A Visual Approach to Data Mining Spatial and Temporal Change

Grant Fraley, Piotr Jankowski, & Cristiano Giovando

Department of Geography, San Diego State University

fraley@rohan.sdsu.edu

A Visual Approach to Data Mining Spatial and Temporal Change

- Introduction
- Clustering Change Values
- Visual Data Mining
- Conclusion
- Discussion

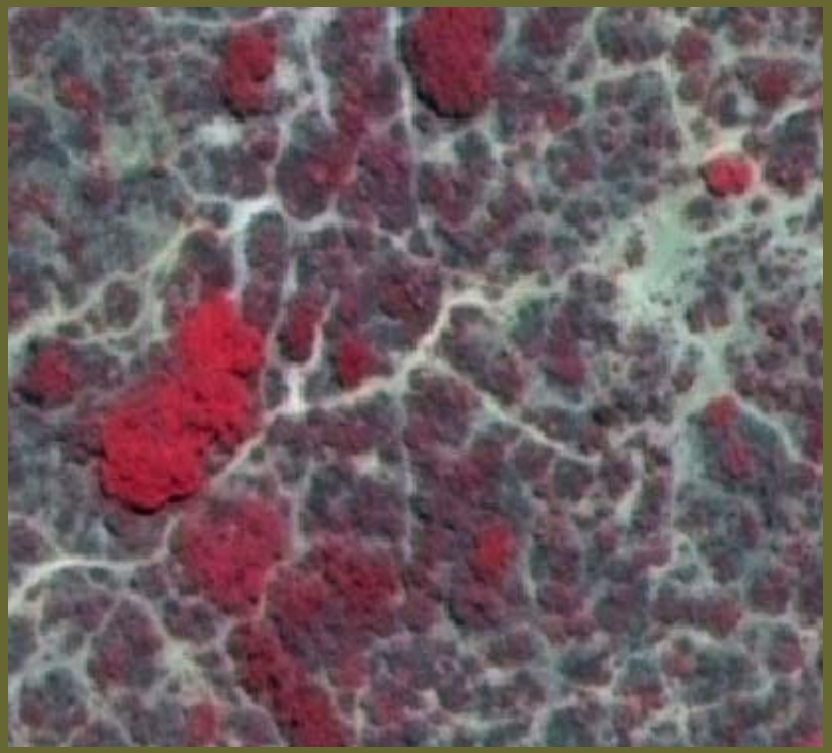
Introduction: Immigration

- Example Analytic Inquiry:
  - What are the salient factors of Mexican emigration (push) to the U.S. this year?
  - Where can we expect an increase or decrease in illegal immigrant apprehensions over the next couple of years?

Introduction: Immigration

May 2006

June 2006



Introduction: Data

- ENFORCE Integrated Database
  - United States Border Patrol (USBP)
  - >12 million illegal apprehension records
    - 1999 to 2006
- Mexico Census
  - Instituto Nacional Estadistica y Geografia (INEGI)

Introduction:
Immigration Propensity

$$MPI = \frac{D_i / \sum D_i}{P_i / \sum P_i}$$

- Migration Propensity Index (MPI)
  - *D*: destination sector apprehension count
  - *i*: municipio of origin
  - *P*: population

Introduction: MPI Change

- MPI of each Mexican municipio, USBP sector pair
- Fiscal year aggregation of apprehensions
- Absolute difference of all year pairs, 1999 - 2006
- 280 MPI change maps

Clustering Change Values: Spatial Cluster Detection

- Spatial statistics to identify significant geographic clusters of immigration propensity change

Clustering Change Values: Global Spatial Autocorrelation

- Moran's *I*
- Getis-Ord General *G*
- Ripley's *K*

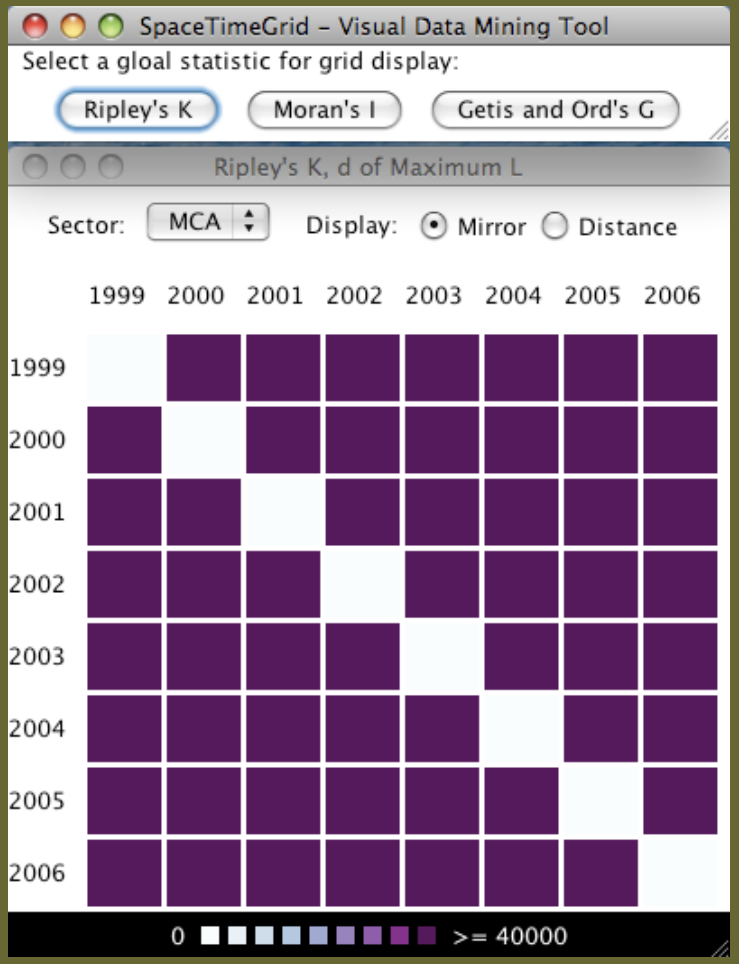
Clustering Change Values: Local Spatial Autocorrelation

- Local Moran's *I*
- Getis-Ord *Gi**

Clustering Change Values: Geographic Computation

- Python + ArcGIS Geoprocessor
- MapServer
  - Python MapScript

Visual Data Mining: Visual Analytic Tool



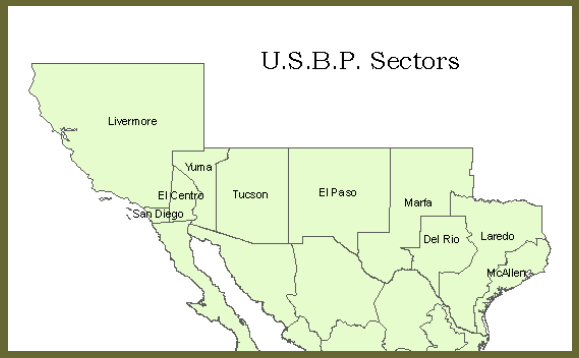
Grant Fraley et al.
Geography, SDSU

A Visual Approach To Data Mining
Spatial and Temporal Change

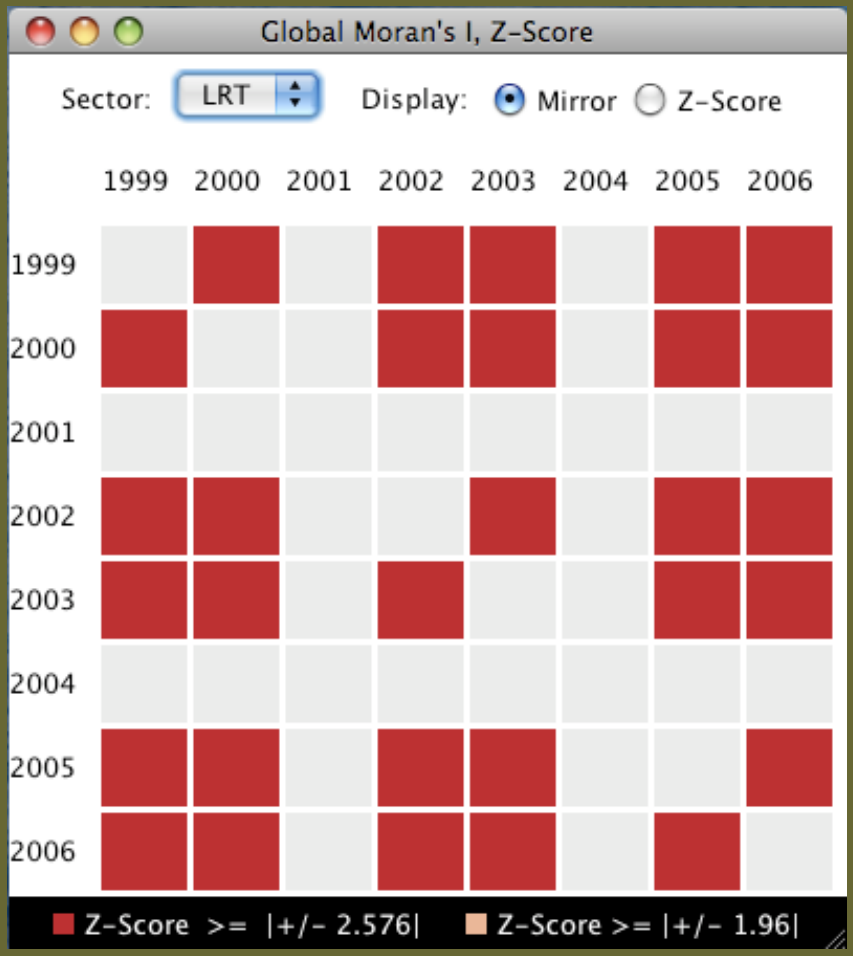
Geospatial Visual Analytics
GIScience 2008

Visual Data Mining: Global Change Matrices

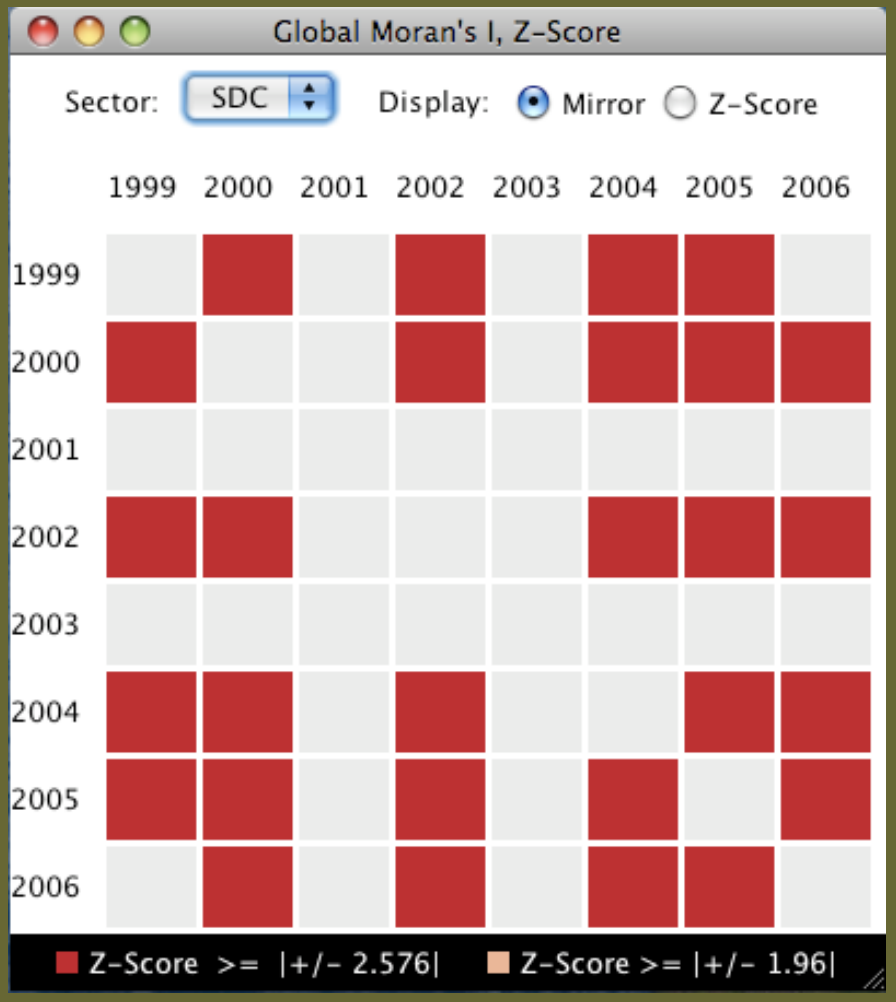
- Temporal Organization
- Grouping Attribute

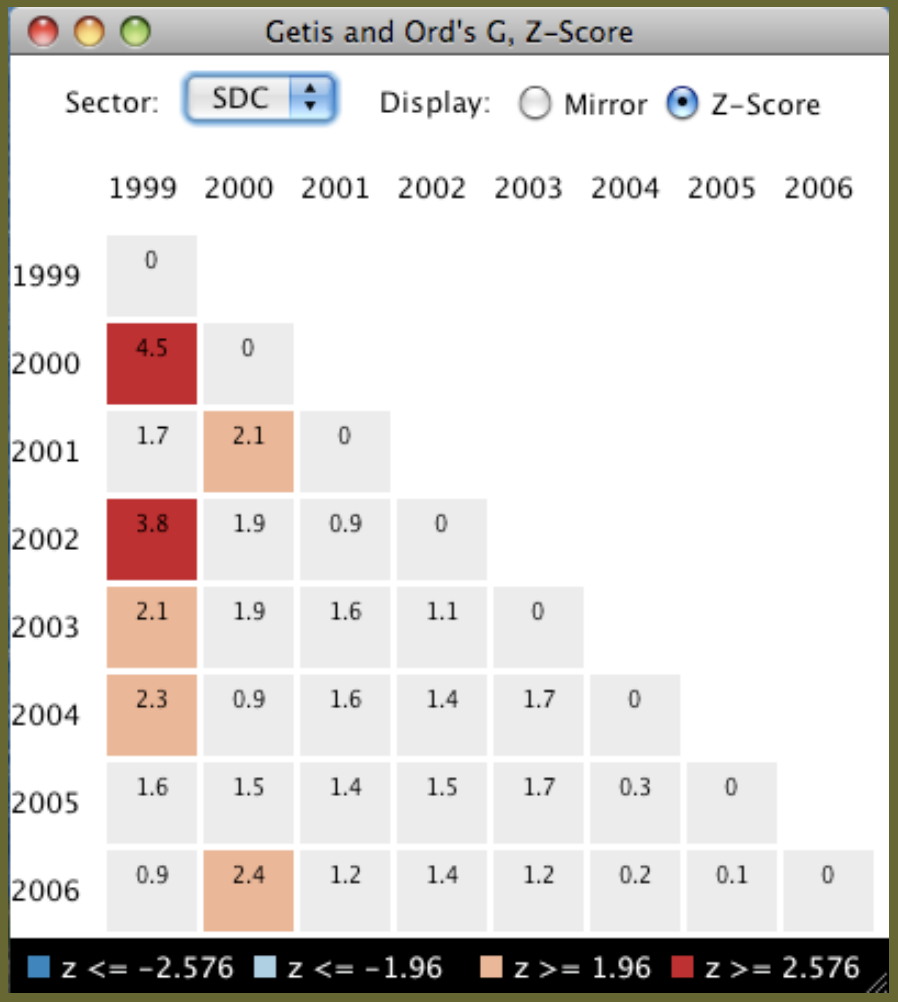


- Data Display
- Linked Map Display



Visual Data Mining: Global Change Matrices



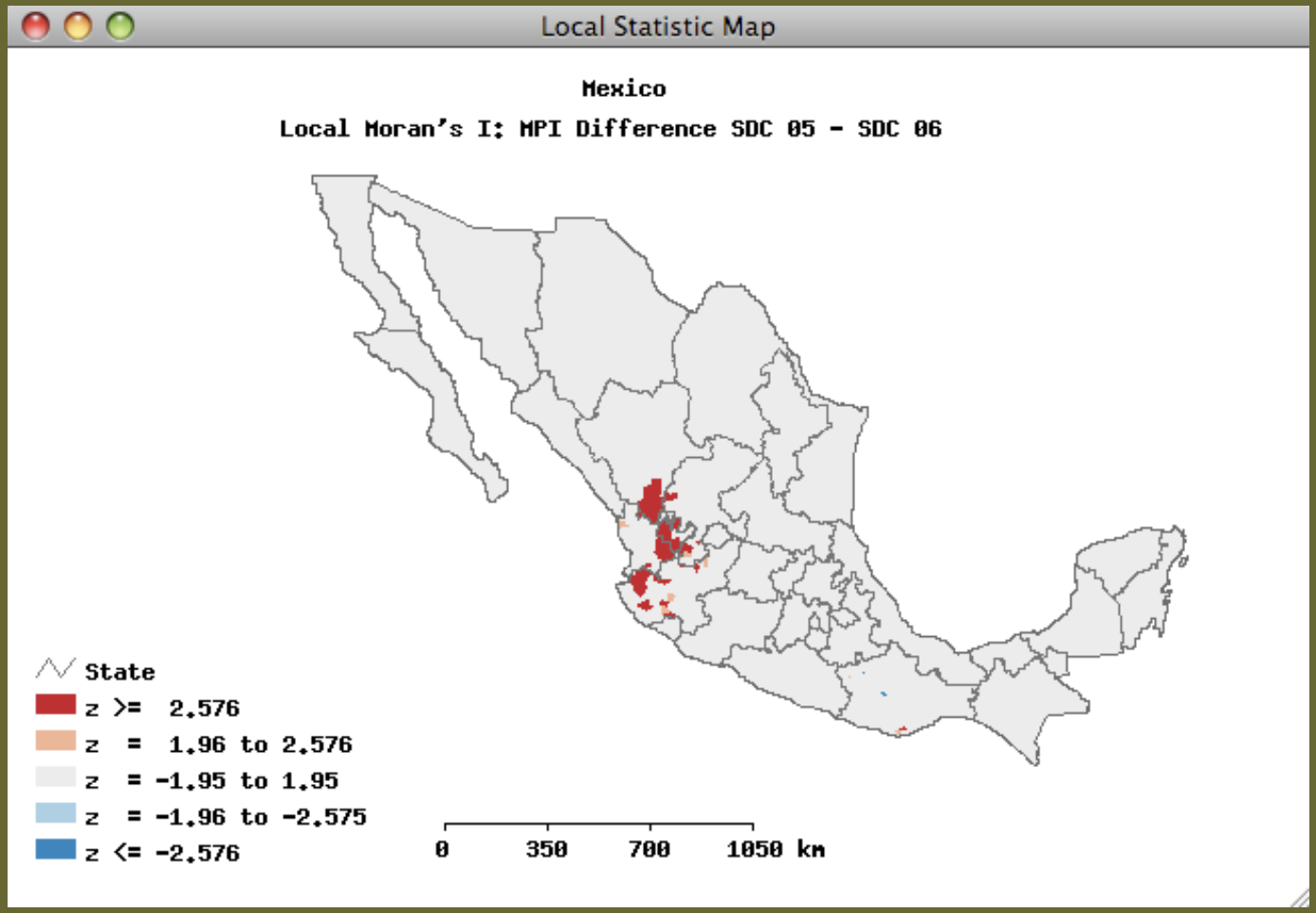


Grant Fraley et al.
Geography, SDSU

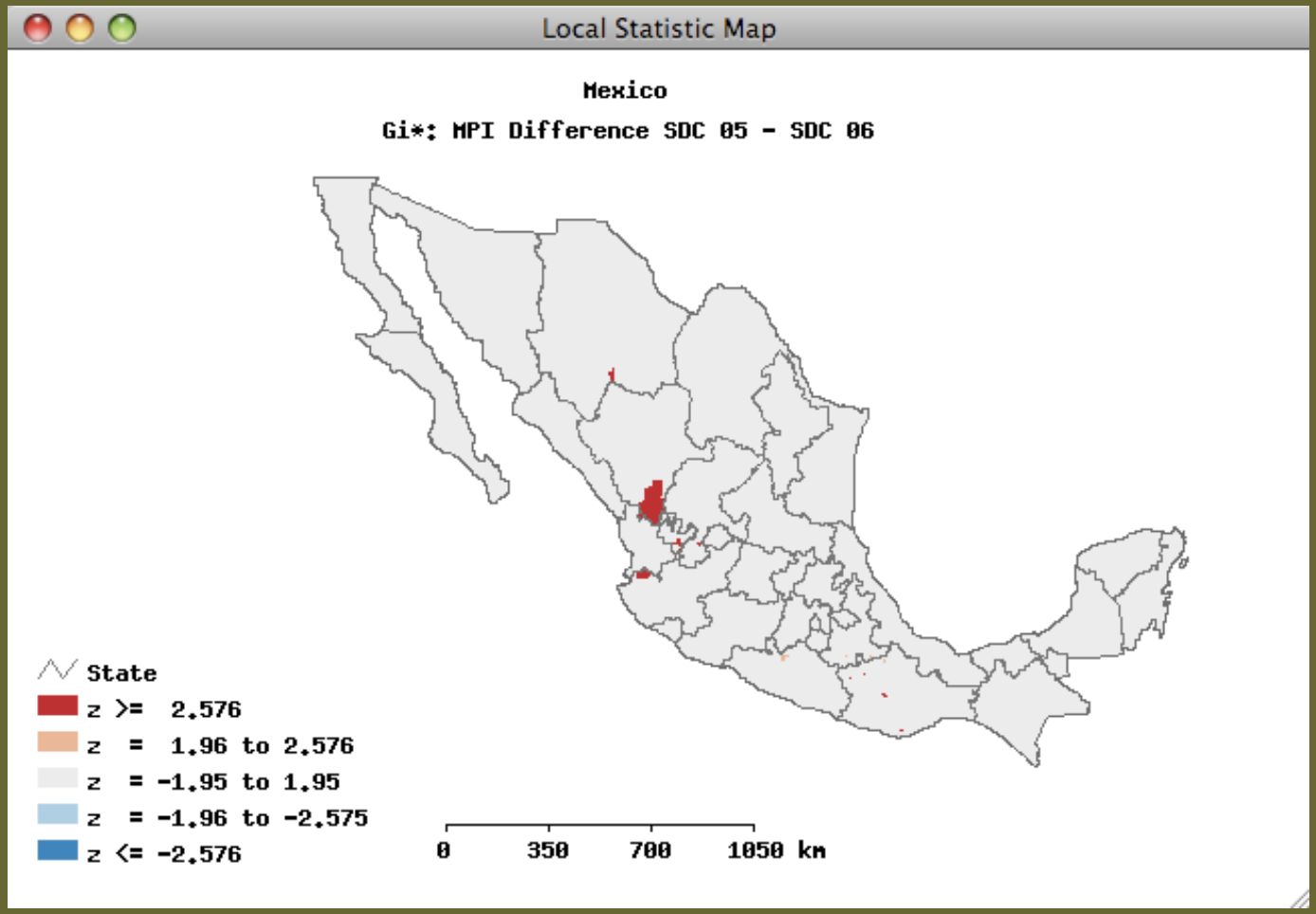
A Visual Approach To Data Mining
Spatial and Temporal Change

Geospatial Visual Analytics
GIScience 2008

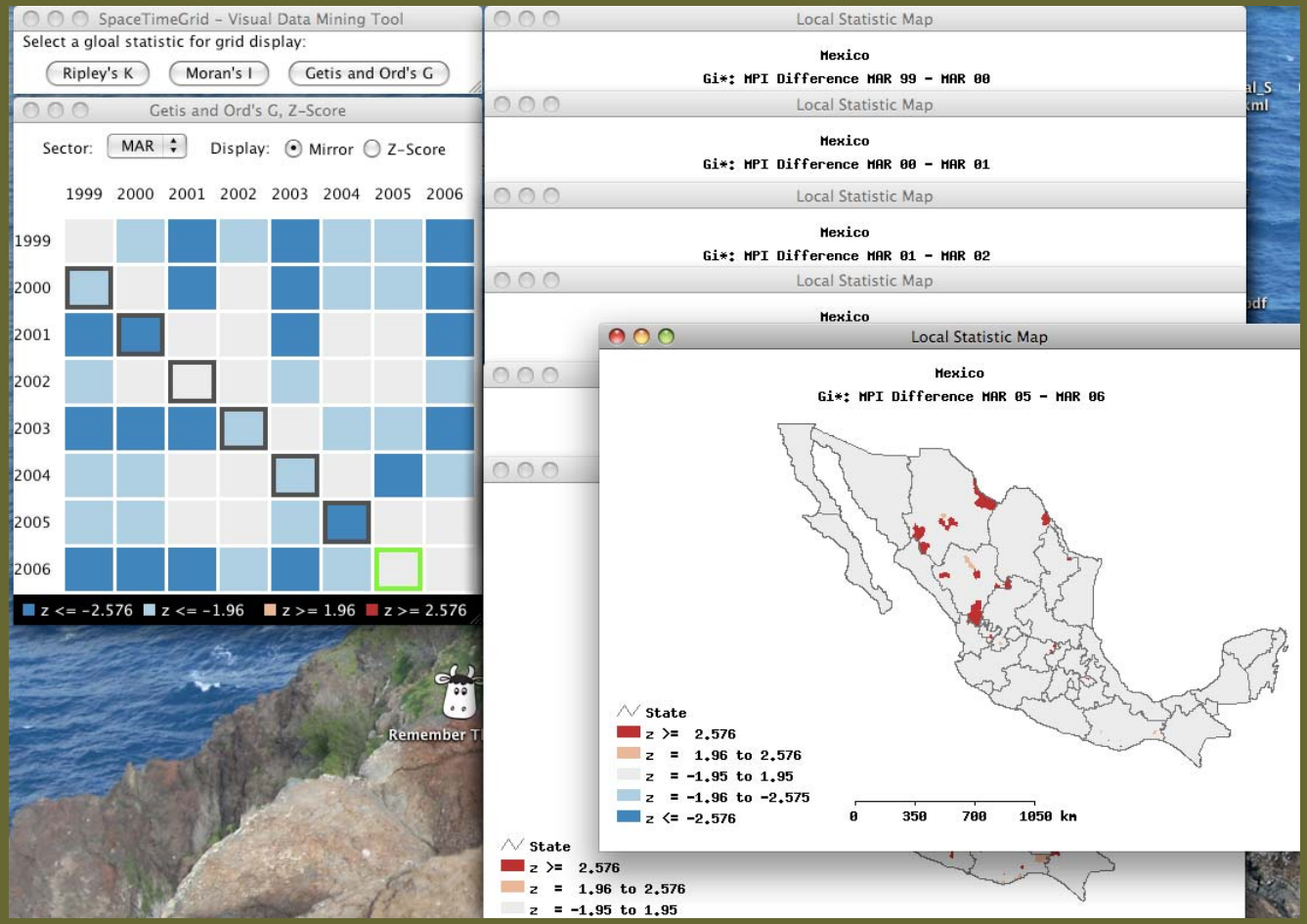
Visual Data Mining: Change Cluster Maps



Visual Data Mining: Change Cluster Maps



Visual Data Mining: Interactive Inquiry



Grant Fraley et al.
Geography, SDSU

A Visual Approach To Data Mining Spatial and Temporal Change

Geospatial Visual Analytics
GIScience 2008

Visual Data Mining: Further Inquiry

- Investigate direction of change
- Proceed with analysis at finer temporal lags
- Contextualize with demographic data or expert knowledge

Discussion

- 1. Permitting temporal lags of less than one year would, for this analysis, result in several more matrices and maps. Questions including within year variability could be addressed, yet computation and interpretation investments would increase. In data mining, we don't necessarily know the appropriate temporal aggregation unit (lag) for analysis. Similar to our choice of a global statistic, Ripley's K, to guide the spatial lag parameter of local spatial statistic calculation, how might we visually or computationally introduce a choice of temporal lag parameter for attribute change calculation in the tool?

- 2. Immigrant destinations are not emphasized in the user interface of the tool, but rather presented as a drop-down menu. Not emphasizing the spatial arrangement of destinations permits the tool to be directly applicable to spatio-temporal phenomenon that do not have an origin-destination component. However, does this design present a significant limitation for data mining patterns of immigration?