Lecture 4: Spatial Statistical Foundations

Spatial Data Mining

Instructor: Yiqun Xie

Attribution: Slides modified based on lecture notes from Dr. Xun Zhou (UI) and Dr. Shashi Shekhar (UMN)

Spatial Statistical Foundations

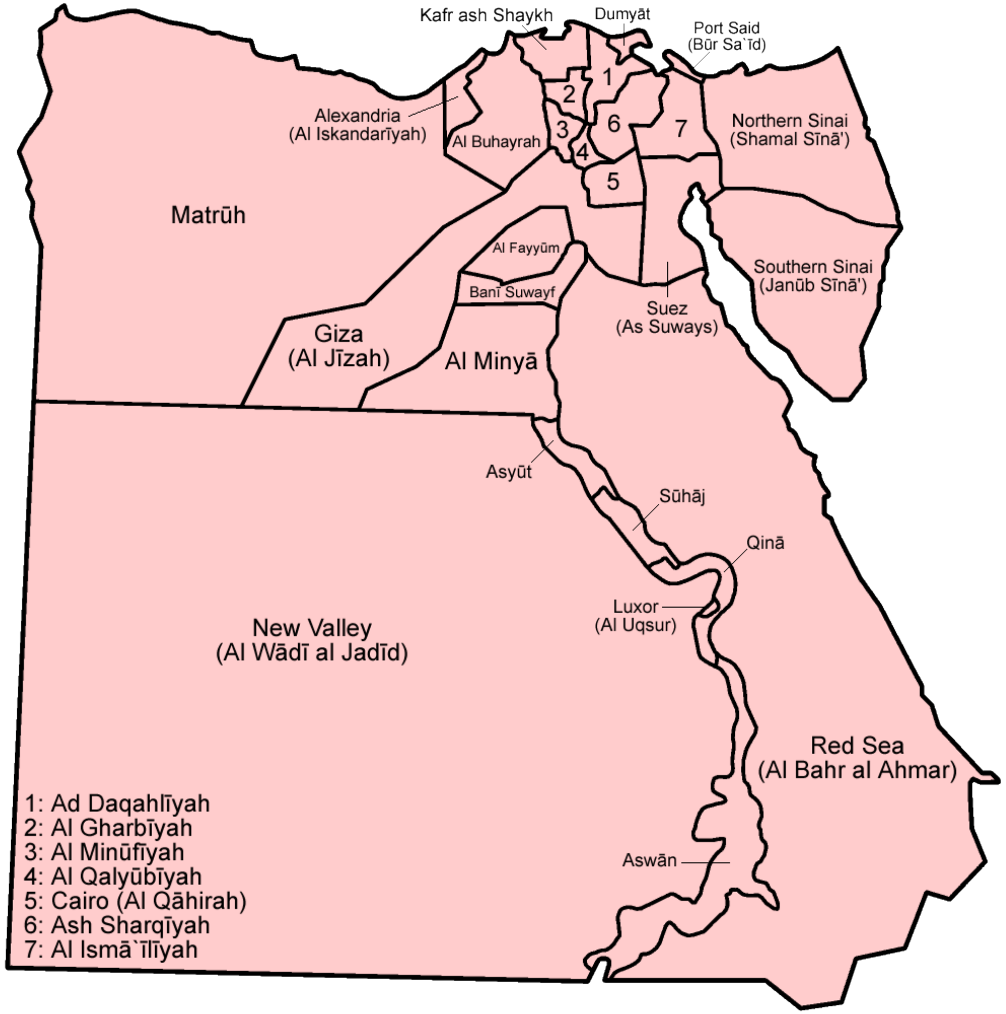
- Statistics
  - The study of collection, analysis, interpretation of data
  - Spatial statistics
    - Statistics for spatial data (point, line, polygon, raster)
    - Unique properties
      - Non i.i.d.
      - Spatial autocorrelation & heterogeneity
      - Isotropy v.s. anisotropy
      - Stationarity v.s. non-stationarity

Overview

- Review of important basic concepts in statistics
  - Know the terminologies and notations
  - There will be some math but don't be scared…
- Discuss different types of spatial statistics
  - Suitable spatial data models, tasks, measures
- In the context of the whole semester
  - More descriptive and exploratory (get the spatial insights)
  - Inform the design & use of data mining techniques

Without Knowing Notations

- Example: Navigation in a foreign language

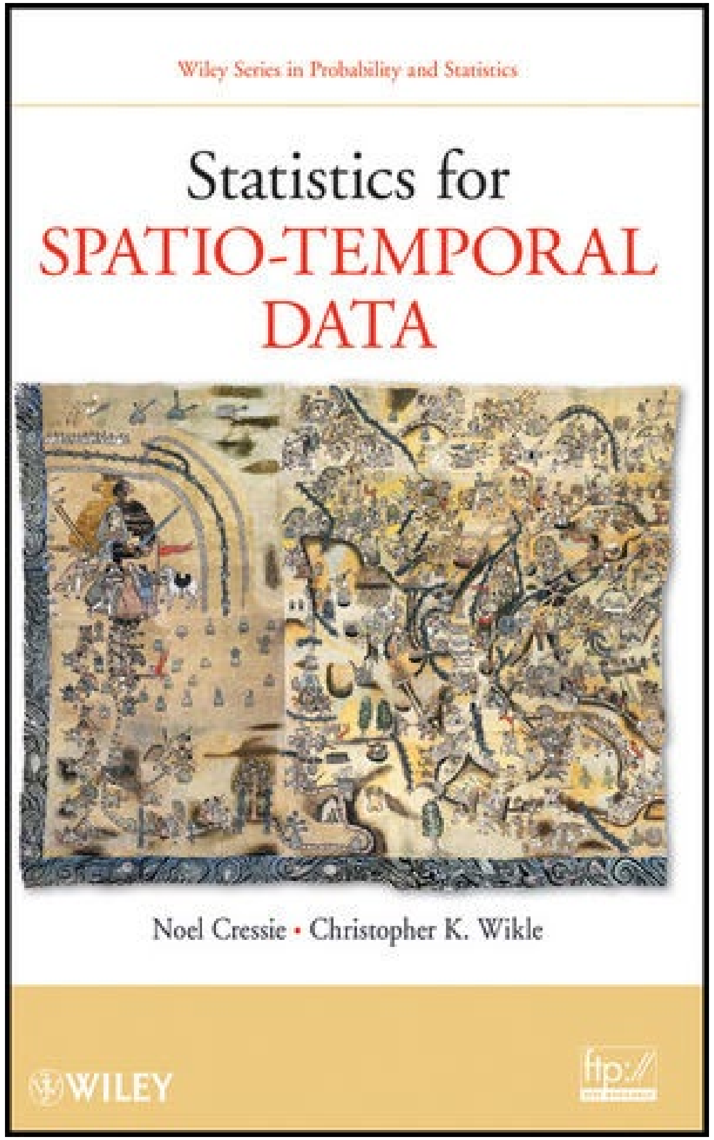


Learning Objectives

- Review of important basic concepts in statistics
  - Know the terminologies and notations
  - There will be some math but don't be scared…
- Know different types of spatial statistics
- In the context of the whole semester
  - More descriptive and exploratory (give spatial insights of data)
  - Inform the design & use of data mining techniques

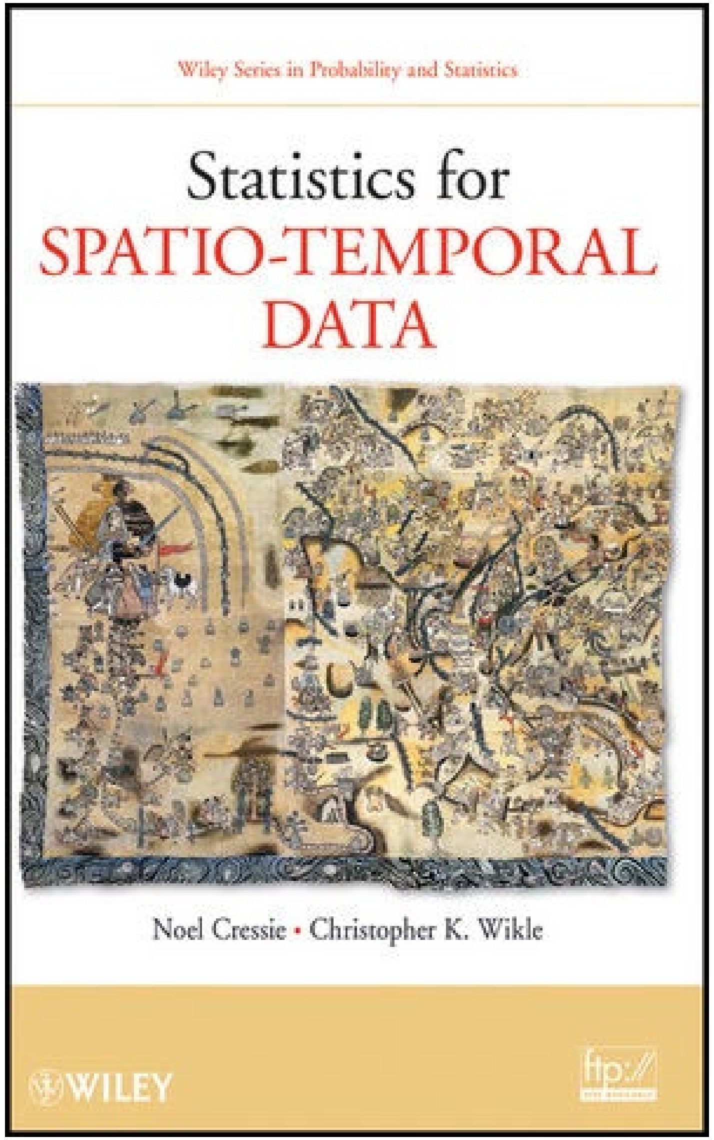
Selection of Topics

- Geostatistics
- Lattice statistics
- Spatial point processes



Selection of Topics

- Measure spatial autocorrelation
  - Relationship between observed values as their spatial relationship (e.g., distance) changes
  - **Geostatistics** → Taking samples from a continuous phenomenon
  - **Lattice statistics** → Discretize (partition) a continuous space
- Measure relative relationship between locations
  - **Spatial point processes**

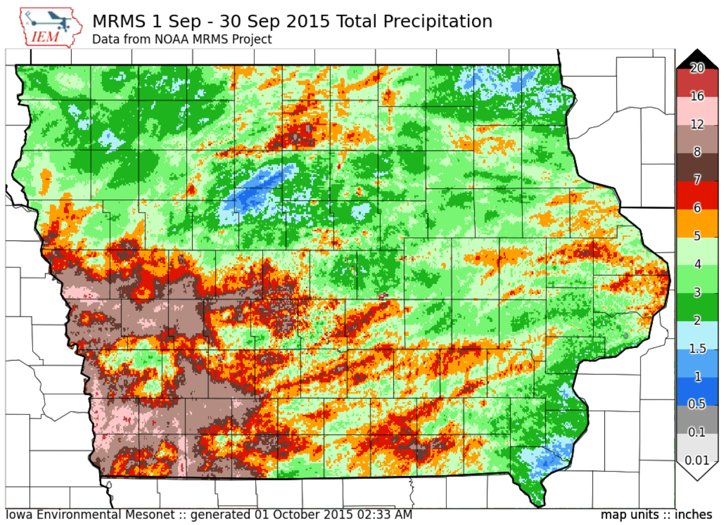


Spatial Autocorrelation

- Tobler's First Law of Geography
  - Everything is related to everything, but nearby things are more relevant than distant things.
- How to model "difference"?
- How to model "nearby" and "distant"?
- Geostatistics and Lattice models use different strategies

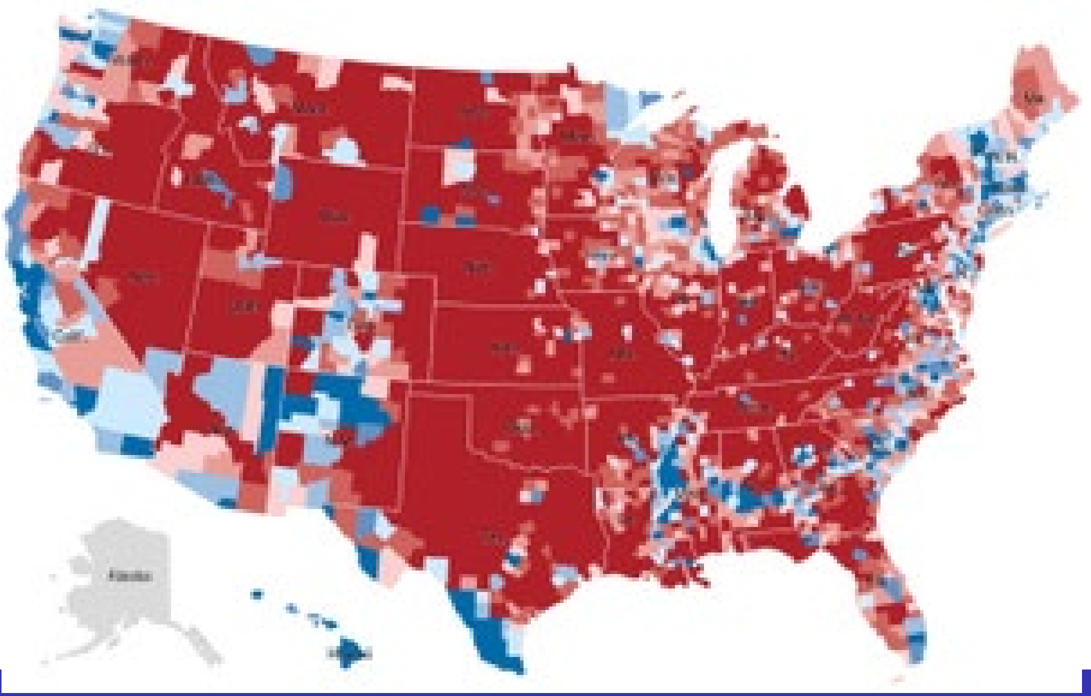
Statistical Models (1): Geostatistics

- Geostatistics
  - A stochastic process Y(s): s ∈ D, D is a r-dimensional Euclidean space
  - Example, the rainfall of the entire Iowa
  - Continuous process over the space
  - Observations at discrete locations
- Used for
  - Exploratory data analysis
  - Spatial interpolation



Statistical Models (2): Lattice Statistics

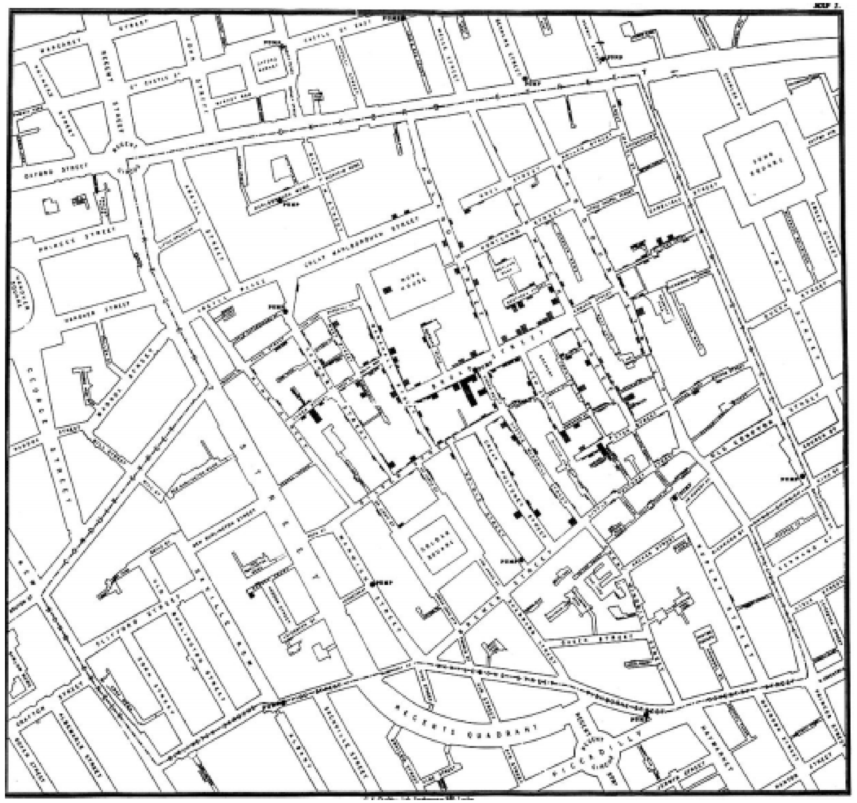
- Areal data model
  - A tessellation of continuous space into (regular or irregular) cells
  - Mapping each unit to a non-spatial attribute value
  - Example: 2016 President Election
  - Y(s): s ∈ D, D is a set of cells
- Used for
  - Spatial pattern discovery
  - Spatial prediction



Statistical Models (3): Point Process

- Spatial point process
  - $\{s_1, s_2, \ldots s_n\}$ s_i are event locations with fixed event type
  - Disease cases
  - Crime locations
  - Traffic accident locations

1854 Broad Street cholera outbreak



Point Process

- Deaths all clustered around a water pump

Tools and Statistical Methods

- Geostatistics
  - Kriging: spatial interpolation
- Lattice model
  - Spatial regression
  - Spatial autocorrelation measures
  - Markov Random Field
- Point Process Model
  - Ripley's K function
  - Spatial scan statistics
  - Deep learning

Summary of Topics

- Geostatistics
  - Stationarity, variogram, Kriging
- Lattice model
  - Moran's I, Geary's C, LISA

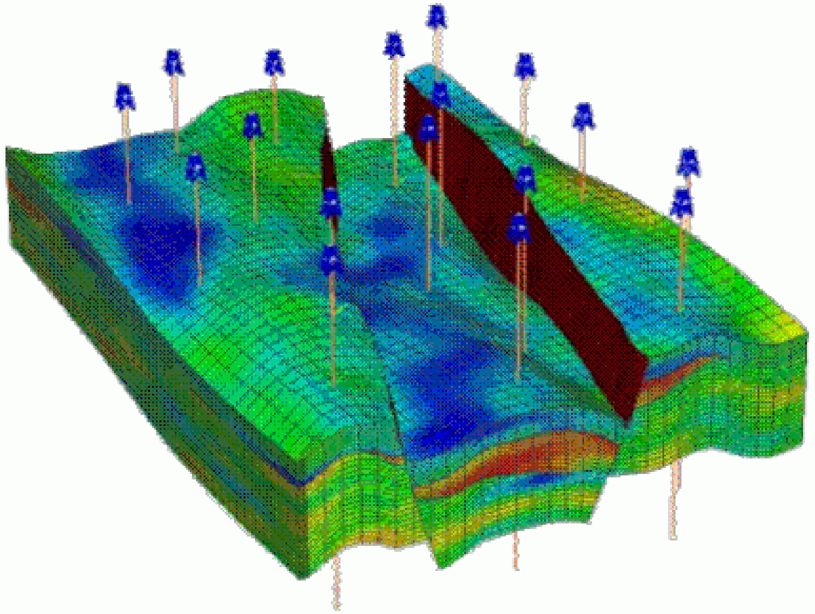
Understand and measure spatial autocorrelation (observed values at locations)

- Point Process Model (preview)
  - Ripley's K, spatial scan statistics

Relationship of locations themselves

Geostatistics

- Also called point-referenced data
  - Estimate precipitation based on records at a set of weather stations
  - Infer ground water level based on sensor readings of a set of gauges
  - Predict mineral resources based on samples at a limited number of sites
- Relationship as a function of location-shift (e.g., distance)
- Statistical assumptions
  - Strict (Strong) stationary
  - Weak stationary
  - Intrinsic stationary



Geostatistics: Stationary

- Strictly Stationary
  - Distribution of value is unchanged with location-shift
  - For $n \geq 1, any\ n\ locations\ \{s_1, s_2, \ldots, s_n\}, h \in R^r,$ r dimensional space of real numbers
    We often have r = 2 or 3 for spatial data
    $\{Y(s_1), Y(s_2), \ldots, Y(s_n)\}\ and\ \{Y(s_1 + h), Y(s_2 + h), \ldots, Y(s_n + h)\}$
    $have\ the\ same\ joint\ distributions$
  - Too strong and unrealistic

- Basic concepts
  - Statistical distribution; joint distribution
  - Probability density function vs. probability mass function
  - Symbol ∈: element of
  - R: real value space; R^r: r dimensional real value space

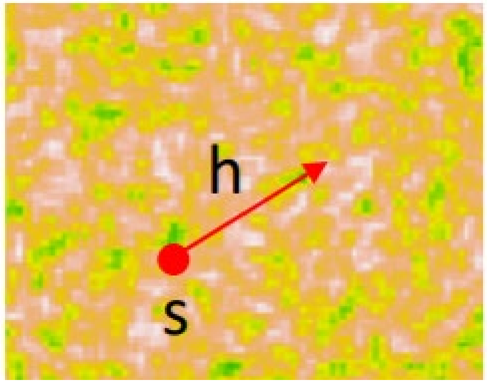
Geostatistics: Stationary

- Weakly stationary
  - Mean unchanged when location shift
  - $E\left(Y(s)\right) = \mu_s =$ constant mean
  - $cov\left(Y(s), Y(s+h)\right) = C(h), h \in R^r$
  - Constant variance: C(0) = Var(s) = constant.
  - The covariance across locations is simply a function of the location shift

- Basic concepts
  - Mean, variance, covariance, covariance matrix
  - Function of …

Variogram

- Intrinsically Stationary
  - The difference between two locations only depends on h
  - $Assuming\ E[Y(s) - Y(s + \boldsymbol{h})] = 0$ (constant mean)
  - $E[Y(s) - Y(s + \boldsymbol{h})]^2 = var[Y(s) - Y(s + \boldsymbol{h})] = 2\gamma(\boldsymbol{h})$
  - $2\gamma(\boldsymbol{h})$ is called variogram. $\gamma(\boldsymbol{h})$ is called **semi-variogram**

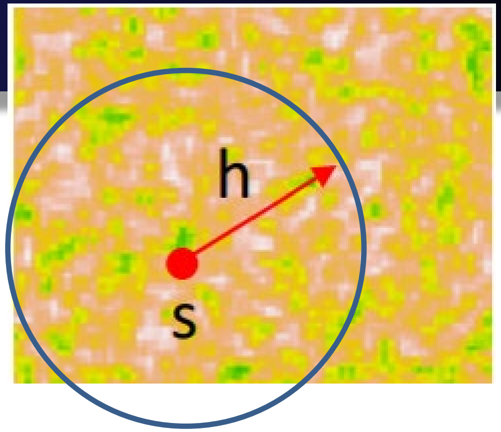


Variogram

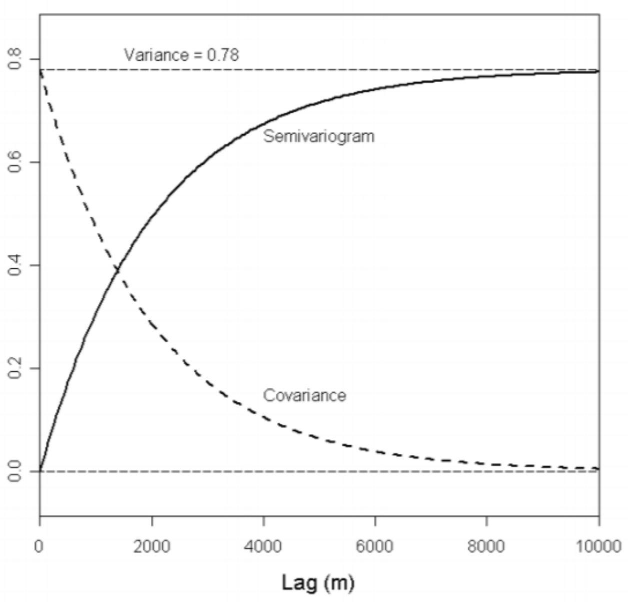
- Intrinsically Stationary
  - The difference between two locations only depends on h
  - $Assuming\ E[Y(s) - Y(s+\boldsymbol{h})] = 0$ (constant mean)
  - $E[Y(s) - Y(s+\boldsymbol{h})]^2 = var[Y(s) - Y(s+\boldsymbol{h})] = 2\gamma(\boldsymbol{h})$
  - $2\gamma(\boldsymbol{h})$ → variogram. $\gamma(\boldsymbol{h})$ → **semi-variogram**.

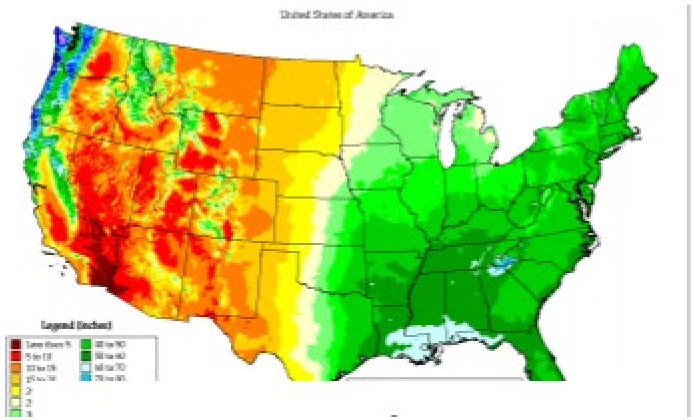
- $2\gamma(h) = var[Y(s) - Y(s+h)]$
  $= Var(Y(s+h)) + Var\big(Y(S)\big) - 2Cov\big(Y(s+h), Y(s)\big) = C(\mathbf{0}) + C(0) - 2C(\boldsymbol{h})$
  $= 2[C(\mathbf{0}) - C(\boldsymbol{h})]$
- $\gamma(h) = C(\mathbf{0}) - C(\boldsymbol{h})$

Semi-Variogram



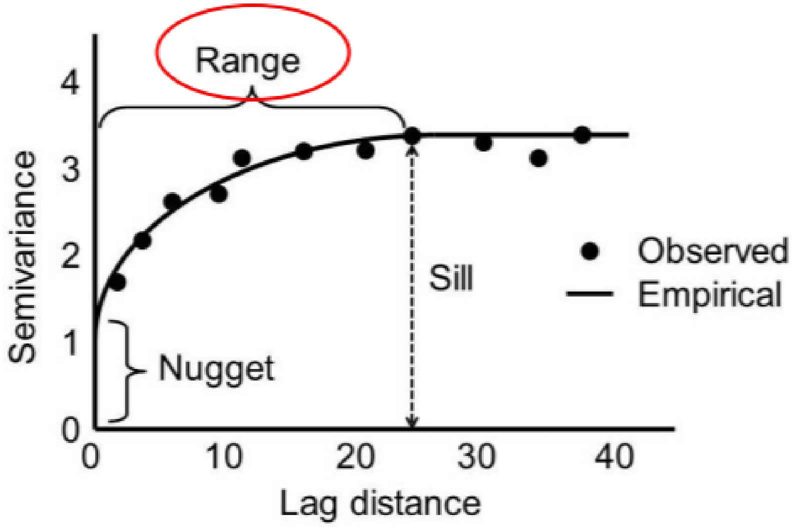
- Isotropy:
  - Assumption: direction does not matter. Only the distance matters
  - Might not be always true.
  - Easy to visualize (draw the curve)
- The semi-variogram has an opposite trend compared to the covariance C(h)
  - Longer lag distance, less correlation (covariance), higher semi-variogram





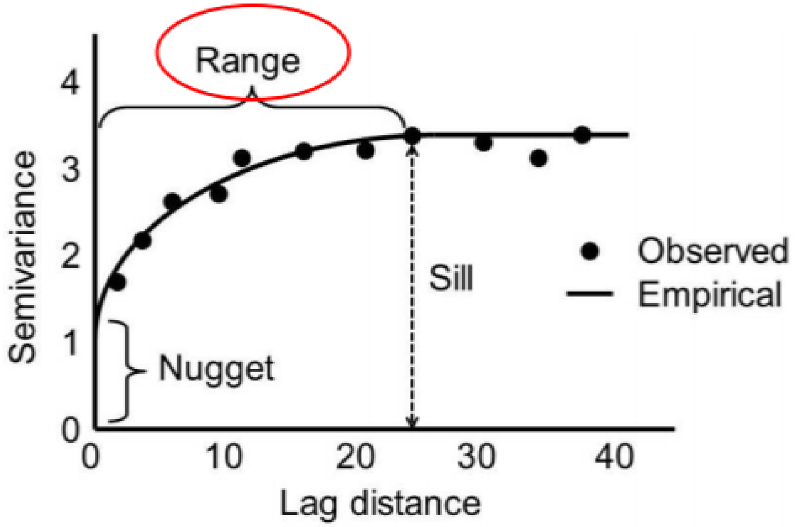
Variogram Plot

- Under the Isotropy and Intrinsic Stationary assumption
  - Smaller h → shorter distance → high covariance (higher correlation), low difference
  - Large h → longer distance → low covariance (lower correlation), high difference
  - Very large h → no effect on the variance or difference. Converged.
- Parameters of the semi-variogram $\gamma(h)$
  - Nugget: the minimum jump close to h = 0. Typical 0.
  - Sill: The $\gamma(h)$ value at which the variogram levels off.
  - Range: the lag h when variogram reaches sill



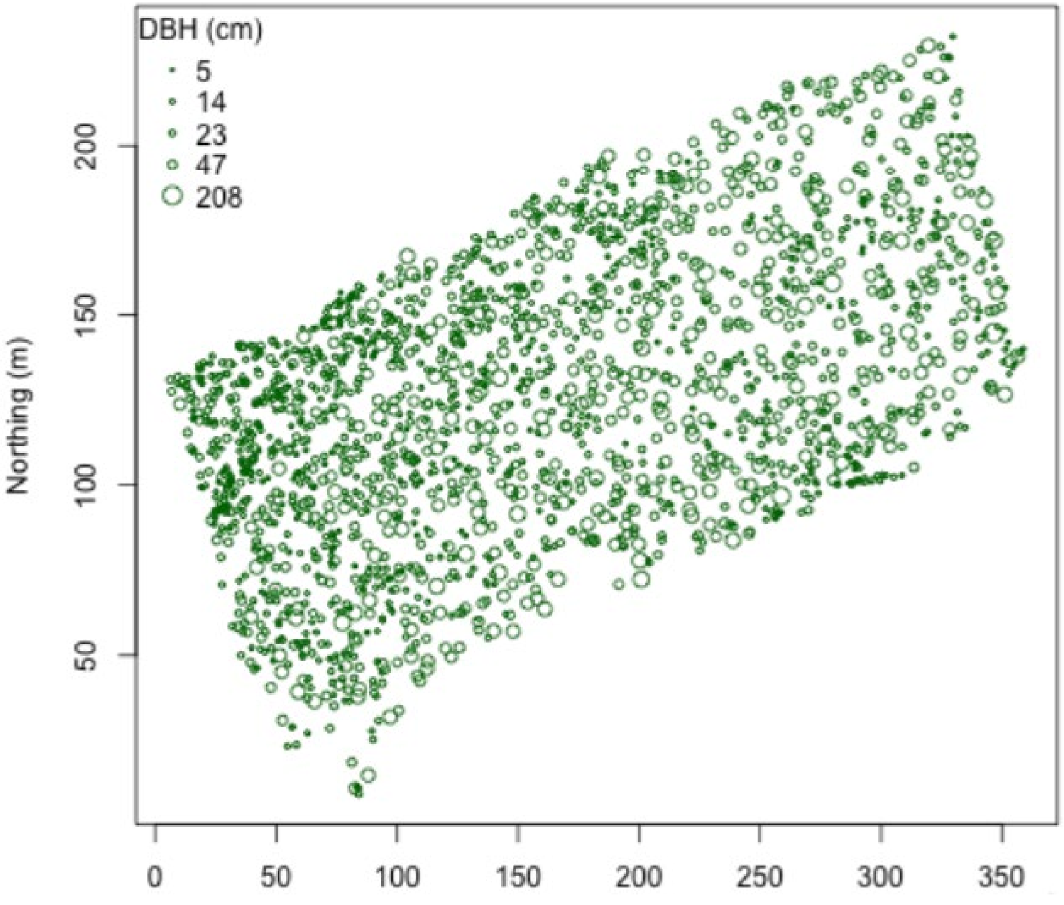
Empirical Semi-variogram

- $\hat{\gamma}(d) = \frac{1}{2|N(h)|} \Sigma_{s_i, s_j \in N(h)} \left[Y(s_i) - Y(s_j)\right]^2$
- For each distance h, calculate the squared difference in value $\left[Y(s_i) - Y(s_j)\right]^2$
  - For every pair of observations in the data with h as their distance (whole set: N(h))
- Plot the points and fit a model (e.g., spherical) with least-squared error
- Get the estimated parameters
  - Nugget
  - Sill
  - Range

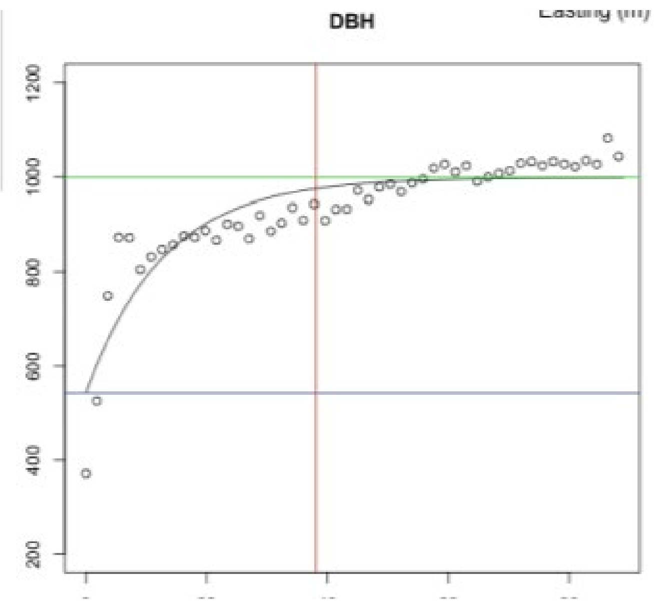


- R package for semi-variogram fitting
  - “gstat” package, “geoR” package
  - fit.variogram()

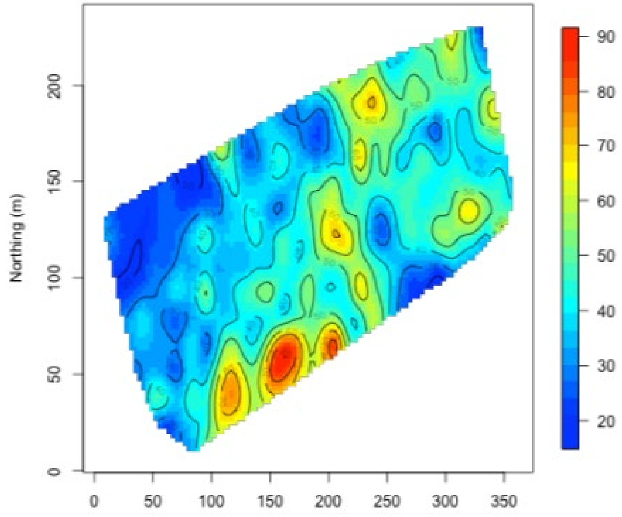
Example



Diameter at breast height on trees



Sem-variogram fitted



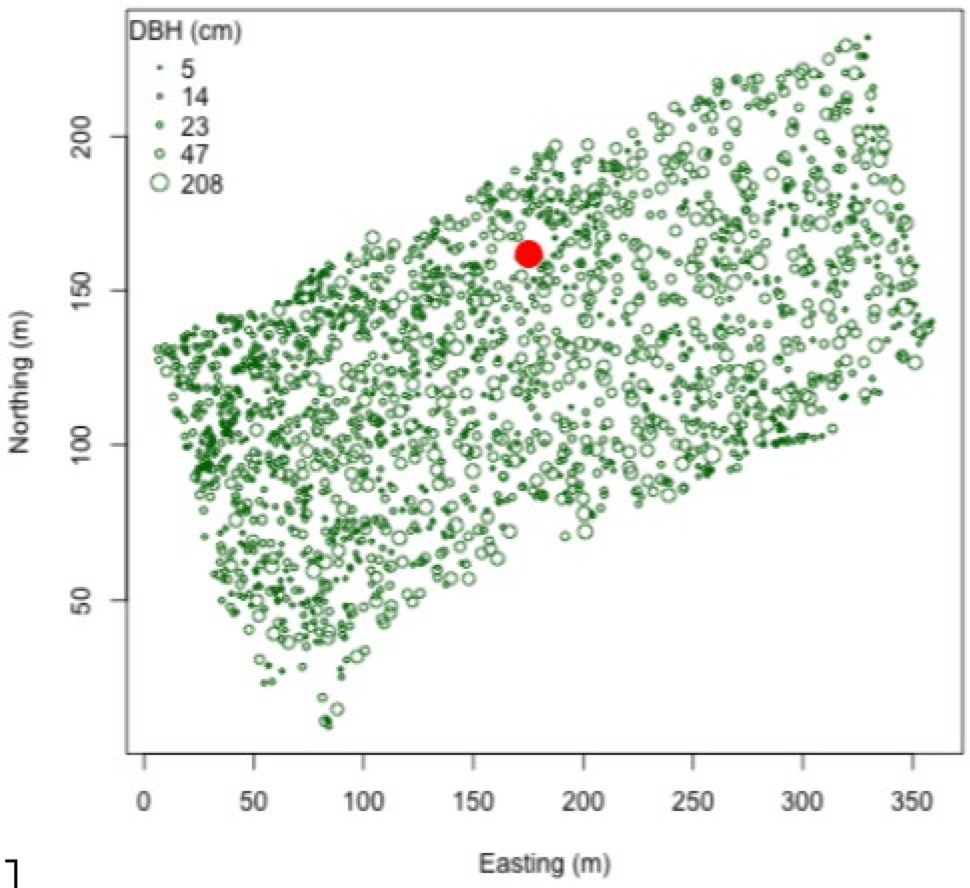
Contour map of DBH

Example Application: Kriging

- Spatial Interpolation / Prediction Model
  - Given observations at a few locations {Y(s1), Y(s2), … Y(sn)}
  - Infer values at a location with unknown value Y(s0)
  - Assumption: intrinsic stationary and a suitable variogram model
- Kriging named after a mining engineer Danie Gerhardus Krige
  - Krige's empirical work to evaluate mineral resources was formalized in the 1960s by French engineer Georges Matheron.
- Ordinary Kriging
  - Only uses the dependent variable Y (temperature) at given locations
- Universal Kriging
  - Uses also covariates X (e.g., rainfall, elevation) at given locations

Ordinary Kriging

- Assumptions
  - Intrinsic Stationary
  - Known covariance C(h)
  - Unknown constant mean E(Y) = u
  - Linear estimation
    - $\hat{y}(s_o) = \Sigma_{i=1}^{n} l_i y(s_i)$
  - Approach:
    - Minimize expected squared loss
    - $E(y(s_0) - \Sigma_{i=1}^{n} l_i y(s_i\ ))^2$



Solution of all l_i (for the new value to predict at a new location s_0)

$$\begin{bmatrix} \gamma(x_1, x_1) & \cdots & \gamma(x_1, x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n, x_1) & \cdots & \gamma(x_n, x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(x_1, x^*) \\ \vdots \\ \gamma(x_n, x^*) \\ 1 \end{bmatrix}$$

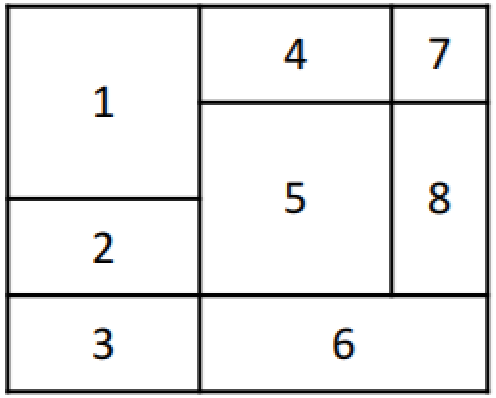
Form of solution (out of the scope of the class; ignore details here and the main goal is to see that the solution is formed by variograms)

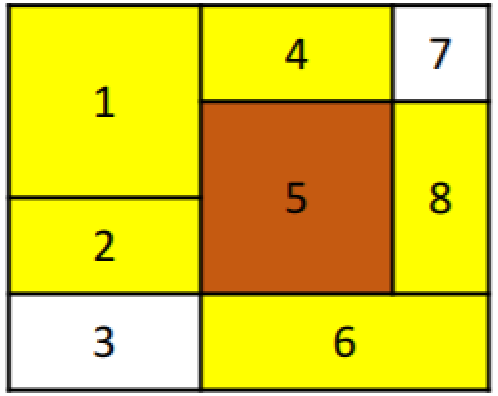
Lattice Statistics

- Also known as the areal model
- Given a complete and disjoint partitioning of the study area and a value for each partition
- Similar to the “Field Model” in Spatial Data Types
- Model spatial autocorrelation and quantify that.
- Spatial Prediction
  - Predict house price of a neighborhood given covariates of the same and nearby neighborhoods

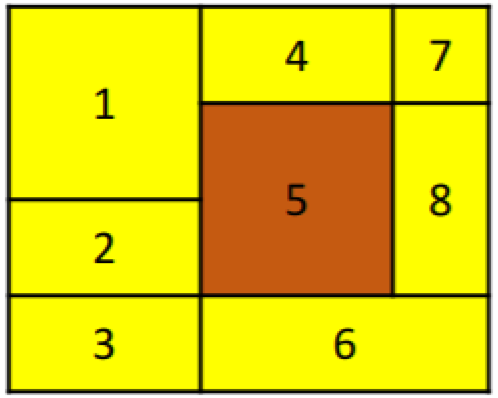
W-Matrix

- Spatial Neighborhood Matrix (W matrix)
  - W_{ij} = 1 if i and j are neighbors
  - W_{ij} = 0 if not neighbors
- Row-normalized W-Matrix
  - Divide each value by row sum





0	1	0	1	1	0	0	0
1	0	1	0	1	0	0	0
0	1	0	0	0	1	0	0
1	0	0	0	1	0	1	0
1	1	0	1	0	0	0	1
0	0	1	0	1	0	0	1
0	0	0	1	0	0	0	1
0	0	0	0	1	1	1	0



0	1	0	1	1	0	0	0
1	0	1	0	1	1	0	0
0	1	0	0	1	1	0	0
1	0	0	0	1	0	1	1
1	1	1	1	0	1	1	1
0	1	1	0	1	0	0	1
0	0	0	1	1	0	0	1
0	0	0	1	1	0	1	0

Spatial Autocorrelation

- Measures the level of global spatial association
- Many numeric measures proposed
  - Moran's I:
  - $I = \frac{n\Sigma_i\Sigma_j w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{(\Sigma_{i\neq j} w_{ij})\Sigma_i (y_i-\bar{y})^2}$ $\quad w_{ij}$ *is the* w-matrix, i and j are locations
  - $I \in [-1, 1]$ 1: strong positive correlation (homogeneous), -1 strong negative correlation
  - Assuming Rook connectivity.

0	1	0	1	0
1	0	1	0	1
0	1	0	1	0
1	0	1	0	1
0	1	0	1	0

$I \approx -1$

1	1	1	0	0
1	1	1	0	0
1	1	1	0	0
1	1	1	0	0
1	1	1	0	0

$I \approx 1$

Note: spatial neighborhood relationship is important (W-matrix)

What is the Moran's I measure if using Queen neighborhood?

Moran's I

- $I = \frac{zWz^t}{zz^t}, W\ is\ a$ row-normalized neighborhood matrix
- $z_i = (y_i - \bar{y})$ (z is a vector)

0	1	0	1	1	0	0	0
1	0	1	0	1	0	0	0
0	1	0	0	0	1	0	0
1	0	0	0	1	0	1	0
1	1	0	1	0	0	0	1
0	0	1	0	1	0	0	1
0	0	0	1	0	0	0	1
0	0	0	0	1	1	1	0

$\rightarrow$

[0, 0.33, 0, 0.33, 0.33, 0,0,0]
[0.33, 0, 0.33, 0, 0.33, 0,0,0]
[0, 0.5, 0, 0, 0, 0.5, 0, 0]
….
….
….

Spatial Autocorrelation

- Geary’s C measure
  - C = $\frac{(n-1)\Sigma_i\Sigma_j w_{ij}\left(y_i-y_j\right)^2}{2\left(\Sigma_{i\neq j} w_{ij}\right)\Sigma_i(y_i-\bar{y})^2}$
  - C ≥ 0, low value has stronger auto-correlation
  - C = 1 means no correlation
  - C = 0 means same value over the space

1	1	1	0	0
1	1	1	0	0
1	1	1	0	0
1	1	1	0	0
1	1	1	0	0

Local Autocorrelation Measures

- LISA
  - Local Moran's I
  - Local Geary's C
  - …
- Original paper
  - http://dces.wisc.edu/wp-content/uploads/sites/30/2013/08/W4_Anselin1995.pdf
  - Luc Anselin: Location Indicators of Spatial Association - LISA

Local Indicators of Spatial Association

- When data is not homogeneous, local behaviors may differ from global behavior (outliers)
- Measures how value at a location is correlated with its neighbors
- For each location (area) we calculate a measure

$$I_i = \frac{Z_i}{m_2} \sum_j W_{ij} Z_j \qquad m_2 = \frac{\sum_i Z_i^2}{N}$$

- W_{ij} is the row-normalized neighborhood matrix
- m_2 = global variance
- $z_i = (y_i - \bar{y})$
- “Z-score multiplied by the average Z-scores of its neighbors”.

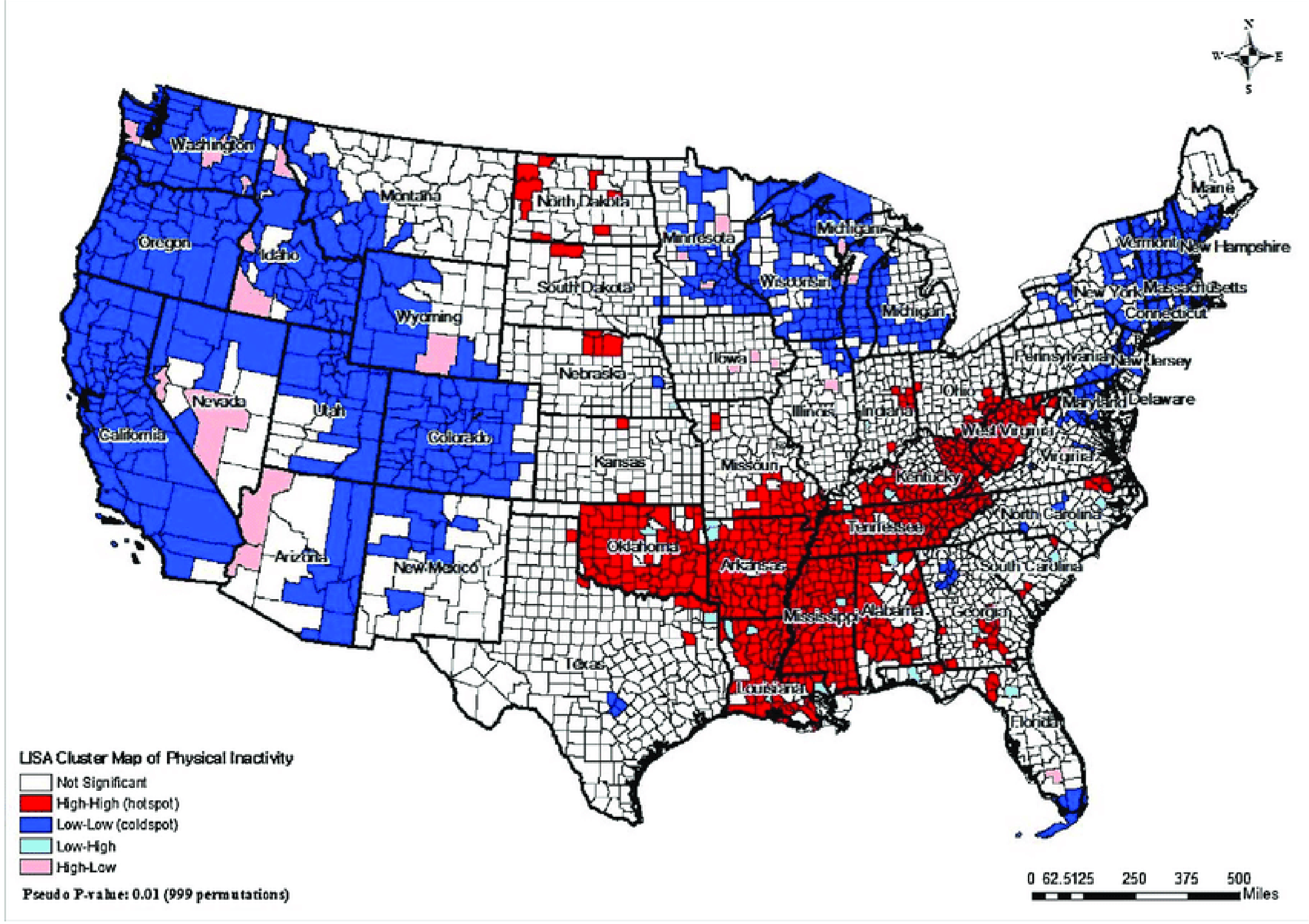
LISA

- Local vs. Global Moran

$$I = \frac{1}{n}\sum_i I_i$$

- High (positive) Local Moran:
  - High value in a high-value neighborhood (hotspot)
  - Low value in a low-value neighborhood (cold spot)
- Low (negative) Local Moran
  - Spatial outlier
- Always done with a Monte-Carlo simulation to assess significance

Local Moran



Physical inactivity of US counties

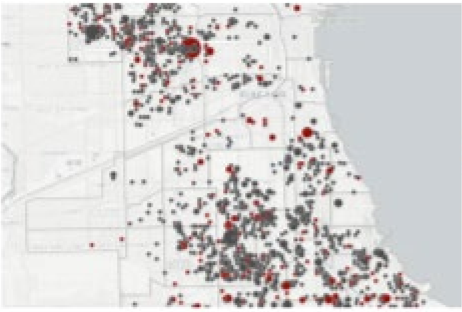
Local Geary'C

Local Geary's C

- $C_i = \frac{1}{m_2} \sum_j w_{ij} (Y_i - Y_j)^2$

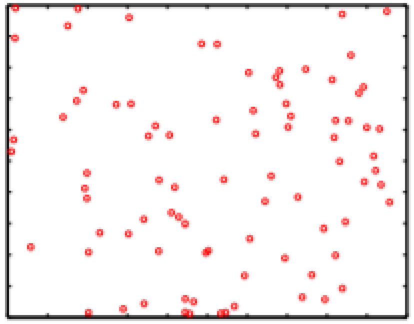
Spatial Point Process Model

- Example
  - Crime event locations
  - Disease event locations

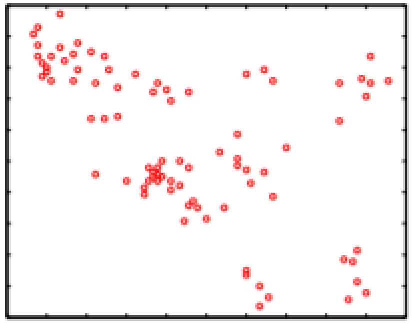


Shooting, Chicago 2010

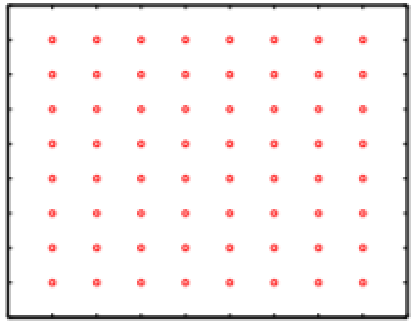
Source: http://assets.dnainfo.com



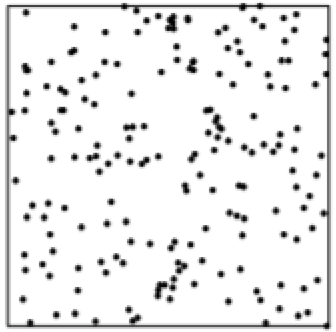
complete spatial random (CSR)



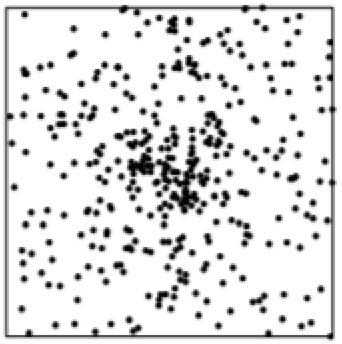
clustering



declustering



CSR



Hotspot

Ripley's K function

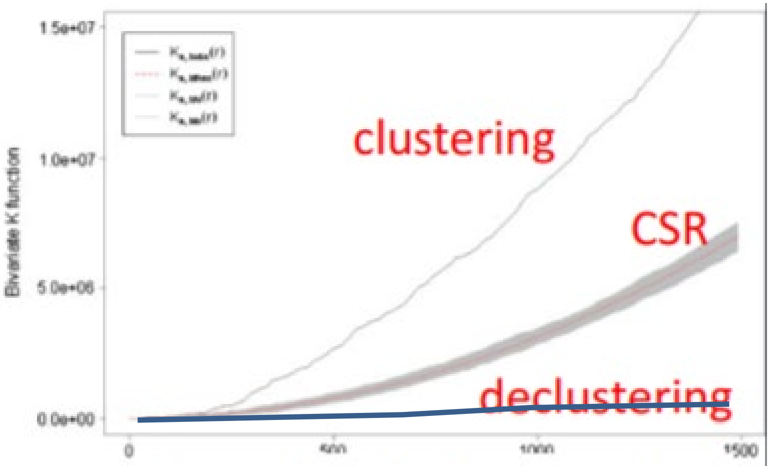
- Hypothesis Testing
  - H0: homogeneous Poisson point process (independent)
  - H1: points tend to cluster with each other
- Test statistic: average point density around each point

Test statistic:

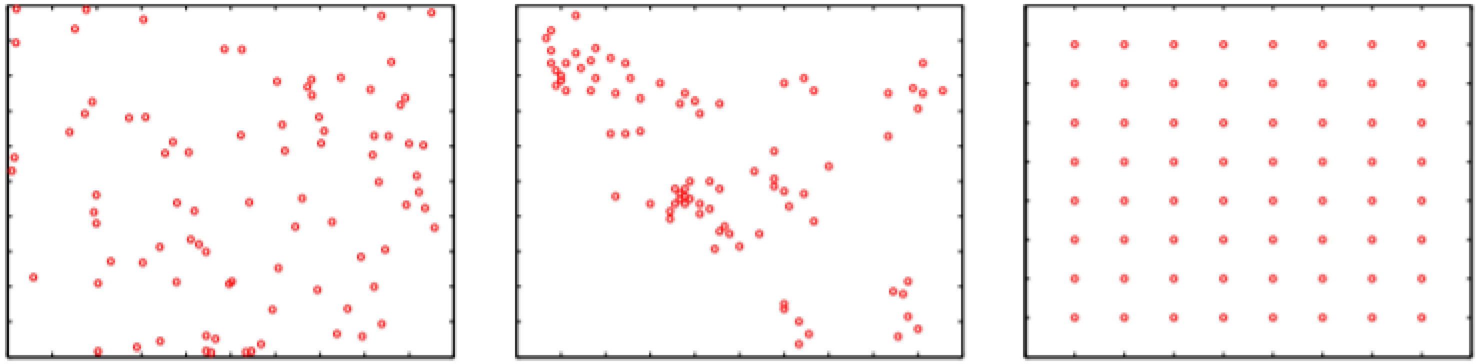
- $K(d) = \lambda^{-1} E(\#\, of\ points\ within\ radius\ d\ of\ a\ point)$
- $\hat{K}(d) = \lambda^{-1} \sum_{i \neq j} I(d_{ij} \leq d)/n$

Under H0, $K(d) = \pi d^2$

$\lambda\ is\ the\ global\ density\ (\frac{points}{area})$



Example of K function plot



CSR Clustering Declustering

Technical Takeaways

- Basic statistical concepts
- Notations
  - A mathematical language (way of easier communication)
- Assumptions
  - Statistical or mathematical models always make simplifying assumptions
    - Spatial data models are always approximations of real-world phenomenon
  - When you try to understand an approach/model, always identify the assumptions first