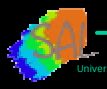
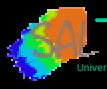
Exploring Spatial Public Health Data with GeoDa

Spatial Analysis Lab
University of Illinois at Urbana-Champaign, USA



Purpose and Structure

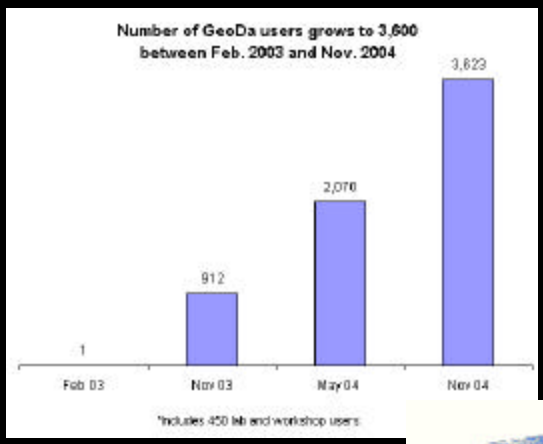
- Overview of how GeoDa can add value to GIS and other analyses using public health examples
- One-on-one technical assistance with your own data available throughout *GeoHealth* conference



What is GeoDa?

- Free, introductory, user-friendly software (http://sal.agecon.uiuc.edu)
- Path from Exploratory Spatial Data Analysis (ESDA) to limited Spatial Regression Analysis tools
- Designed for lattice data (polygons and points)
- GIS expertise not assumed

GeoDa Users



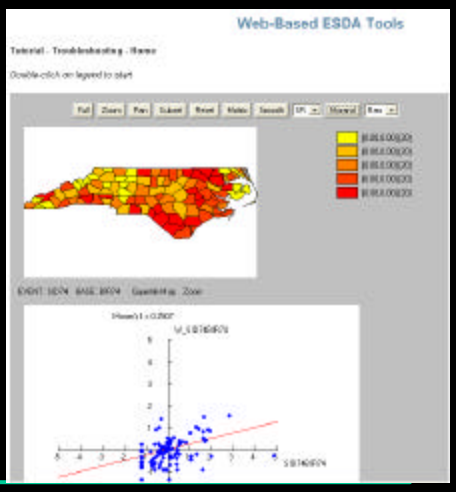
17 downloads from NZ

Who develops GeoDa?

- The Spatial Analysis Lab (SAL) at UIUC: http://sal.agecon.uiuc.edu
- Directed by Dr. Luc Anselin
- Mission
- GeoDa Team

SAL Projects

- CSISS Tools
- Geovisualization and Spatial Analysis of Cancer Data
- Prostate Cancer ESDA and Spatial Statistics
- Web-Based Tools for the Exploration of Spatial Data



SAL Funders

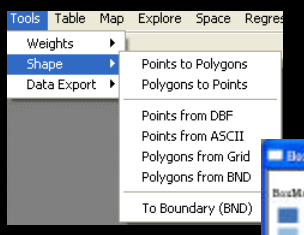
GeoDa & Other Free Software

- Stand-alone, Windows platform
- Philosophy: Supplement, not Substitute
- Based on ideas in SpaceStat and DynESDA but built new in C++
- Other free packages: CrimeStat, SaTScan, R-Geo (see Luc Anselin's review at *http://www.naaccr.org/*)

GeoDa Overview

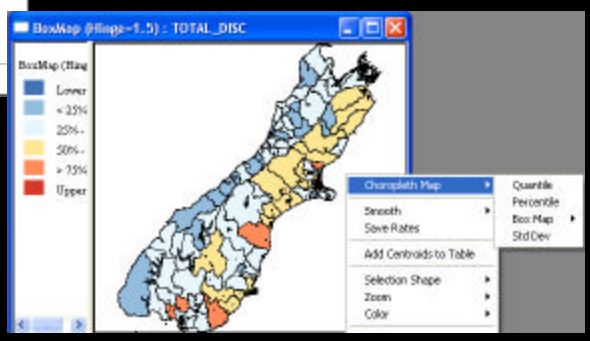
1 Data Manipulation
2 Mapping
3 Weights Matrices
4 Multivariate Exploratory Data Analysis (EDA)
5 Spatial Autocorrelation
6 Spatial Regression

Overview: Data Manipulation & Mapping

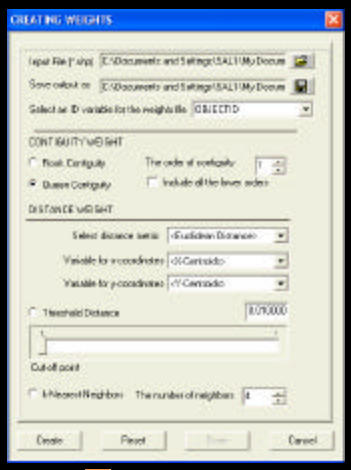


1 Data Manipulation

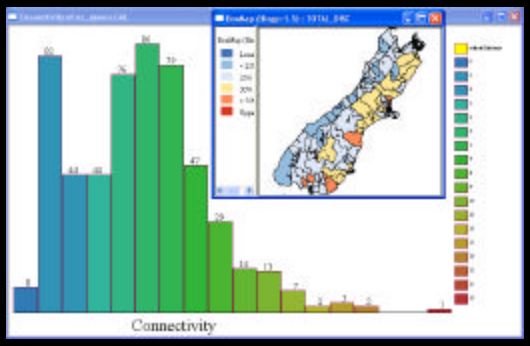
2 Mapping



Overview: Weights Matrices

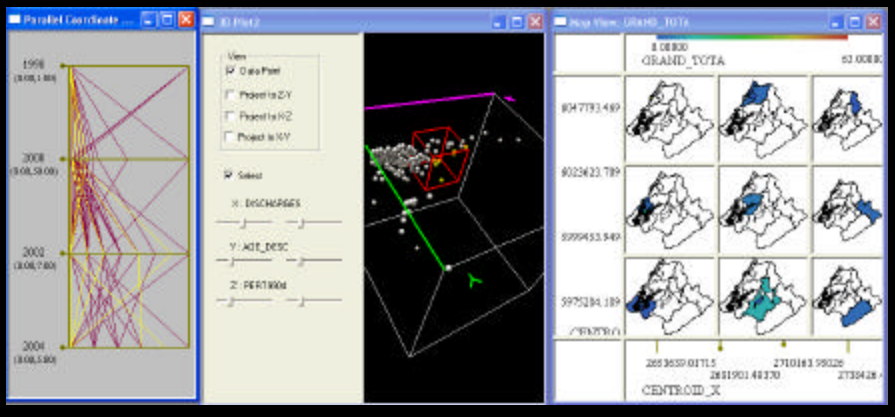


3 Create Weights Matrices



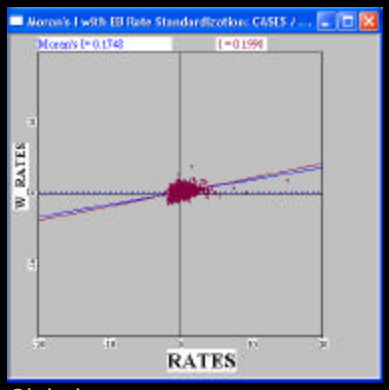
Weights Characteristics

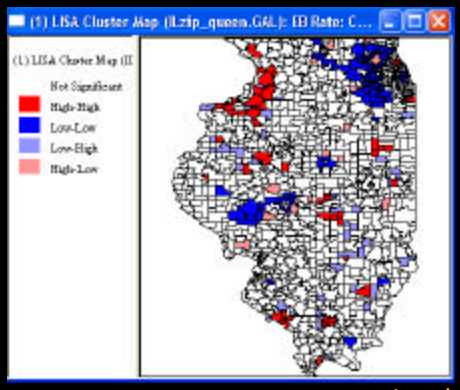
Overview: Multivariate EDA



4 Multivariate Exploratory Data Analysis (EDA) DEMO

Overview: Spatial Autocorrelation

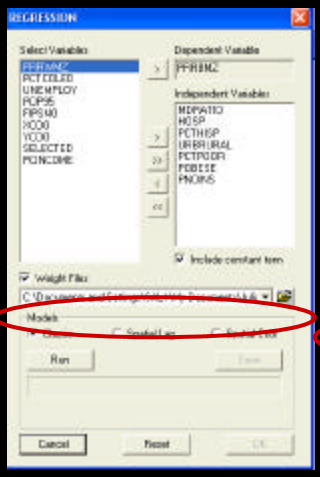


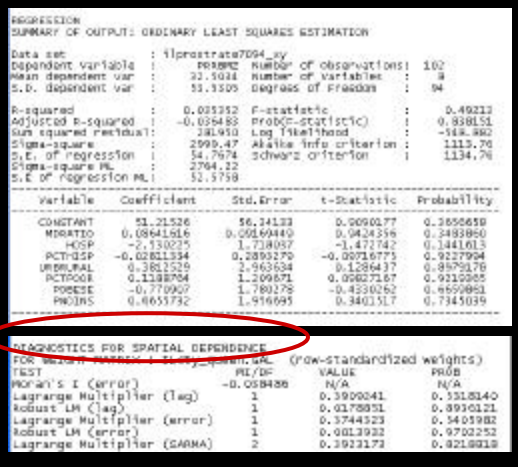


Global Local

5 Spatial Autocorrelation

Overview: Spatial Regression



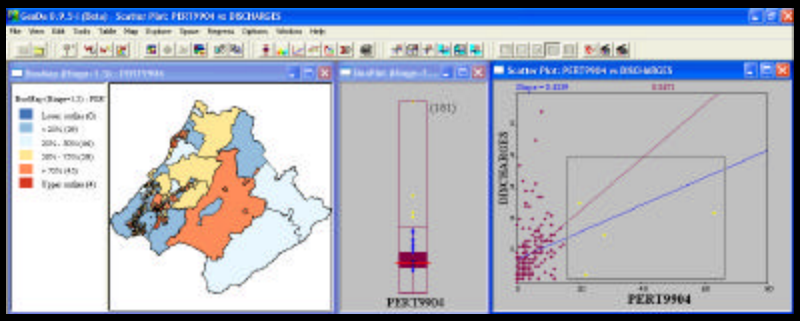


6 Spatial Regression

Key Features for Health Data

- Linking and Brushing
- Rate Smoothing
- Global and Local Clustering

Linking & Brushing



- All windows are dynamically linked
- Observations can be selected through brushing
- On-the-fly re-computation of scatter plot results

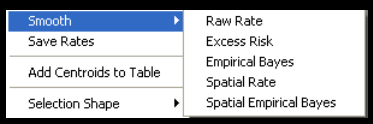
DEMO

Rate Smoothing

- Raw rates used to estimate underlying disease risk
- Differences in population size related to problem of variance instability and spurious outliers
- Rate smoothing addresses variance instability by borrowing strength from other spatial units

GeoDa Smoothing Options

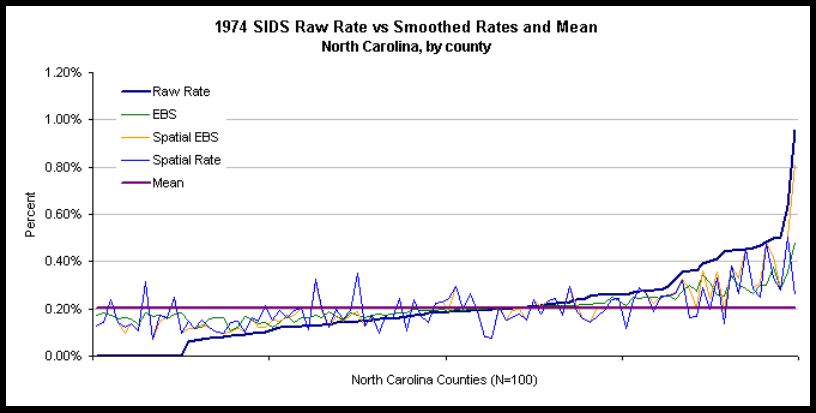
GeoDa offers four smoothers:



Where Strength is Borrowed From

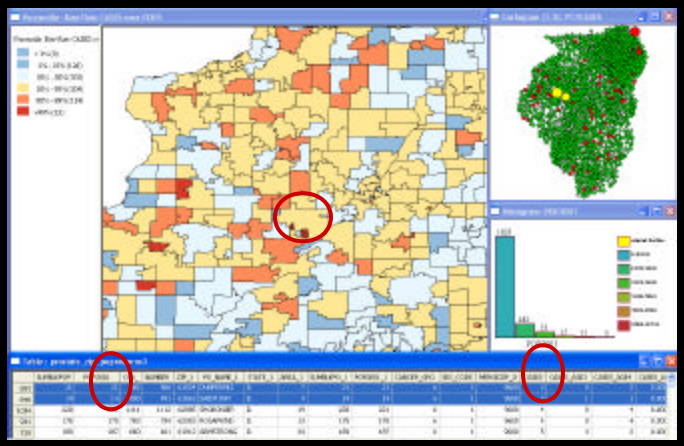
- **Excess Risk**: Expected risk based on product of raw rate and average overall risk of all observations
- **Empirical Bayes (EBS)**: Overall mean of the underlying risk distribution of all observations
- **Spatial Rate**: Neighbors, as defined in spatial weights matrix
- **Spatial Empirical Bayes**: Same as EBS but strength not borrowed from all observations, only regional subset

Rate Smoothing: Comparison



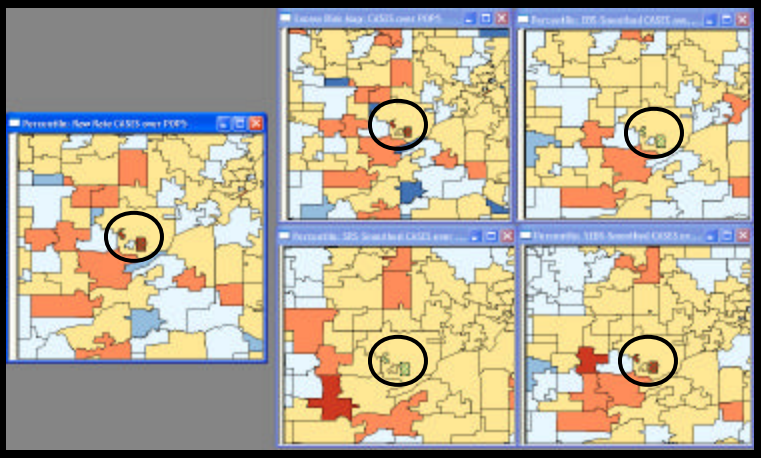
Raw rate compared to smoothed rates and mean

Rate Smoothing: Illustration of Variance Instability



Raw rates for prostate cancer in Illinois, by zip code

Rate Smoothing: Outlier Comparison



Raw and smoothed rates for prostate cancer in Illinois, by zip code

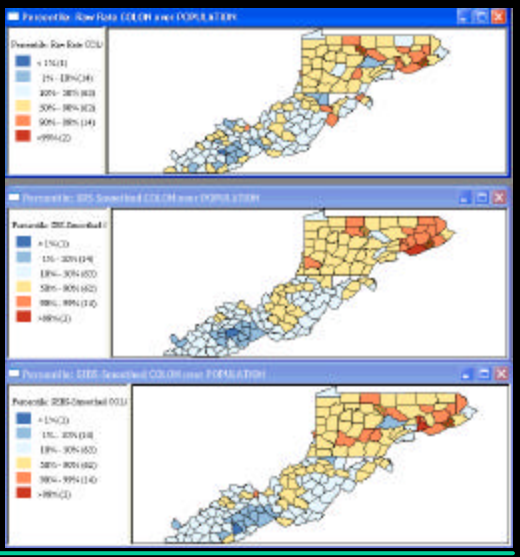
Rate Smoothing with Weights: Comparison

Raw rates for colon cancer in Appalachia, by county

Spatial Rates with Queen Matrix

Spatial Empirical Bayes Rates with Queen Matrix

DEMO



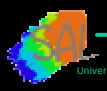
Global and Local Clustering

Global Moran's I

- What is the extent of clustering in the total area?
- Is this clustering significantly different from a random spatial distribution?

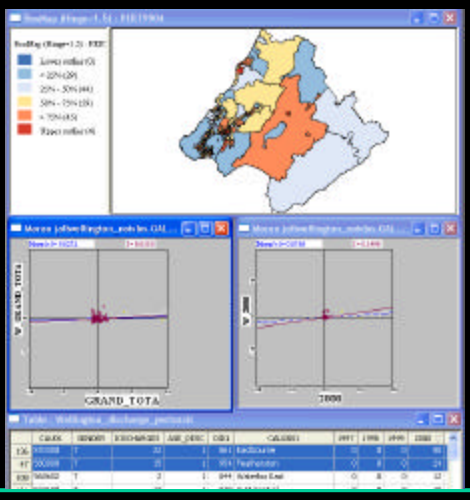
Local Moran's I

- Do local clusters (high-high or low-low) or local spatial outliers (high-low or low-high) exist?
- Are these local clusters and spatial outliers statistically significant?



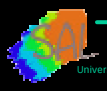
Global Clustering

1997-2004 Pertussis Cases in Wellington

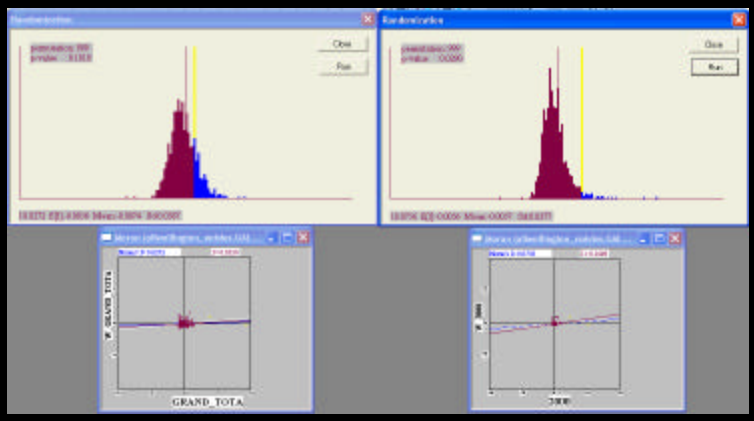


Global Moran's I scatter plot for 1997-2004 and 2000 cases

$$I = \left(\frac{N}{S_0}\right)\frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}$$



Global Clustering: Significance Tests

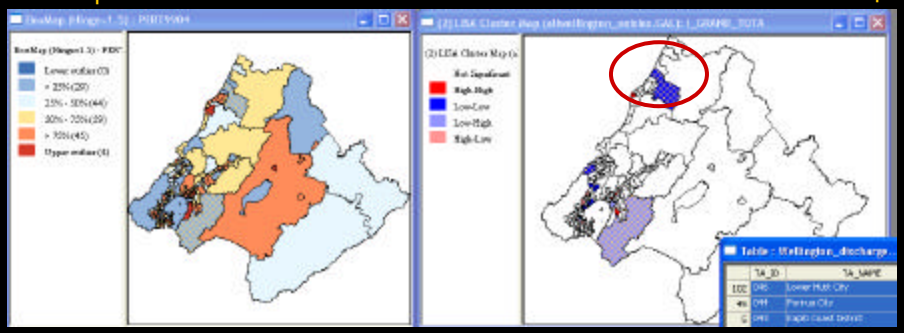


Global Moran's I scatter plots and significance tests
for 1997-2004 & 2000 Pertussis cases in Wellington

Local Clusters

Box map

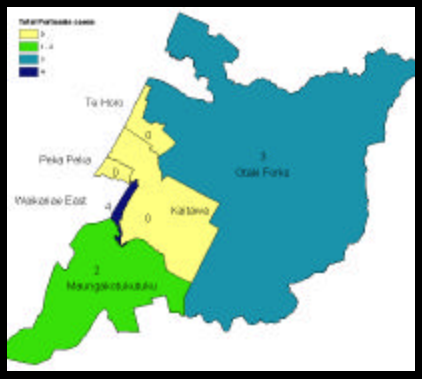
LISA map



Local Indicators of Spatial Association (LISA) maps can be used to identify spatial clusters and outliers at different significance levels

Local Cluster Computation

Illustration of low-low Kaitawa cluster



$$I_i = \frac{(x_i - \mu_x)\sum_j w_{ij}(x_j - \mu_x)}{\sum_i (x_i - \mu_x)^2 / n}$$

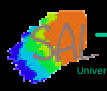
Cross-product of standardized value for area i and average standardized values of neighbors j

Numerator for Kaitawa (mean=5): (0-5)*[(0-5)+(3-5)+(2-5)+(4-5)+(0-5)]

Denominator: Sum of squared standardized values for each area i in total study area, divided by N

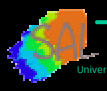
GeoDa uses row-standardized weights (rows sum to one)

DEMO



Conclusion

- GeoDa offers spatial analysis tools that can be used separately or in addition to GIS and statistical analyses
- GeoDa results can be exported and integrated in other packages
- Rate smoothing and global/local clustering are particularly useful for health data



Current & Future Developments

- OpenGeoDa
  - Cross-platform, open-source
  - Including full Help system
- Python modules for cancer-related
  - smoothing
  - cluster analysis
  - spatial regression

More Information

- http://sal.agecon.uiuc.edu/geoda_main.php
- Subscribe to Openspace listserv for GeoDa support
- One-on-one technical assistance with your own data at *GeoHealth 2004*