Introduction to Exploratory Spatial Data Analysis

Local Spatial Autocorrelation

Sergio Rey

School of Geographical Sciences and Urban Planning
Arizona State University

http://geodacenter.asu.edu/

Outline

- Principles
- Local Moran
- Gi Statistic

Global vs. Local Analysis

- Global Analysis
  - one statistic to summarize pattern
  - clustering
  - homogeneity
- Local Analysis
  - location-specific statistics
  - clusters
  - heterogeneity

LISA Definition

- Anselin (1995)
  - Local Spatial Statistic
  - indicate significant spatial autocorrelation for each location
- Local-Global Relation
  - sum of LISA proportional to a corresponding global indicator of spatial autocorrelation

LISA Forms of Global Statistics

- Every Decomposable Statistic
  - if global = a. [$\sum_i$ component(i)]
  - then local = component(i)
- General Case: Local Gamma
  - $\Gamma = \sum_i (\sum_j w_{ij} a_{ij})$
  - local $\Gamma_i = \sum_j w_{ij} a_{ij}$
  - $\Gamma = \sum_i \Gamma_i$

Examples

- Local Moran
  - $I_i = m \sum_j w_{ij} y_j$
- Local Geary
  - $c_i = m \sum_j w_{ij} (y_i - y_j)^2$
- Local Sokal
  - $s_i = m \sum_j w_{ij} |y_i - y_j|$

Local Moran

Local Moran

- Local Moran Statistic
  - $I_i = (z_i / m_2) \sum_j w_{ij} z_j$
  - $m_2 = \sum_i z_i^2$ does not vary with i
  - $\sum_i I_i = n.I$ link local-global
  - $I = \sum_i I_i / n$ global is mean of locals

Inference

- Analytical
  - equal probability assumption
  - normal approximation to statistic
  - poor in small samples: do not use
- Computational
  - conditional permutation
  - hold value at i fixed, permute others

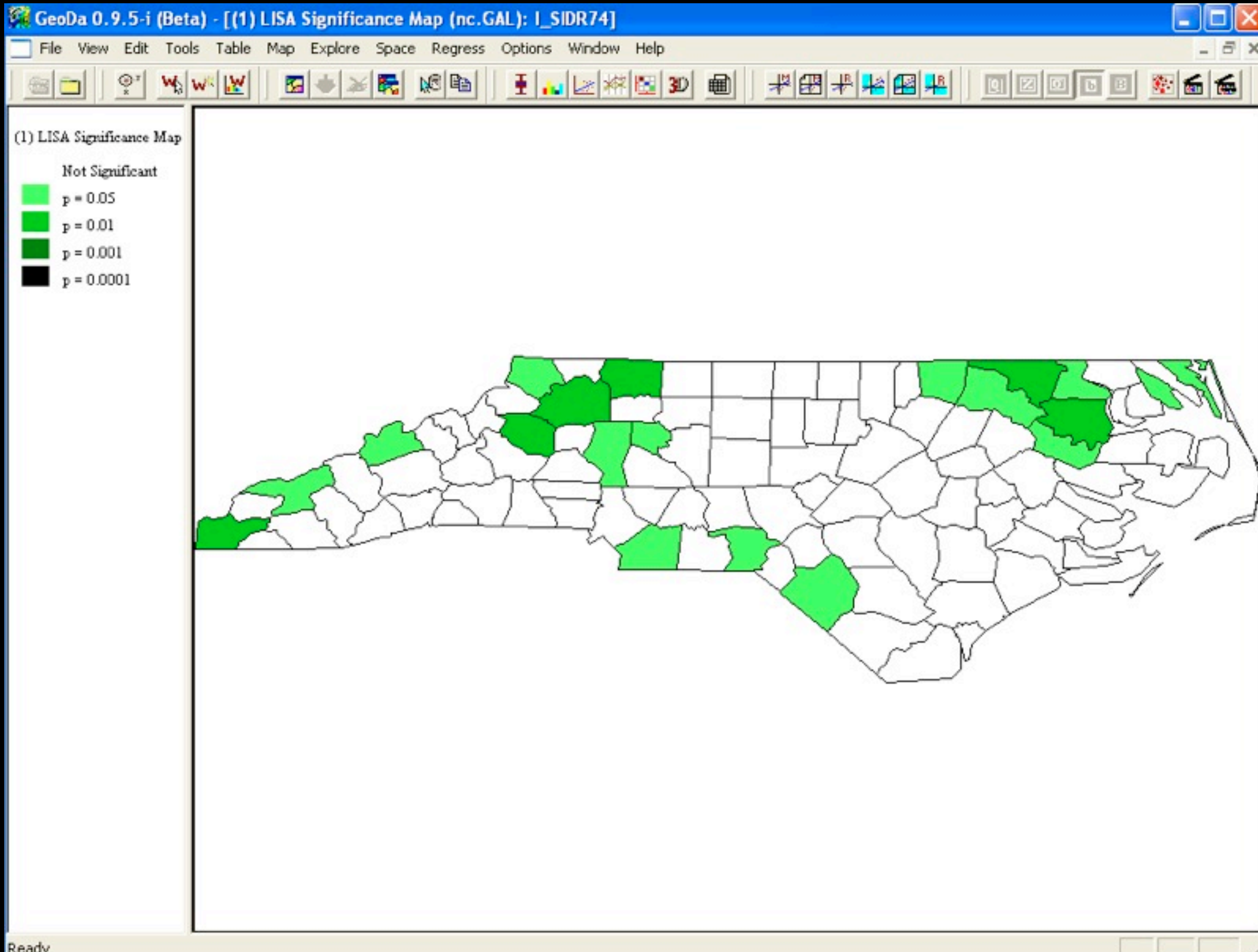
LISA Significance MAP

- Locations with Significant Local Statistics
    - multiple comparison problem
    - sensitivity analysis to p-value
- Choropleth Map
    - shading by significance
    - non-significant locations not highlighted

Multiple Comparisons

- $\alpha = 1 - (1 - \alpha_o)^{1/k}$
- α_o = overall (joint) significance level (e.g. 0.05)
- k = number of comparisons
- α = marginal significance level
- If k=48 and $\alpha_o = 0.05$

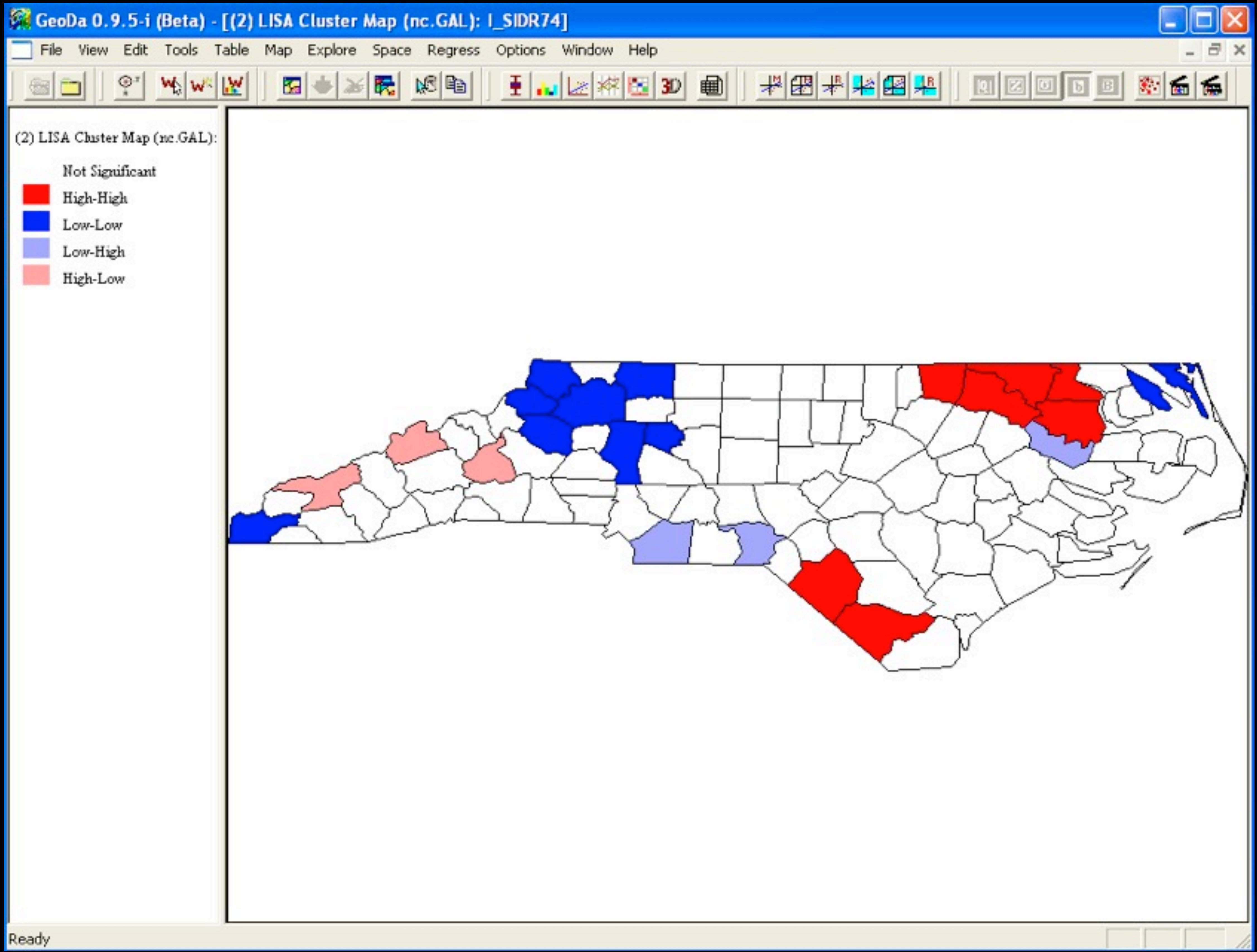
  $\alpha = 1 - (1 - 0.05)^{1/48} = 0.00107$



LISA Significance Map for NC SIDS Rate

LISA Cluster Map

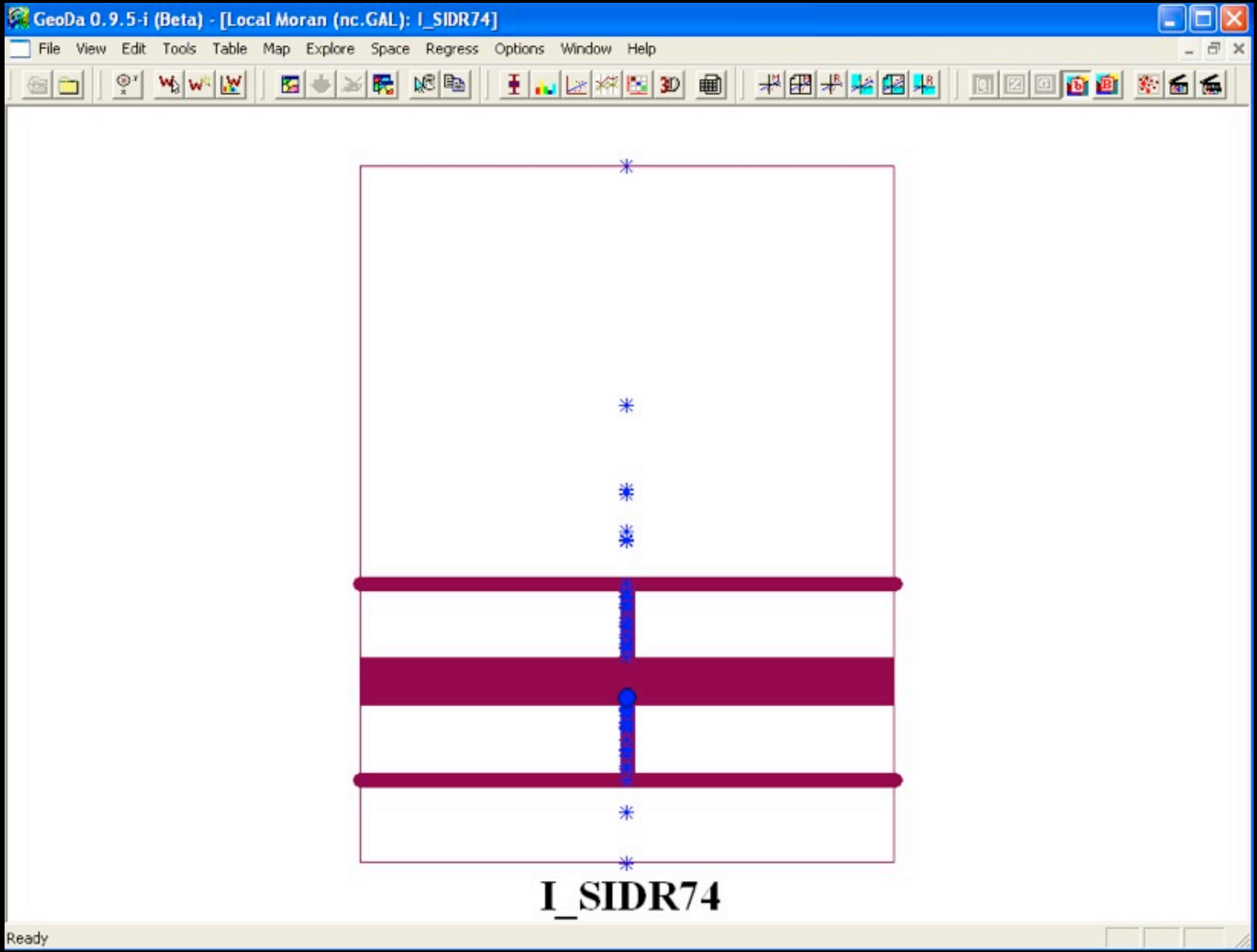
- Only the Significant Locations
  - matches significance map
- Types of Spatial Autocorrelation
  - spatial clusters
    - high-high (red), low-low (blue)
  - spatial outliers
    - high-low (light red), low-high (light blue)



LISA Cluster Map for NC SIDS Rate

LISA Box Plot

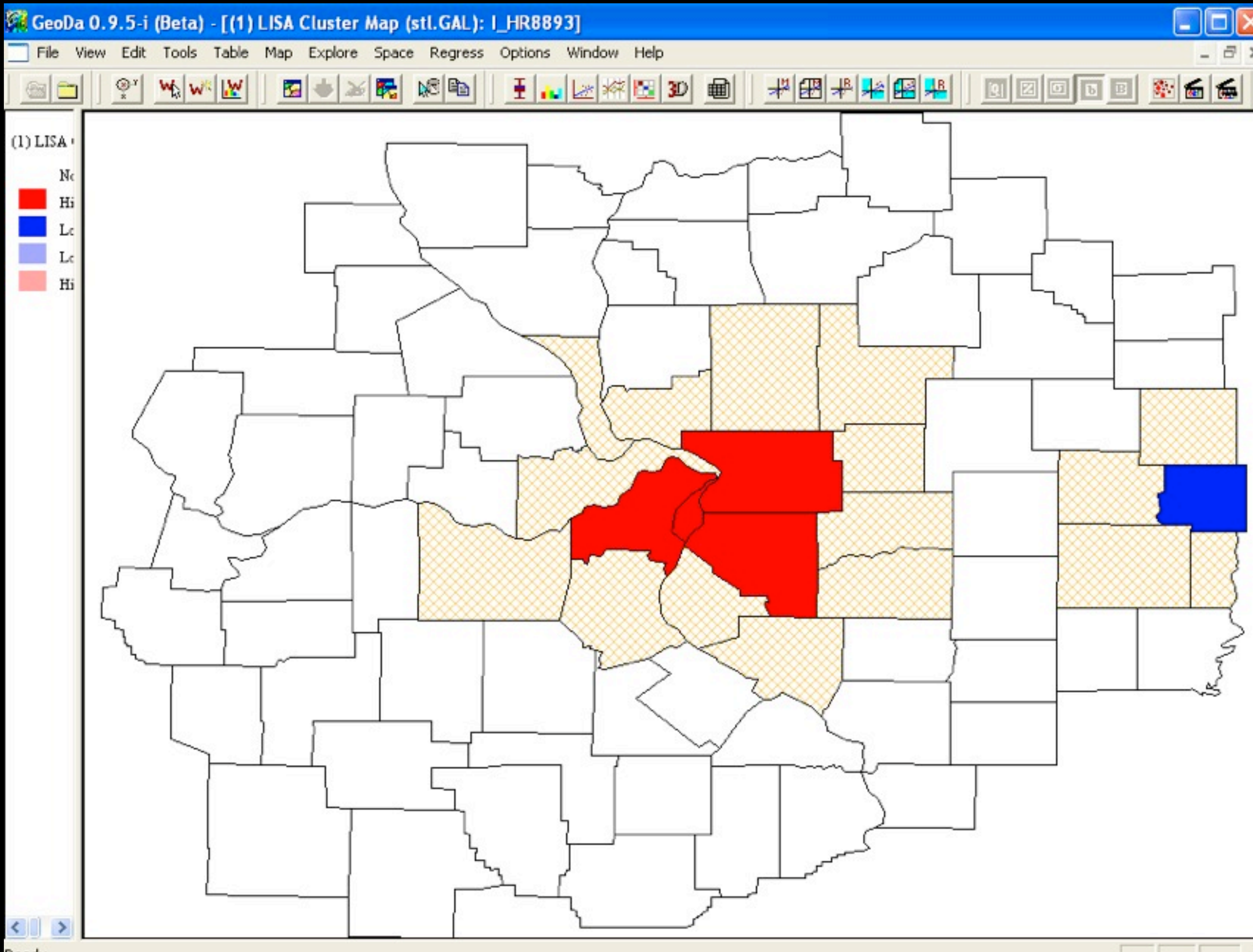
- Distribution of Local Statistics
  - global is related to average of locals
  - check for asymmetry
- Outliers
  - individual locations that may unduly influence the global statistic



LISA Box Plot with Outliers

Spatial Clusters and Spatial Outliers

- Spatial Outliers
  - individual locations
- Spatial Clusters
  - core of the cluster in LISA map
  - cluster itself also includes neighbors
  - use $p < 0.001$ to identify meaningful cluster cores and their neighbors



Spatial Cluster Cores and Neighbors

Caveats

- LISA Clusters and Hot Spots
  - suggest interesting locations
  - suggest significant spatial structure
  - do not explain
- Need to Account for Multivariate Relations
  - univariate spatial autocorrelation due to other covariates
  - scale mismatch

Gi Statistics

Distance Statistics

- Getis-Ord G_i and G_i^*
  - contiguity as distance bands
  - focused: on a given location
  - one statistic for each location
- Not LISA in strict sense
  - does not add up to global

G Statistics

- G_i Statistic
  - $G_i = \sum_j w_{ij}(d)y_j \,/\, \sum j\ y_j$
  - Note: denominator is constant over i
  - $w_{ii} = 0$
- G_i^* Statistic
  - $G_i^* = \sum_j w_{ij}(d)y_j \,/\, \sum j\ y_j$
  - same as G_i but $w_{ii} \neq 0$, i is included in the numerator

Interpretation

- Local Spatial Autocorrelation
- Positive
  - significant: clustering of high values
- Negative
  - significant: clustering of low values
- Inference
  - analytical: randomization assumption
  - computational: permutation

Visualization

- Map of Significant Locations