GE2215 Lecture 11
Spatial Statistical Analysis

Dr. Yan Yingwei
Department of Geography
National University of Singapore

Slides for education purpose only

Recap: Local operations

- Local operations are cell-by-cell operations
- Local operations can be operated on single or multiple rasters
  - Local operations on a single raster
    - Reclassification: one-to-one mapping and one-to-many mapping
  - Local operations with multiple rasters
    - Mathematical operations: $+$, $-$, $\times$, $\div$
    - Summary statistics: maximum, minimum, range, sum, mean, median, and standard deviation, which apply to numeric data only
    - Tabulation of measures: majority, minority, combine and number of unique values, which apply to both numeric and categorical data

Slides for education purpose only

Recap: Focal operations

- A focal operation, also called a neighbourhood operation, involves a focal cell and a set of its surrounding cells.
- A neighbourhood is defined by shape and size
- Focal operations are usually performed on a single layer
  - Summary statistics: maximum, minimum, range, sum, mean, median, and standard deviation
  - Tabulation of measures: majority, minority, and variety

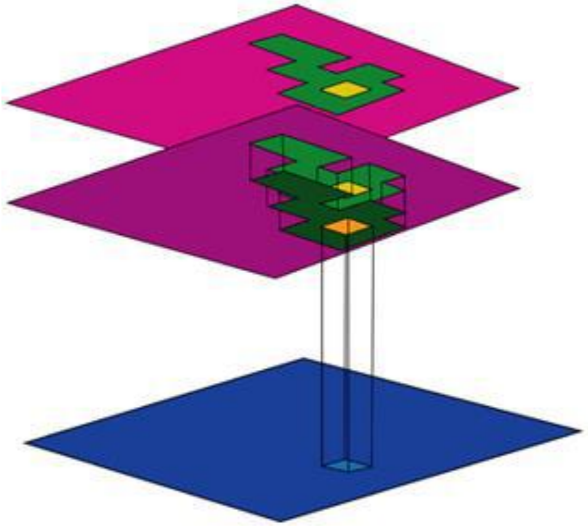
Slides for education purpose only

Recap: Zonal operations

- A zonal operation works with groups of cells of same values or like features.

  These groups are called zones

- Zones may be contiguous or non-contiguous



Slides for education purpose only

Recap: Global operations

- Global operations are similar to zonal operations whereby the entire raster dataset's extent represent a single zone
- Global operations apply a bulk change to all cells in a raster

Slides for education purpose only

Recap: Global operations

- Euclidean distance: calculating the closest distance away from a source
- Allocation: The cell value in an allocation raster corresponds to the closest source cell for the cell
- Direction: The cell value in a direction raster corresponds to the direction in degrees that the cell is from the closest source cell

1.0	**2**	1.0	2.0
1.4	1.0	1.4	2.2
1.0	1.4	2.2	2.8
1	1.0	2.0	3.0

2	**2**	2	2
2	2	2	2
1	1	1	2
1	1	1	1

90	**2**	270	270
45	360	315	287
180	225	243	315
1	270	270	270

Slides for education purpose only

Recap: Raster data clip

- There are two methods to clip a raster
  - Specify an analysis mask (clip by mask)
  - Specify the minimum and maximum x-, y-coordinates (clip by rectangle)




(d) A rectangle

(a) Input raster (b) An analysis mask (c) Clip-by-mask result (e) Clip-by-rectangle result

Slides for education purpose only

Recaps: Map algebra

- Map algebra basically involves doing math with maps
- The key difference with Algebra is that it only applies to rasters
- Map algebra can be local, focal, zonal and global
- Commonly used mathematical functions
  - Arithmetic operations
  - Statistical operations
  - Relational operations
  - Trigonometric operations
  - Exponential and logarithmic operations

Slides for education purpose only

Outline

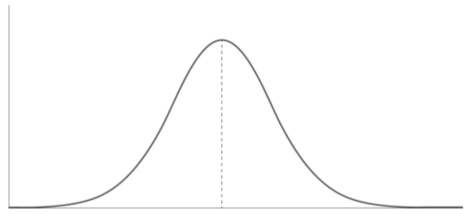
- Basics of Statistics
  - Descriptive Standard Statistics
  - Inferential Standard Statistics
- Introducing Spatial Statistics
  - Descriptive – Centrographic Statistics
  - Inferential – Point Pattern Analysis

    – Spatial Autocorrelation

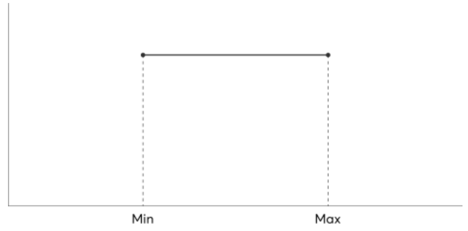
Slides for education purpose only

Standard Statistical Analysis

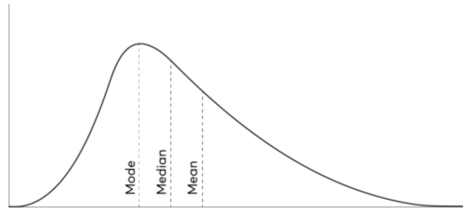
- Two parts of Standard Statistics (A quick view)
  1. **Descriptive statistics**
     - Obtain <u>summary</u> measures to <u>describe</u> a set of data
     - Three types of measures:
       - **<u>Central tendency</u>**: mean, median values
       - **<u>Variability</u>**: standard deviation
       - **<u>Frequency distributions</u>**



Normal Distribution



Uniform Distribution



Skewed Distribution

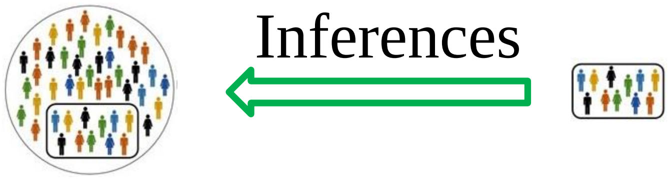
Slides for education purpose only

Standard Statistical Analysis

- Two parts of Standard Statistics (A quick review)

  2. Inferential statistics

  - Make inferences from a sample about a population
  - Some definitions
    - **Population**: all occurrences
    - **Sample**: a subset of the population
    - **Parameters**
    - **Statistics**



Population

- ***Parameters***
  - Values from the population

Sample

- ***Statistics***
  - Values from the sample

Statistics are estimates of parameters

An example: Are left-handed people more intelligent than right-handed people?

Slides for education purpose only

Outline

- Basics of Statistics
  - Descriptive Standard Statistics
  - Inferential Standard Statistics

- Introducing Spatial Statistics
  - Descriptive – Centrographic Statistics
  - Inferential – Point Pattern Analysis

    – Spatial Autocorrelation

Slides for education purpose only

What Is Spatial Statistics

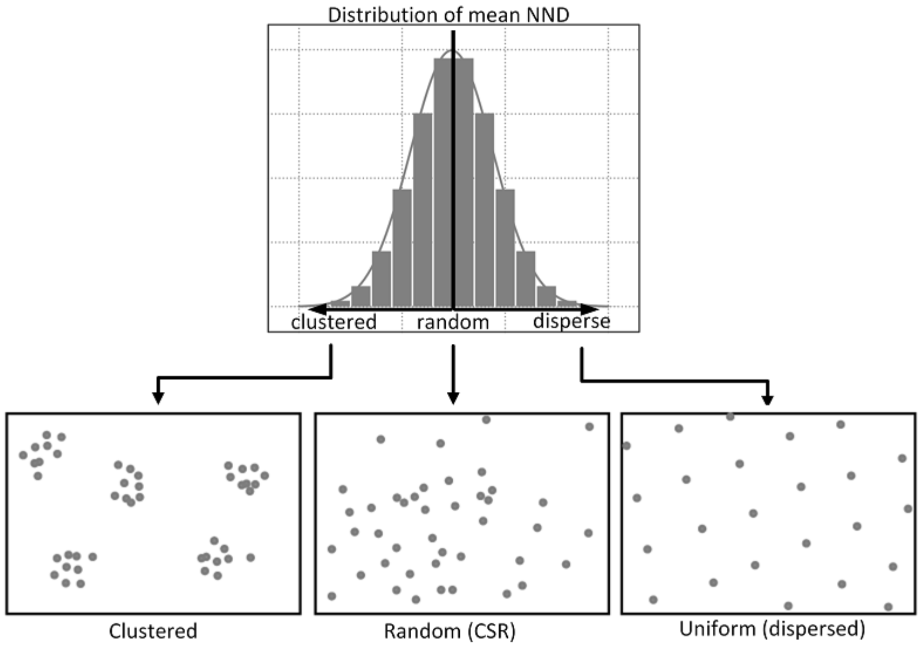
- Definition of Spatial Statistics
  - The field of study concerning statistical methods that use **space** and **spatial data**
  - Space and location matters

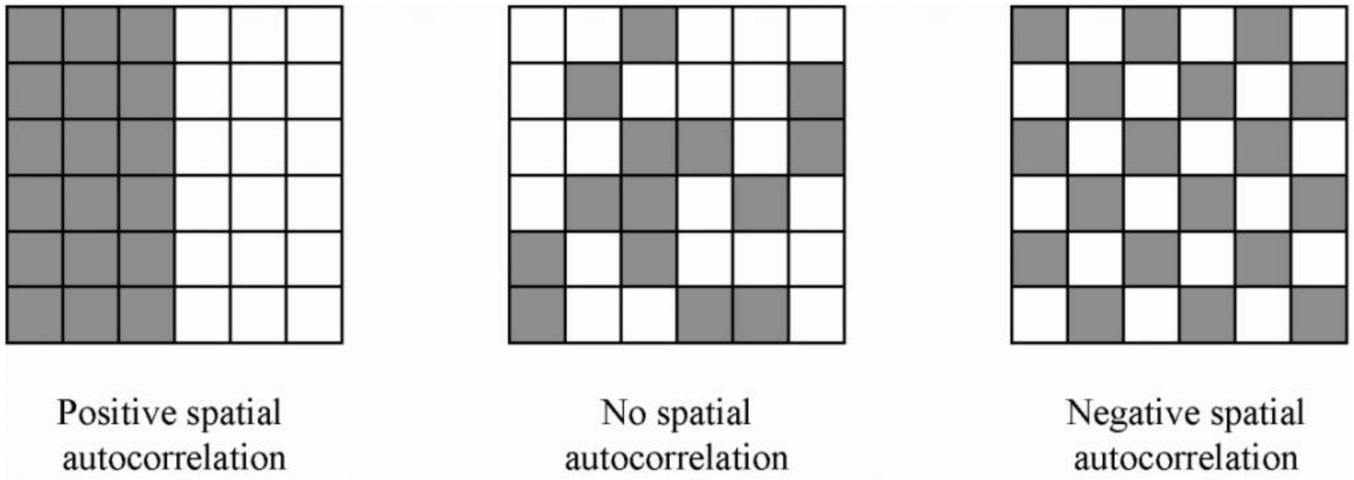
- **Spatial data**
  - Point: bus stops, grocery stores
  - Polyline: transportation networks
  - Polygon: parks, campus
  - Continuous data: temperature, precipitation

Slides for education purpose only

What Is Spatial Statistics

- **Similarly, spatial statistics also have two parts:**
  - **Descriptive** spatial statistics: Centrographic statistics
  - **Inferential** spatial statistics: e.g., Point Pattern Analysis, Spatial Autocorrelation





Nearby cases are similar

Neighbor cases have no particular relationship

Nearby cases are dissimilar

Slides for education purpose only

Outline

- Review (Refreshing your memory) of Statistics
  - Descriptive Standard Statistics
  - Inferential Standard Statistics

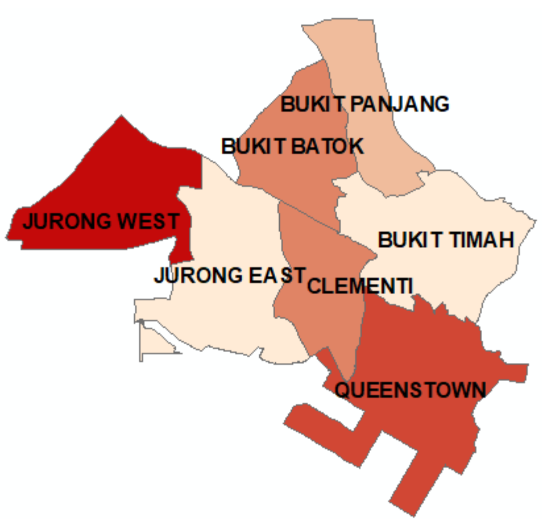
- Introducing Spatial Statistics
  - Descriptive – Centrographic Statistics
  - Inferential – Point Pattern Analysis

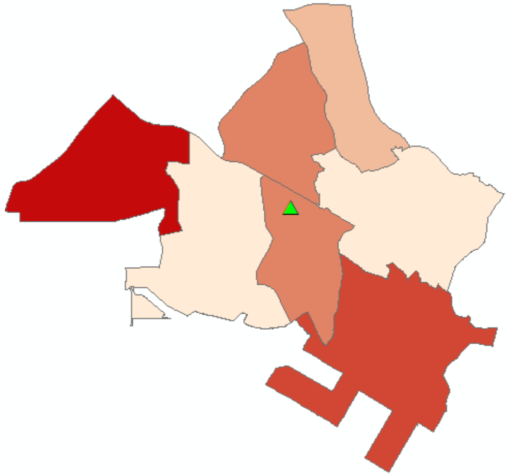
    – Spatial Autocorrelation

Slides for education purpose only

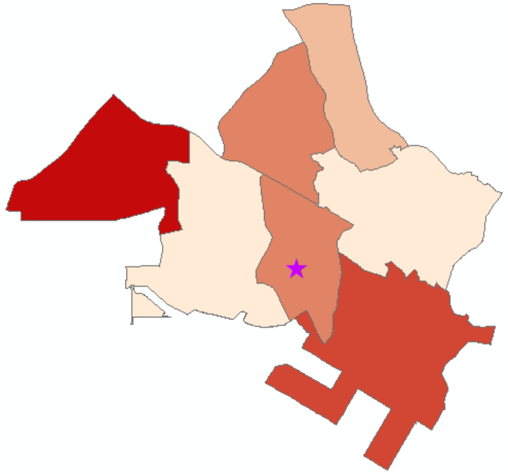
Descriptive Spatial Statistics: Centrographic Statistics

- Measures of Centrality
  - Centroid (Polygon data)
  - Weighted mean center





Geometric Center (Centroid)



Weighted mean center
(elder population center)

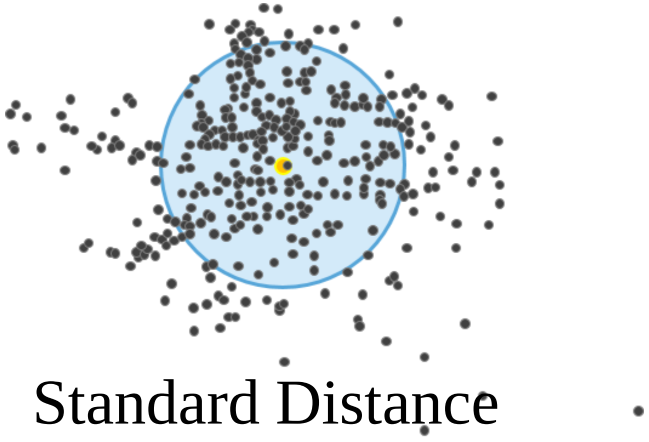
Slides for education purpose only

Descriptive Spatial Statistics: Centrographic Statistics

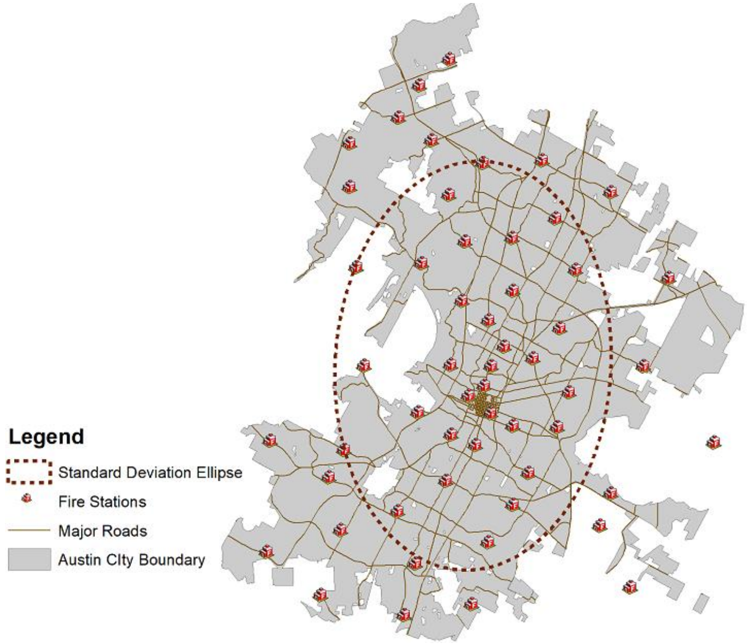
- Measures of Variance
  - Standard Distance
    - Measure the <u>degree</u> to which features are concentrated or dispersed around the geometric mean center
    - Does not capture the <u>shape</u> of distribution

  - Standard Deviational Ellipse (SDE)
    - Describe the <u>direction</u> of the dispersion
    - A typical method for measuring the activity space

A perfect circle here.
Size of circle depends on how well distributed the points are



Standard Distance



Standard Deviational Ellipse

https://gistbok.ucgis.org/bok-topics/point-pattern-analysis

only

Outline

- Review (Refreshing your memory) of Statistics
  - Descriptive Standard Statistics
  - Inferential Standard Statistics

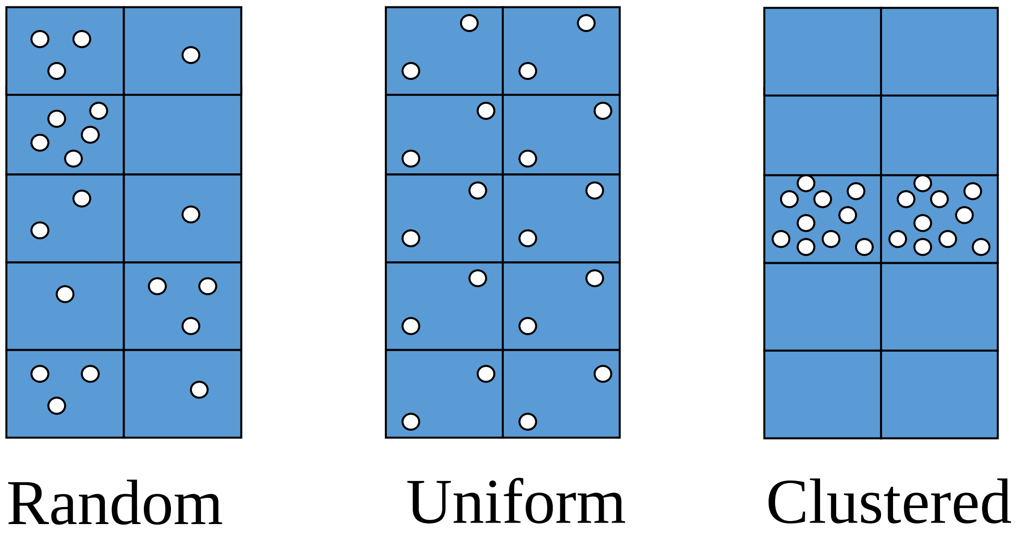
- Introducing Spatial Statistics
  - Descriptive – Centrographic Statistics
  - **Inferential – Point Pattern Analysis**

    – Spatial Autocorrelation

Slides for education purpose only

Inferential Spatial Statistics: Point Pattern Analysis

- **What is the purpose of Point Pattern Analysis?**
  - Test hypotheses about spatial patterns of points
- **What do we mean by spatial patterns?**
  - Mainly three types of spatial patterns
- **Why spatial patterns matter?**
  - Often, we cannot observe the spatial process, so we have to infer the process by observing the spatial patterns
  - Nonrandom ← Spatial processes
    - First order effect
    - Second order effect



Three types of spatial patterns

Slides for education purpose only

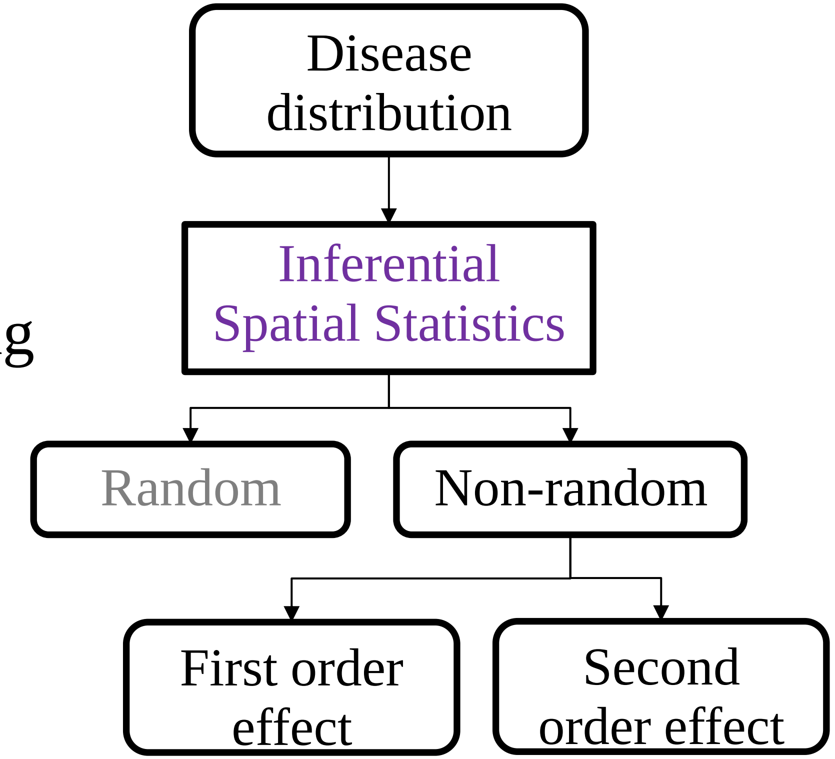
First Order Effect

- **Diseases cluster (e.g., cancer, cold, coronavirus)**
- **First order effect describes Spatial Heterogeneity**
  - Variation in the study area (uneven distribution)
  - The probability of a spatial phenomenon occurring at each location varies throughout a study region
    - Diseases cluster
    - Robberies cluster

Slides for education purpose only

Second Order Effect

- **Diseases cluster (e.g., cancer, cold, coronavirus)**
- **Second order effect describes Spatial Interdependence**
- The occurrence of an event at one location increases the probability of the same event occurrence at the neighboring location, sometimes resulting in clusters across a study region.
  - Interdependence between points themselves
    - Diseases cluster because people catch them from others who have the disease (e.g., cold)

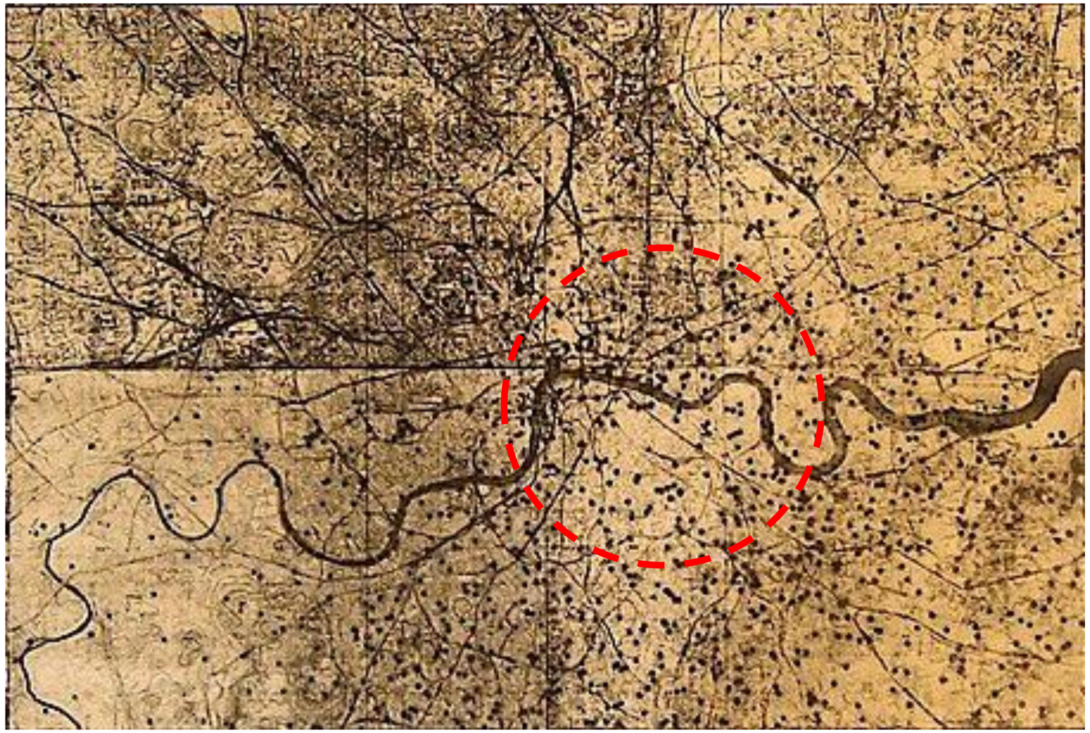


Exploring the cause of a disease

Slides for education purpose only

Point Pattern Analysis

- German launched 1,358 V-2 Rockets at London during World War II
- At first, its guidance systems were thought to be too primitive to hit specific targets
- However, according to the bomb damage maps, the impact sites seemed (visually) to be clustered
- If the rocket strikes were spatially clustered, the guidance systems must have been improved
- It was essential to assess if the impact sites were spatially clustered quantitatively



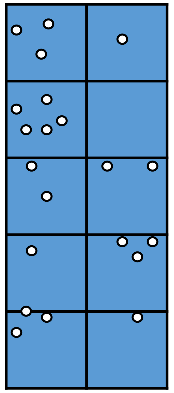
Distribution of V-2 Rocket Strikes on Central London, 1944

Exploratory Spatial Data Analysis (ESDA) method

Slides for education purpose only

Point Pattern Analysis

- Approaches to point pattern analysis
  - Density-based approaches
  - Distance-based approaches
- **Density-based approaches**
  - Quadrat Analysis
  - Kernel Density Estimation



Construct grids

3	1
5	0
2	2
1	3
3	1

Variance
Mean
Var/Mean

Calculate variance and mean, and create the variance-mean ratio

Threshold values:

Uniform : Var/Mean ≈0
Random : Var/Mean ≈1
Clustered : Var/Mean >1

Determine spatial distribution

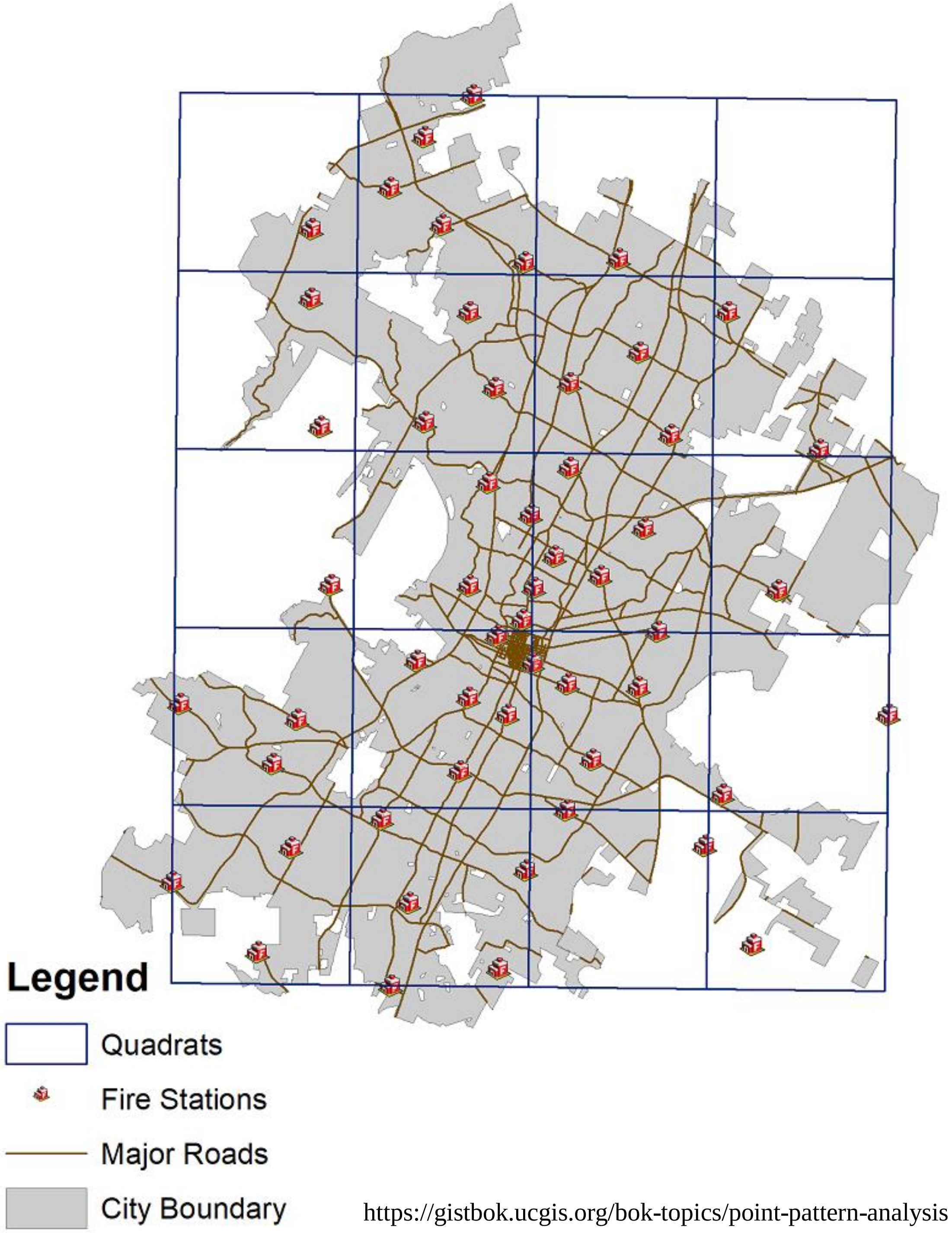
- **Quadrat Analysis**

The biggest problem is selecting the suitable quadrat size **(Modifiable Area Unit Problem)**

Slides for education purpose only

Point Pattern Analysis

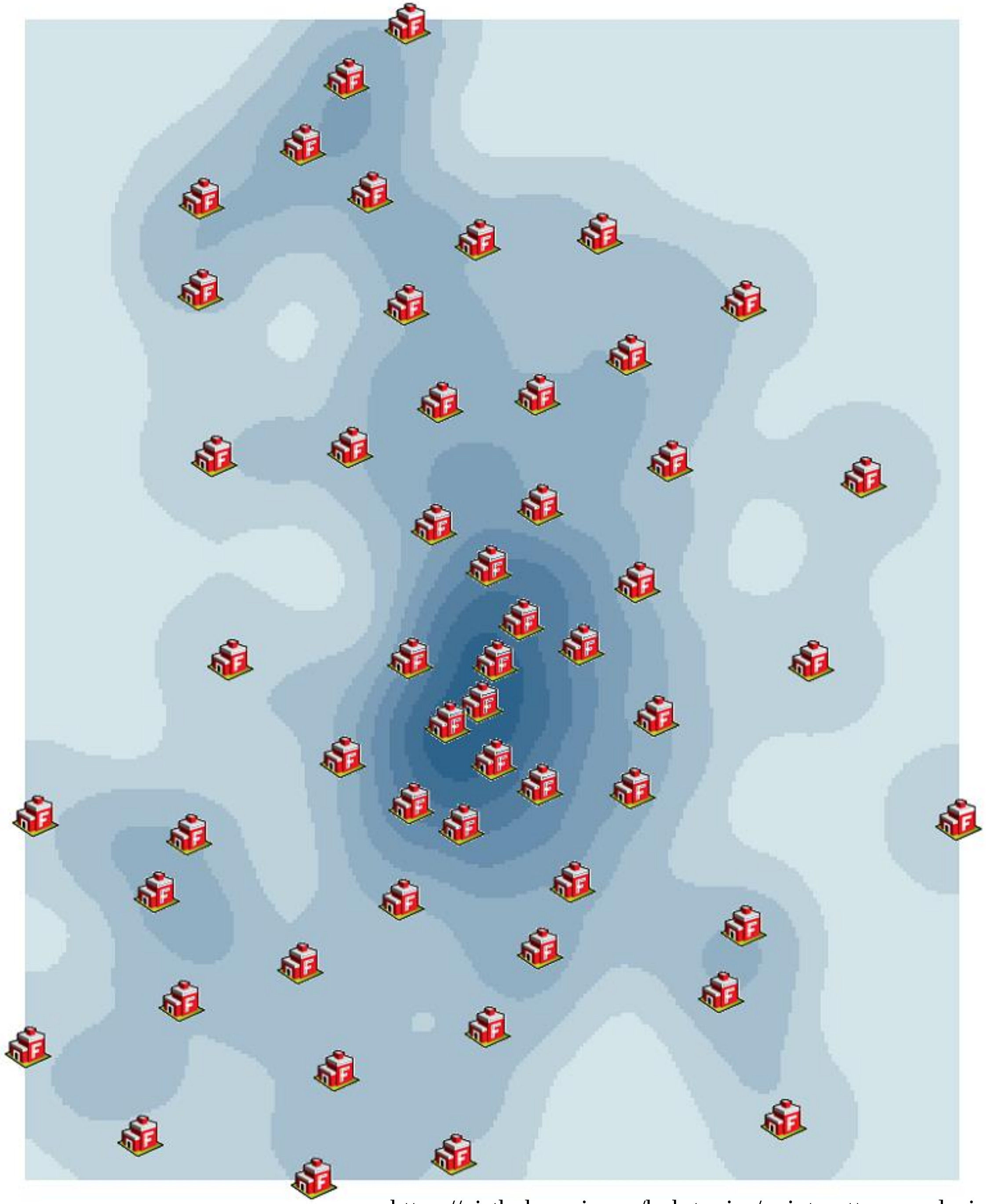
- Approaches to point pattern analysis
  - <u>Density</u>-based approaches
  - <u>Distance</u>-based approaches
- **Density-based approaches**
  - Quadrat Analysis
  - Kernel Density Estimation



https://gistbok.ucgis.org/bok-topics/point-pattern-analysis

Point Pattern Analysis

- Approaches to point pattern analysis
  - Density-based approaches
  - Distance-based approaches
- **Density-based approaches**
  - Quadrat Analysis
  - Kernel Density Estimation



https://gistbok.ucgis.org/bok-topics/point-pattern-analysis

Point Pattern Analysis

- Approaches to point pattern analysis
    - Density-based approaches
    - Distance-based approaches
- **Distance-based approaches**
    - **Nearest Neighbor Index (NNI)**
    - G function
    - F function
    - Ripley's K function

Slides for education purpose only

Point Pattern Analysis

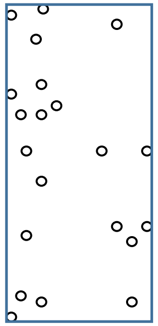
- Distance-based approaches
  - **Nearest Neighbor Index (NNI)**

- **Calculate NNI**

1. The mean of the distance <u>observed</u> between each point and its nearest neighbor
2. The <u>expected</u> mean distance <u>if</u> the distribution is random

$$NNI = \frac{Observed\ mean\ distance}{Expected\ mean\ distance}$$

$$Expected\ mean\ distance = 0.5\sqrt{\frac{Area}{Number\ of\ points}}$$



Point	Nearest Point	Distance
1	2	1
2	3	0.1
3	2	0.1
…	…	
20	21	1
21	19	0.4

Calculating the <u>nearest neighbor</u> distance

Threshold values:

- Random pattern, NNI ≈ 1
- Clustered pattern, NNI ≈ 0
- Uniform pattern, NNI ≈ 2.149

Slides for education purpose only

Outline

- Review (Refreshing your memory) of Statistics
    - Descriptive Standard Statistics
    - Inferential Standard Statistics

- Introducing Spatial Statistics
    - Descriptive – Centrographic Statistics
    - Inferential – Point Pattern Analysis

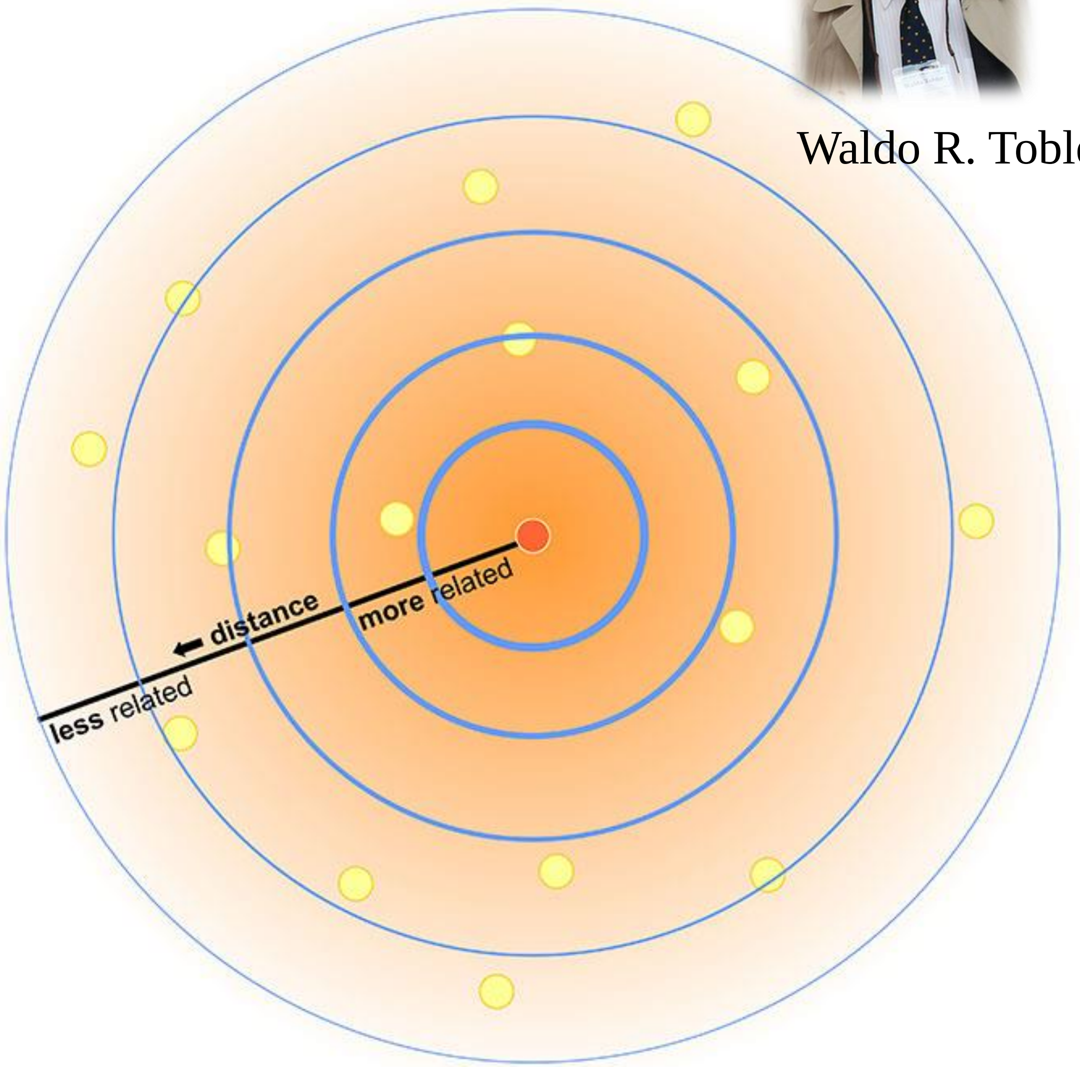
– Spatial Autocorrelation

Slides for education purpose only

Spatial Autocorrelation

Waldo R. Tobler

- Spatial Autocorrelation (SA) is the most important concept in Spatial Statistics
- Definition of Spatial Autocorrelation
  - The <u>degree</u> to which characteristics at one location are similar (or dissimilar) to those nearby
- **It is the confirmation of Tobler's first law of geography**
  - *Everything is related to everything else, but near things are more related than distant things*



Slides for education purpose only

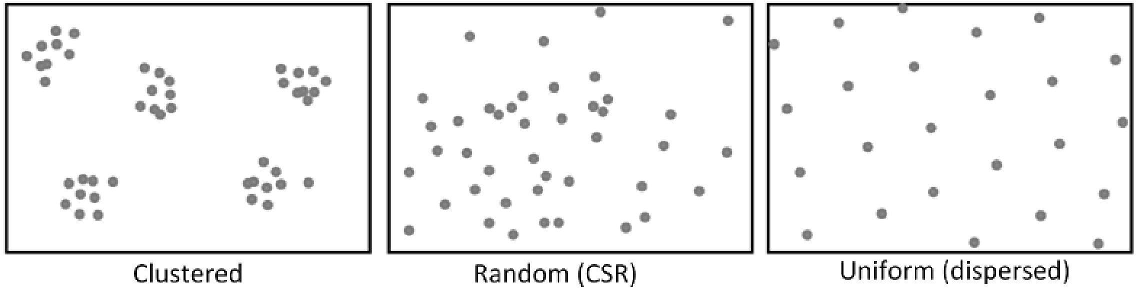
Spatial Autocorrelation

- **Differences and similarities between Point Pattern Analysis and SA**

➢ *Point Pattern Analysis*

✓ Points only

✓ Only location, no "magnitude" value

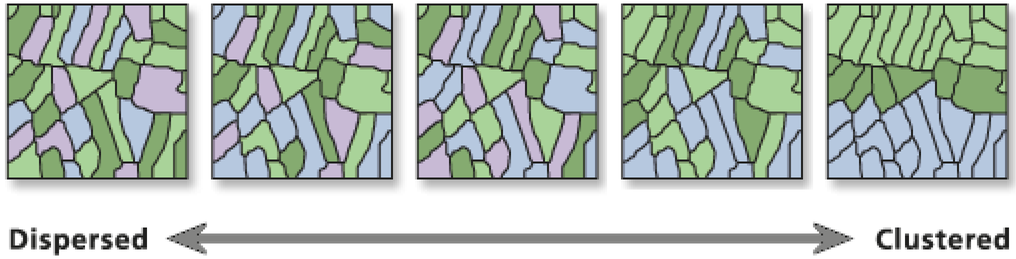


➢ *Spatial Autocorrelation*

✓ Points and polygons

✓ Location and attribute variable

❖ Income, GDP, rainfall, crime rate etc.



Slides for education purpose only

Spatial Autocorrelation

The name explains itself

Spatial:

Location-related,

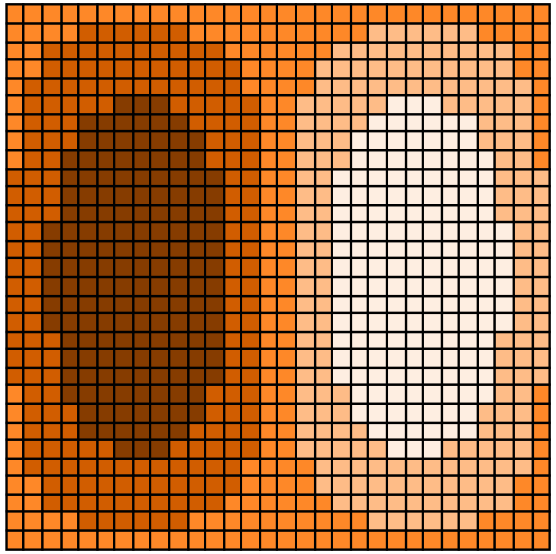
On a map

Auto:

Self (Interdependence)

Correlation:

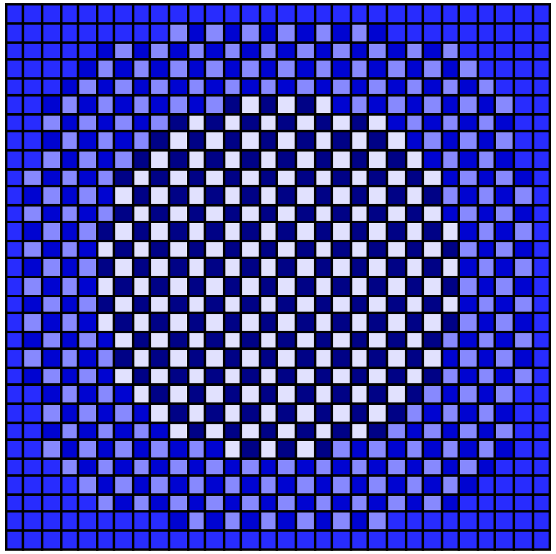
Degree of relative similarity



Positive Spatial Autocorrelation
Similar values cluster together on a map

White: High value

Dark: Low value



Negative Spatial Autocorrelation
Dissimilar values cluster together on a map

Source: Dr Dan Griffith

Slides for education purpose only

Why Is Spatial Autocorrelation Important?

- **Two reasons**

1. **It <u>implies the existence of spatial process</u>**
   - Why are near-by things similar to each other?
   - Why do high-income people live “next door” to each other?
   - Why are rich cities surrounded by other rich cities?
   - All these **geographical** questions are related to **locations**, and can be explained by **Spatial Autocorrelation**

Other examples?

1. **Nearby things are similar to each other**
2. **Nearby things are dissimilar to each other**

Slides for education purpose only

Why Is Spatial Autocorrelation Important?

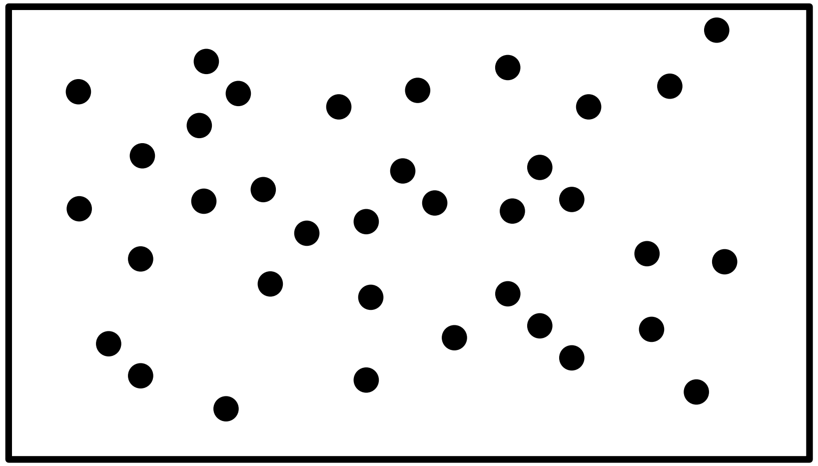
- **Two reasons**

1. **Spatial Autocorrelation <u>invalidates many traditional statistical inference assumptions</u>**
    - Traditional statistical inference often assumes that values of observations in each sample are <u>**independent**</u> of one another
    - However, if SA exists, the assumption is wrong
    - In this case, <u>spatial statistics</u> should be used

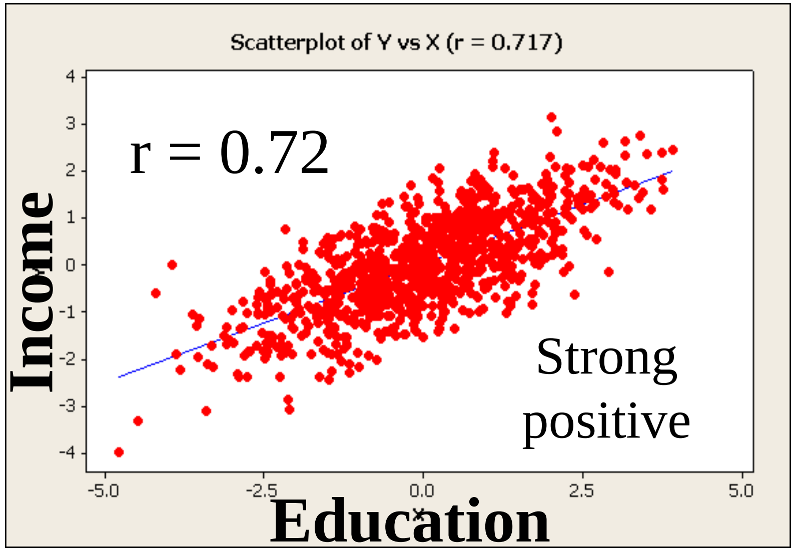
Slides for education purpose only

How Does SA Impact Standard Statistics?

- Correlation analysis between Income and Education by **Standard Statistics**



Sample collection locations



Correlation Analysis

Slides for education purpose only

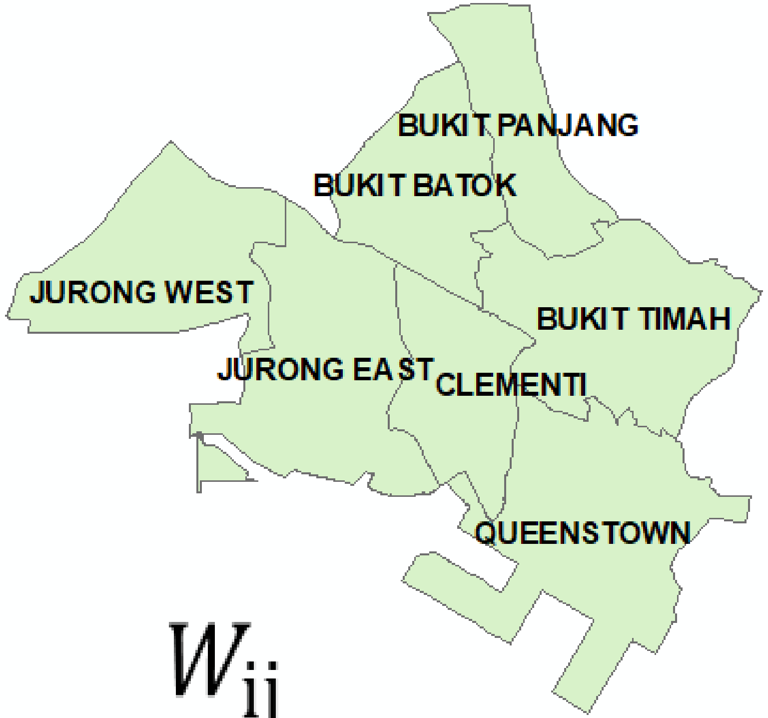
Measuring Spatial Autocorrelation

- Spatial Autocorrelation is measured based on the *Spatial Weight Matrix (SWM)*
- The *SWM* measures the relative location between each pair of points (or polygons)
- The *SWM* is a representation of the spatial structure of your data
- Two methods for measuring *Spatial Weight Matrix*
  - Weights based on contiguity – **binary** (0, 1)
  - Weights based on distance – **continuous** variable

Slides for education purpose only

Contiguity-based SWM

	BUKIT BATOK	BUKIT PANJANG	BUKIT TIMAH	CLEMENTI	JURONG EAST	JURONG WEST	QUEENSTOWN
BUKIT BATOK	*						
BUKIT PANJANG	1	*					
BUKIT TIMAH	1	1	*				
CLEMENTI	1	0	1	*			
JURONG EAST	1	0	0	1	*		
JURONG WEST	0	0	0	0	1	*	
QUEENSTOWN		0	1	1	0	0	*



W_{ij}

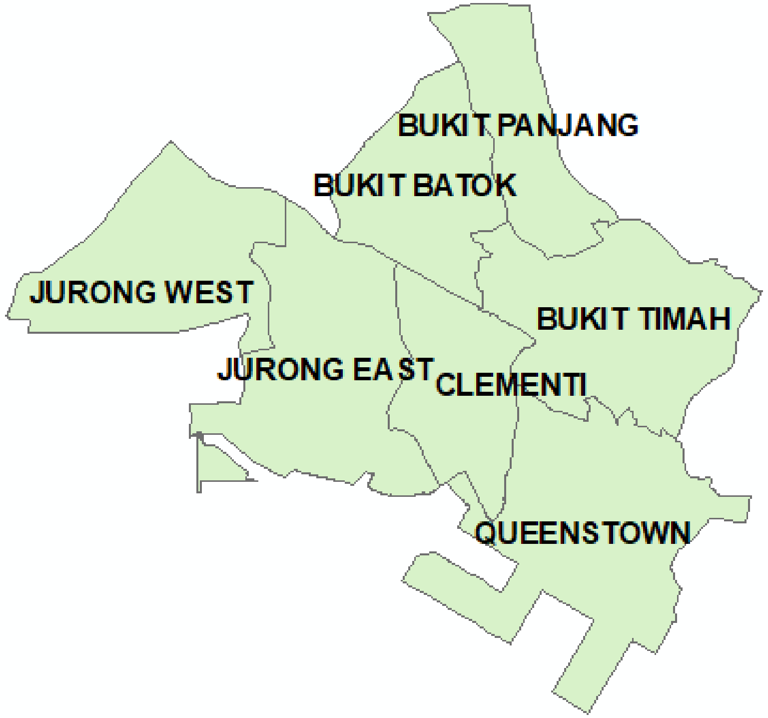
Contiguity-based Spatial Weight Matrix

If in adjacency relation 1, else 0

Slides for education purpose only

Distance-based SWM

	BUKIT BATOK	BUKIT PANJANG	BUKIT TIMAH	CLEMENTI	JURONG EAST	JURONG WEST	QUEENSTOWN
BUKIT BATOK	*						
BUKIT PANJANG	0.42	*					
BUKIT TIMAH	0.20	0.22	*				
CLEMENTI	0.23	0.18	0.28	*			
JURONG EAST	0.22	0.15	0.16	0.34	*		
JURONG WEST	0.17	0.12	0.10	0.15	0.25	*	
QUEENSTOWN	0.12	0.12	0.22	0.24	0.15	0	*



Distance-based Spatial Weight Matrix

The weight between two polygons is inversely proportional to the distance between them

Slides for education purpose only

Global Measures of Spatial Autocorrelation

- Moran's I is the most common measure of Spatial Autocorrelation

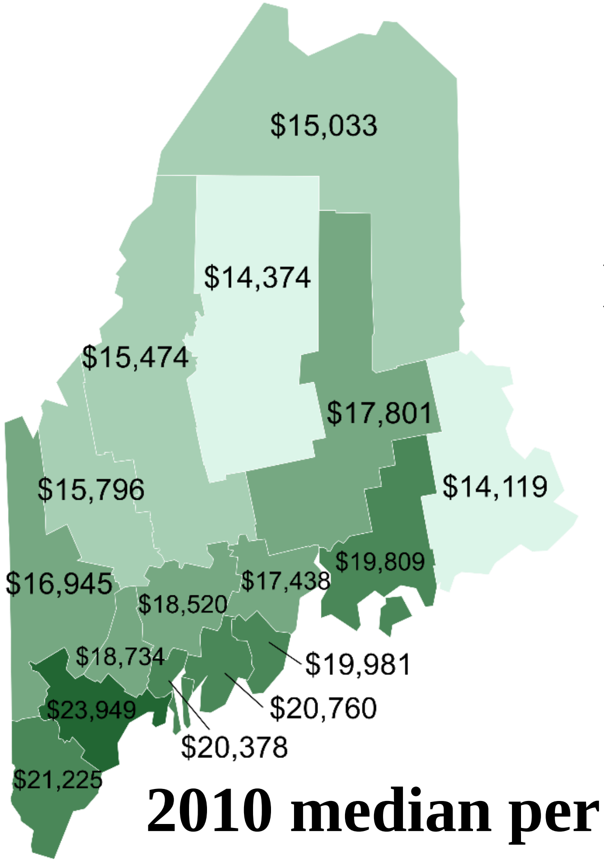
+1 High positive spatial autocorrelation

0 No spatial autocorrelation

-1 High negative spatial autocorrelation

Patrick Alfred Pierce Moran

Australian



Moran's I = 0.377

Positive SA

2010 median per capita income for Maine, USA

Slides for education purpose only

Global Measures of Spatial Autocorrelation

- Geary's C is similar to Moran's I

2 — High negative spatial autocorrelation

1 — No spatial autocorrelation

0 — High positive spatial autocorrelation

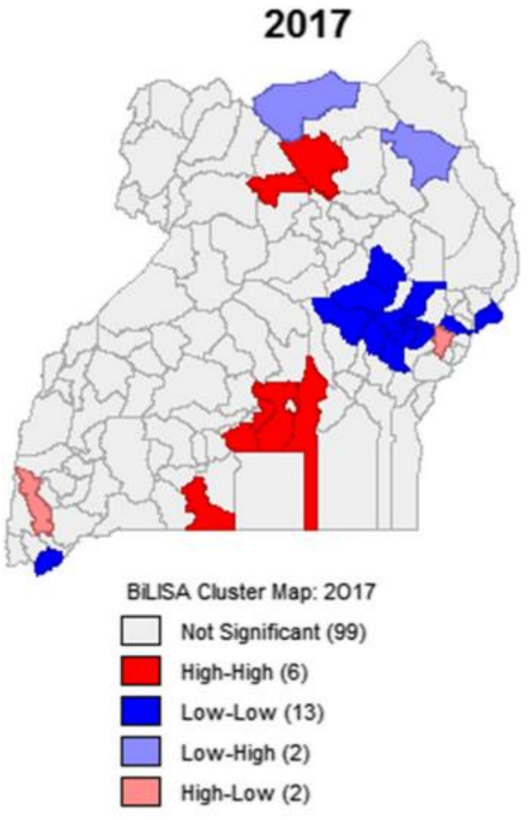
Roy C. Geary

- **Comparison of Moran's I and Geary's C**
  - ➢ In general, they result in similar conclusions
  - ➢ Moran's I is often preferred due to the greater general stability

Slides for education purpose only

Hot Spots and Cold Spots

- Four types of <u>Spatial Clustering Patterns</u>
  - Positive Spatial Autocorrelation
    - High-High clustering (<u>High</u> values cluster together, **Hot Spots**)
    - Low-Low clustering (<u>Low</u> values cluster together, **Cold Spots**)
  - Negative Spatial Autocorrelation
    - Low-High clustering (<u>Low</u> values surrounded by <u>high</u> values)
    - High-Low clustering (<u>High</u> values surrounded by <u>low</u> values)



Clustering of TB and HIV across Uganda

- **Both Moran's I and Geary's C can <u>only distinguish</u> positive or negative spatial autocorrelation**
- **<u>Cannot tell</u> if these are hot spots or cold spots**

Slides for education purpose only

Getis-Ord General G-Statistic

- The general G statistic can distinguish between hot spots and cold spots
  - G is relatively <u>large</u> → potential <u>hot spots</u>
  - G is relatively <u>low</u> → potential <u>cold spots</u>

Arthur Getis

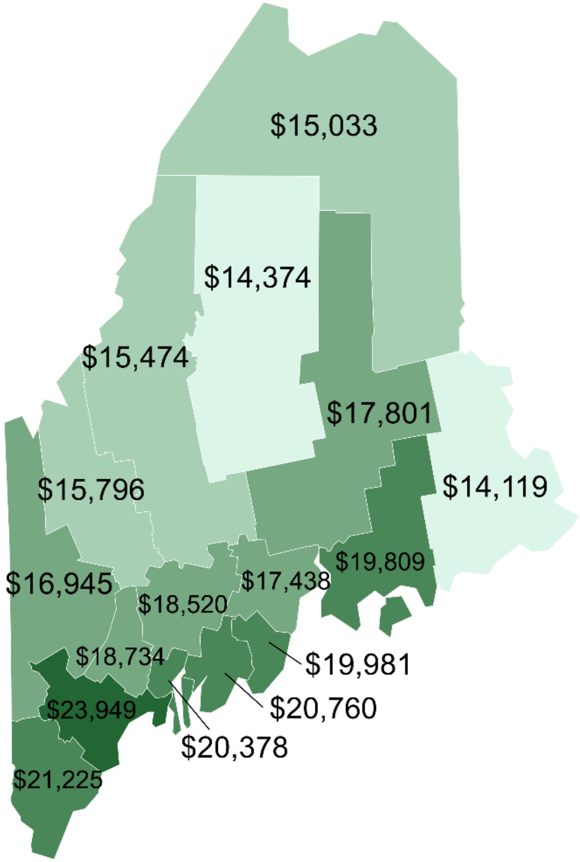
- General G will **NOT** show <u>negative</u> spatial autocorrelation
- General G can <u>only</u> use <u>distance-based</u> Spatial Weight Matrix
- General G should <u>only</u> be calculated for <u>ratio scale</u> data
  - It allows you to compare two values using ratio
    - E.g., crime rates & birth rates

J. Keith Ord

Slides for education purpose only

Local Indicators of Spatial Association (LISA)

- Global measures of Spatial Autocorrelation are a measure for the entire area
- Local measures are needed when we want to look at the SA at each location or observation unit
- Why local measures matter?
  - Global measures indicate **No** SA, but **local** SA exists
  - **Positive** and **negative** SA exist **simultaneously** on the same map
  - The above issues are caused by **spatial heterogeneity**



What are the SA effects at each county of Maine?

Slides for education purpose only

Local Indicators of Spatial Association (LISA)

- LISA was proposed by Luc Anselin in 1995, therefore, LISA is often called Anselin's LISA
  - Luc Anselin 1995. *Local Indicators of Spatial Association-LISA*. Geographical Analysis 27: 93 -115

Luc Anselin

Local indicators of **spatial association**—LISA
L Anselin - Geographical analysis, 1995 - Wiley Online Library
… where A is a global **indicator** of **spatial association** and 7 is a scale factor. In other words, the sum of the **local indicators** is proportional to a global **indicator**. For the latter, a statement …
Save Cite Cited by 16906 Related articles All 9 versions

(Latest accessed 28 Oct 2024)

Slides for education purpose only

THE UNIVERSITY OF CHICAGO | DEPARTMENT OF SOCIOLOGY

Luc Anselin

Stein-Freiler Distinguished Service Professor of Sociology and the College; Director, Center for Spatial Data Science; Committee on Quantitative Methods in Social, Behavioral, and Health Science; Senior Fellow, NORC

Luc Anselin is a native of Belgium, where he did undergraduate work and a master's degree in economics and econometrics at the Free University of Brussels (VUB). His PhD is from Cornell University in the interdisciplinary field of Regional Science.

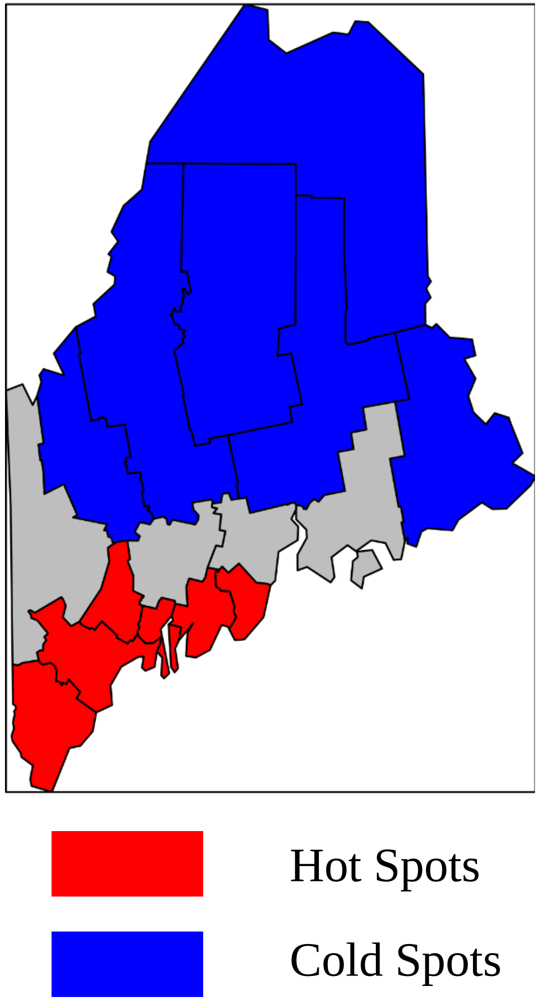
Anselin comes to the University of Chicago from Arizona State University where he was a Regents' Professor and held the Walter Isard Chair. He was the founding Director of the School of Geographical Sciences and Urban Planning. He also started and directed the GeoDa Center for Geospatial Analysis and Computation. Before ASU, he was a Professor at the University of Illinois, Urbana-Champaign, where he directed the

https://sociology.uchicago.edu/directory/Luc-Anselin

Slides for education purpose only

Identifying Hot and Cold Spots with LISA

- Global measures can only tell us there are Hot or Cold Spots in the whole area
- But LISA can tell us where the Hot and Cold Spots are



Slides for education purpose only

Local Indicators of Spatial Association (LISA)

- Moran's I, Geary's C, and Getis-Ord General G are the global measures, and they all have their local versions
  - Local Moran's I (https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm)
  - Local Geary's C
  - Local Getis-Ord G (https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/hot-spot-analysis.htm)

Slides for education purpose only

Demo: Identifying Hot Spots

- **Exploring 911Call data using Hot Spot Analysis**

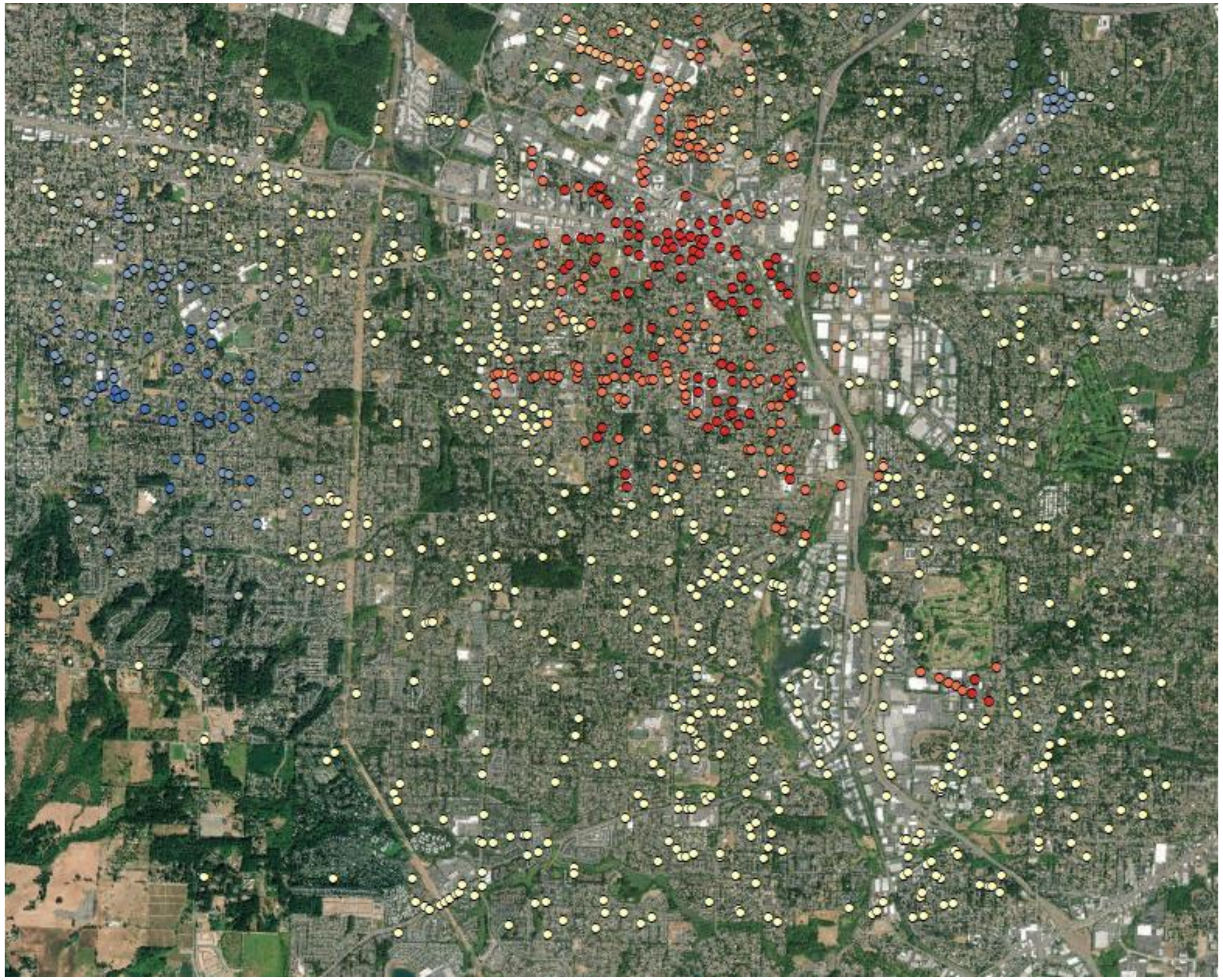
Where are the Hot Spots of 911Calls

Distributions and numbers of 911Calls in Portland, Oregon

Slides for education purpose only

Demo: Identifying Hot Spots

Gi_Bin
- Cold Spot - 99% Confidence
- Cold Spot - 95% Confidence
- Cold Spot - 90% Confidence
- Not Significant
- Hot Spot - 90% Confidence
- Hot Spot - 95% Confidence
- Hot Spot - 99% Confidence



Red: Hot Spots

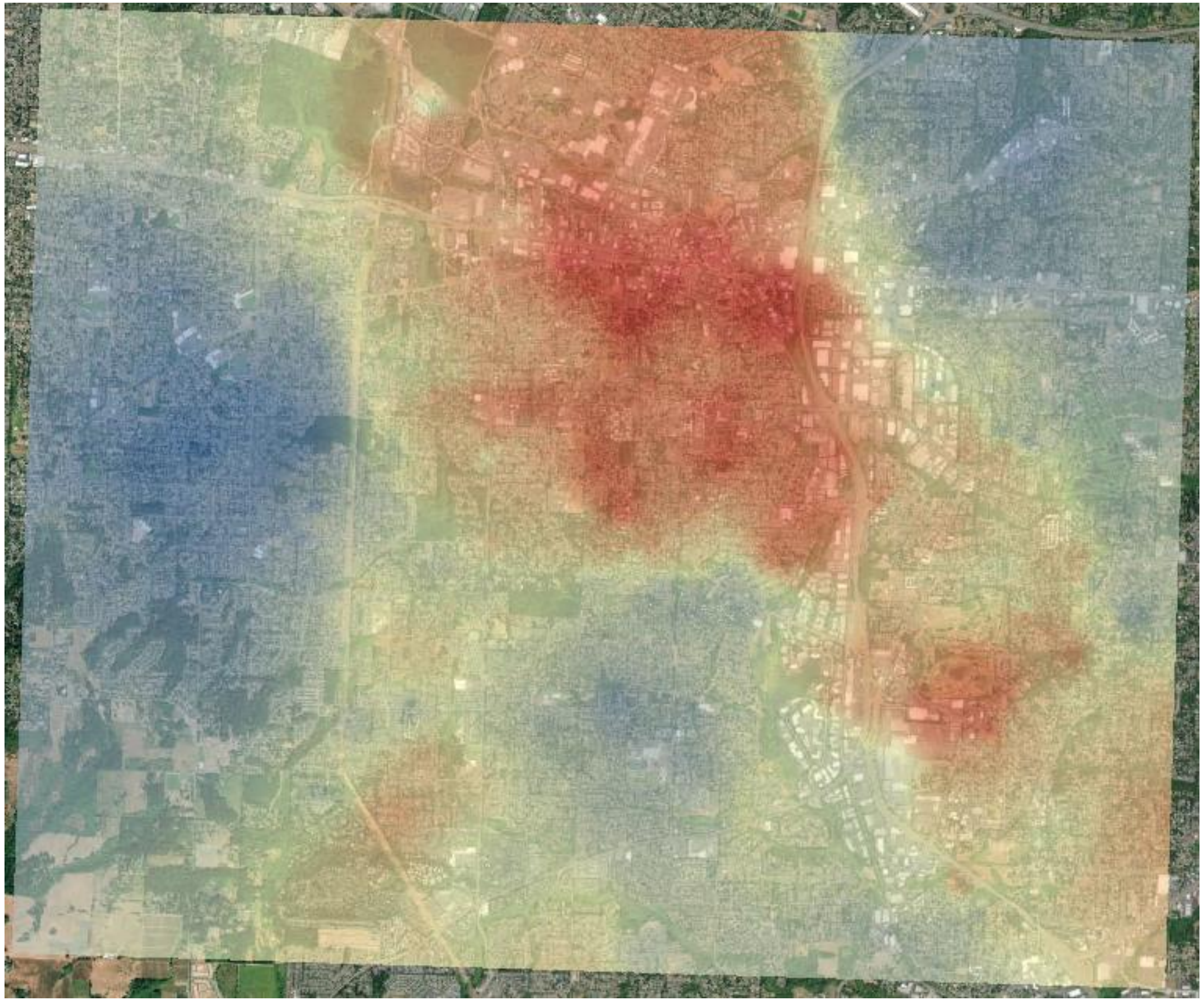
Blue: Cold Spots

Use IDW to estimate values between points

Slides for education purpose only

Demo: Identifying Hot Spots

IDW Interpolation



Red indicates potential hot spots

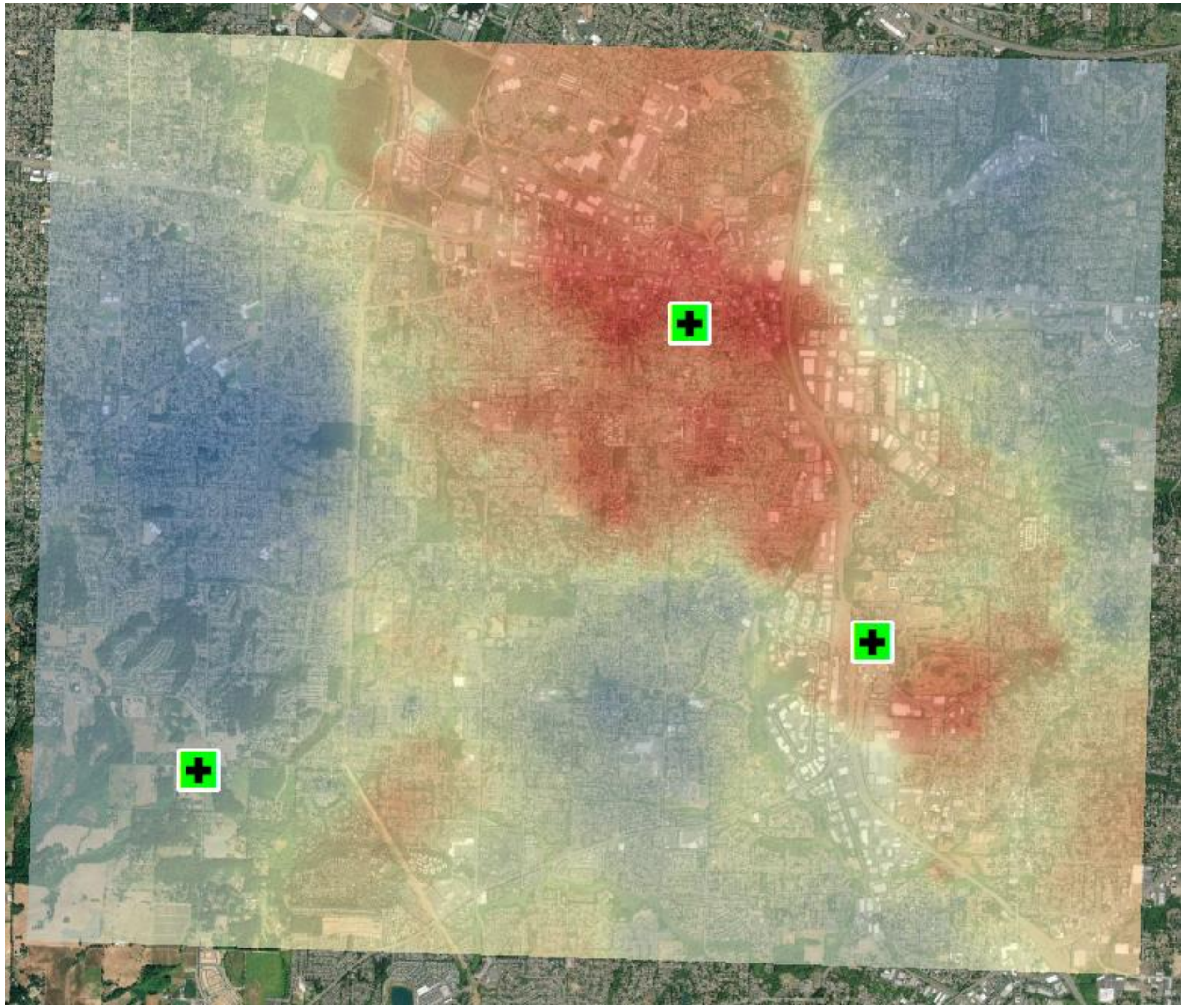
Blue indicates potential cold spots

- **What does the map imply?**
- **How can we apply the Hot Spot map?**

Slides for education purpose only

Demo: Identifying Hot Spots

Where should the Response Stations be located?



Response Stations

Please watch the full demo here: https://storymaps.arcgis.com/stories/42313404667341c3a6b33ce5f73573e6

The relevant data and lab handout are available upon request

Slides for education purpose only

Software for Spatial Statistics

- ArcGIS (licensed)
- QGIS
- GeoDA (free)
  - Designed by Luc Anselin, Arizona University
- R software (free)
- Python
  - PySAL: Python Spatial Analysis Library

Slides for education purpose only

Summary

- **Descriptive and Inferential Standard Statistics**
- **What is spatial statistics**
  - Descriptive and Inferential Spatial Statistics
    - Spatial data: point, polyline, polygon and continuous data
- **Point pattern analysis**
  - Spatial patterns: Random, Uniform and Clustered distributions
  - Spatial processes: First and second order effects (Spatial heterogeneity and interdepedence)

Slides for education purpose only

Summary

- **Spatial autocorrelation**
  - Global measures: Global Moran's I, Geary's C, and Getis-Ord general G
    - Spatial Weight Matrix: contiguity-based and distance-based
  - Local Indicators of Spatial Association (LISA)
    - Identifying Hot Spots and Cold Spots

Slides for education purpose only

THANK YOU

Slides for education purpose only