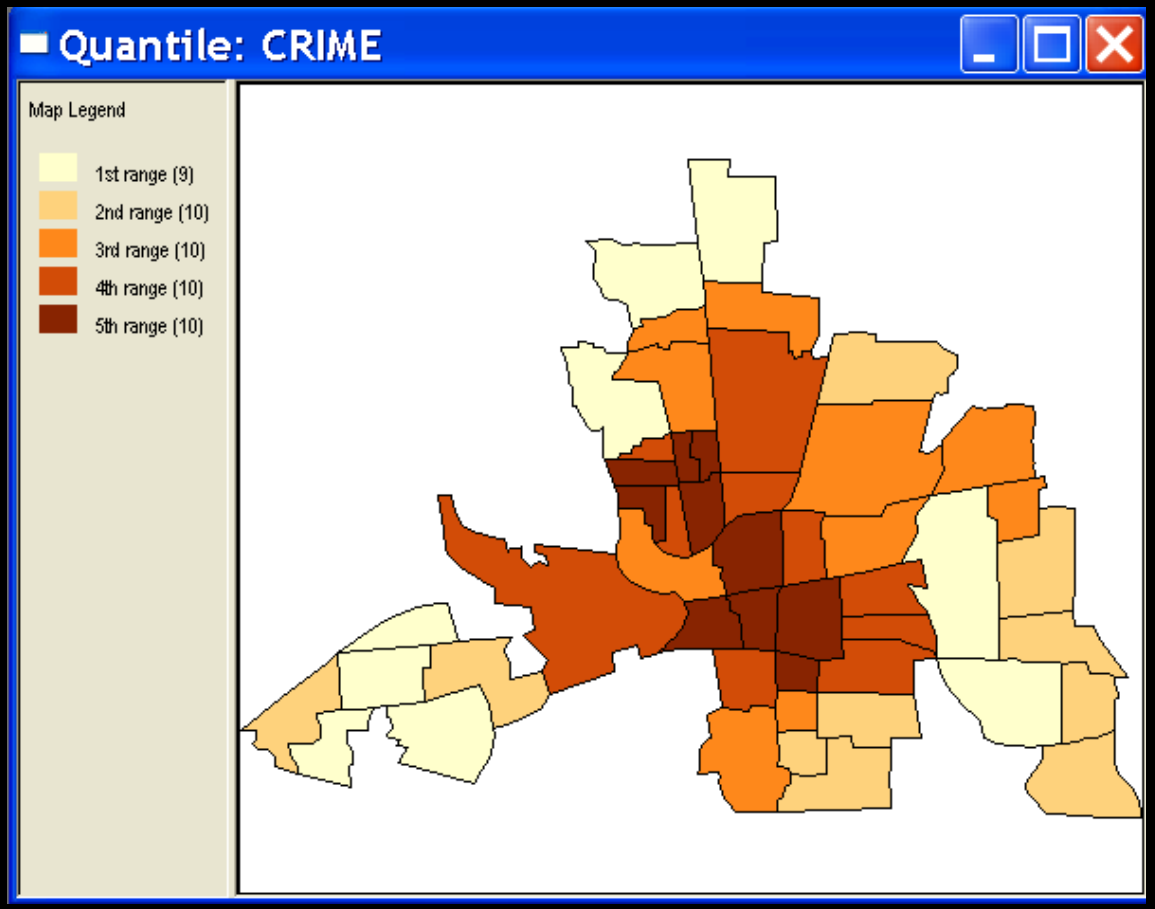
GeoDa Demonstration

Mapping in GeoDa

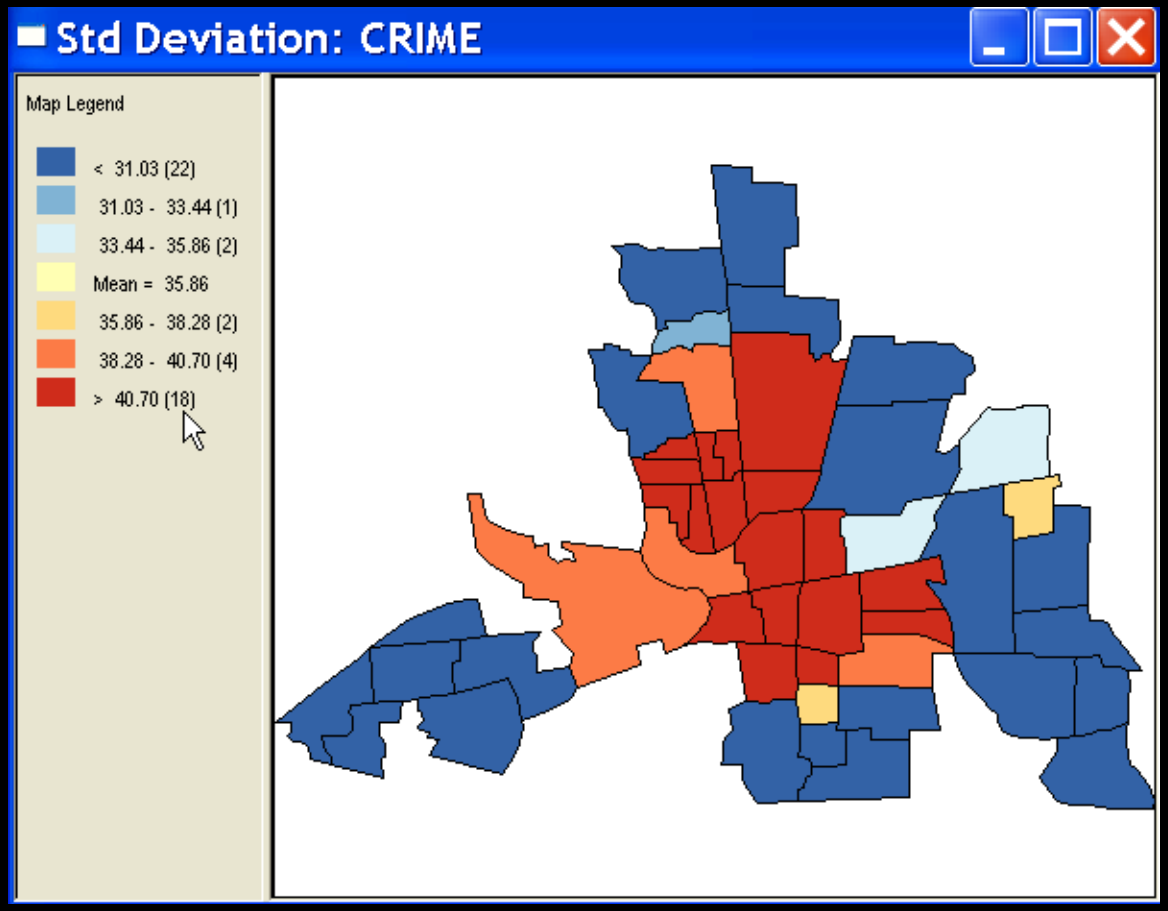




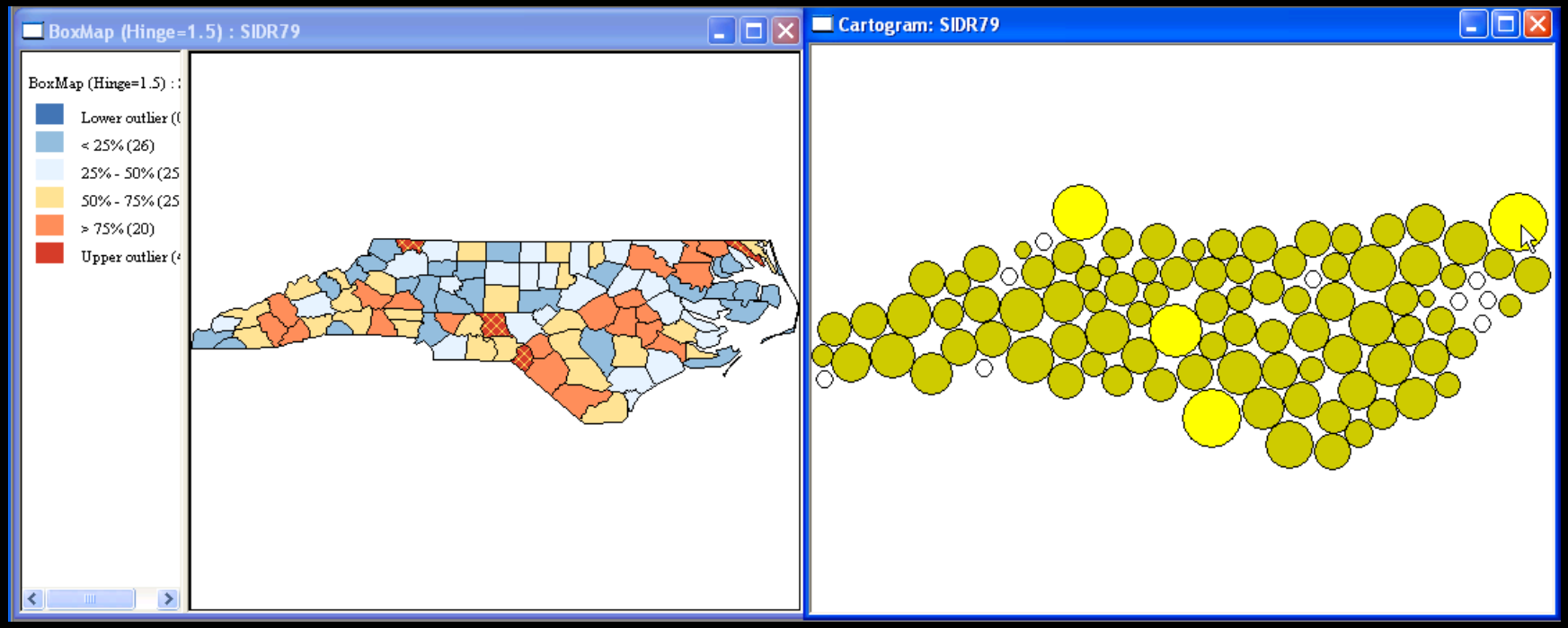
Quantile Map



Standard Deviational Map

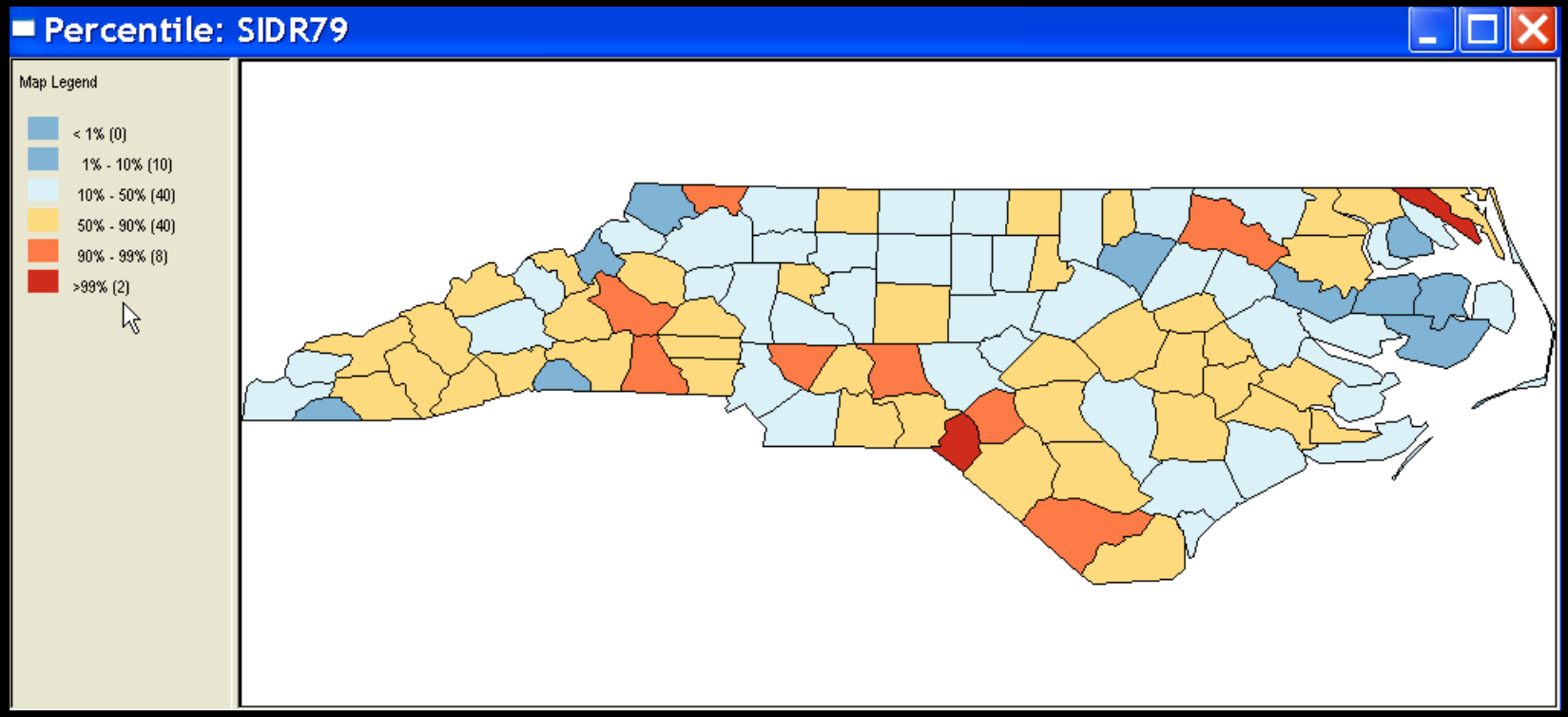


Cartogram in GeoDa

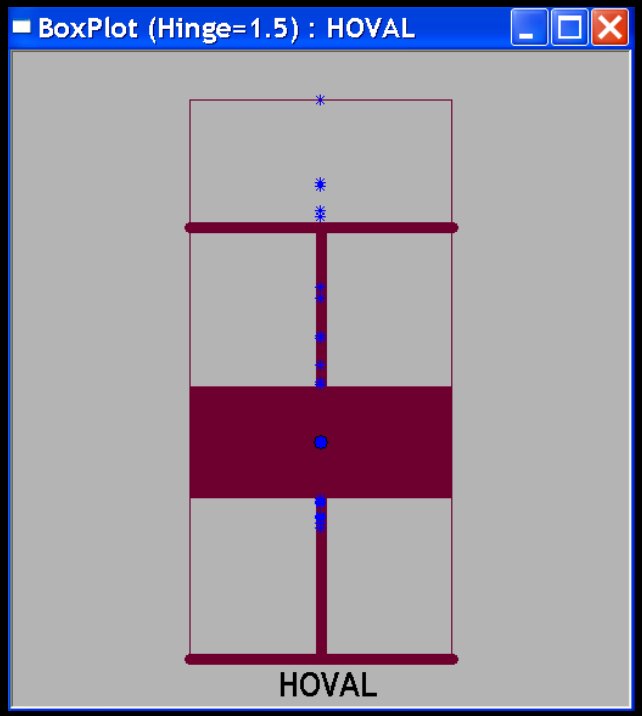


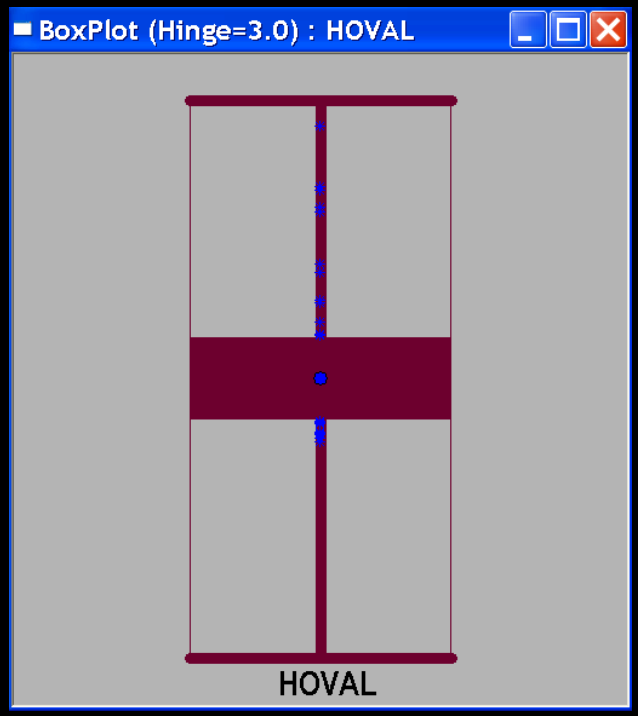
SIDS Rates (79) NC Counties

Percentile Map



Outliers in Box Plot

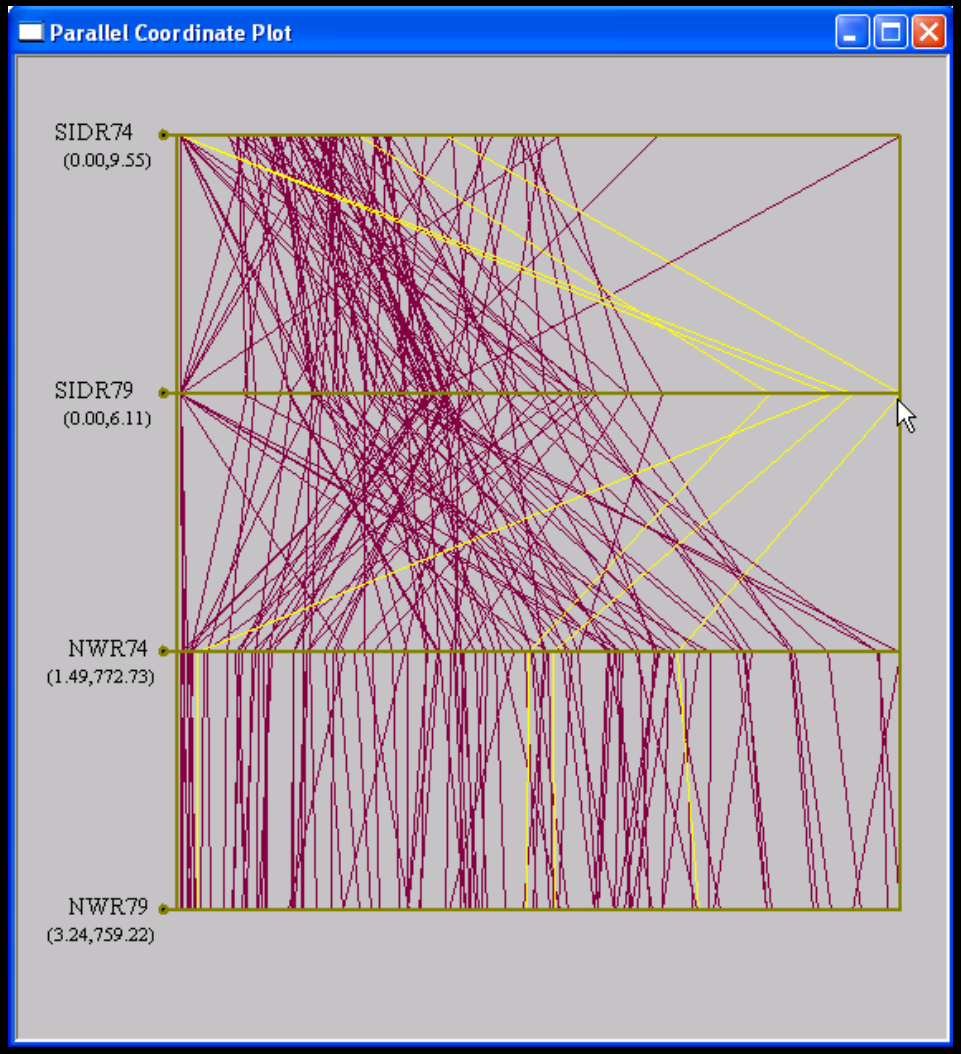




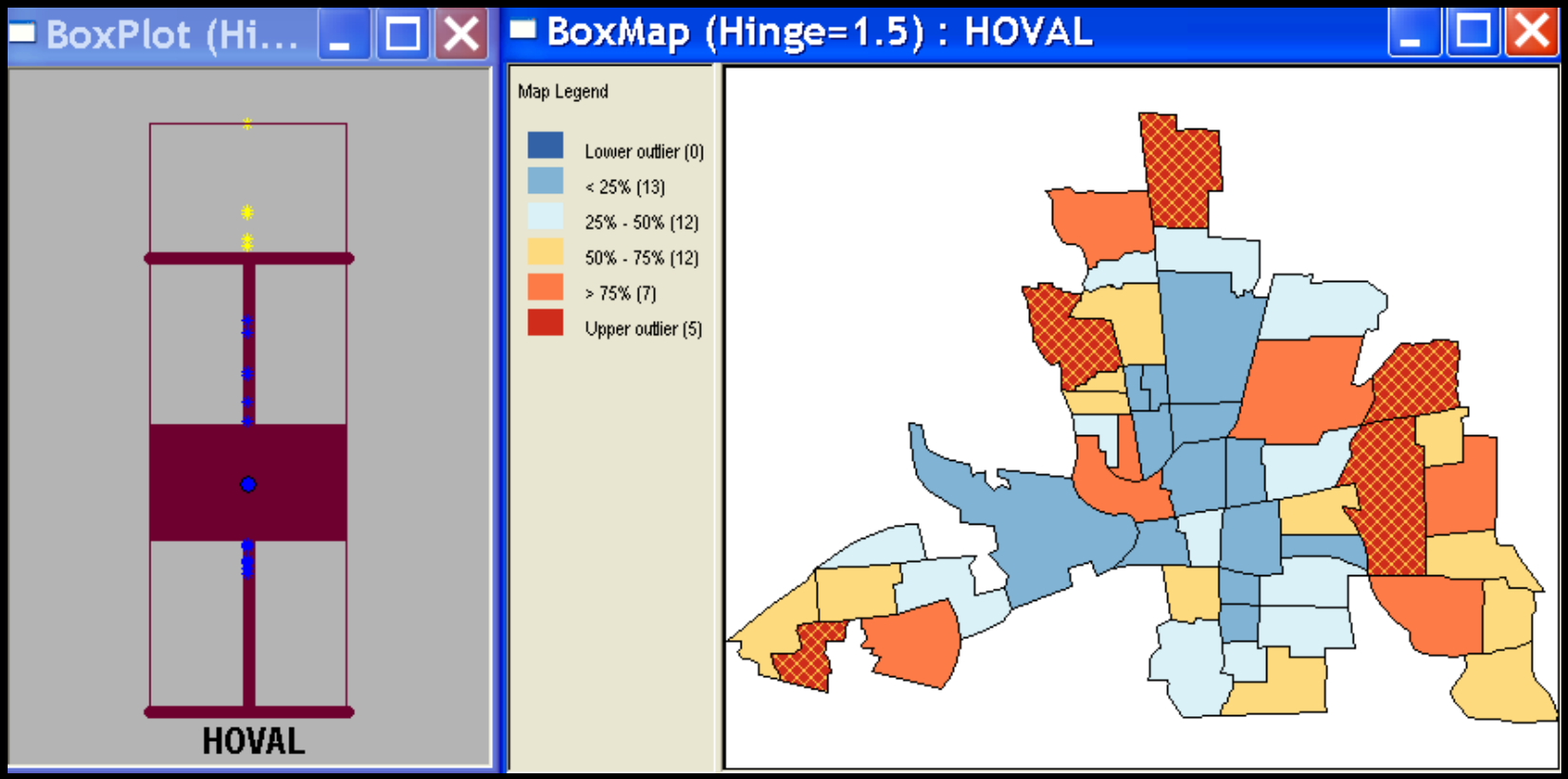
outlier = outside “fence”

fence = Q3 + 1.5 times IQR inter-quartile range 75%-25%)

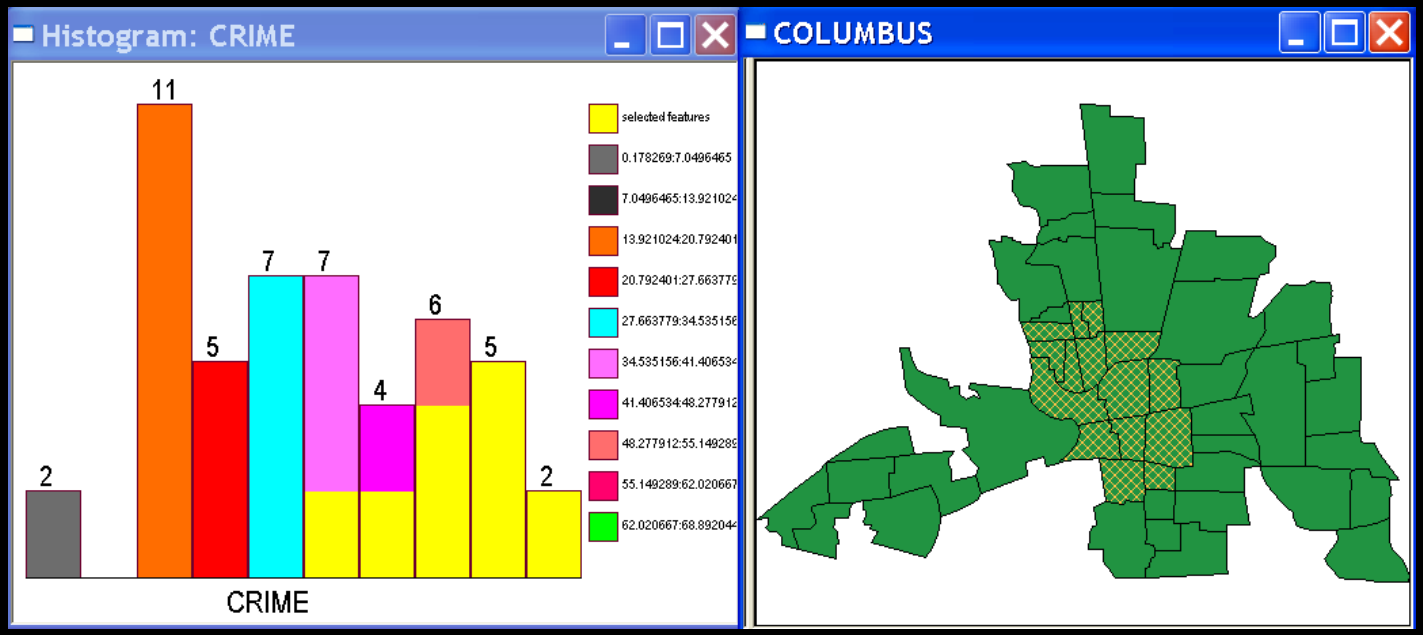
Outliers in Parallel Coordinate Plot



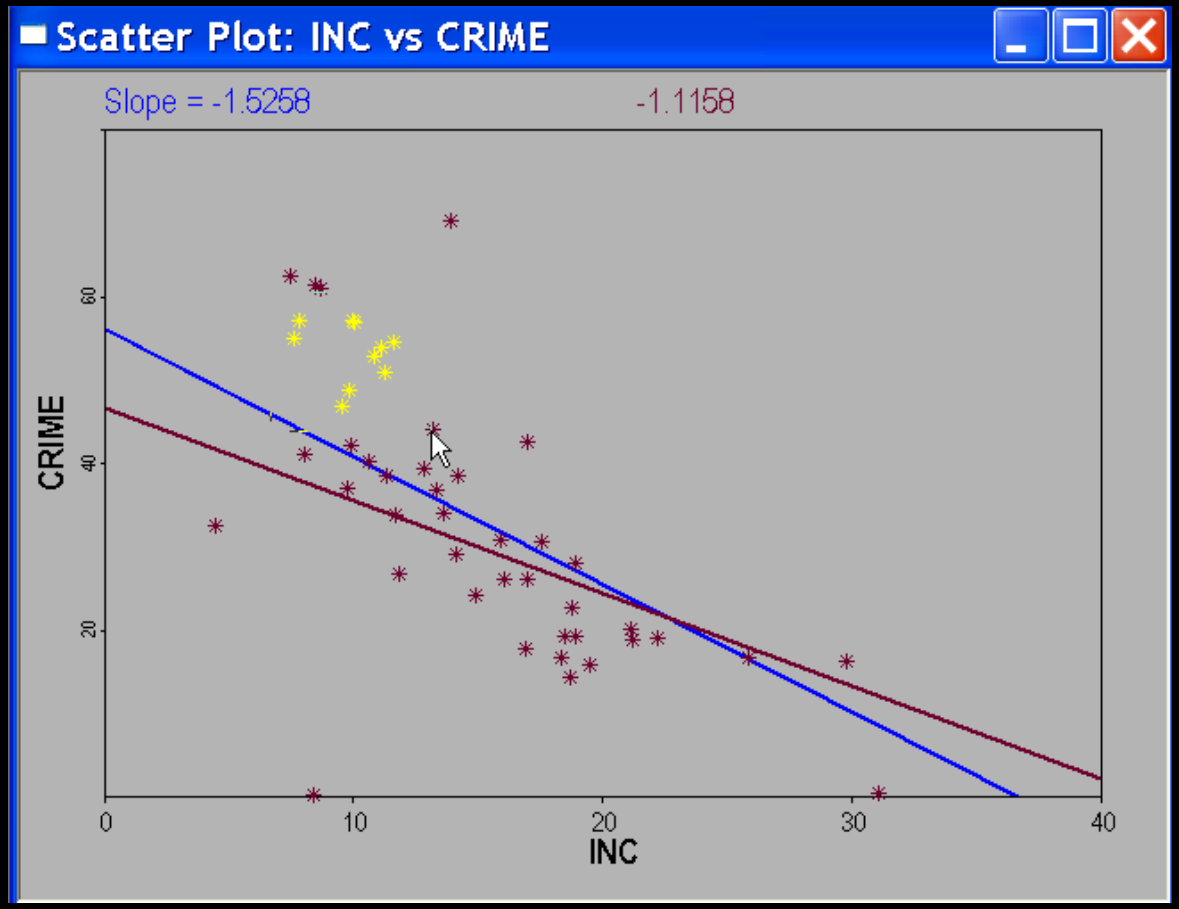
Linked Box Map in GeoDa



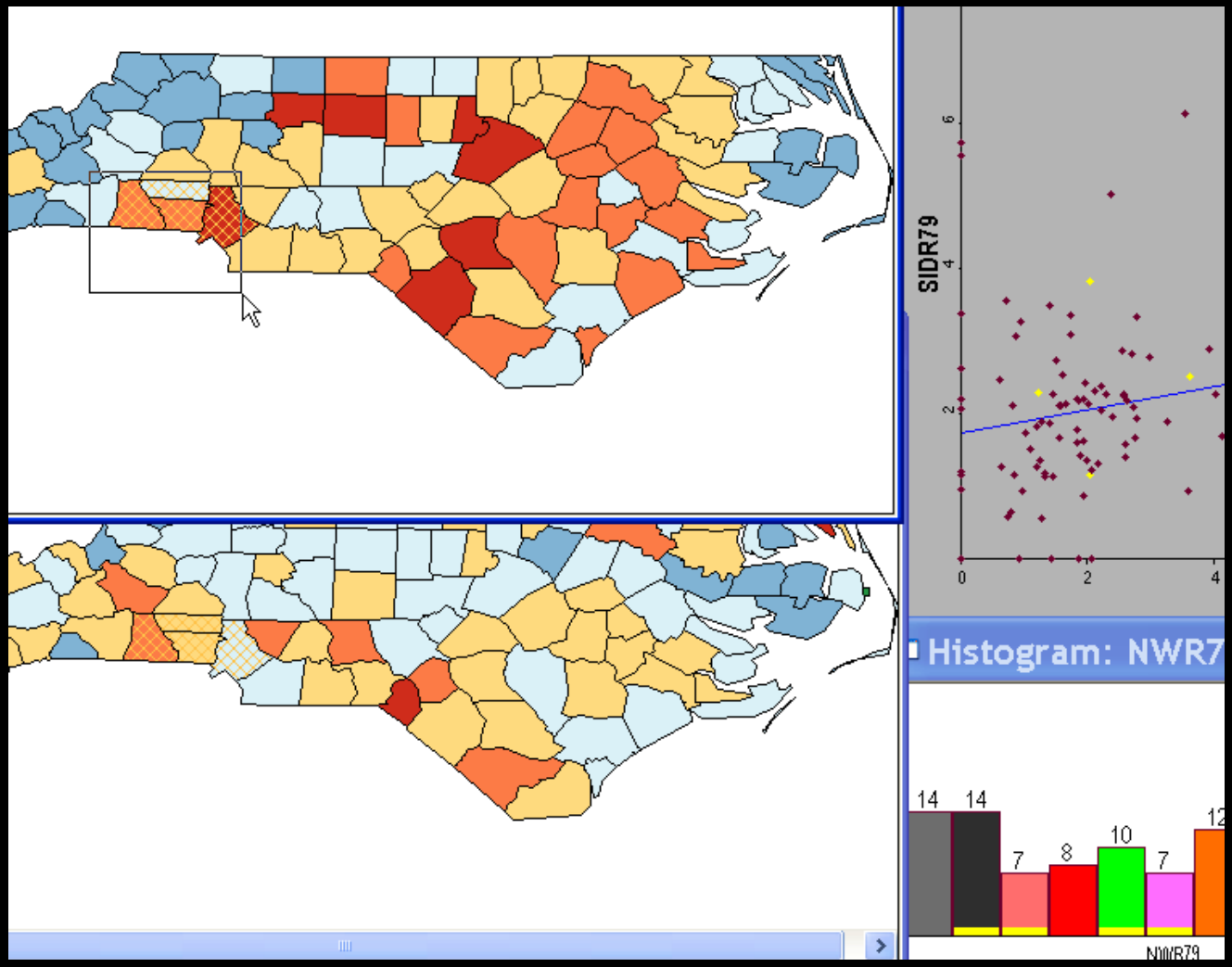
Regional Histogram



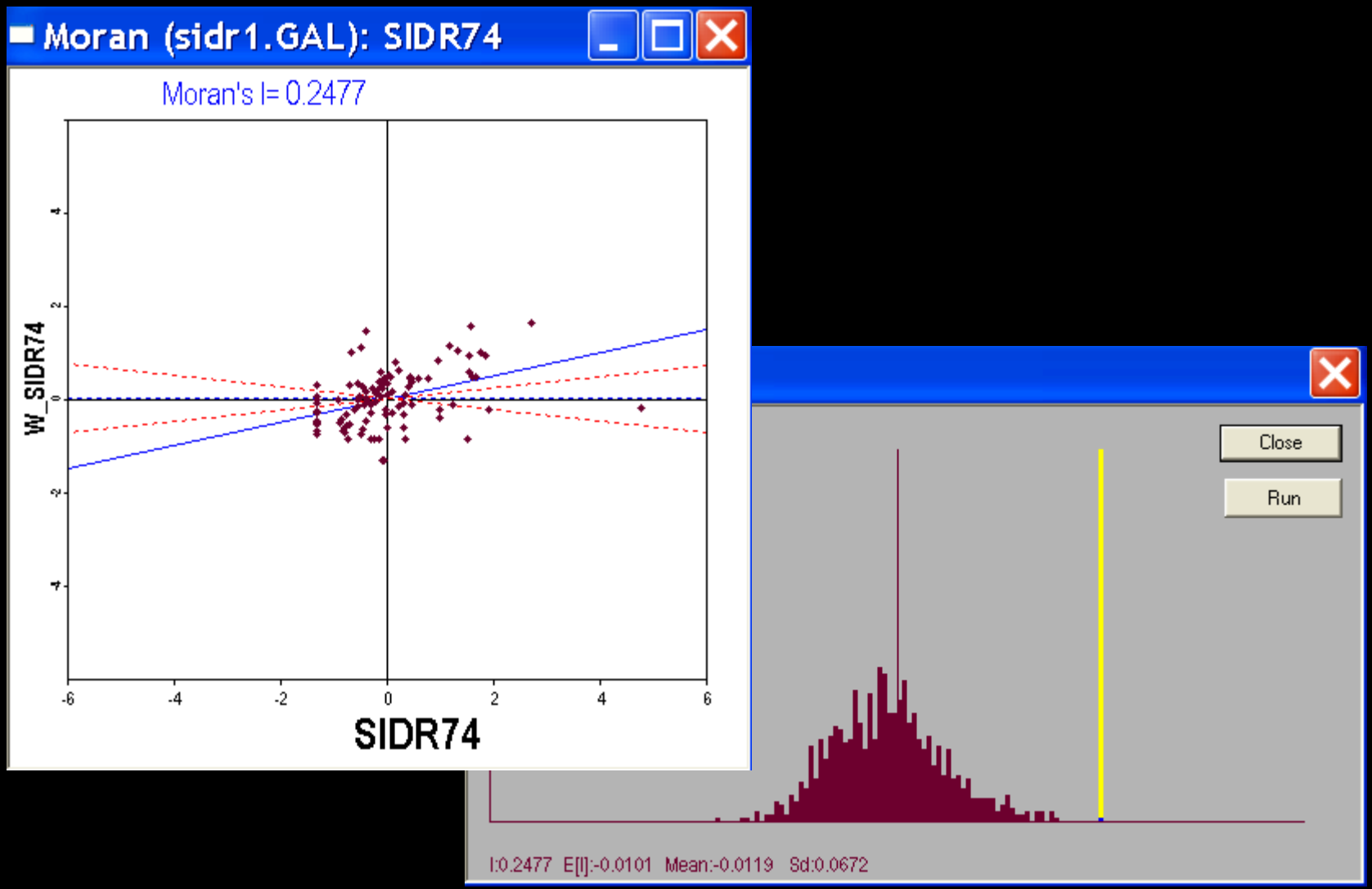
Selection in Scatterplot



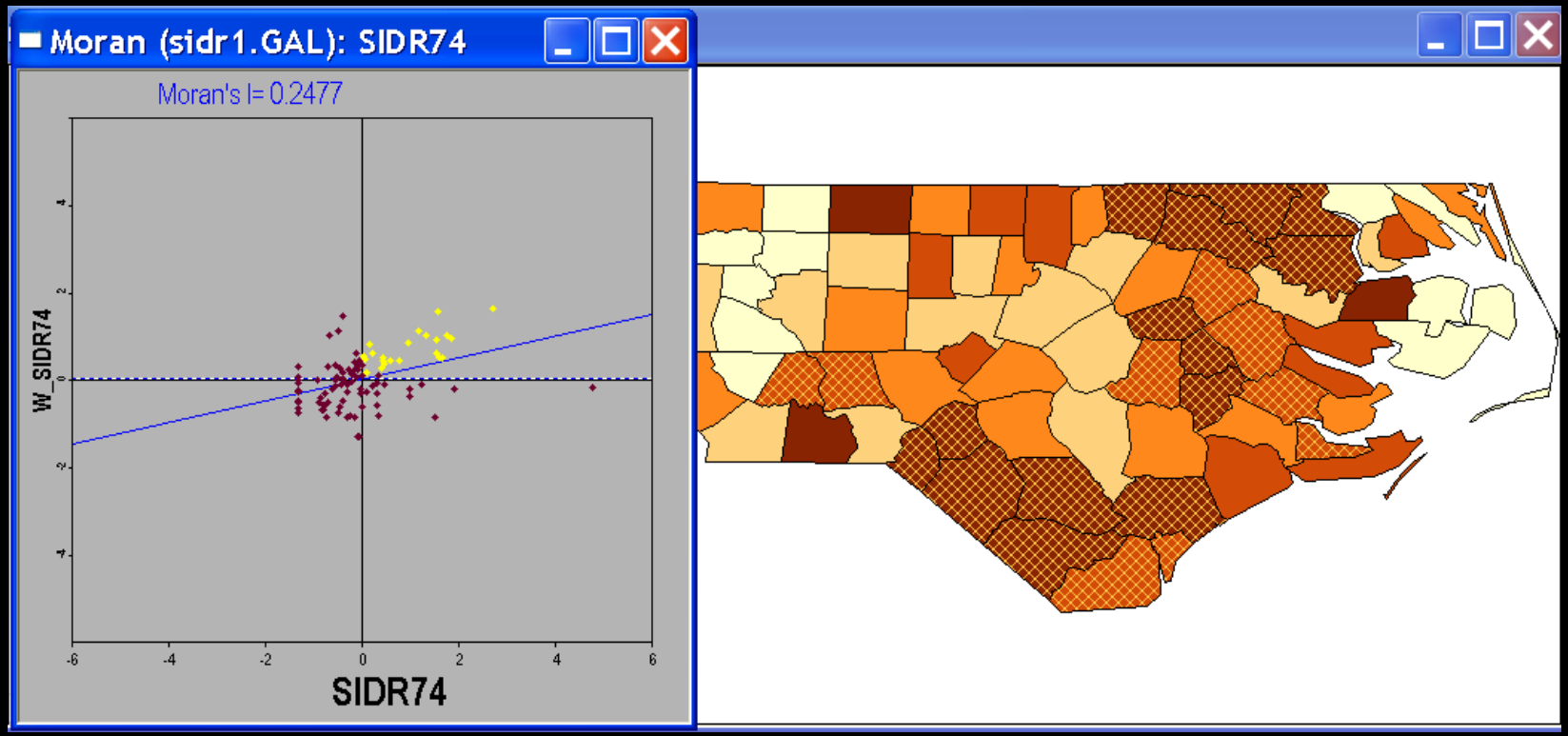
Map Brushing in GeoDa



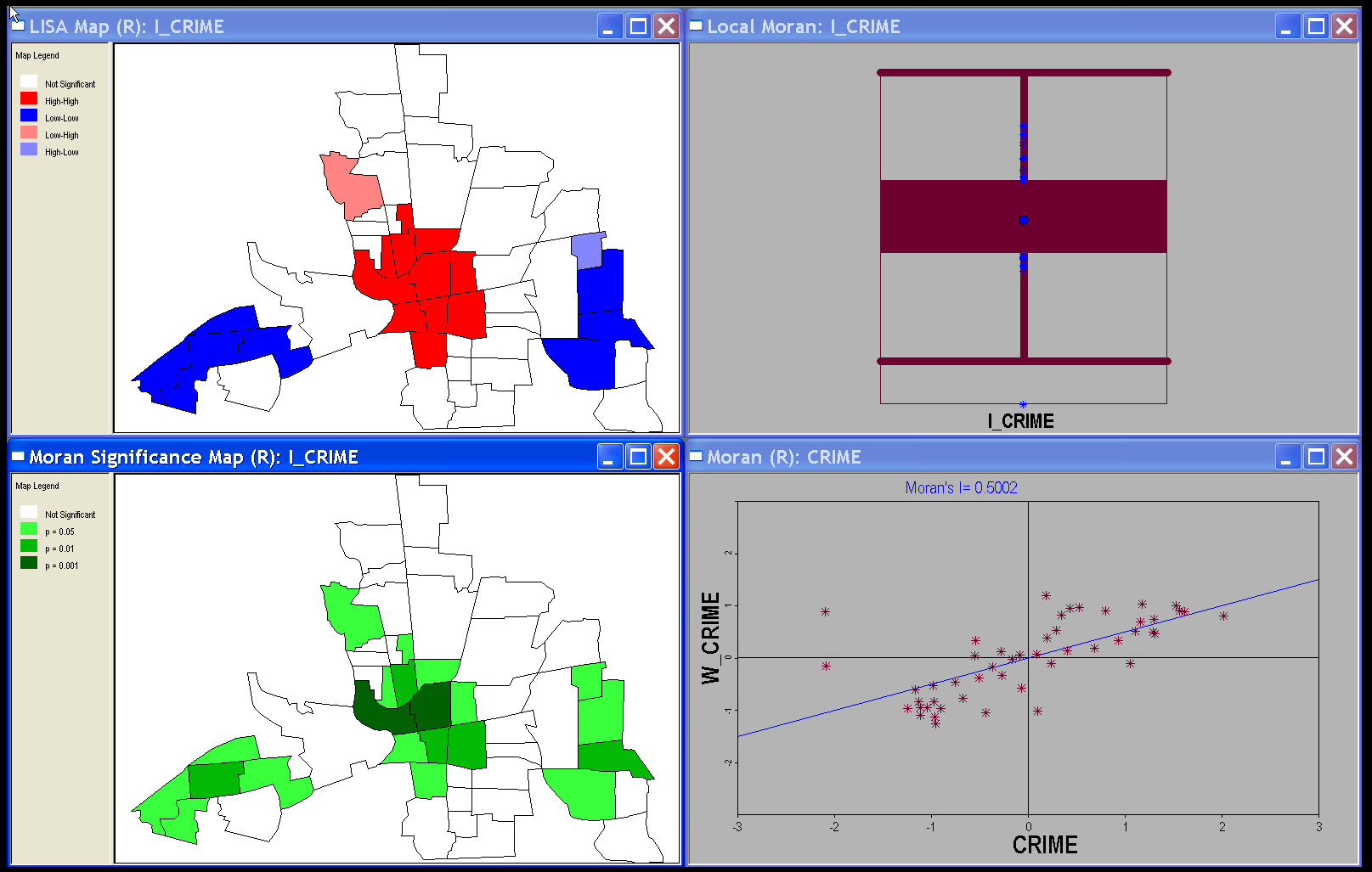
Significance Envelope



Spatial Clusters



LISA MAPS


LISA Map (R): I_CRIME
Map Legend
Not Significant
High-High
Low-Low
Low-High
High-Low
Local Moran: I_CRIME
I_CRIME
Moran Significance Map (R): I_CRIME
Map Legend
Not Significant
p = 0.05
p = 0.01
p = 0.001
Moran (R): CRIME
Moran's I= 0.5002
W_CRIME
CRIME
-3
-2
-1
0
1
2
3

Space-Time Scatterplots