Center for Spatially Integrated Social Science

Mapping and Analysis for Spatial Social Science

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

Outline

- Introduction
- Geovisualization
- Statistical Maps
- Map Smoothing
- Linking and Brushing
- Visualizing Spatial Autocorrelation
- Space-Time Correlation

Introduction

Spatial Models

- Growing Interest in Space, Spatiality and Spatial Interaction among Theoretical Social Sciences
- Overarching Concepts
  - social interaction
  - context, neighborhood effects
  - interacting agents, strategic interaction
  - spatial externalities, agglomeration
  - geography as a proxy

Spatial Data

- Growing Interest in the use of Spatial Data in Empirical Social Science
  - georeferenced data
    - addresses, lat-lon (GPS survey)
  - distance and accessibility measures
    - access to infrastructure, spatial mismatch
- Role of GIS
  - affordable and transparent spatial data manipulation

Geographic Information Systems

- GIS as a Set of Tools
  - Burrough: "set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes"
  - a GIS, GISes (= systems)
- GIS as Science (the "new" geography)
  - Goodchild: Geographic Information Science
    - generic scientific questions pertaining to geographic data
    - central role of spatial analysis
  - GIScience

What is Spatial Analysis

- From Data to Information
  - beyond mapping: added value
  - transformations, manipulations and application of analytical methods to spatial (geographic) data
- Lack of Locational Invariance
  - analyses where the outcome changes when the location of the objects under study changes
    - median center, clusters, spatial autocorrelation

spatial analysis avant la lettre Dr. Snow's map of cholera deaths in London

Components of Spatial Analysis

- Visualization
  - Showing interesting patterns
- Exploratory Spatial Data Analysis
  - Finding interesting patterns
- Spatial Modeling, Regression
  - Explaining interesting patterns

Implementation of Spatial Analysis

- Beyond GIS
  - Analytical functionality not part of typical commercial GIS
    - Analytical extensions, DynESDA2
  - Exploration requires interactive approach
  - Spatial modeling requires specialized statistical methods
    - Explicit treatment of spatial autocorrelation
    - Space-time is not space + time
- Methods of Geovisualization, ESDA and Spatial Regression Analysis

(Limited) Illustration of Techniques

- Geovisualization
  - specialized maps
- ESDA
  - dynamically linked windows
- Spatial Correlation Analysis
  - global and local spatial autocorrelation
  - space-time correlation

Geovisualization

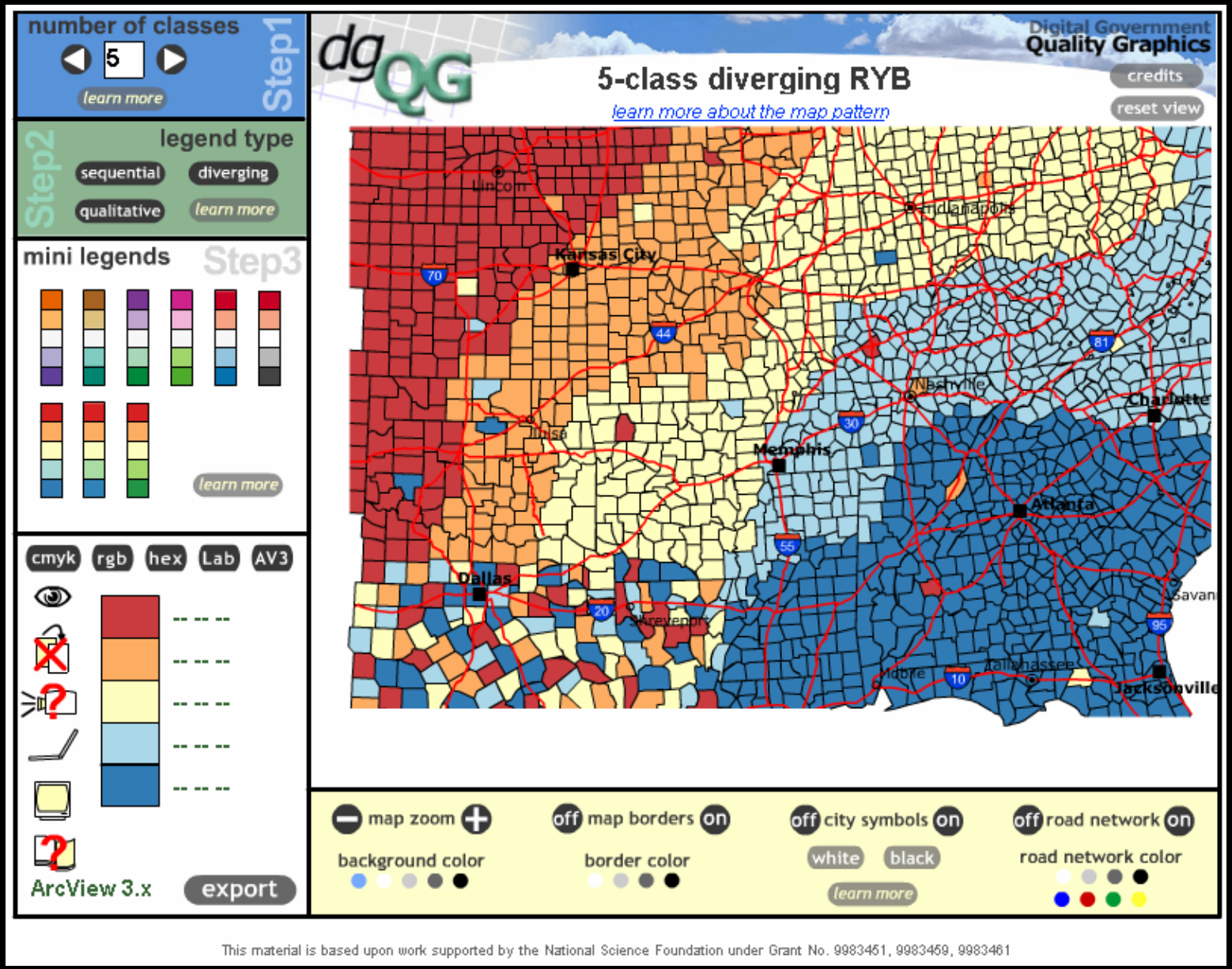
Beyond Mapping

- Map
  - "a collection of spatially defined objects" (Monmonier)
- Geovisualization
  - combination of map and scientific visualization methods (computer science)
  - exploit human's pattern recognition abilities
- How to lie with maps
  - many design issues
  - human perception can be tricked

Choropleth Maps

- Map Counterpart of Histogram
  - values/attributes for discrete spatial units
  - choro from choros (colors) NOT chloro
- Practical Issues
  - choice of intervals
    - equal interval, equal share (quantiles), standard deviational, ...
  - choice of colors
    - important for perception of patterns
  - misleading role of area
    - larger areas “seem” more important

Color Brewer www.colorbrewer.org

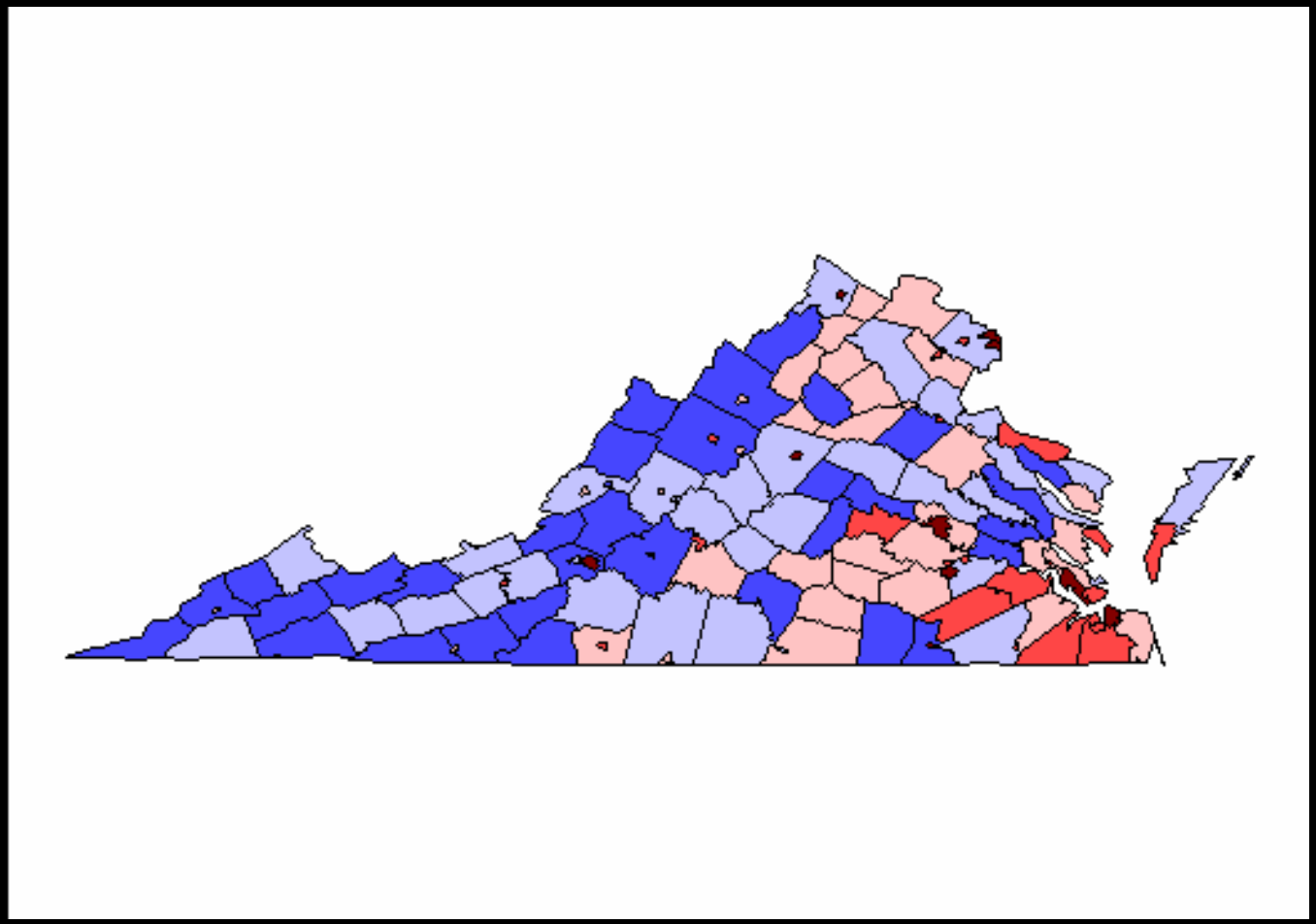


This material is based upon work supported by the National Science Foundation under Grant No. 9983451, 9983459, 9983461

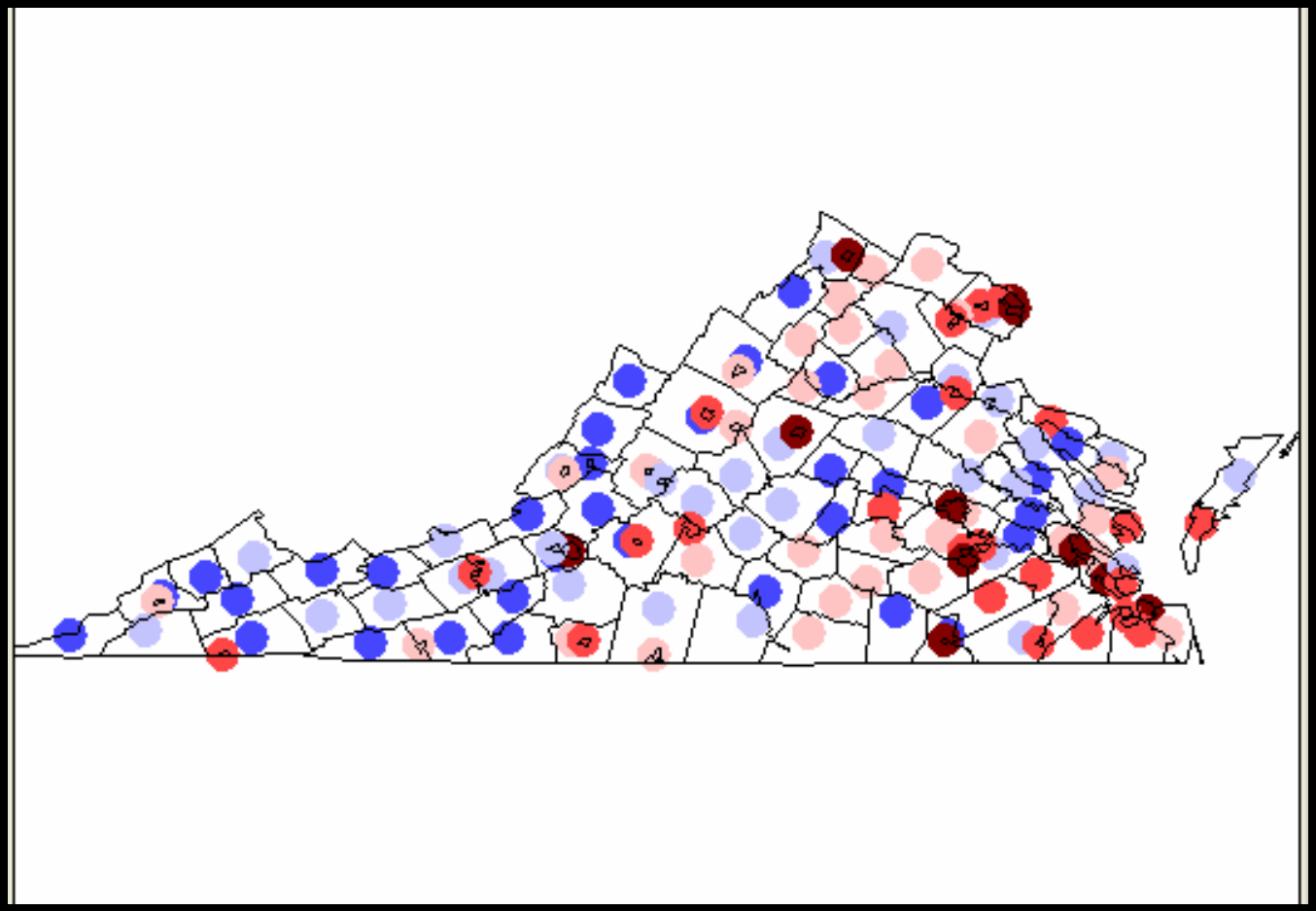
Cartograms

- Misleading Effect of Area
  - large areal units draw attention
- Symbol Maps
  - symbols (bars, circles) superimposed on actual areas
- Cartogram
  - change the layout to reflect size other than area
    - population size, variable magnitude
  - respect topology (spatial arrangement)

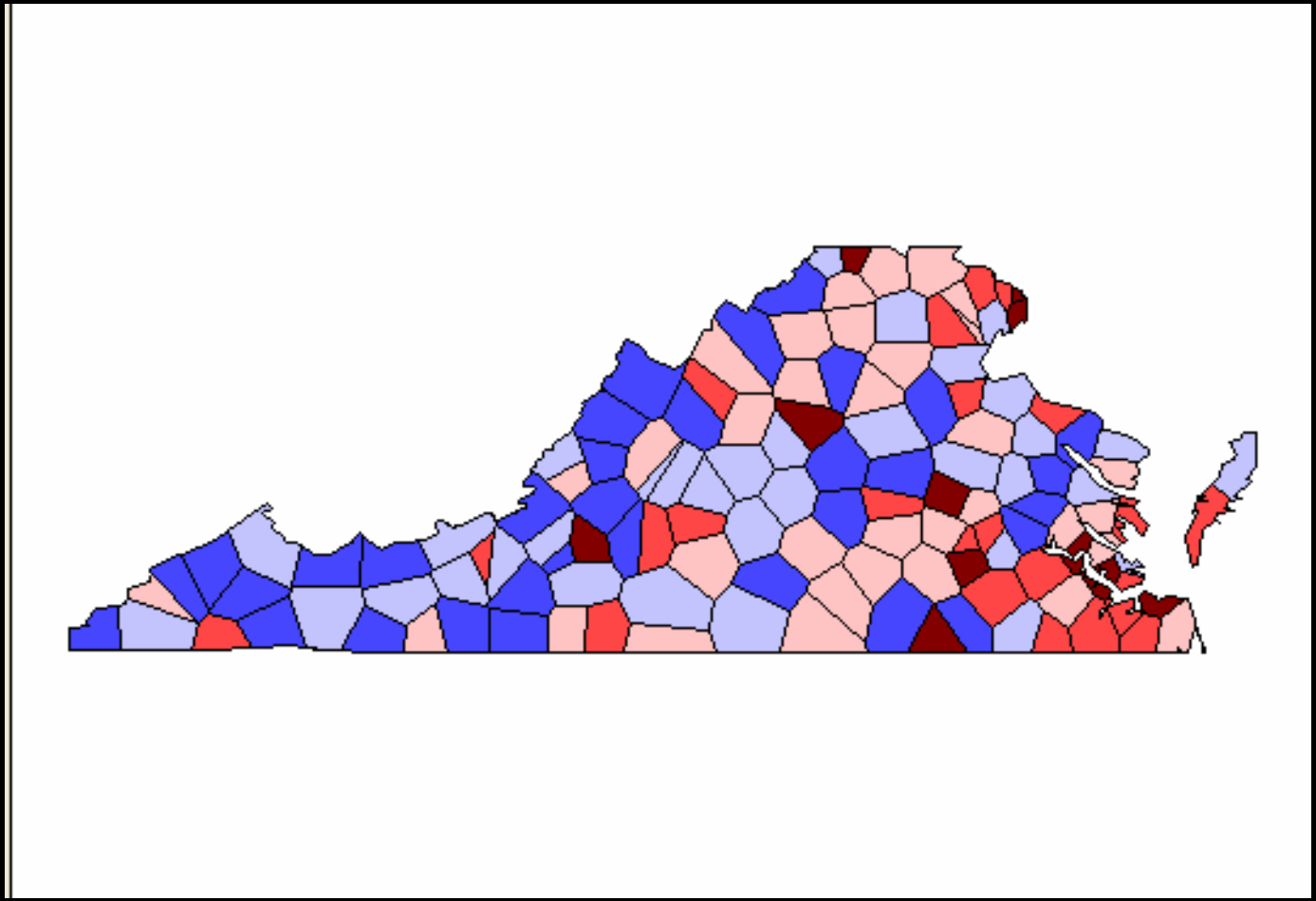
Choropleth Map (Juvenile Crime VA)



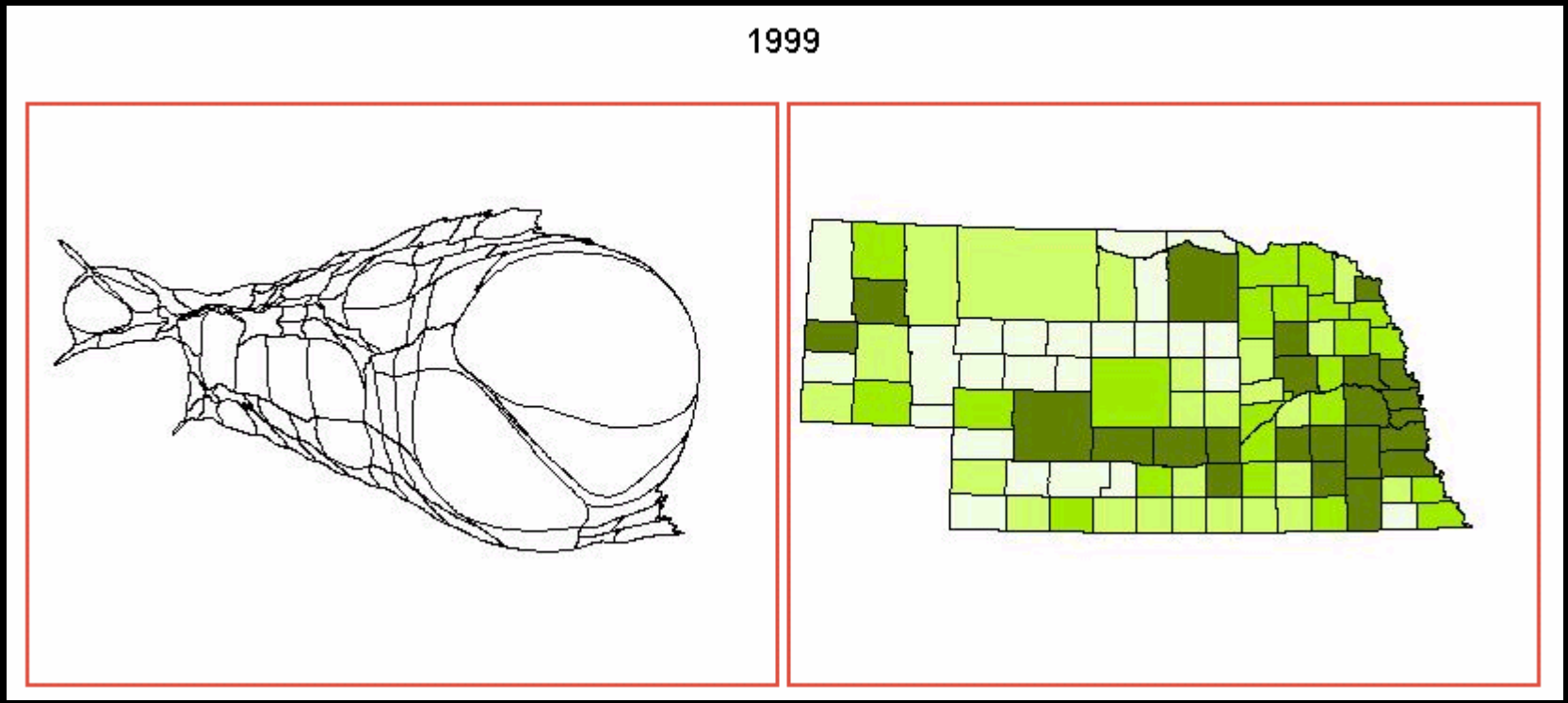
Point Map



Thiessen Polygons

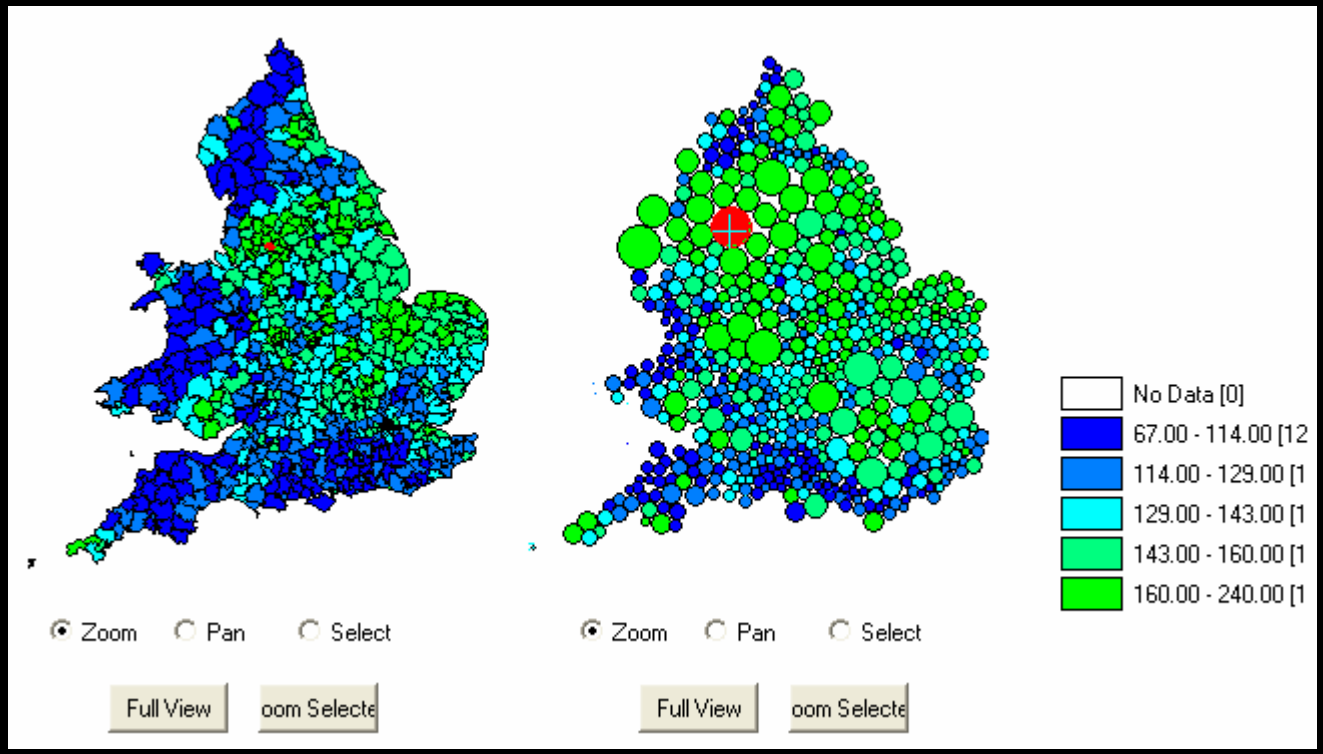


Contiguous Cartogram



Nebraska county population - http://www.bbr.unl.edu/cartograms/pop.html

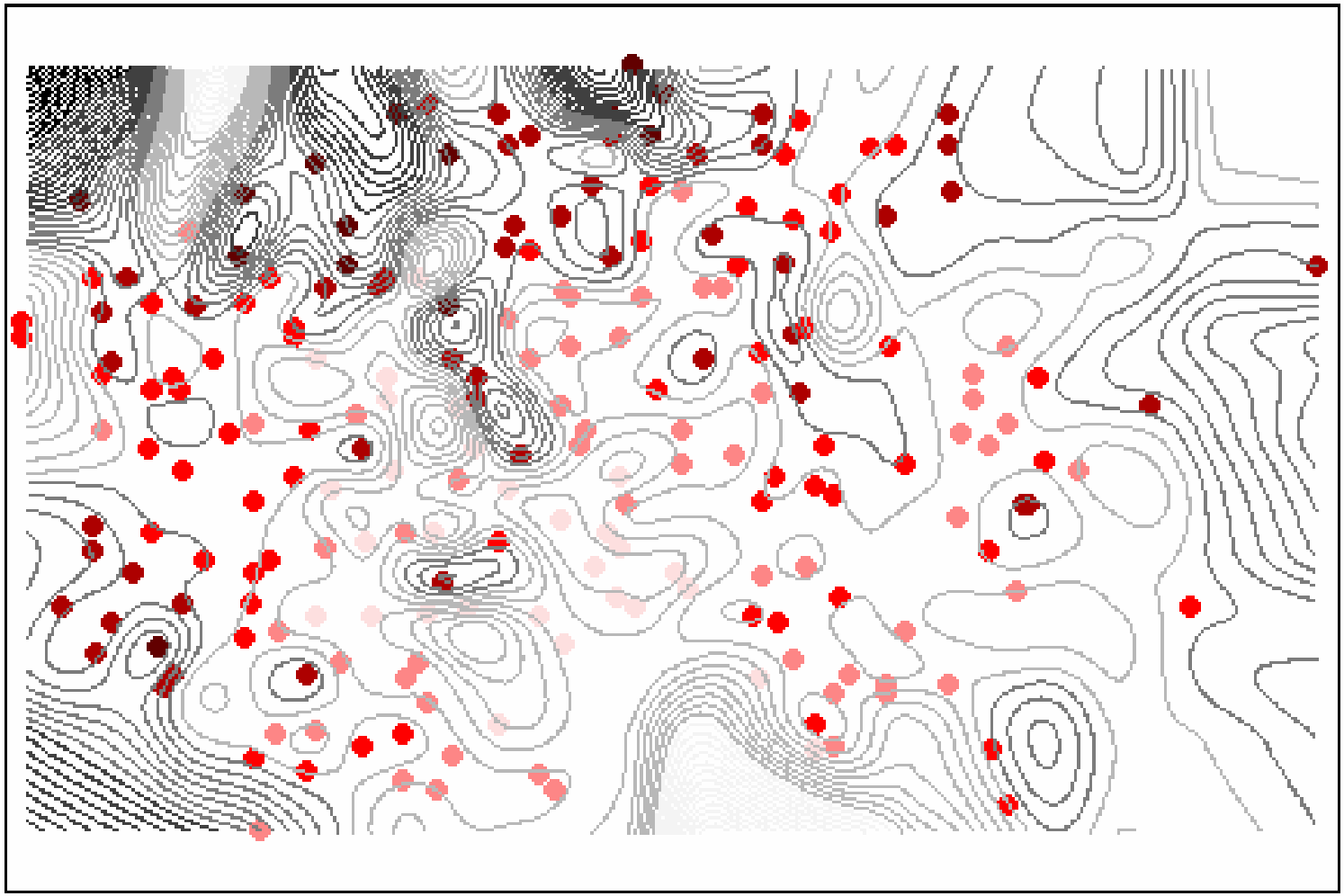
Cartogram (GeoTools)



Infant Mortality England and Wales, 1851

Isopleth Maps

- Map Counterpart of Density Plot
  - values/attributes for continuous fields
  - lines of equal value, contour lines
  - 3-d surface plots
- Practical Issues
  - choice of intervals
  - spatial interpolation
    - construct “observations” for locations that are not observed
    - statistical problem = spatial prediction

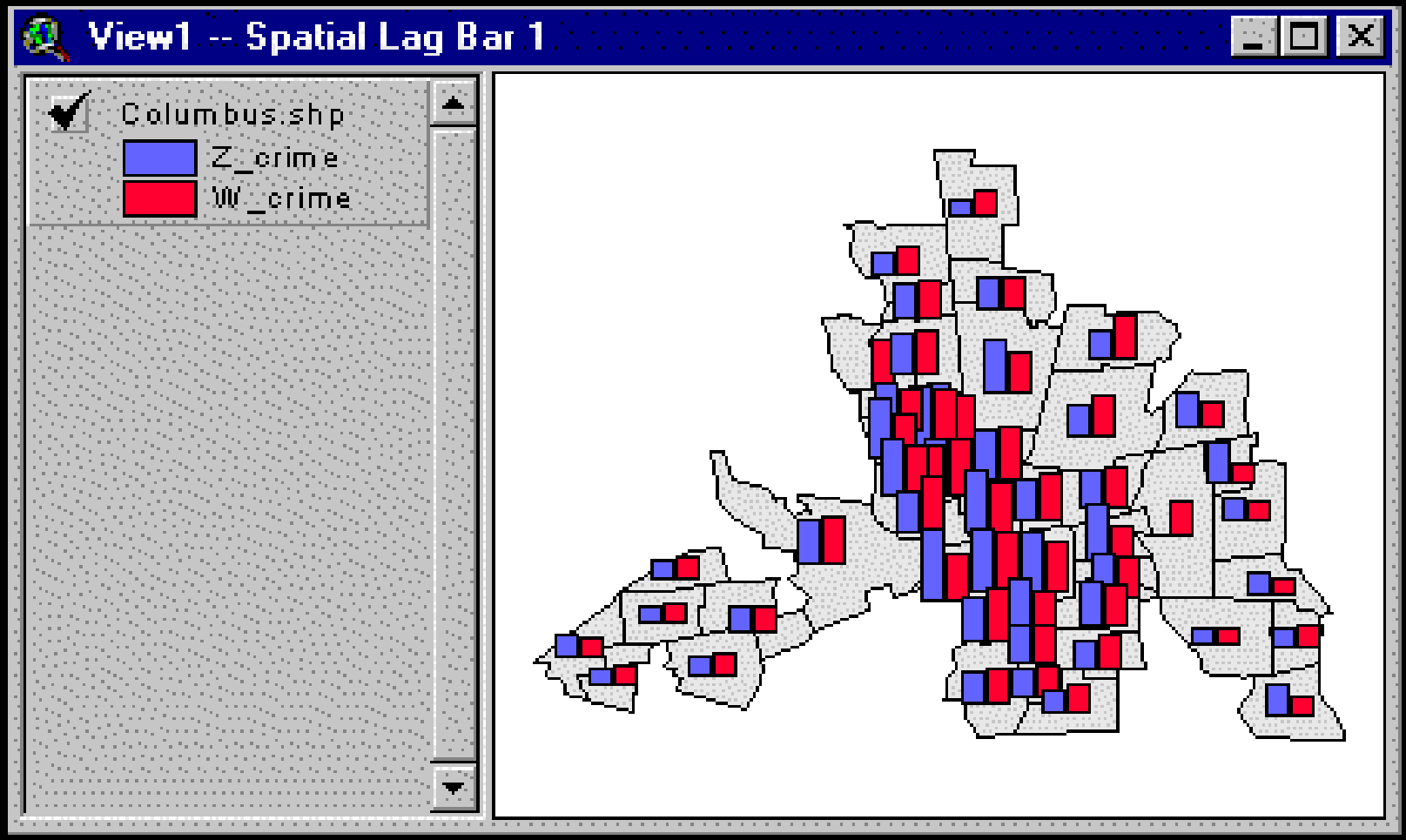


Residential Sales Price, Baltimore MD (1980)
sample points (darker is higher) and contours

Statistical Maps

Visualizing Spatial Distributions

- Spatialized EDA
  - icons and glyphs matching locations
  - special case of symbol maps
- Box Map
  - outlier map
  - visual popout, both magnitude and location
- Regional Box Plots
  - spatial heterogeneity
  - different distributions in spatial subsets



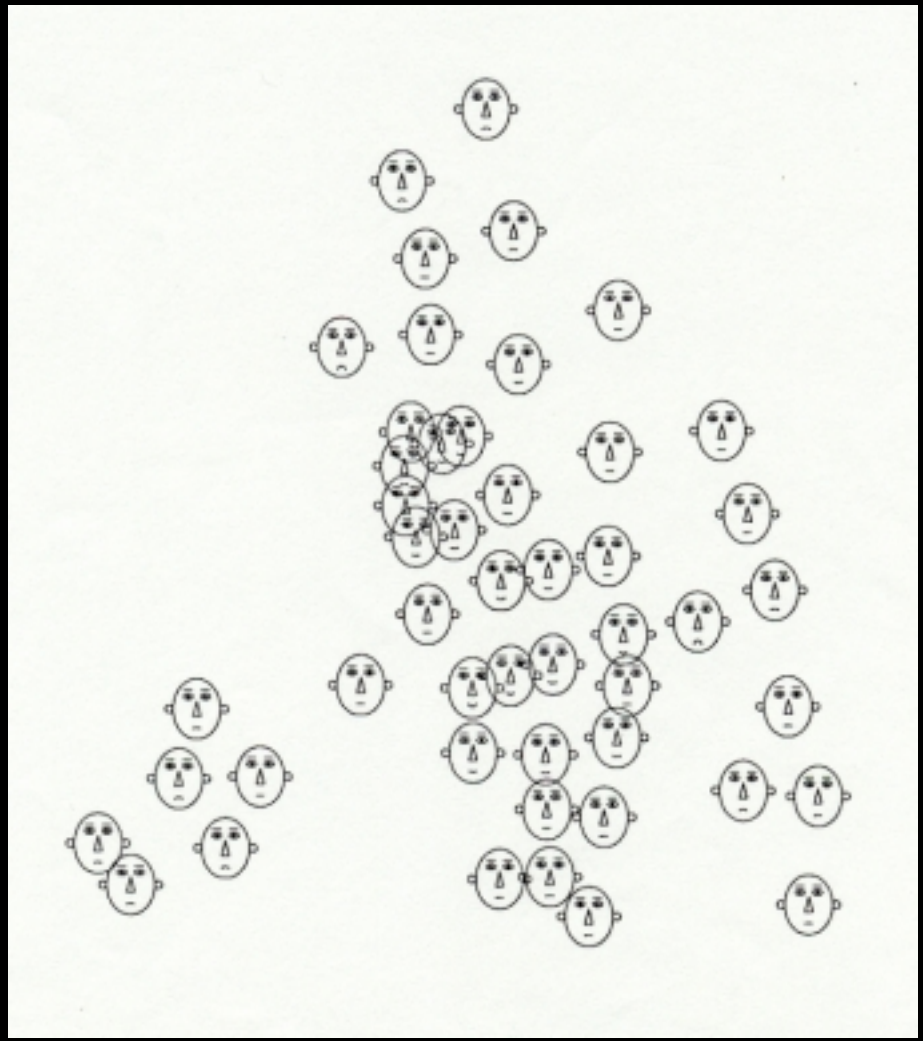
spatial lag bar chart

blue = crime at i, red = spatial lag, average crime for neighbors

Spatialized EDA

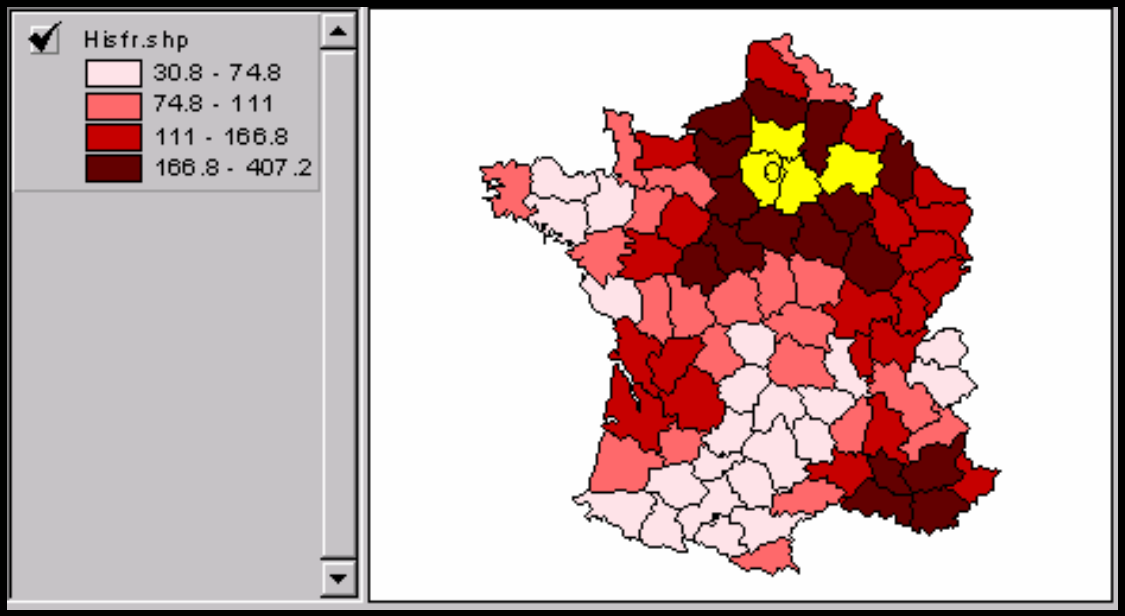
Spatial Chernoff Faces

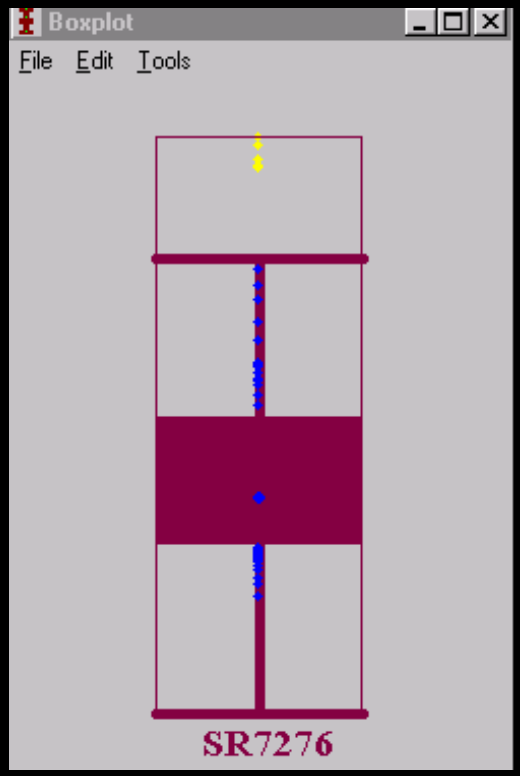
the burglar's view
of crime clusters in
Columbus



Box Map

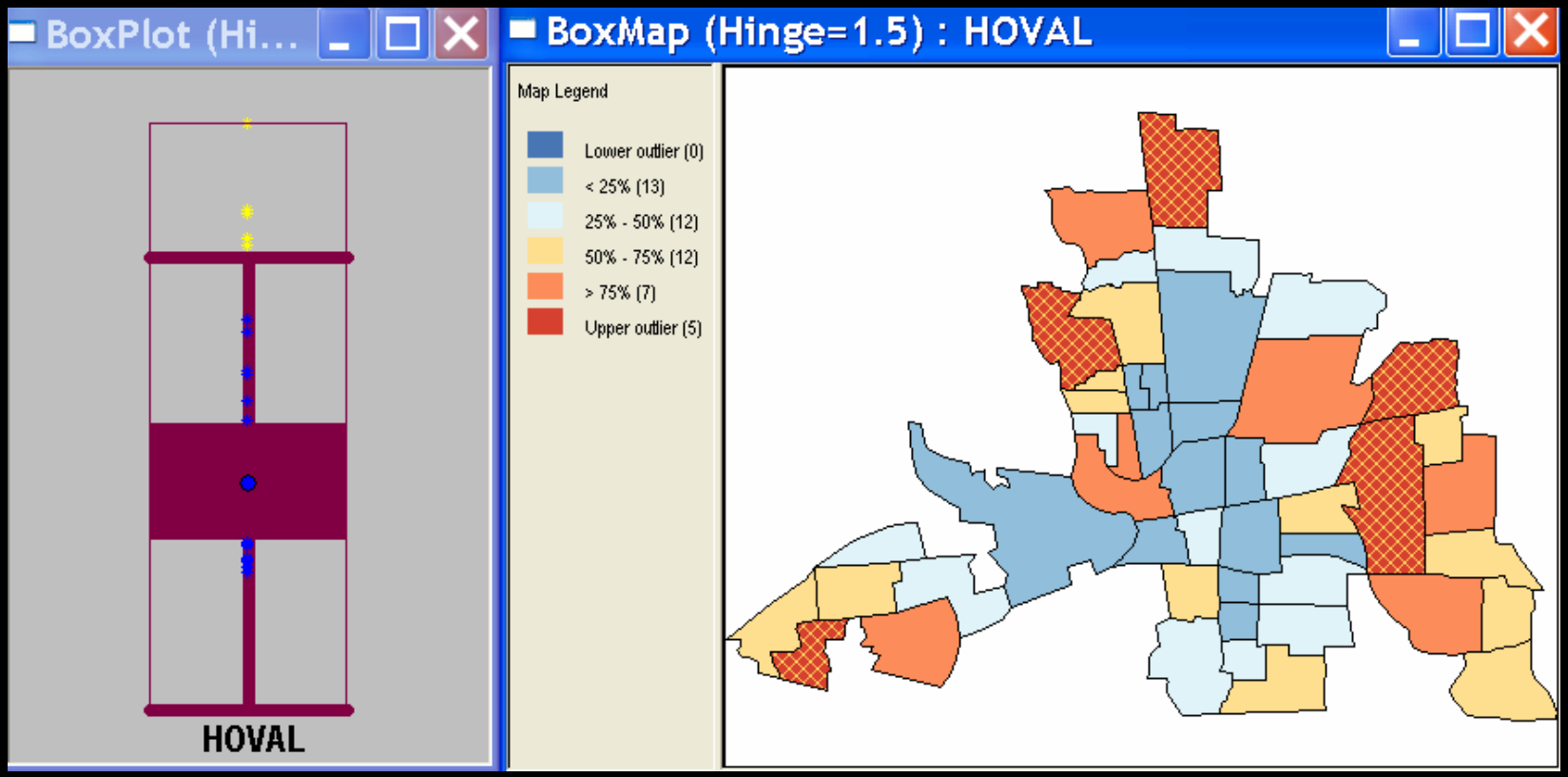
- quartile map with outliers highlighted



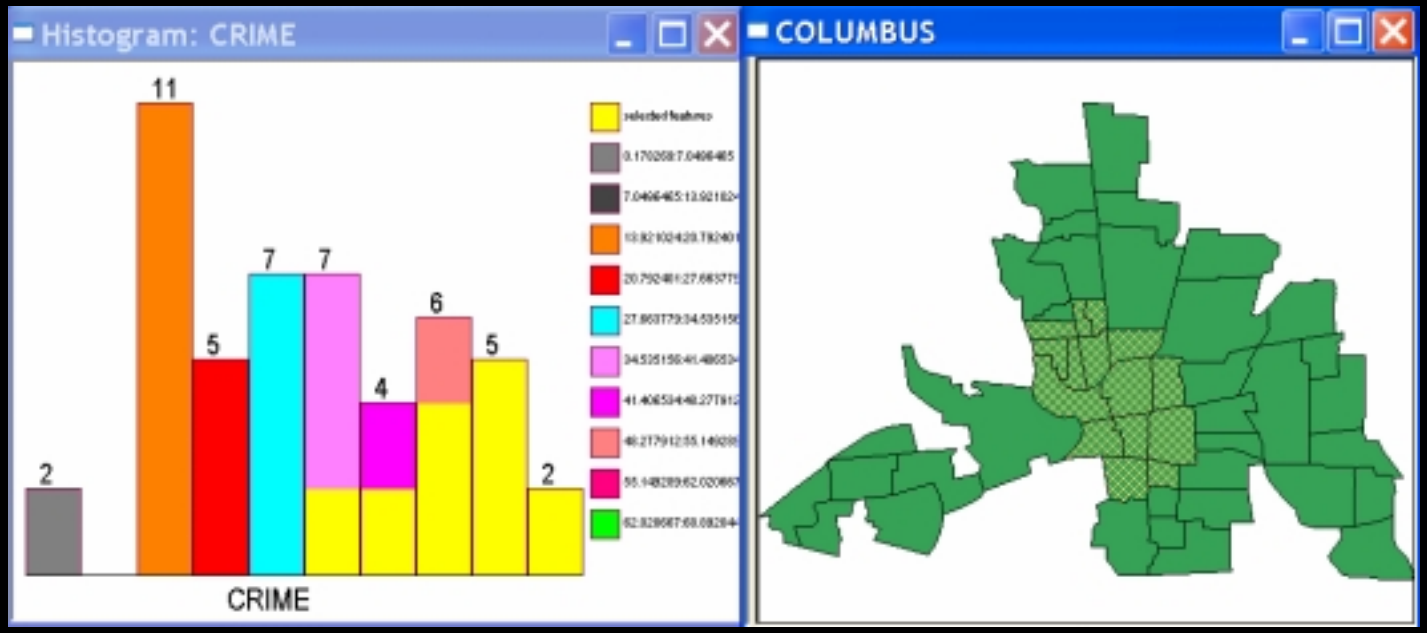


suicide rates in France (Durkheim 1897)

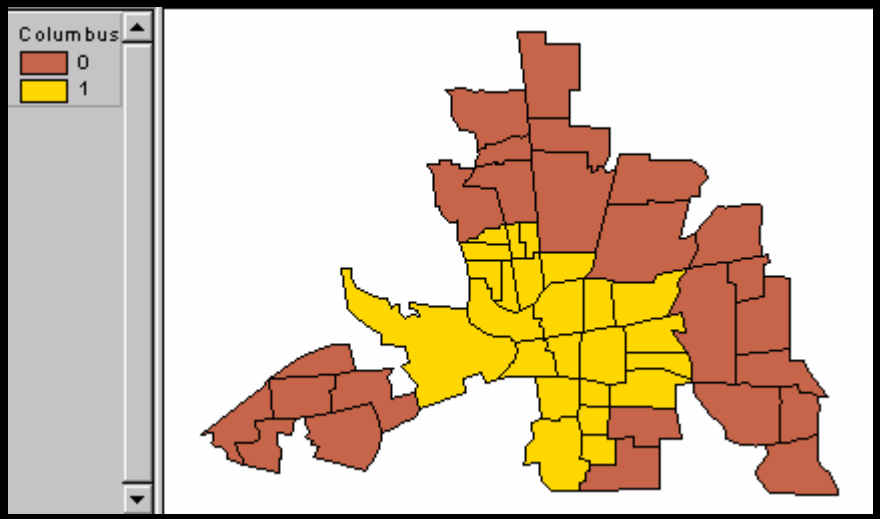
Linked Box Map in DynESDA2

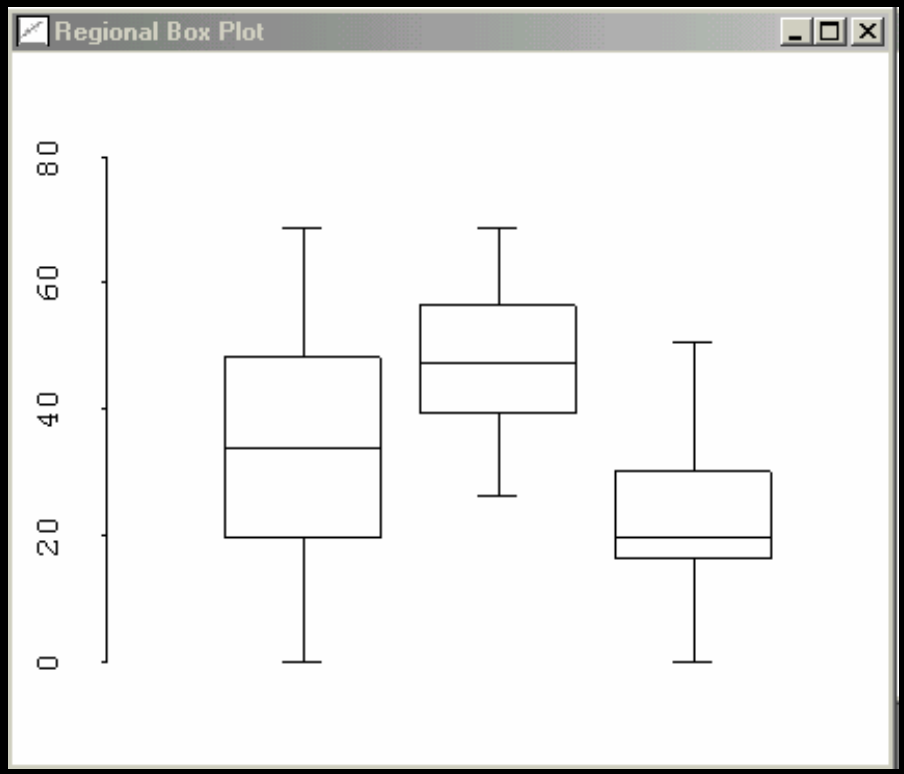


Regional Histogram



Regional Box Plot



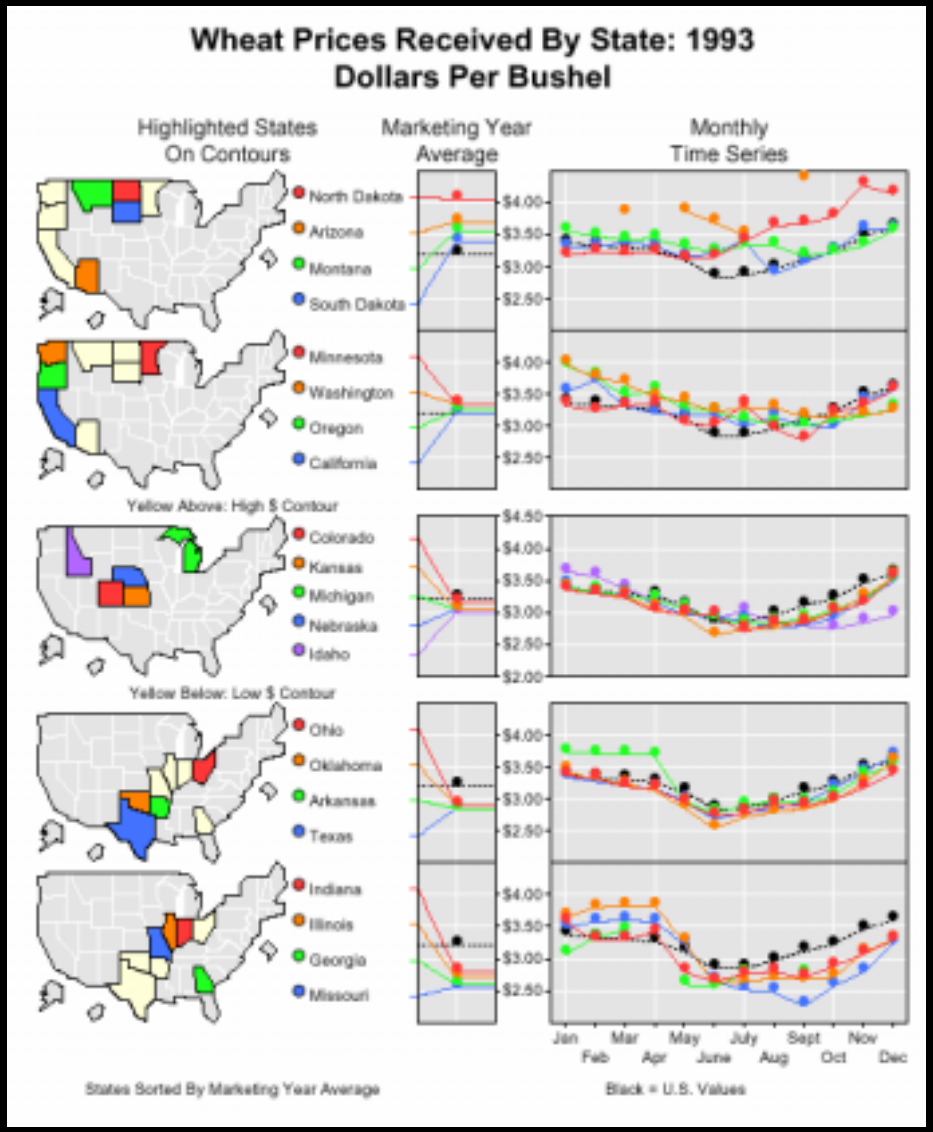


Columbus crime: core vs periphery

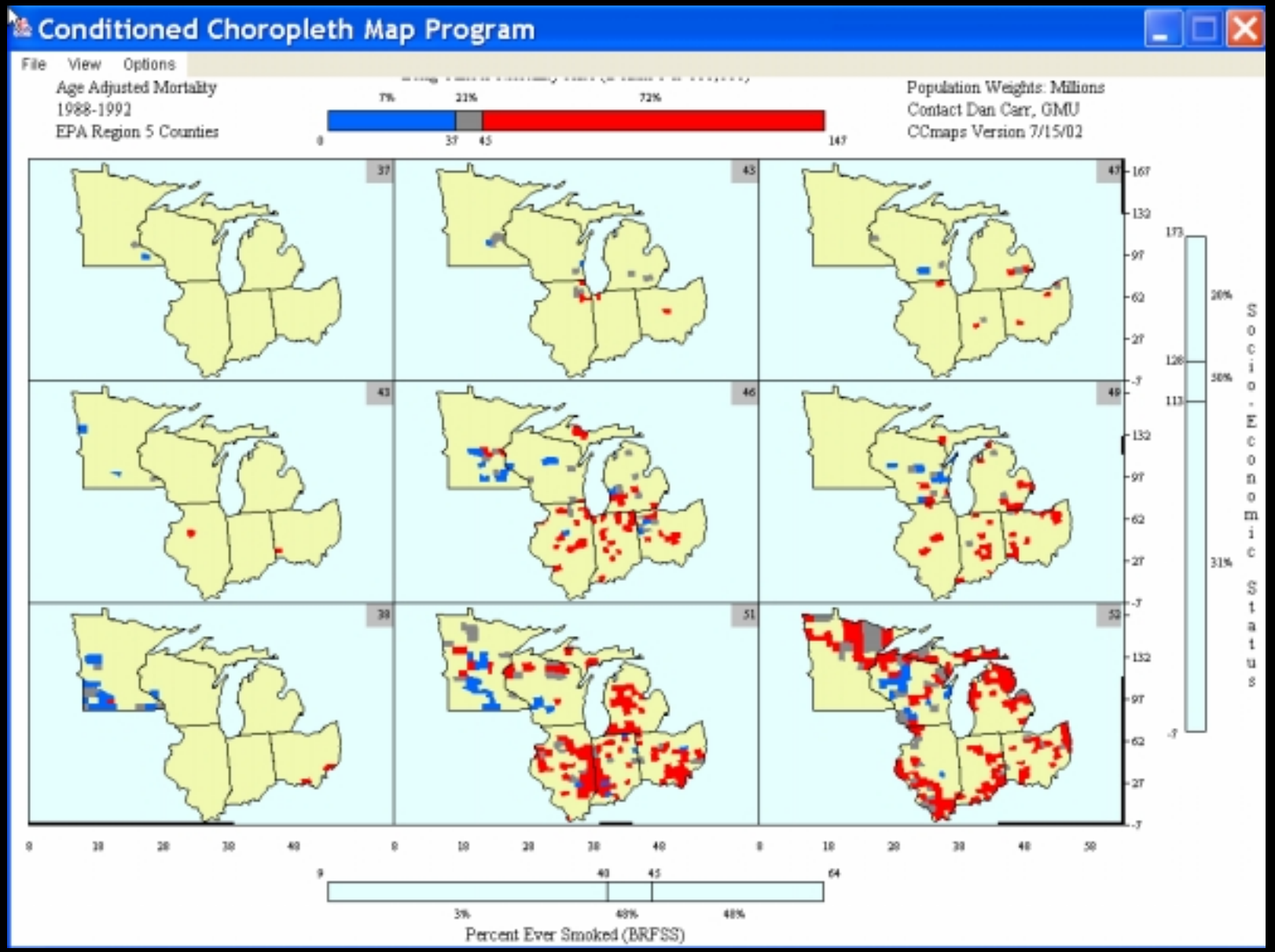
Map Plots and Plot Maps

- Linked Micromap Plots - LM plots
  - a micromap for each quantile
  - micromaps linked to other statistical graphs
- Conditioned Choropleth Maps - cc maps
  - choropleth maps on dependent variable
  - micromap matrix
  - conditioning along two dimensions

Linked Micromap Plots (Carr)



Conditioned Choropleth Map (Carr)



Map Smoothing

Mapping Events

- Events as Locations
  - individual points
    - point pattern analysis
- Events as Rates
  - areal aggregates
    - counts of events
    - rate = # events / # population at risk
    - raw rate is ML estimate of “risk”

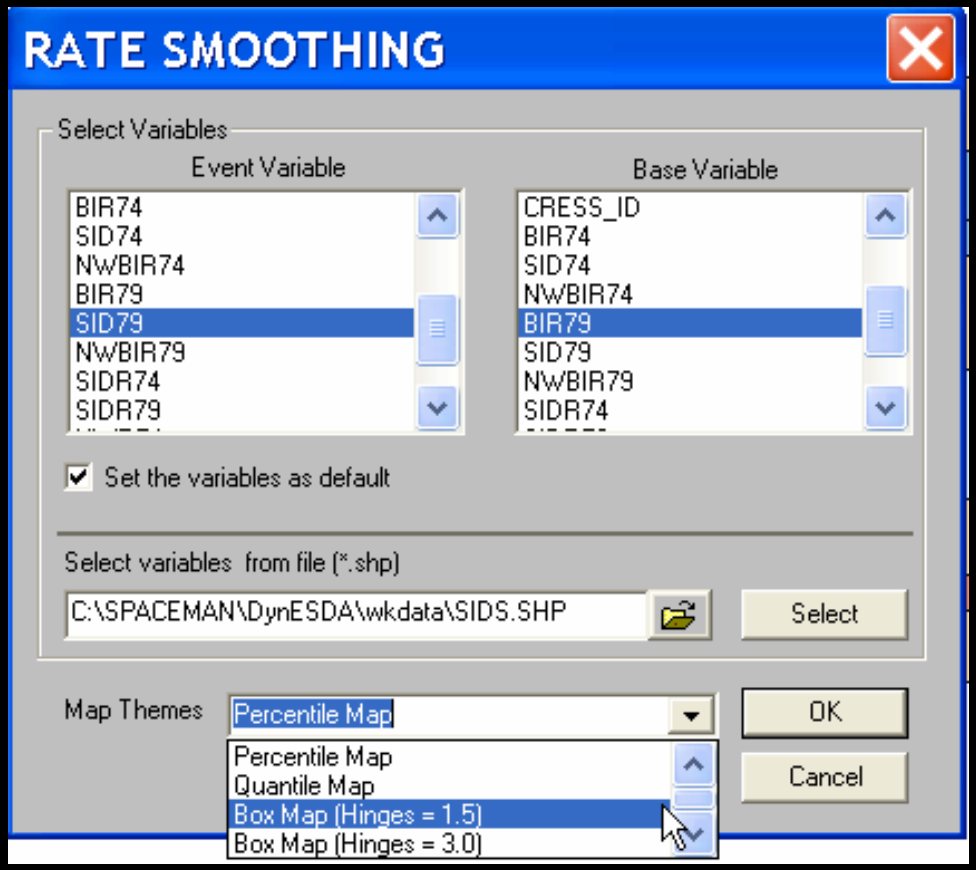
Problems with Rate Maps

- Intrinsic Heterogeneity
  - variance depends on mean
  - variance depends on base
- Variance Instability
  - spurious outliers
- Excess Risk is Non-Spatial
  - does not account for spatial autocorrelation

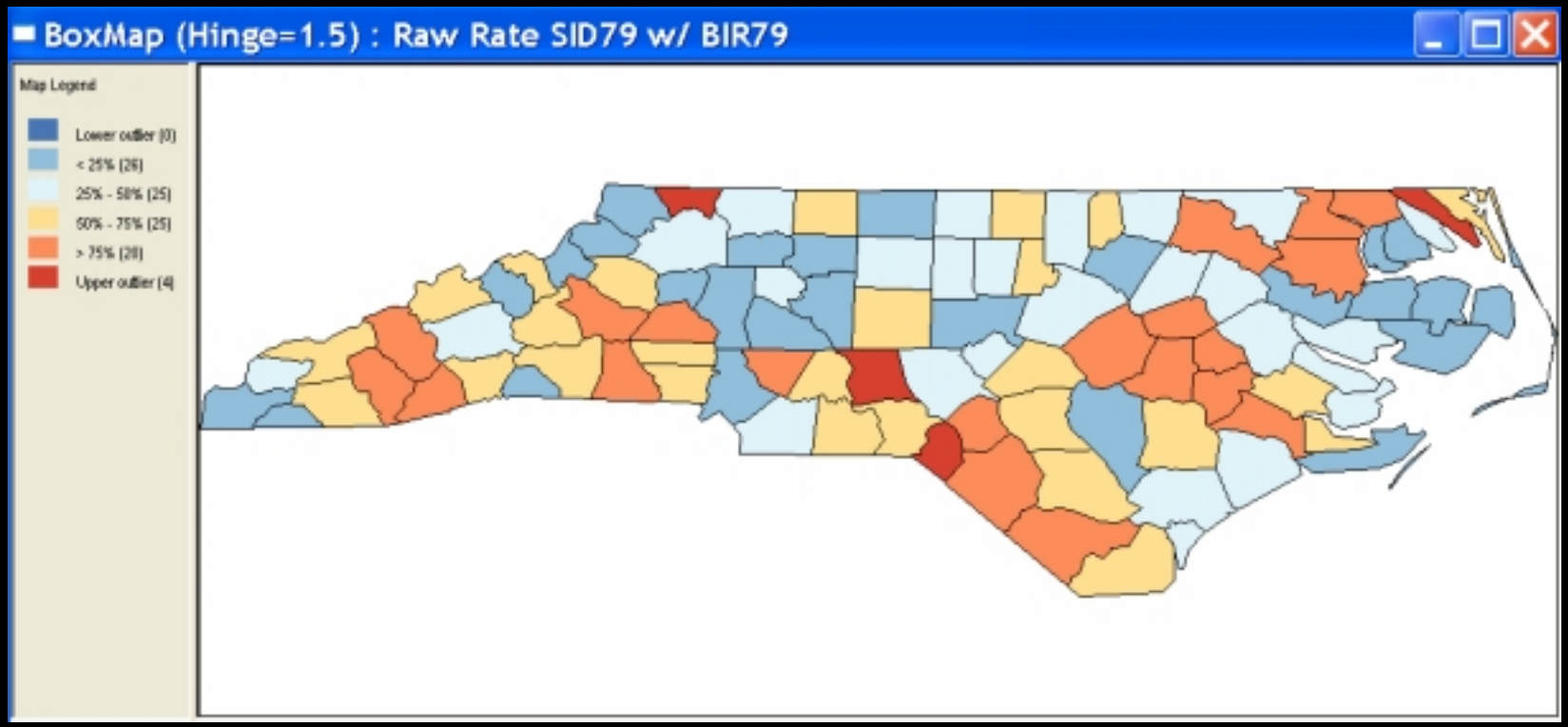
Map Smoothing

- Empirical Bayes
  - shrink rates to reference
  - national average
  - regional average = subset average
- Spatial Rate Smoother
  - spatial moving average
  - spatial range defined by spatial weights

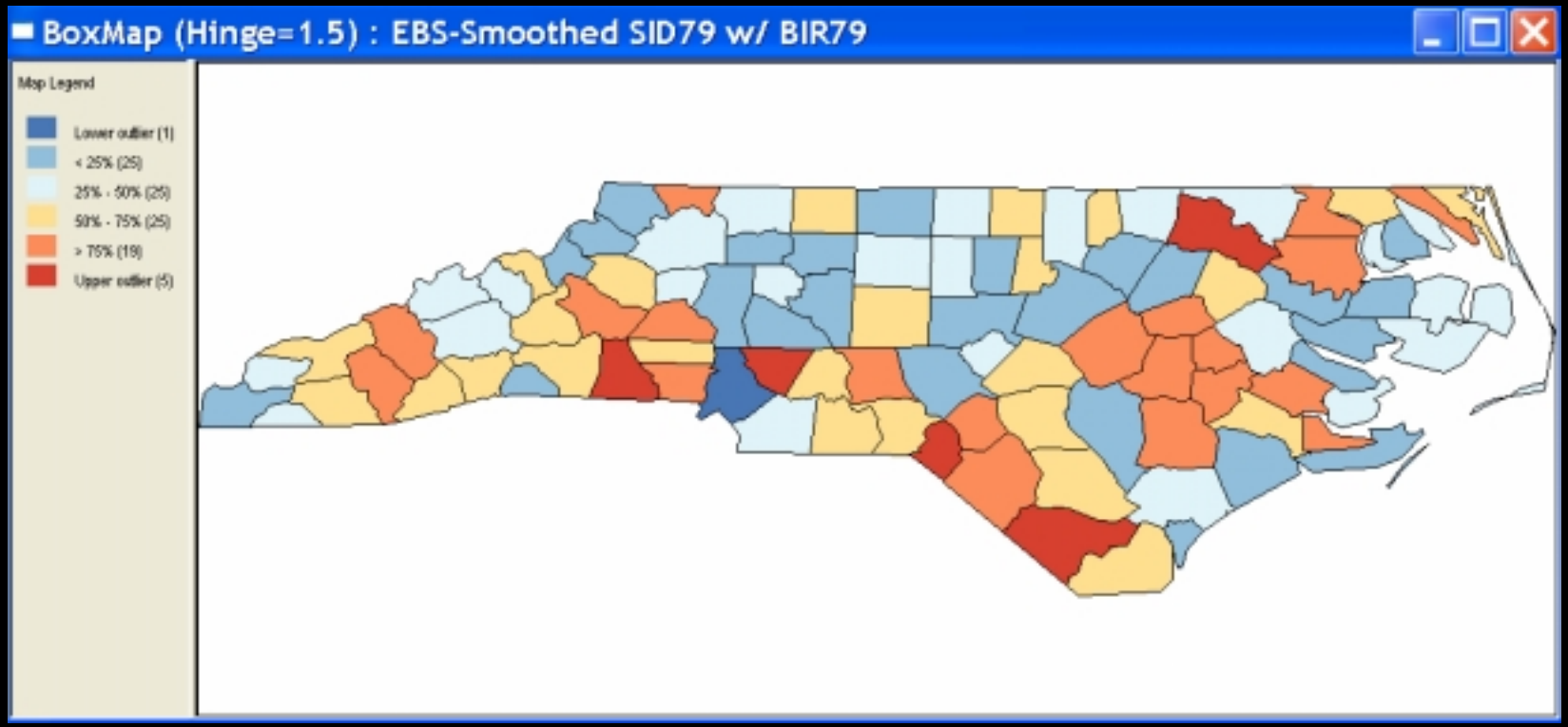
Event and Base



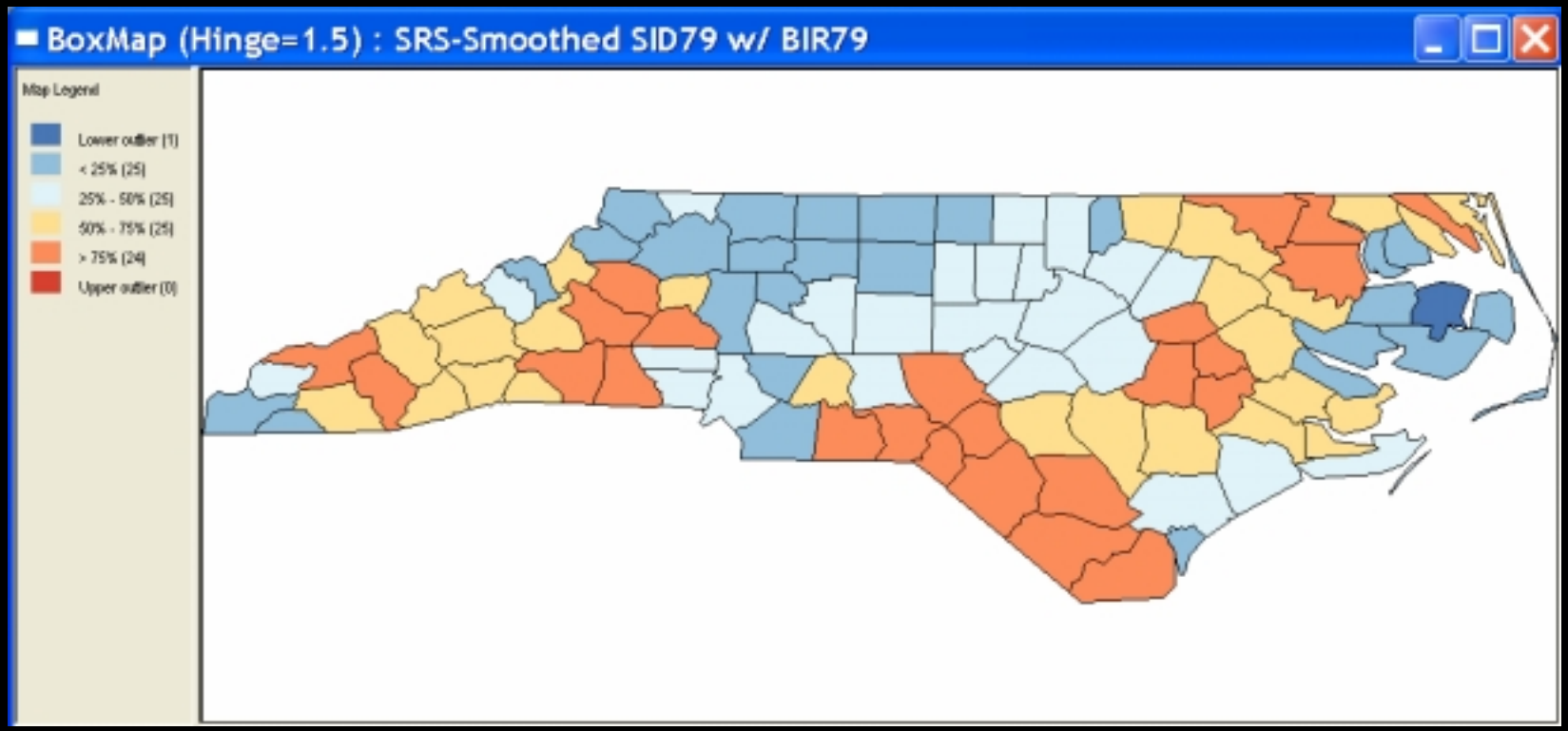
Raw Rate Map



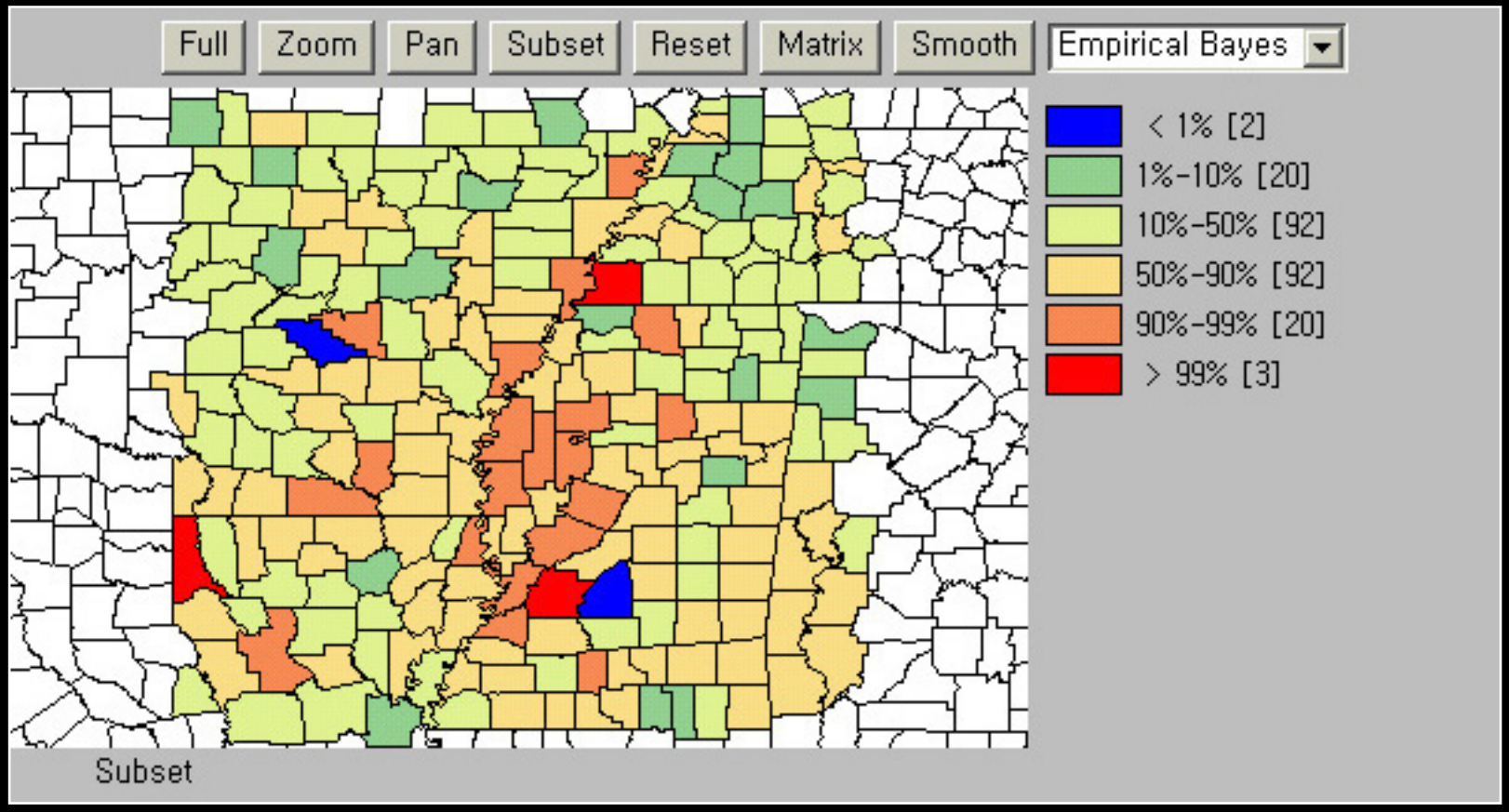
EB Smoothed Map



Spatial Rate Smoother



Regional EB Smoothing

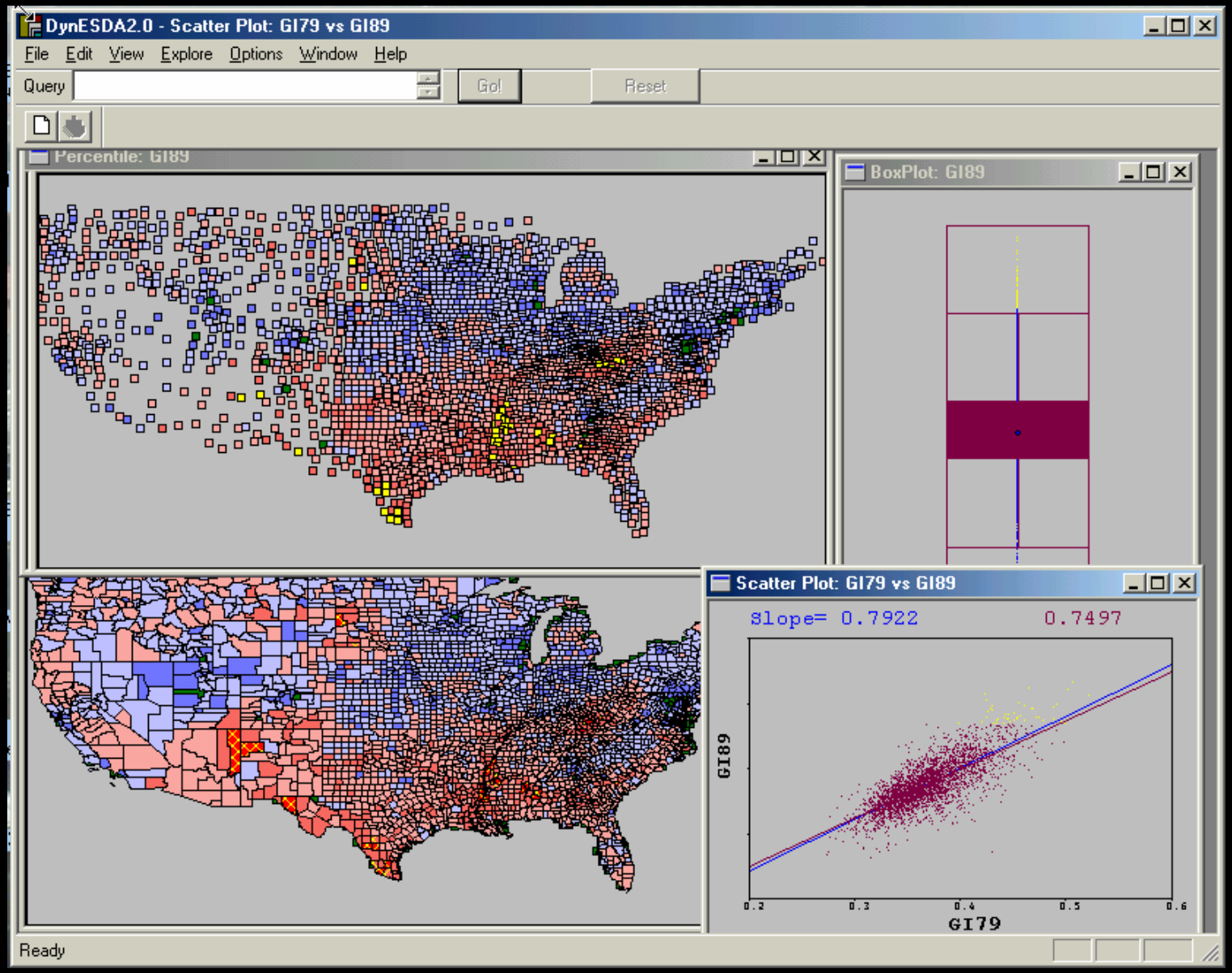


Linking and Brushing

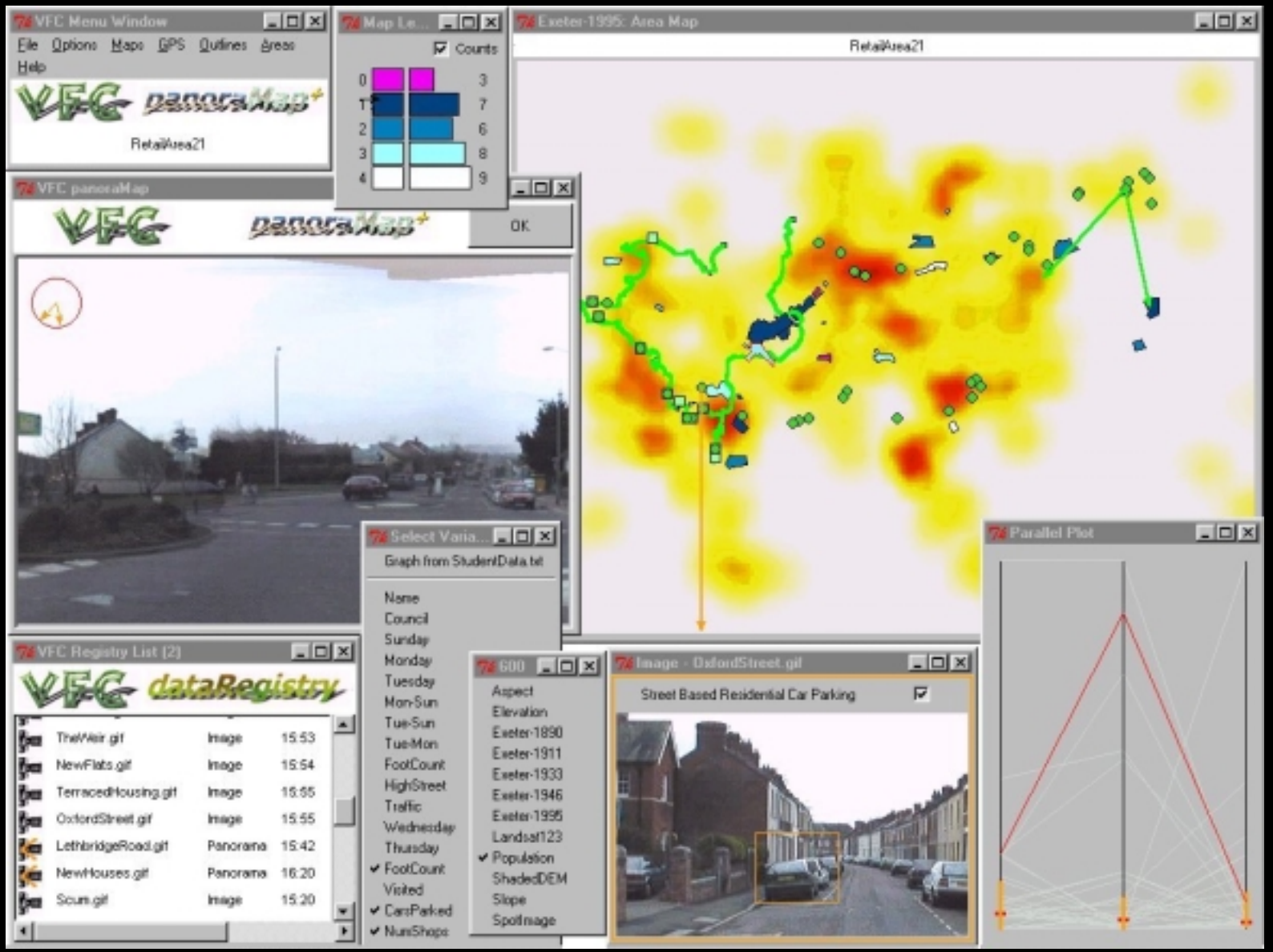
Linking

- Views
  - different “views” of data
    - statistical graphs: histogram, box plot, scatterplot
    - map
    - Table (list)
- Dynamic Linking
  - views dynamically linked
    - click on one view and corresponding observations (points, areas) on other views are highlighted

Linking Point and Polygon Maps



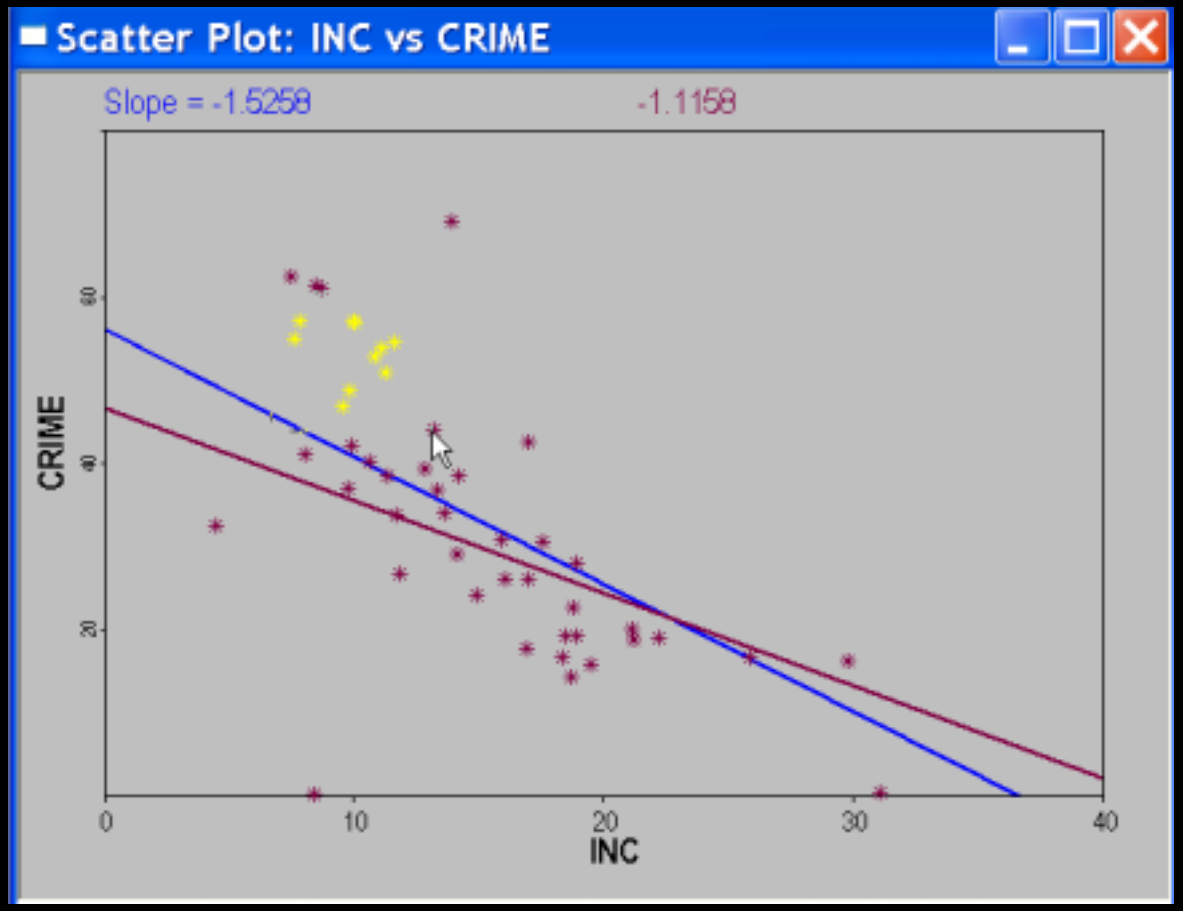
Dynamic Linking and Multimedia - panoraMap



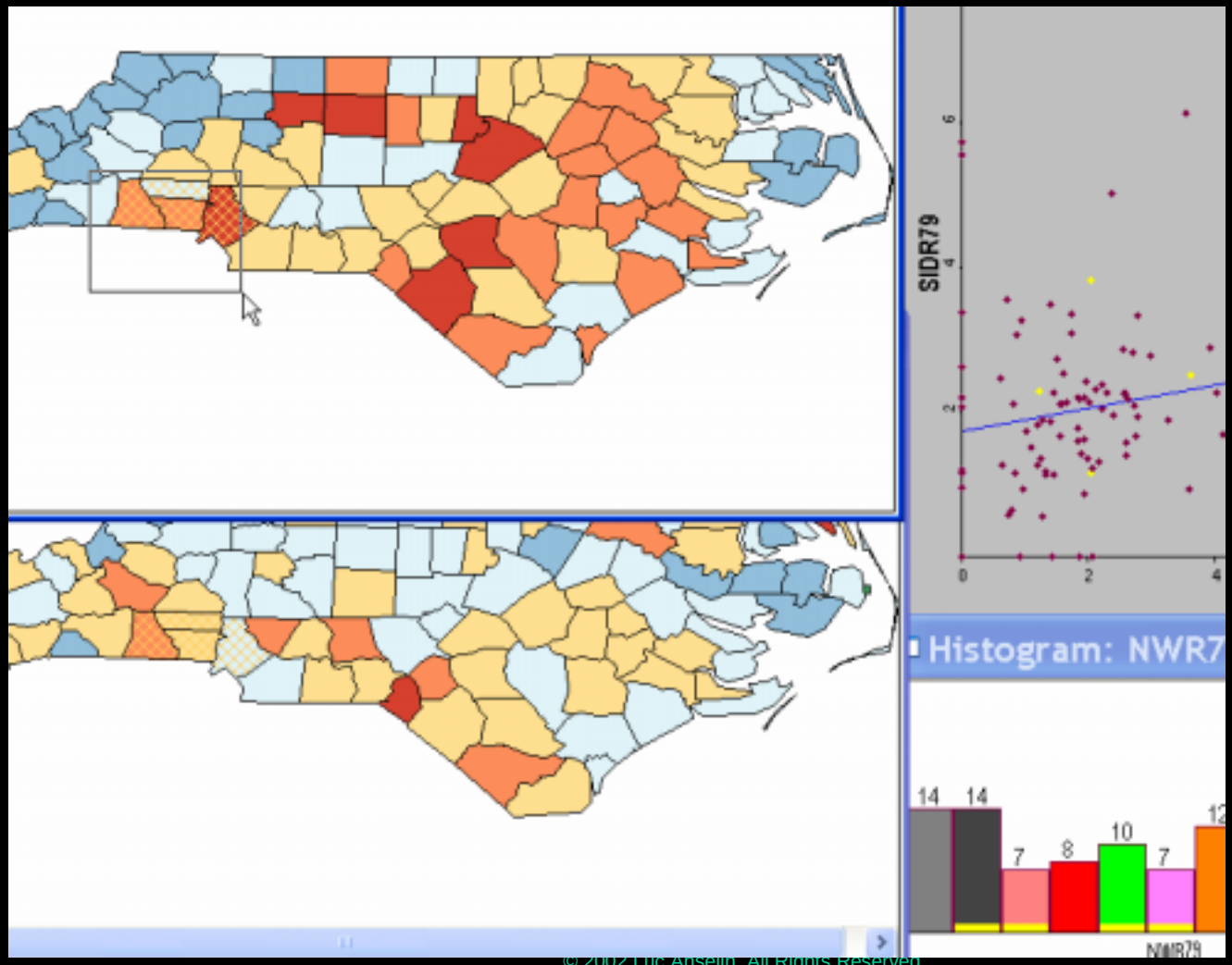
Brushing

- Brushing
  - moving “brush” over map or graph highlights matching observations in other statistical graphs and vice versa
- Brushing Scatterplots
  - recalculates slope of regression line
- Geographic Brushing
  - simultaneous selecting on multiple maps

Selection in Scatterplot

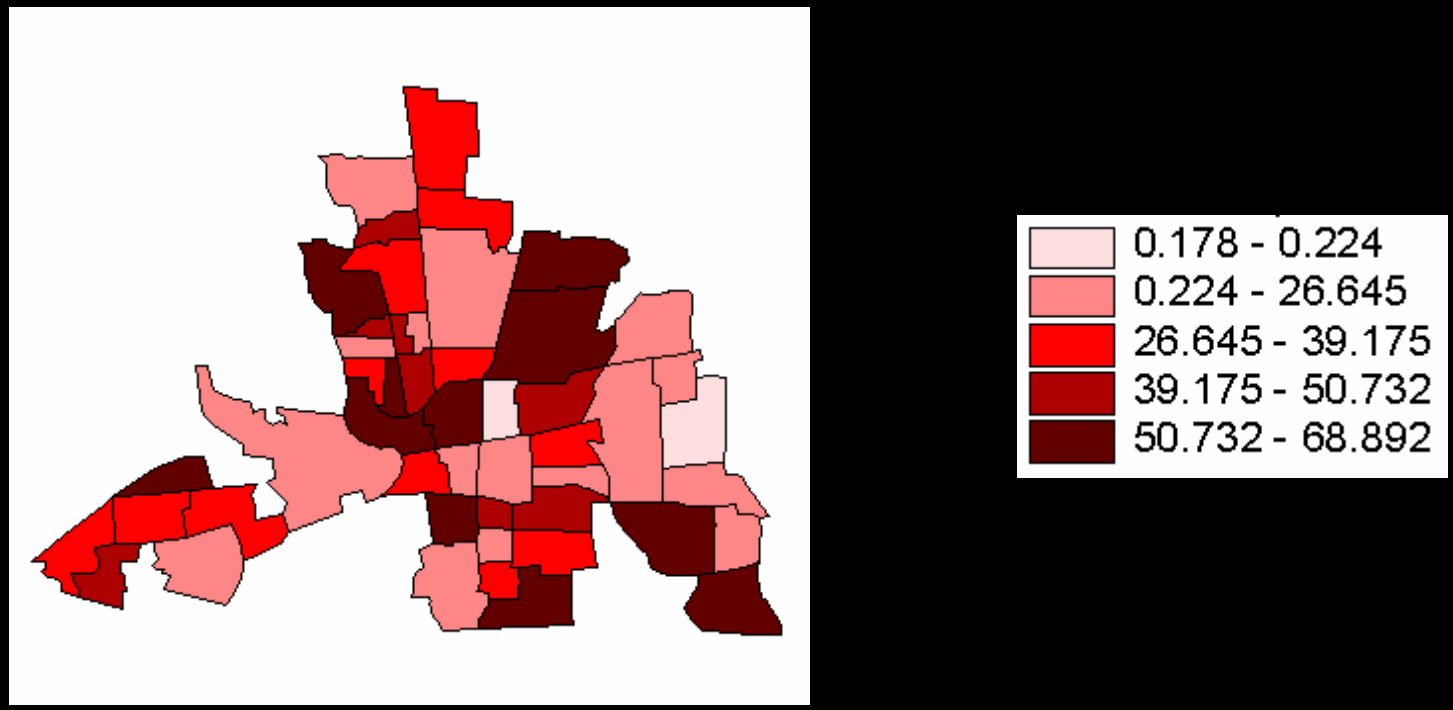


Map Brushing in DynESDA2

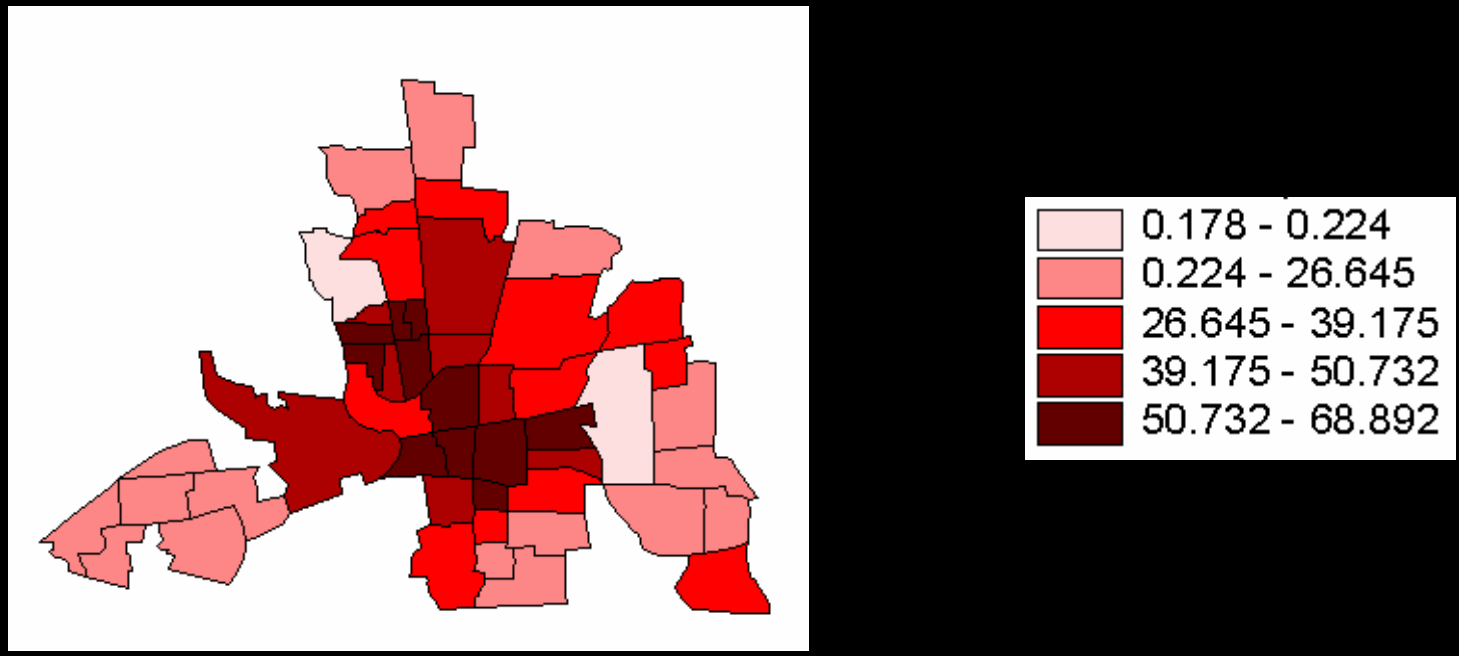


Visualizing Spatial Autocorrelation

Random or Clustered?



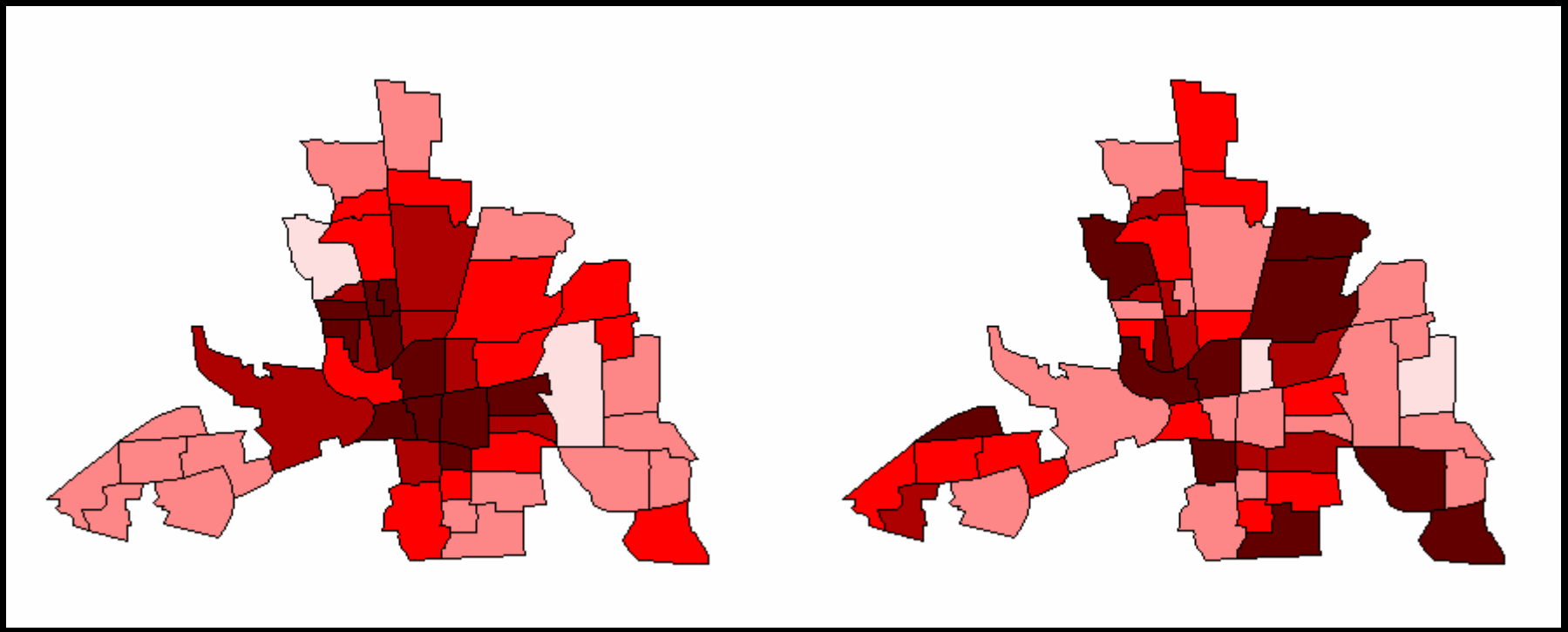
Random or Clustered?



Moran's I

- Moran's *I* Spatial Autocorrelation Statistic
  - cross-product statistic
    $I = (N/S_0) \Sigma_i \Sigma_j w_{ij}. z_i.z_j / \Sigma_i z_i^2$
    with $z_i = x_i - \mu$ and $S_0 = \Sigma_i \Sigma_j w_{ij}$
- Inference
  - normal distribution
  - randomization
  - permutation

Observed (left) and randomized (right) distribution for Columbus Crime



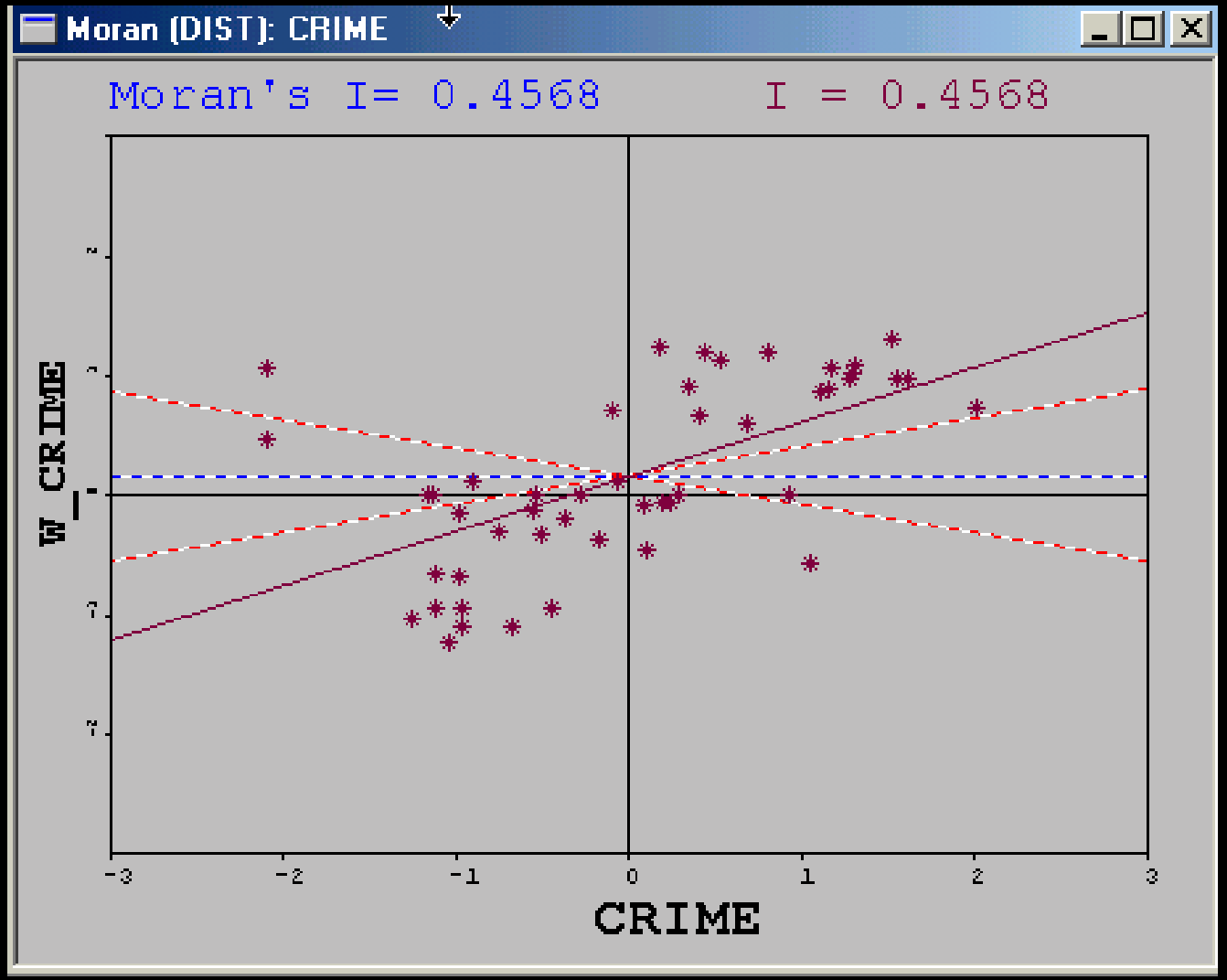
Moran's I = 0.486

Moran's I = -0.003

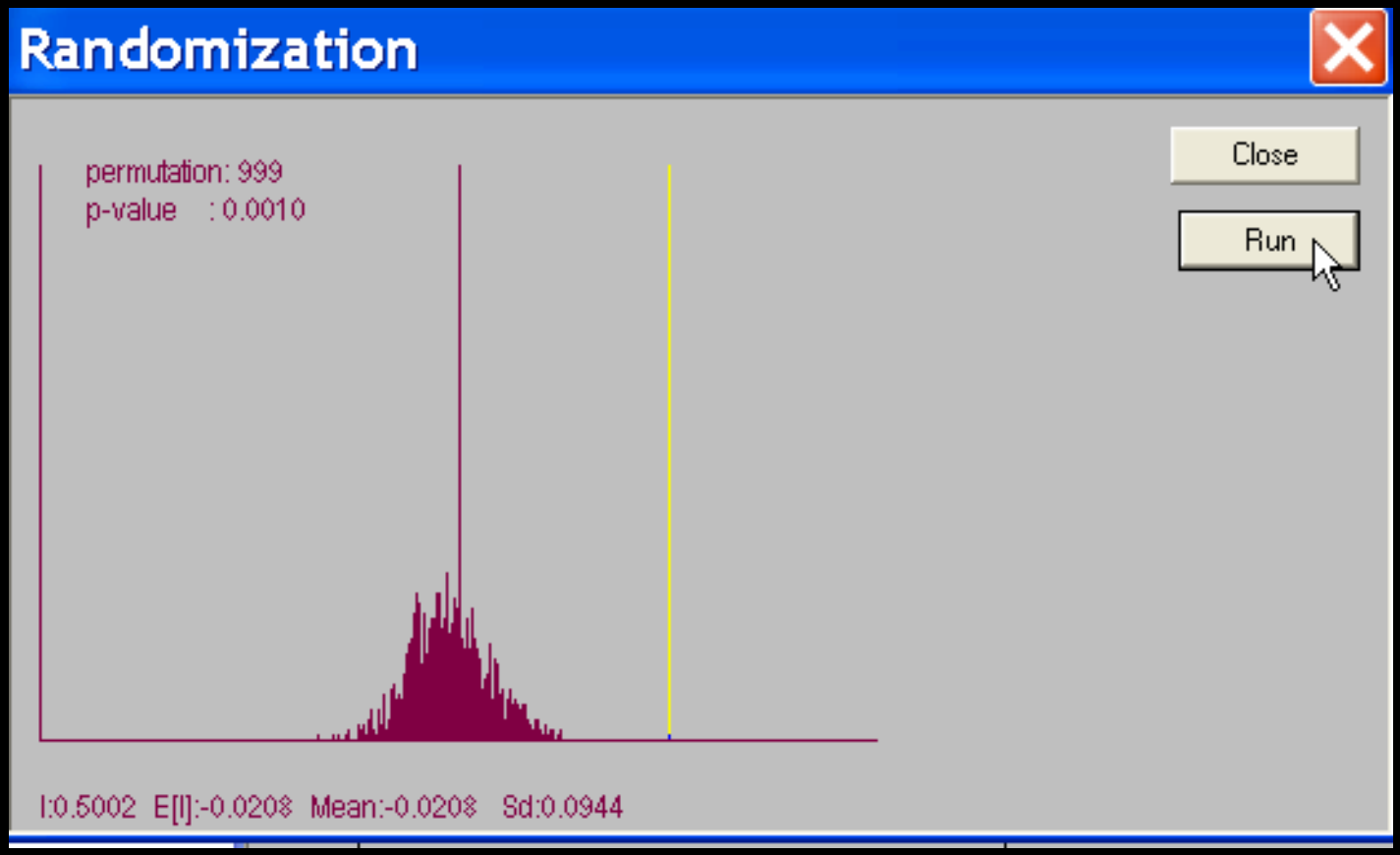
Moran Scatterplot

- Linear Spatial Autocorrelation
  - linear association between value at i and weighted average of neighbors: $\Sigma_j w_{ij} y_j$ vs. y_i , or Wy vs y
  - four quadrants
    - high-high, low-low = spatial clusters
    - high-low, low-high = spatial outliers
- Moran's I
  - slope of linear scatterplot smoother
  - I = z'Wz / z'z

Significance Envelope

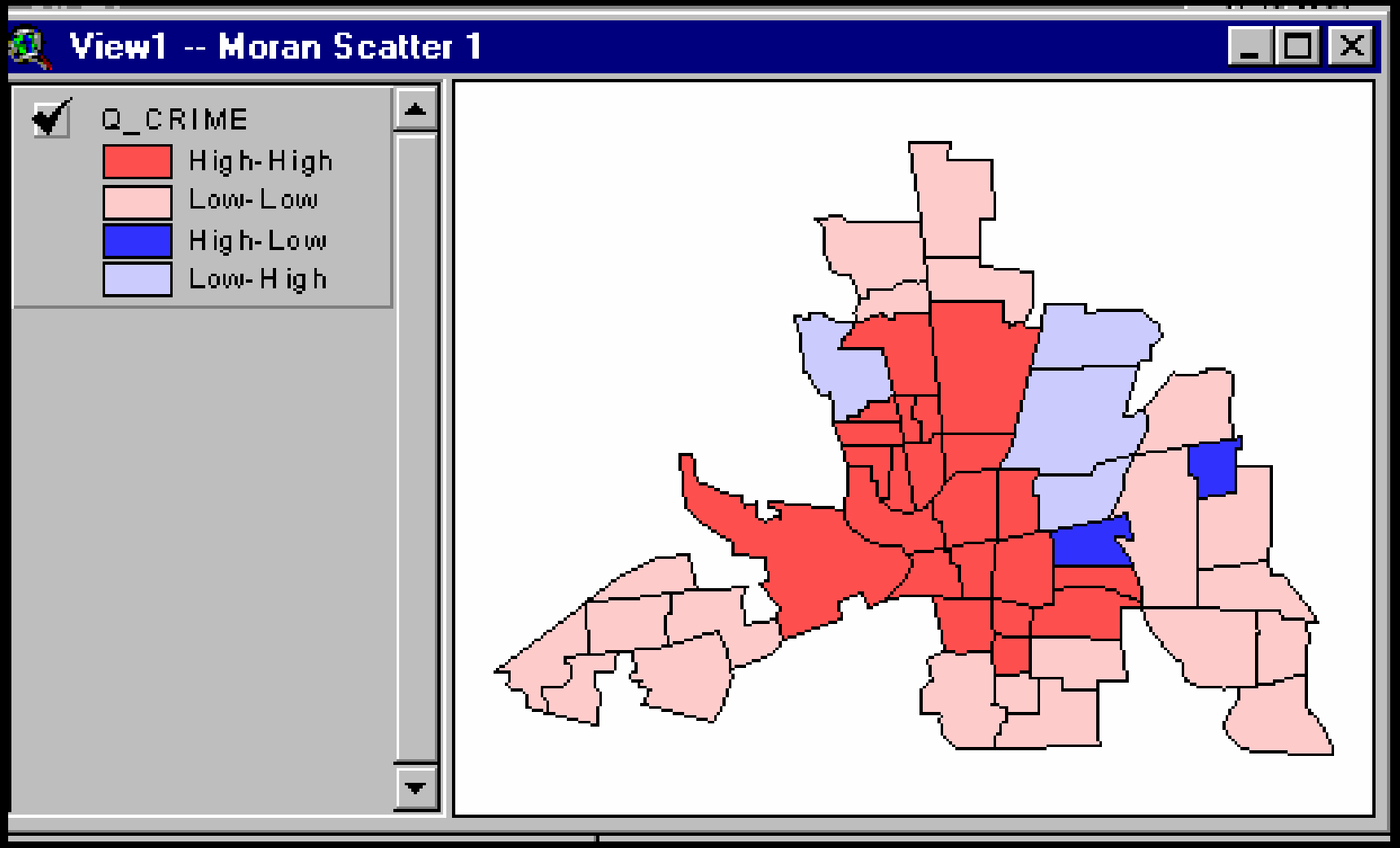


Reference Distribution (CRIME)



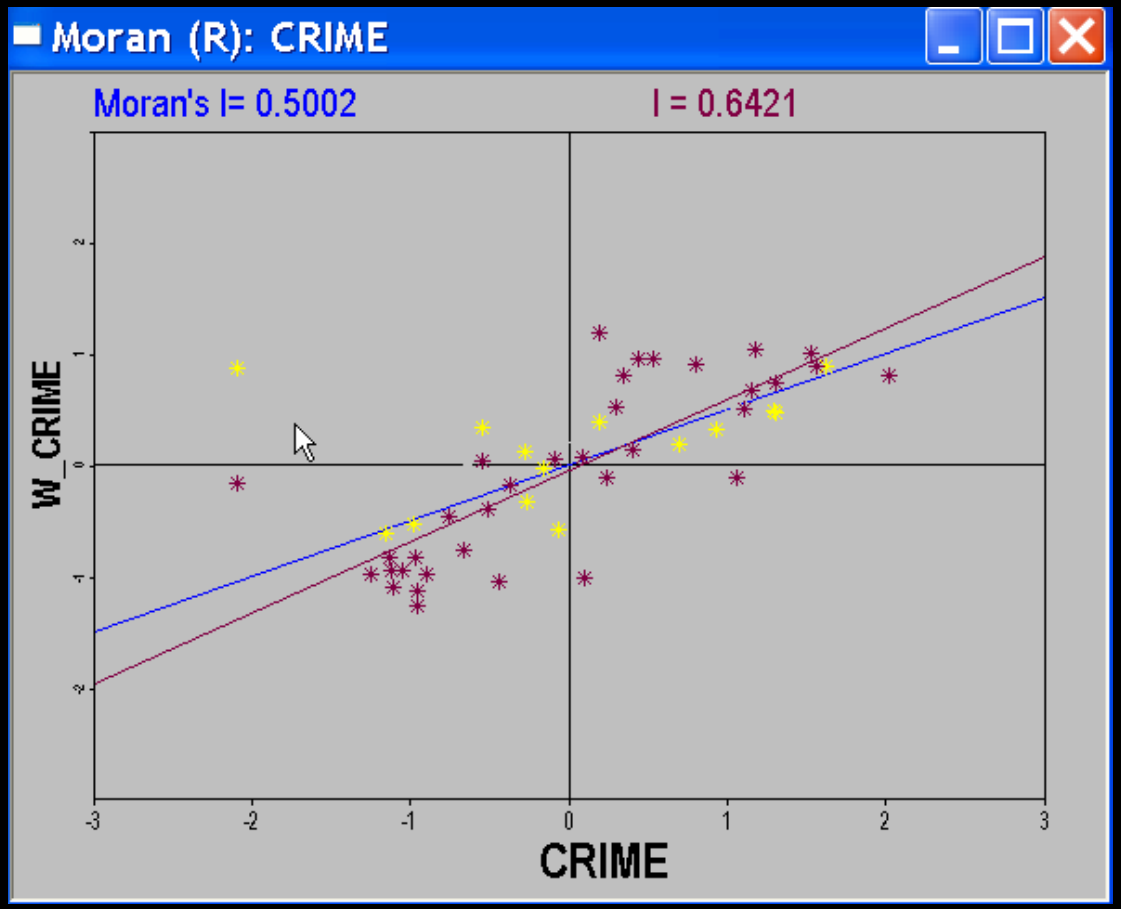
Use of Moran Scatterplot

- Classification of Spatial Autocorrelation
- Local Nonstationarity
  - outliers
  - high leverage points
  - sensitivity to boundary values
- Regimes
  - different slopes in subsets of the data



Moran Scatterplot Map for Columbus crime
four quadrants of the scatterplot (not “significant”)

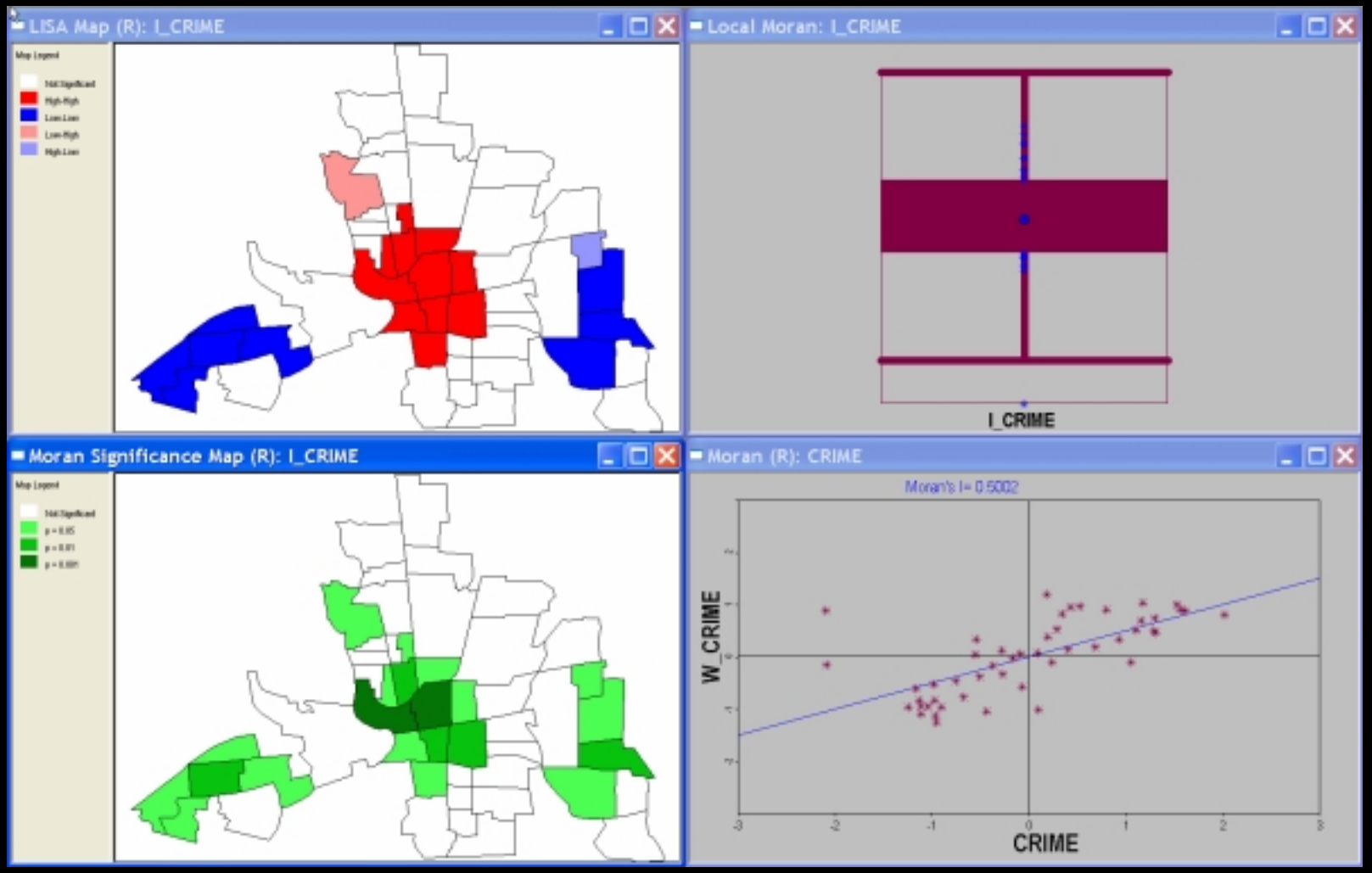
Moran Scatterplot - Regimes



Local Moran

- Local Moran Statistic
  - $I_i = (z_i / m_2) \Sigma_j w_{ij} . z_j$
  - $\Sigma_i I_i = N . I$
- Inference
  - randomization assumption
  - conditional permutation
  - local dependence or heterogeneity?
- Visualization
  - LISA map and Moran Significance Map

LISA MAPS

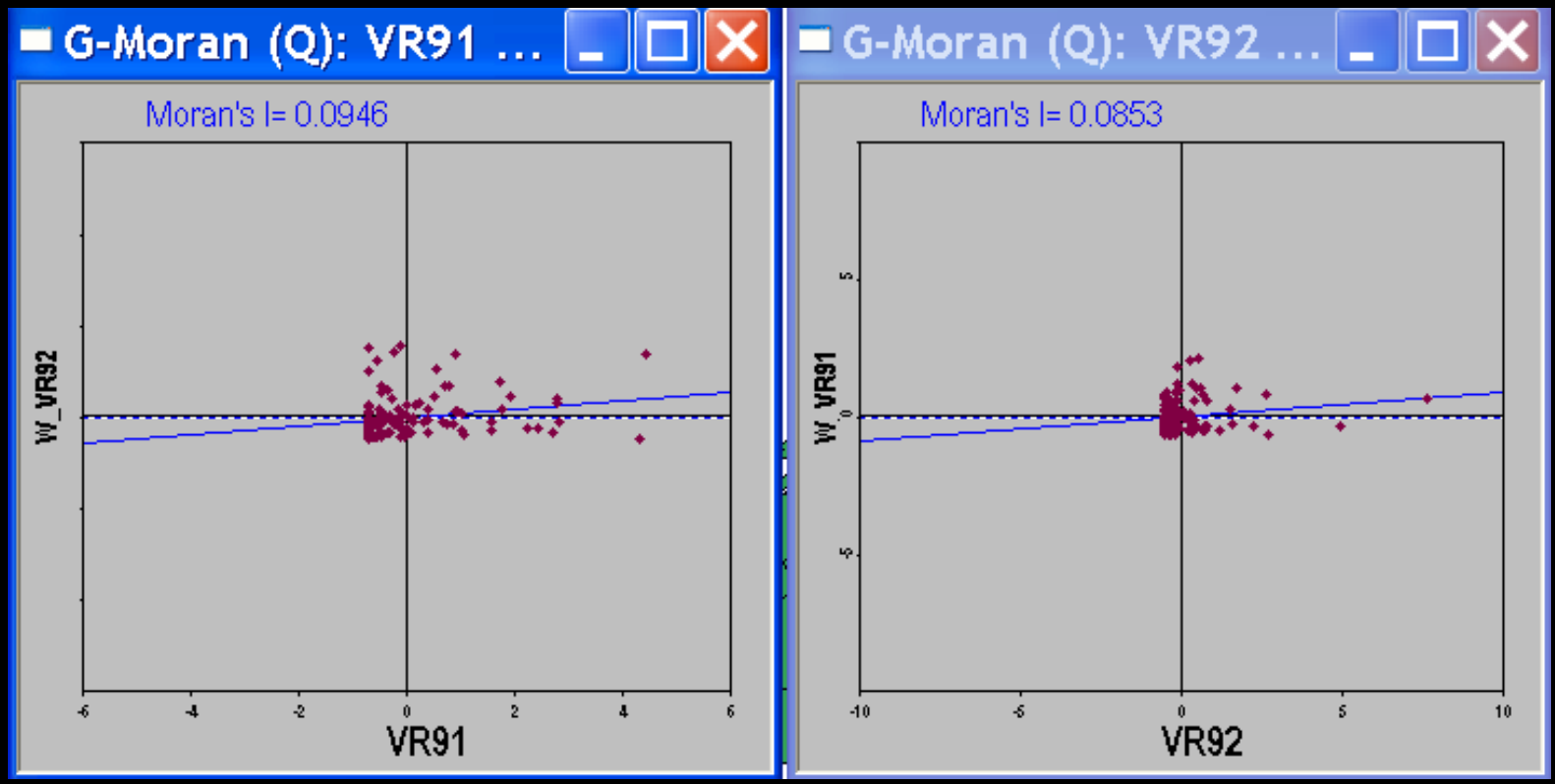

LISA Map (R): I_CRIME
Local Moran: I_CRIME
I_CRIME
Moran Significance Map (R): I_CRIME
Moran (R): CRIME
Moran's I= 0.5002
W_CRIME
CRIME

Space-Time Correlation

Space-Time Moran Scatterplot

- Generalized Moran Scatterplot
  - Regression slope of Wz_t on z_{t-1}
    - both variables standardized
    - = visualization of Wartenberg multivariate Moran statistic
  - Significance testing
    - permutation
    - permutation envelope (2.5% and 97.5% from permutation reference distribution)
- Four Types of Association
  - High-high, Low-low; High-low, Low-high

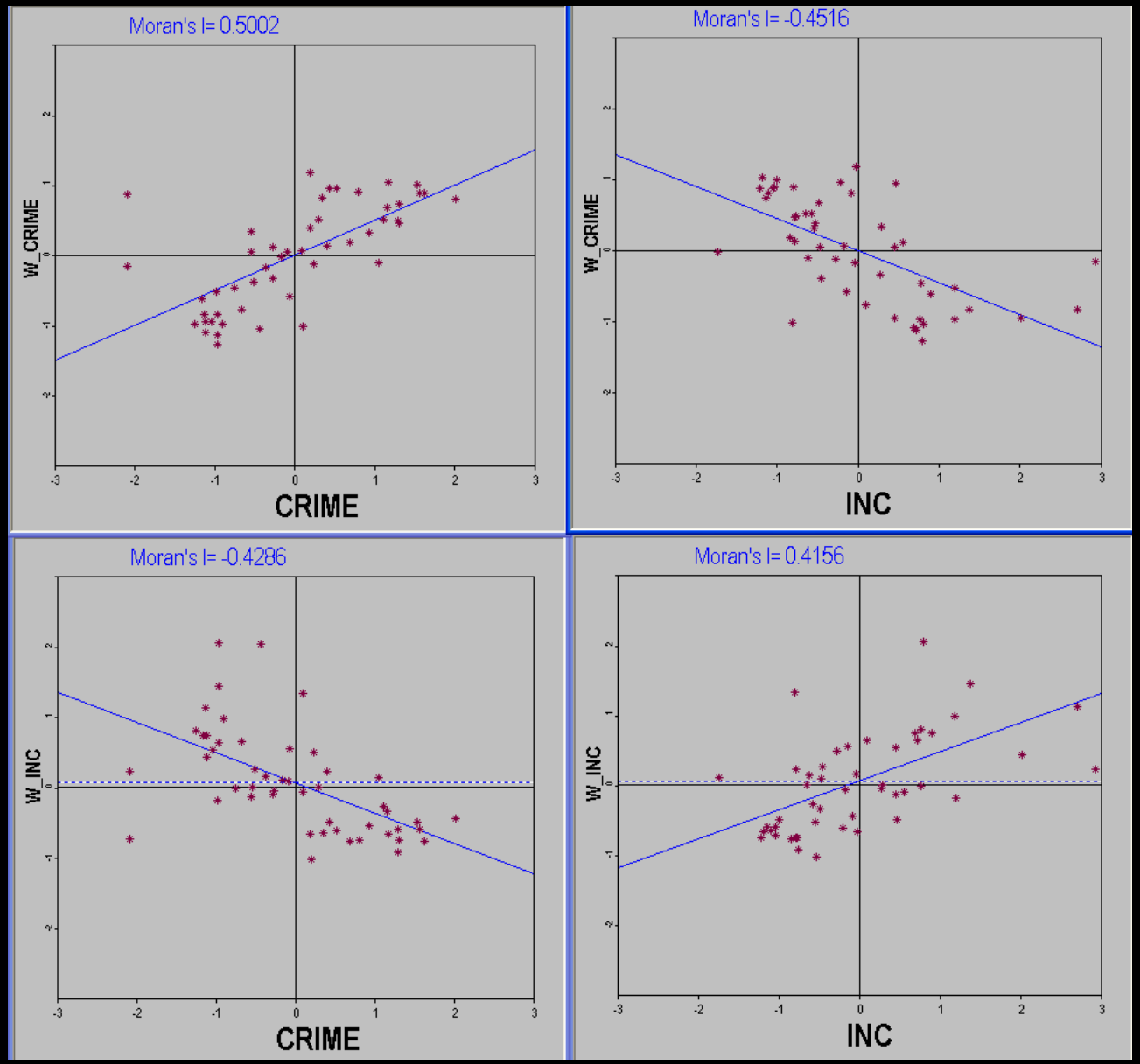
Space-Time Moran Scatterplot



p = 0.001

p = 0.002

Moran Scatterplot Matrix



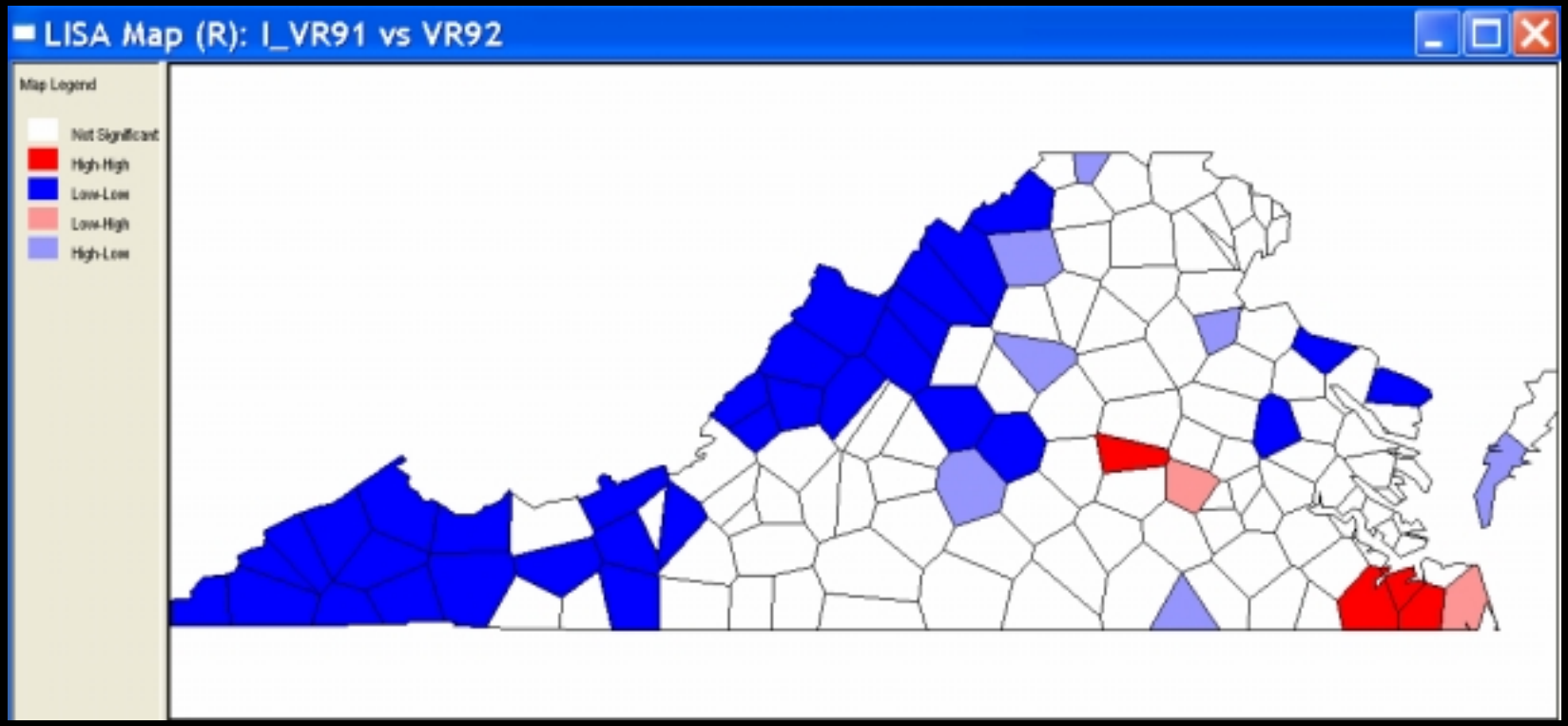
Generalized LISA

- Generalization of Local Moran
  - $z_{1i} \times \sum_j w_{ij} z_{2j}$
    - z_1 and z_2 different variables
    - same variable at different times
- Inference
  - Null hypothesis
    - random assignment between value of z_1 at i, t and "neighboring" values of z_2

Space-Time Patterns

- Space-Time Cluster = Contagion
  - High (above avg) values at a location surrounded by High values at different time
    - compare to high-high same time
  - Similar for Low-Low
- Space-Time Outlier = Change
  - High (above avg) surrounded by Low (below avg) at different time
  - Similar for Low-High
- Significance based on permutation

Space-Time LISA Maps



Interpretation and Limitations

- Most Important
  - assessing lack of spatial randomness
  - suggests “significant” spatial structure
- Multivariate Association
  - univariate spatial autocorrelation may result from
    - multivariate association
    - scale mismatch
  - need to control for other variables = spatial regression
- LISA Clusters and Hot Spots
  - suggest interesting locations
  - do not explain