GEOG 5330: Applied Spatial and Spatiotemporal Data Analysis

Introduction

Guofeng Cao

http://www.spatial.ttu.edu

Department of Geosciences
Texas Tech University

guofeng.cao@ttu.edu

Fall 2017

Components of Geospatial Analysis and Modeling

- Data do not equal information
- Components of spatial analysis (geospatial data in particular)
  - Visualization: Showing interesting patterns (mapping, geovisualization)
  - Exploratory spatial data analysis: Finding interesting patterns
  - Spatial modeling, regression: Explaining interesting patterns

Topics

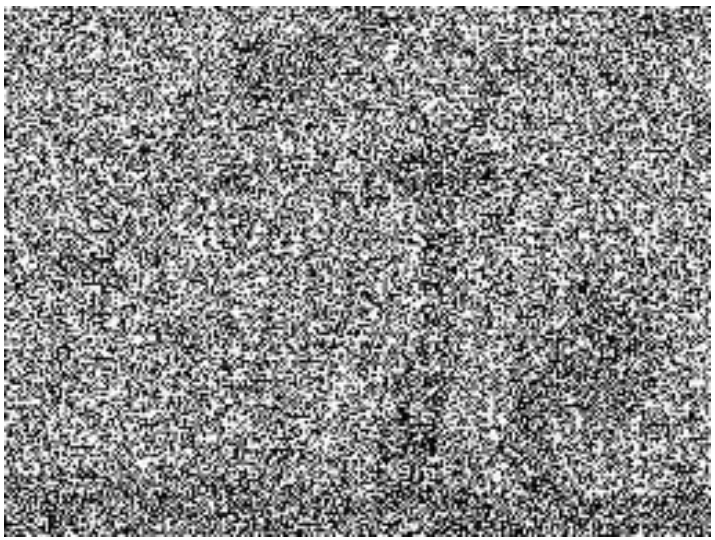
- Spatial data representation and manipulation
  - R Basics
  - GIS using R
- Point pattern analysis
  - Species distribution modeling (e.g., MaxEnt)
- Areal data analysis
  - Exploratory analysis for cluster detection
  - Bayesian geospatial model
  - Change-of-support
- Geostatistics
  - Kriging family of methods
  - Model-based Geostatistics
  - Space-time Kriging
- Time series analysis
  - Remote sensing imagery
  - Google Earth Engigne
- Characterization and quantification of geospatial uncertainty

Characteristics of Geospatial Data

1. Spatial (and temporal) Context: "Everything is related to everything else, but near things are more related than distant things"
   - Waldo Toblers First Law (TFL) of geography
   - nearby things are more similar than distant things
   - phenomena vary slowly over the Earth's surface
   - Compare time series

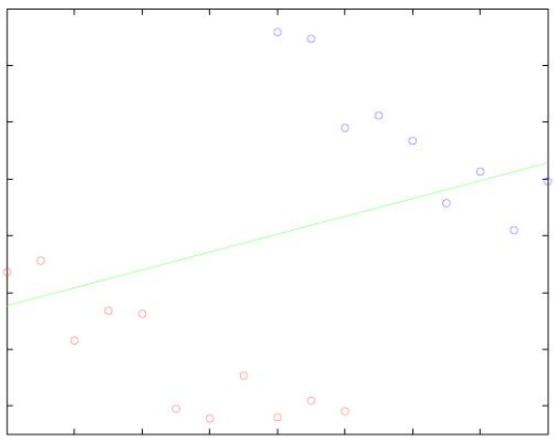
Characteristics of Geospatial Data

- Implication of Tobler's First Law (TFL)
  - We can do samplings and fill the gap using estimation procedures (e.g. weather stations)
  - Spatial patterns
  - Image a world without TFL:
    - White noise
    - No lines, polygons or geometry (how to draw a polygon on a white noise map?)

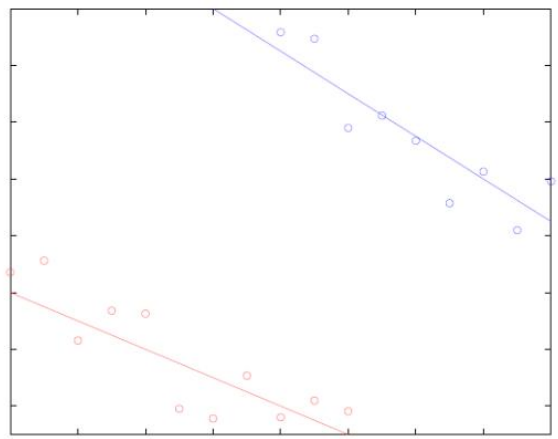


Characteristics of Geospatial Data

2. Spatial heterogeneity
   - "Second law of geography" (Goodchild, UCGIS 2003)
   - Earths surface is non-stationary
   - Laws of physical sciences remain constant, virtually everything else changes
     - Elevation,
     - Climate, temperatures
     - Social conditions
   - Implications
     - Global model might be inconsistent with regional models
     - Spatial Simpsons Paradox (a special case of modified areal unit problem, which we will discuss more in the later of this class)



(a) Global Model



(b) Regional Models

Characteristics of Geospatial Data

Side note: example of Simpson's paradox

- Simpson's paradox usually fools us on tests of performance in real life
- The following is a real life example. Comparison of recovery rates between a new treatment and a traditional treatment for kidney stones.

	New Treatment	Traditional Treatment
Small Stones	93%(81/87)	87%(234/270)
Large Stones	73%(192/263)	69%(55/80)
All	78%(273/350)	83%(289/350)

- Comparison of batting average of two baseball players:

	1996	1997	Combined
Derek Jeter	25.0%(12/48)	31.4%(183/582)	31.0%(195/630)
David Justice	25.3%(104/411)	32.1%(45/140)	27.0%(149/551)

Characteristics of Geospatial Data

- In a spatial settings, it is related to modified areal unit problem (MAUP) or omitted variable problem, which will discuss more in the later of this class

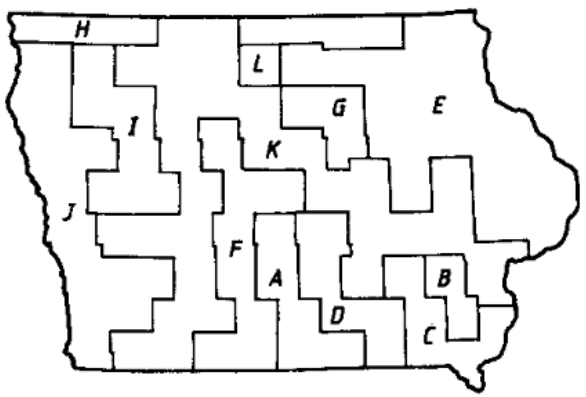


Figure 2a. Zoning system that minimises the regression slope coefficient (-24, r = -.25)

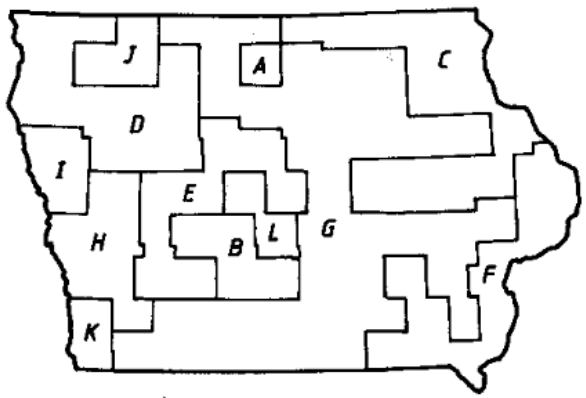


Figure 2b. Zoning system that maximises the regression slope coefficient (12, r = .87)

Figure: Image Courtesy of OpenShaw

Characteristics of Geospatial Data

3. Fractal behavior
   - What happens as scale of map changes?
   - Coast of Maine
- Implications
  - Scale is critical for the problem of study
  - Volume of geographic features tends to be underestimated
    - length of lines
    - area of polygons
  - Think of the difference of distances that an ant and elephant needed to travel from where I stand to the center of memorial circle

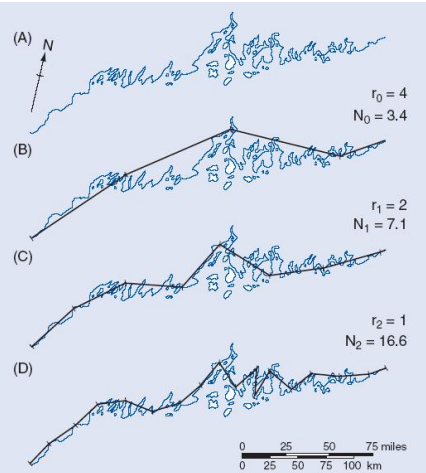


Figure 4.18 The coastline of Maine, at three levels of recursion: (A) the base curve of the coastline; (B) approximation using 100 km steps; (C) 50 km step approximation; and (D) 25 km step approximation

Characteristics of Geospatial Data

Summary: three interrelated characteristics of geospatial data

- spatial context/pattern/structure/dependence/texture..
- spatial heterogeneity/locality
- Fractal behaviors/scaling effects

Elements of Geospatial Data

Elements

- Georeferenced measurements (point or area/region specific samples) Spatial arrangement: regular or irregular (gridded or scattered sampling locations)
- variables/attributes: continuous or discrete (e.g., chemical concentration, soil types, disease occurrences)
- auto- and cross-correlation endemic to spatial data (Toblers first law of Geography)

Types of spatial data

- Point pattern data
- Areal data
- Geostatisticla data

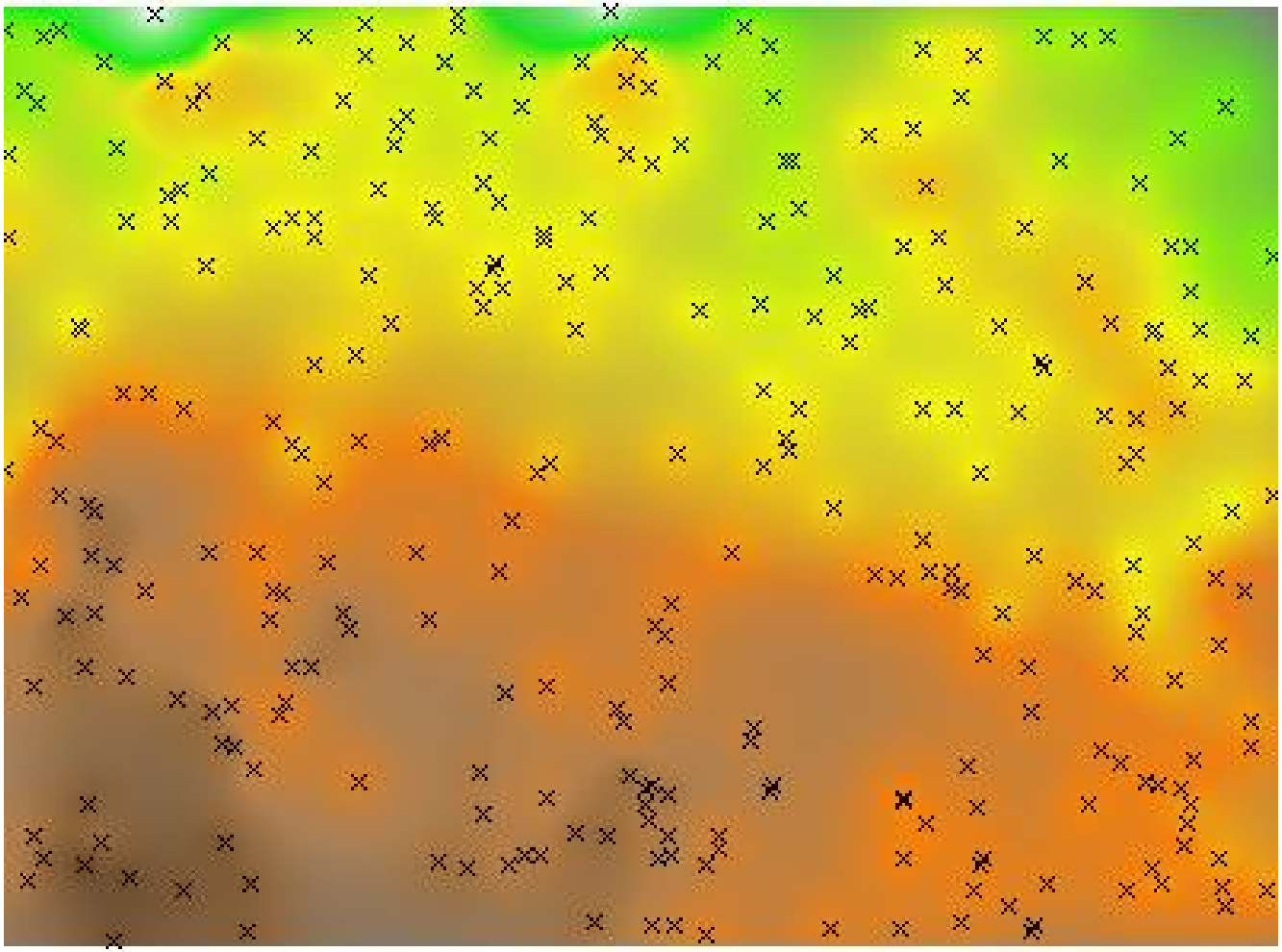
Types of Geospatial Data: Geostatistical Data

Geostatistical data

- Attributes vary continuously in space, e.g., temperature, rainfall, elevation
- Measurements of nominal scale (e.g., soil types), or interval/ratio scale (e.g., depth of boreholes)
- Sampling only at fixed set of locations
- Occurs often in physical-related sciences

Types of Geospatial Data: Geostatistical Data

Example: *300 randomly placed points*

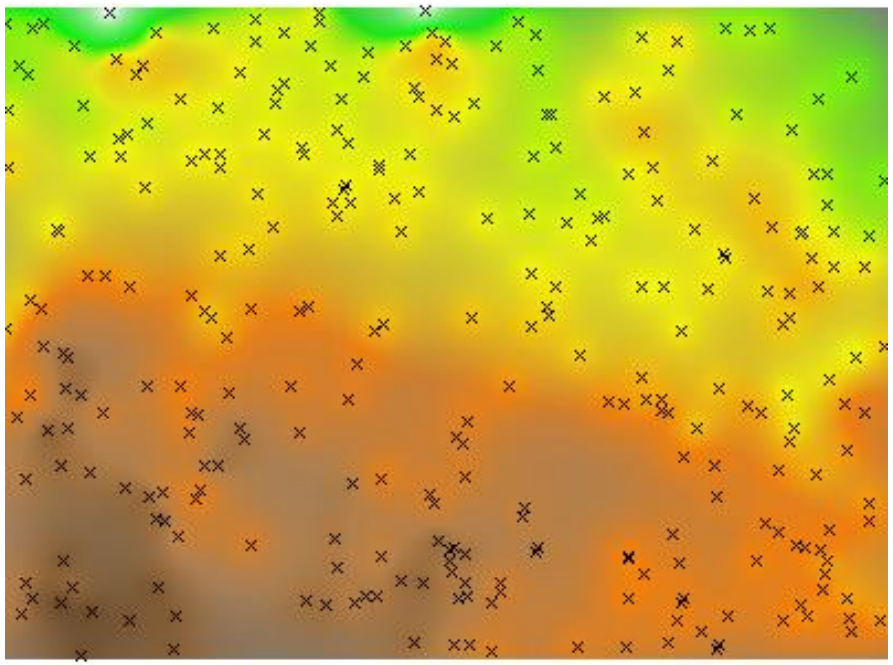


Types of Geospatial Data: Geostatistical Data

Objective

- Mapping spatial variations of regional variables
- Make estimation at unsampled locations

Example: elevation surface generated from 300 points

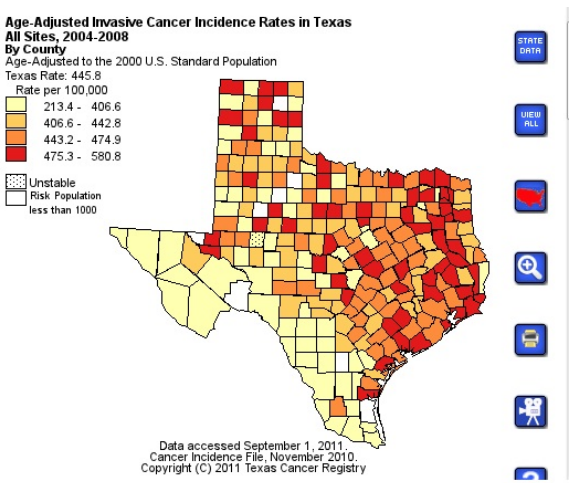


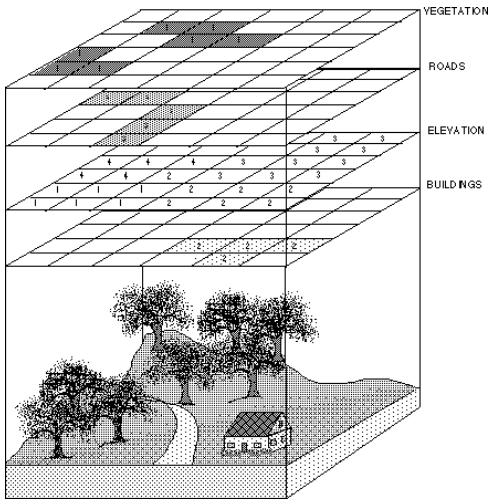
Types of Geospatial Data: Areal Data

Areal (lattice) data

- attributes take values only at fixed set of areas or zones, e.g., administrative districts, pixels of satellite images
- Attributes distribute homogeneously within a region
- Lattice or uniform raster data could be taken as a special case of this type of data

Example:



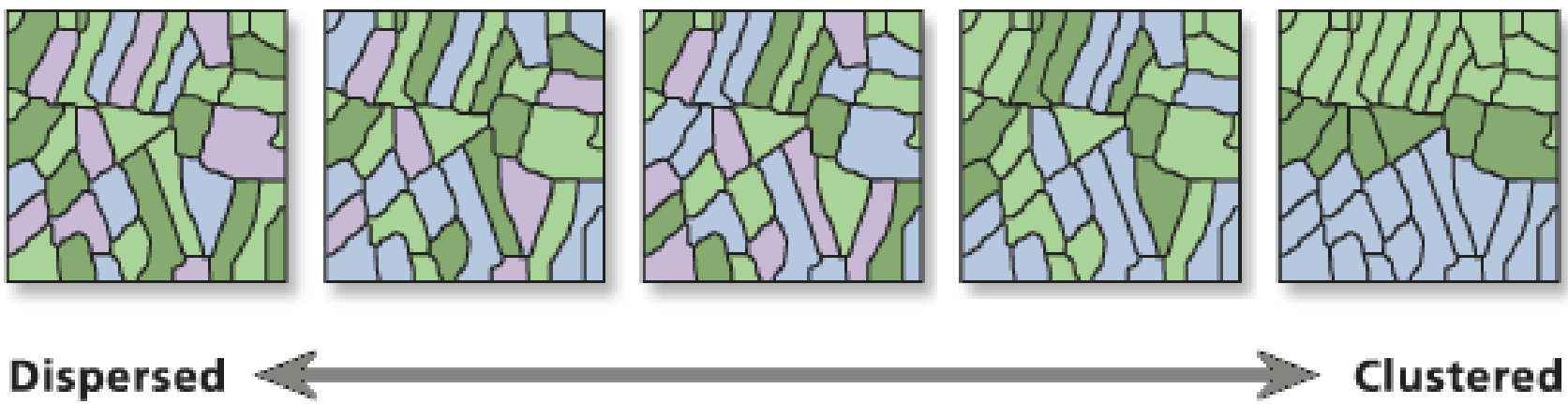


Types of Geospatial Data: Areal Data

Objective

- Detect and model spatial patterns or trends in areal values
- Use covariates or relationships with adjacent areal values for inference (e.g., disease rates in light of socioeconomic variables)

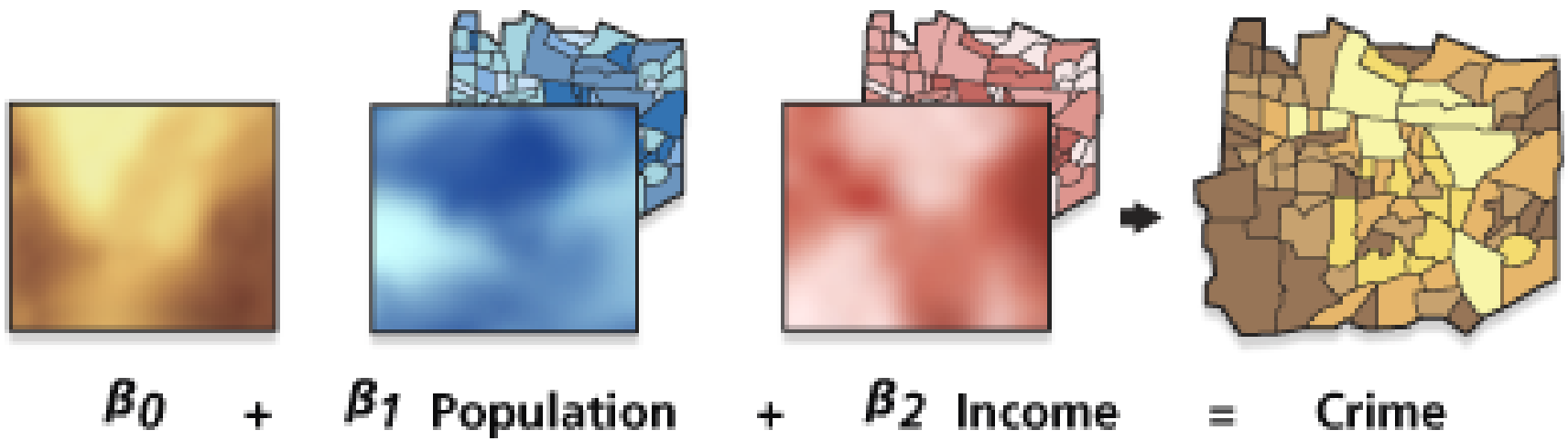
Example:



Types of Geospatial Data: Areal Data

Example 2: *find the correlation among maps*

- It is analog to the cases in traditional statistics, but each variable is (multidimensional) 'maps' instead of single 'numbers'



Types of Geospatial Data: Point Pattern Data

Point pattern data

- series of point locations with recorded events, e.g., locations of trees, epic centers, disease or crime incidents
- attribute values also possible at same locations, e.g., tree diameter, magnitude of earthquakes (marked point pattern)

Example

Types of Geospatial Data: Point Pattern Data

Objective

- detect clustering or regularity, as opposed to complete randomness, of event locations (in space and time)
- If abnormal clustering detected, investigate possible relations with potential factors, e.g., density of disease occurrences with socio-economic status
- Difference with geostatistical point data

Types of Geospatial Data: Summary

Summary

- Geostatistical data
- Spatial point pattern
- Areal (lattice) data