Spatial Variation and Clustering

Lecture #14 | GEOG 510
GIS & Spatial Analysis in Public Health

Varun Goel

Outline

- Spatial (geographic) variation
- Probability Mapping
- Smoothing
- Kernel Density

Geographic Variation

- Variation in some phenomenon across space or from place to place
  - We can observe this in tables, but view it in maps
    - Events (e.g., disease cases)
    - Locations (e.g., hospitals)
    - Values (e.g., average income)

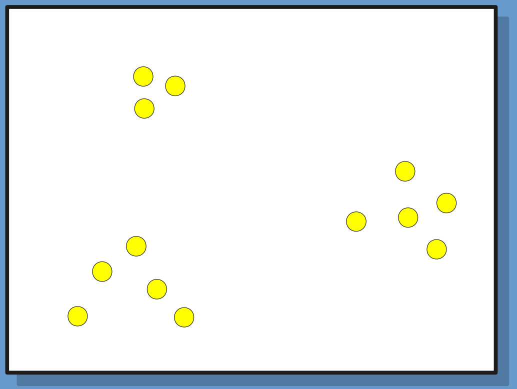
Geographic Variation

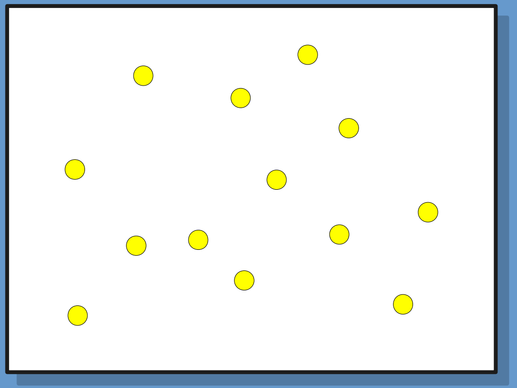
- Visual observation can provide additional information not included in a table
  - Spatial relationships
  - Useful for hypothesis generation
    - Especially when moving from descriptive to explanatory research, Pattern to Process
      - e.g., integrate other layers (colocation) or measure distance to features

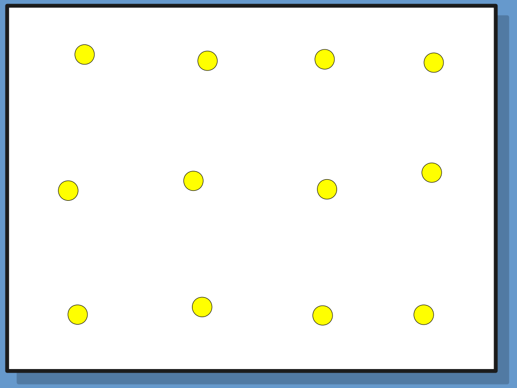
Spatial Pattern

Events

- Basic Concepts
  - Clustered
    - Events are located or distributed near to one another
  - Random
    - Events are located or distributed such that there is no regular pattern
  - Ordered (dispersed)
    - Events are located or distributed in a regular or repeating fashion



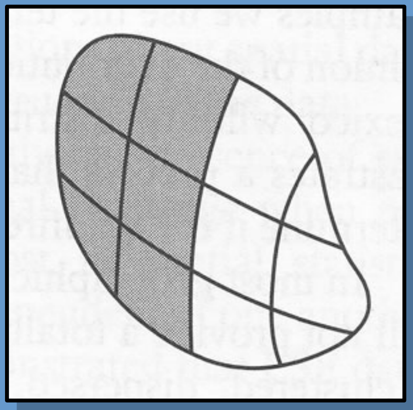


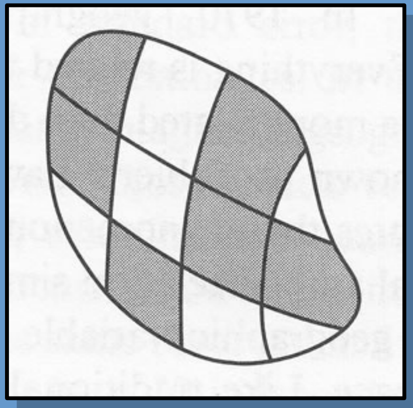


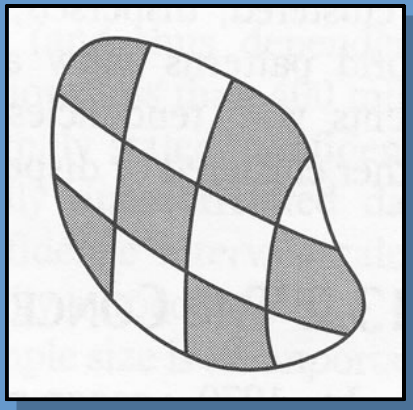
Spatial Pattern

Attribute Values

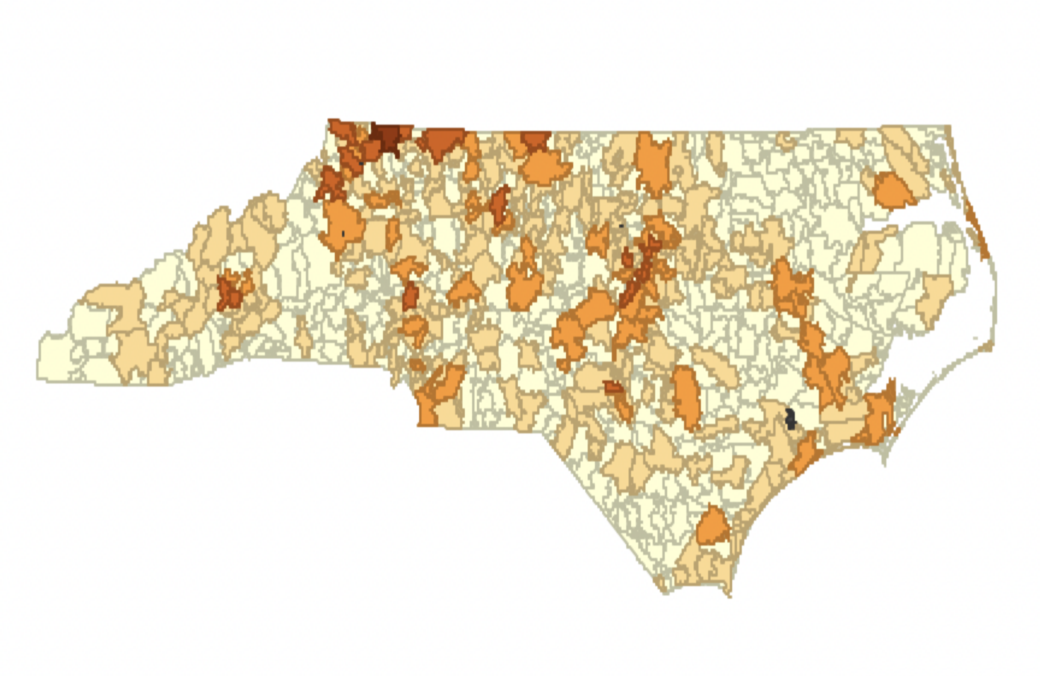
- Basic Concepts
  - Clustered
    - Values are configured or distributed near to one another
  - Random
    - Values are configured or distributed such that there is no regular pattern
  - Ordered (dispersed)
    - Values are configured or distributed in a regular or repeating fashion

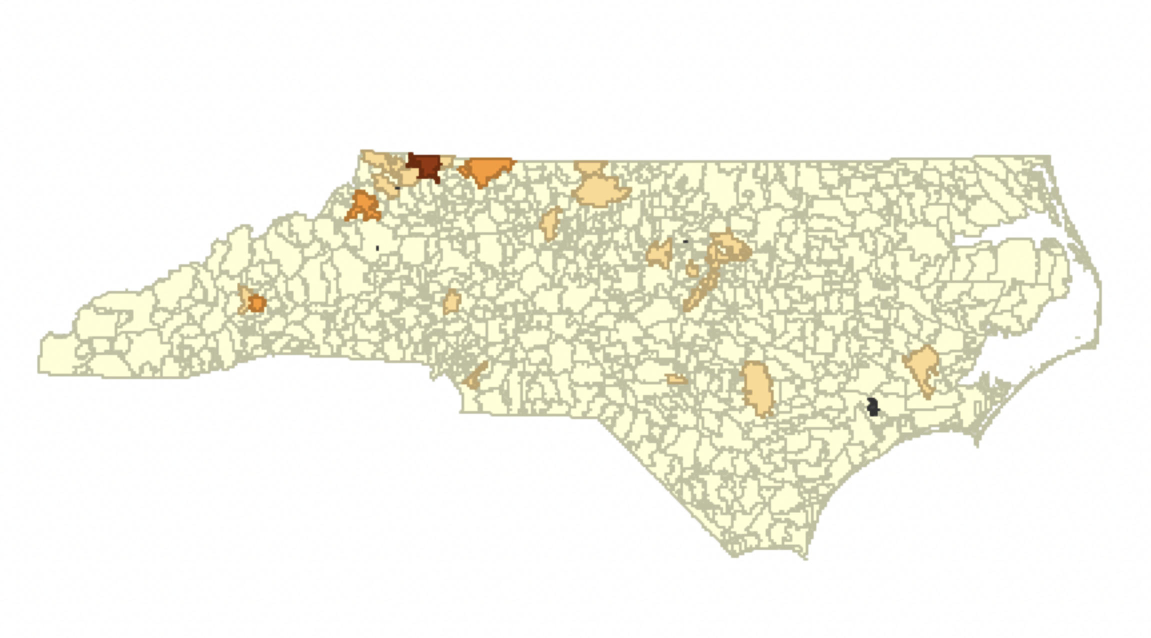




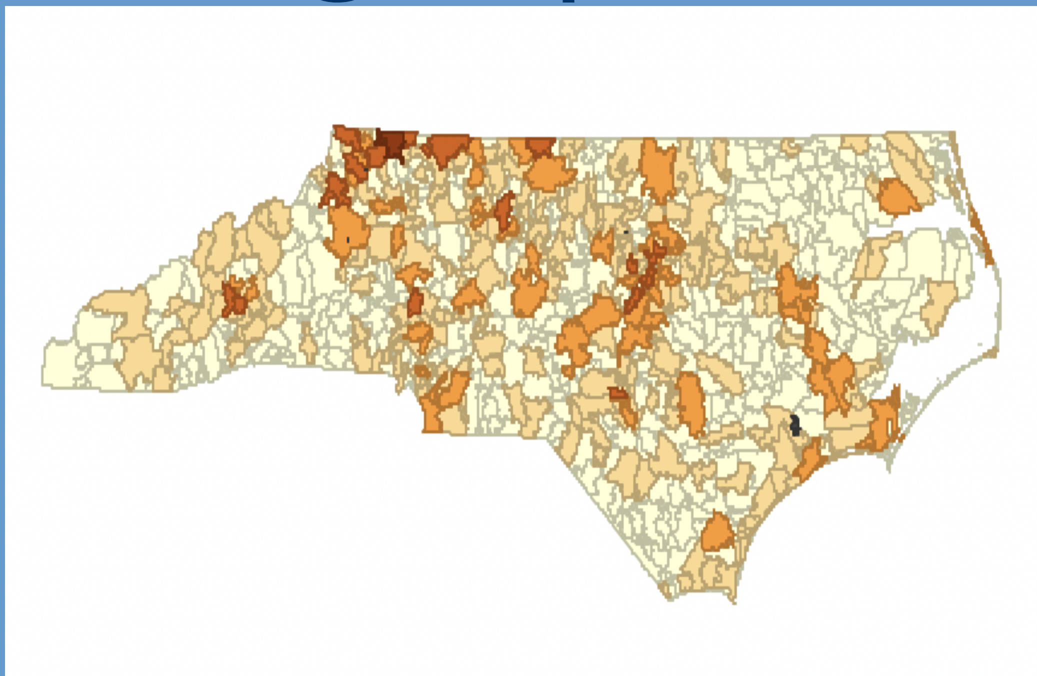


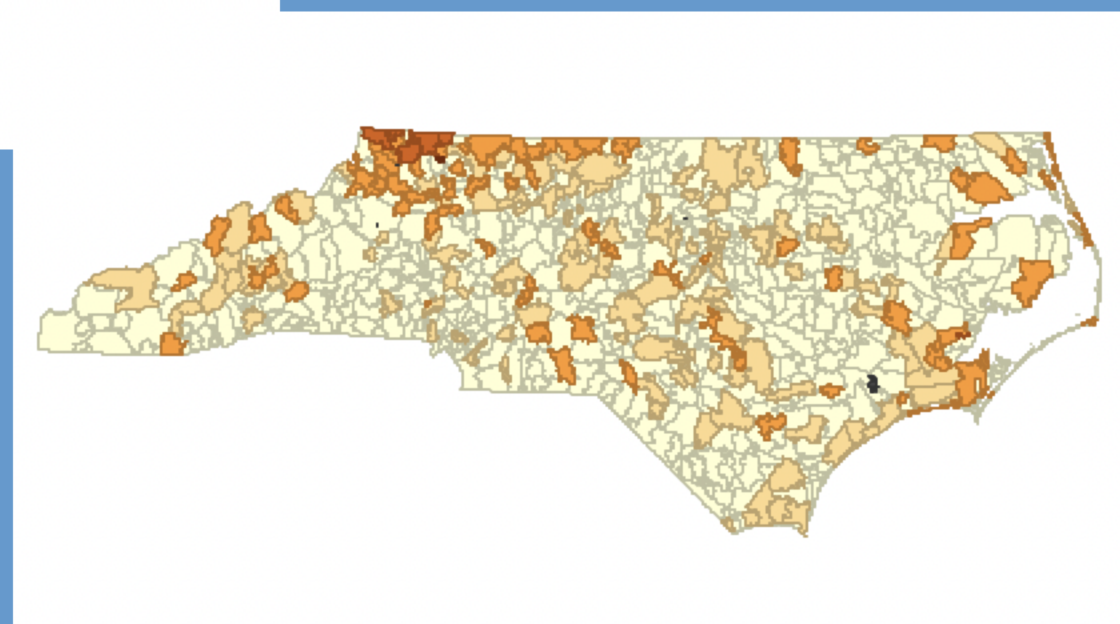
Geographic Variation



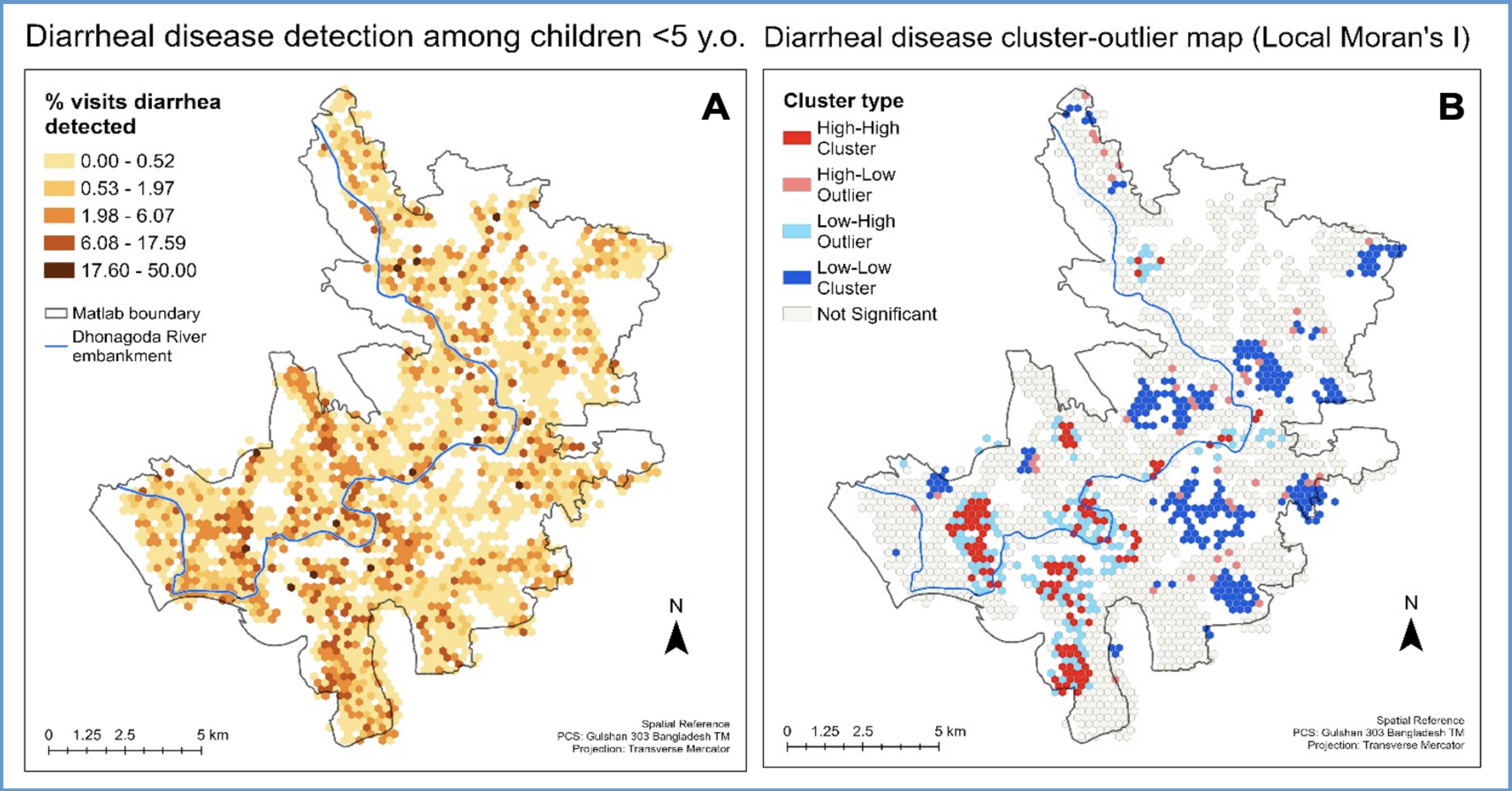


Geographic Variation





Geographic Variation



Geographic Variation

- Visual observation is subjective
  - While patterns may be visible, our eyes do not provide an objective test
  - We have techniques that assist us in interpreting spatial patterns
    - Autocorrelation
    - Clustering

Geographic Variation

- Variation in number of cases (per areal unit)
- Initial concern
  - Population size
    - Even if the disease incidence/prevalence rate is constant, the number of cases will vary geographically based on population size

Geographic Variation

- Calculating incidence/prevalence rates (for areal units)
- Initial concern
  - At risk population
    - Correctly identifying appropriate underlying population for rate calculations (denominator)
      - e.g., age, gender

Geographic Variation

- Calculating incidence/prevalence rates (for areal units)
- Range (meaning) of values
  - When we create choropleth maps, the data will be binned based on the range of the data
    - We still have light-dark colors, even though the data range may be small!
    - Percentage, Box, and Standard Deviation maps in GeoDa

Geographic Variation

- Calculating incidence/prevalence rates (for areal units)
- Initial concern
  - Geographic variation in the composition of the population
    - Especially for variations in age structure!
  - Age standardization
    - Allows for comparison among populations with differing age structures

Geographic Variation

Table - Distribution of the US Population in 1988

Age Group	Population (% of Total)
<5	18,300,000 (7%)
5-19	52,900,000 (22%)
20-44	98,100,000 (40%)
45-64	46,000,000 (19%)
>64	30,400,000 (12%)
Total	245,700,000 (100%)

	Florida	Alaska
Crude mortality rate	1069/100,000	399/100,000
Age-adjusted mortality rate	797/100,000	760/100,00

"What would the comparable death rate be in each state if both populations had identical age distributions?"

Age Standardization

- Direct and Indirect age standardization
  - Both produce a “new” value for each areal unit
  - Can be compared to observed value
    - Calculate a ratio
  - Likely need to be performed outside of QGIS

Age Standardization

- Direct Standardization
  - Requires
    - Age-specific cases and populations for each areal unit
    - A chosen “standard” population
  - Answers the question:
    - What would the overall morbidity/mortality rates be if the areal units all had the exact same population structure

Age Standardization

- Direct Standardization
  - Calculate age specific rates for each areal unit
  - Multiply age specific rates by the number of people in each age group (of the "standard" population)
    - Sum (or calculate new "total" rate for observation units)
    - Ratio: Divide individual areal unit values (calculated rates) by the overall population rate

Age Standardization

- Indirect Standardization
  - Similar to Direct, but you do not have the age specific cases/rates for each areal unit
    - But, we must have them for the overall study area
  - Answers the question:
    - What would the morbidity/mortality rates be if individual areal units experienced that m/m at the same rate as the overall population

Age Standardization

- Indirect Standardization
  - Calculate age specific rates for entire study area
  - Multiply age specific rates by the observed number of people in each age group for each areal unit
    - Sum age group counts for total
    - Divide the observed rate/count for each areal unit by its expected rate/count (calculated)

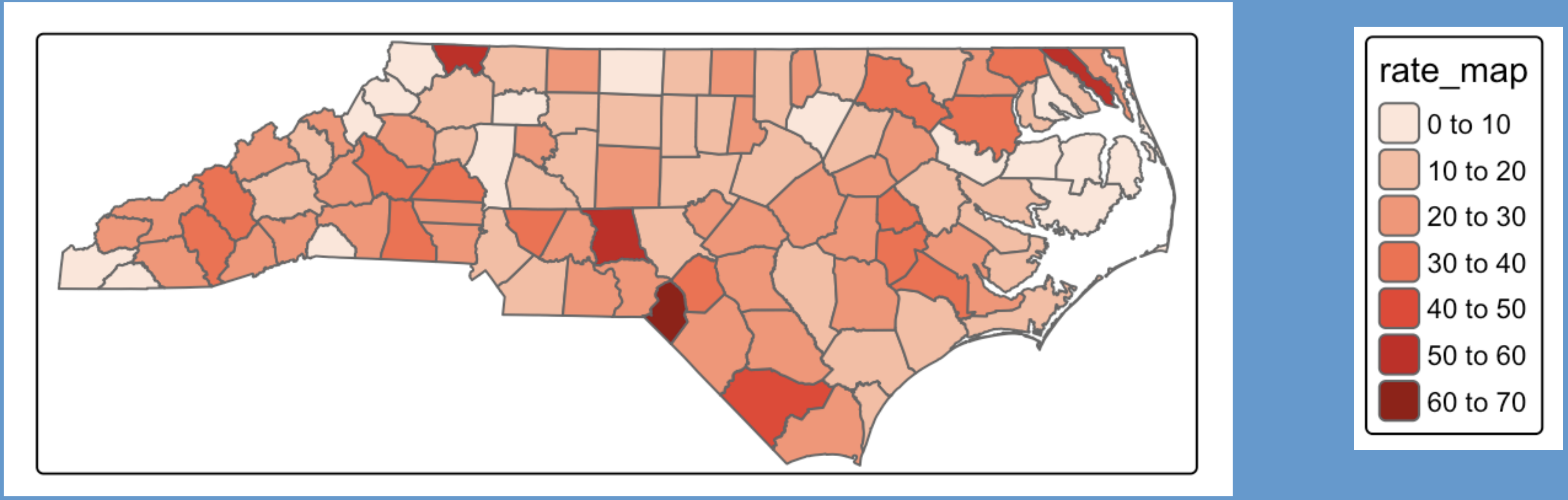
Geographic Variation

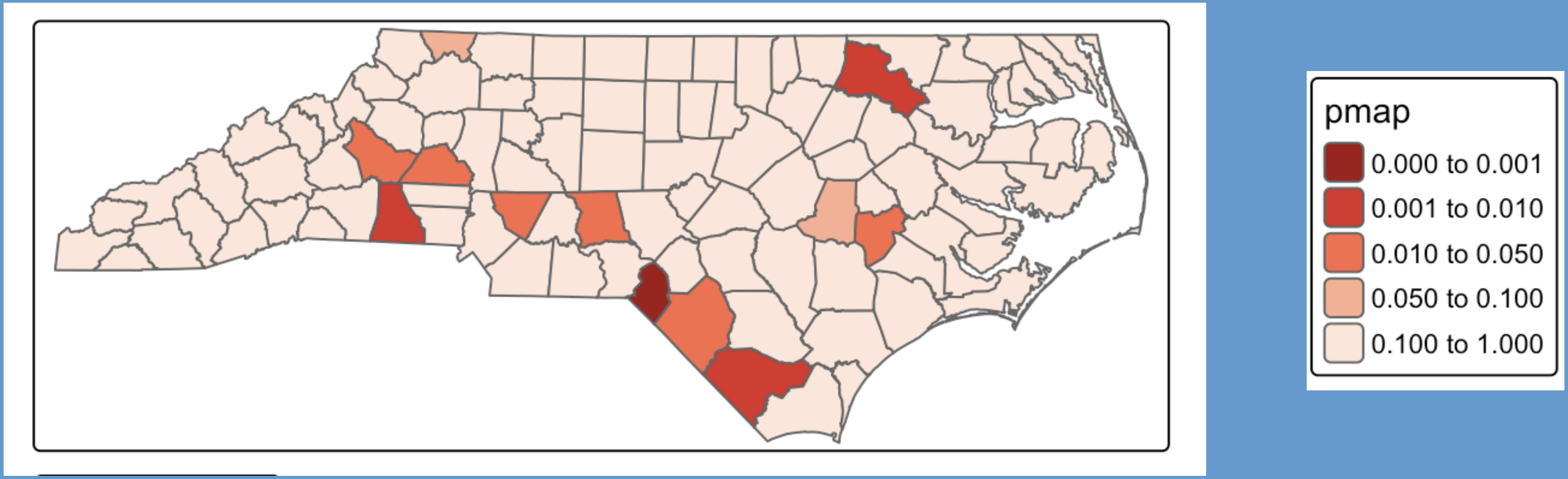
- Calculating incidence/prevalence rates (for areal units)
- Another concern
  - Small numbers problem
    - When areal units have few people, a difference in one case can make a huge difference in the calculated rate
      - Poisson Probabilities
      - Empirical Bayes smoothing
      - Geographic aggregation

Poisson Probabilities

- Map the probability that observed number of cases (or more) would randomly occur (by chance) in a region, given the overall rate of occurrence in the study area

Poisson Probabilities





Poisson Probabilities

- Start with the overall number of cases/events (n) and overall number of people (pop)
- Calculate rate, p, for study area
- For each areal unit calculate the expected number of cases,
- Calculate probability, based on the observed number of cases, k

$$p = \frac{n}{pop}$$

$$P(x \geq k) = 1 - \sum_{x=0}^{k} \frac{\left(e^{-\lambda} \lambda^{x}\right)}{x!}$$

Poisson Probabilities

- Output is a probability, which is then mapped
  - Low probability values are the most likely to be true “high” values
    - However, cannot be interpreted as magnitude!
      - High population regions
      - Small deviations from expected will produce low probabilities

Poisson Probabilities

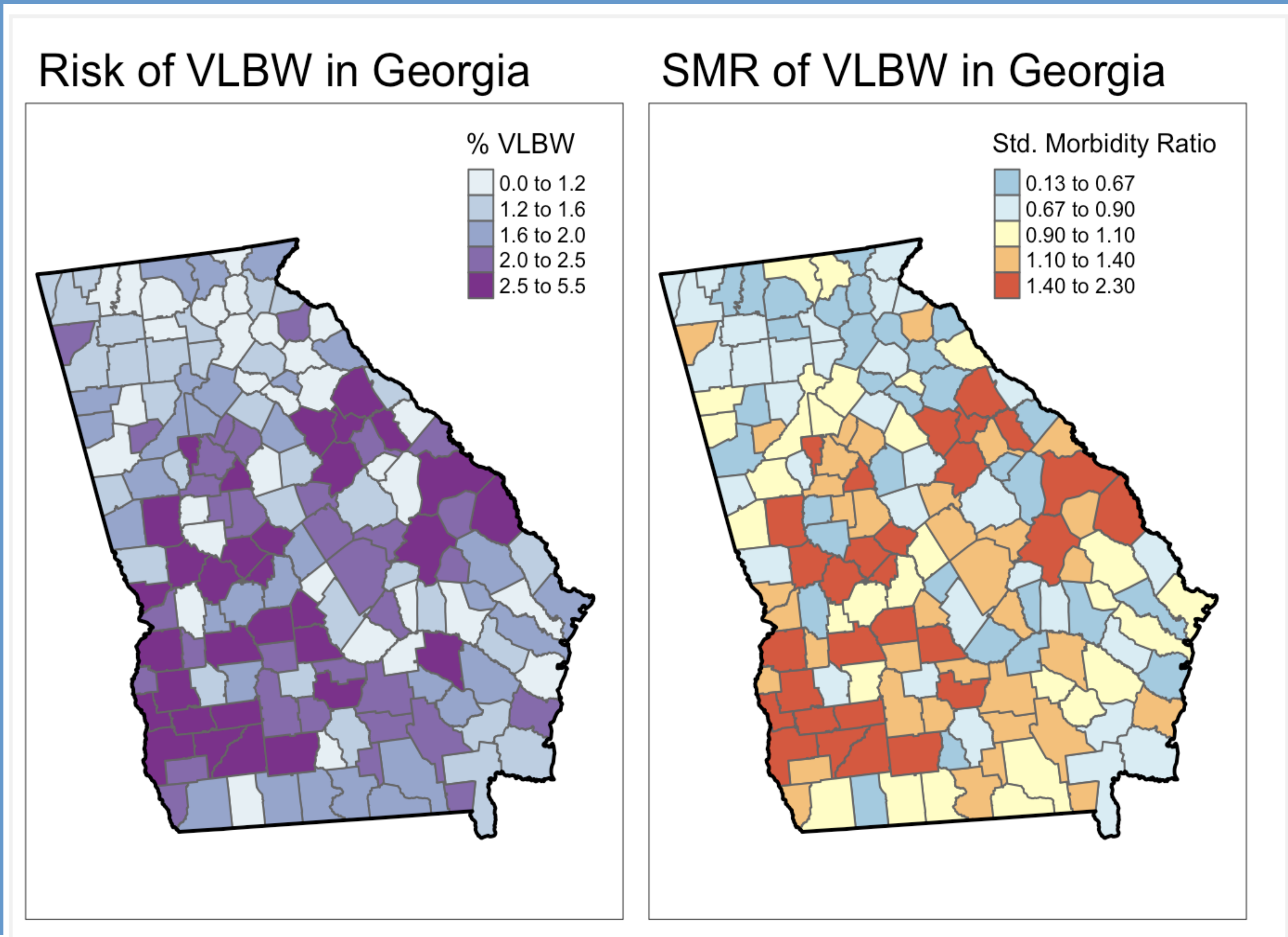
- Must be calculated outside of three software packages we will use in this course
  - QGIS, GeoDa, SaTScan
  - May have to be done in other statistical software
    - I've created an R script for reference

Excess Risk in GeoDa

- Interesting terminology
  - Similar to Poisson, but without the significance
  - Simply a rescaling of rate values based on overall population rate

$$Excess\ Risk = \frac{\dfrac{k_i}{pop_i}}{\dfrac{n}{pop}}$$

Excess Risk in GeoDa



Smoothing

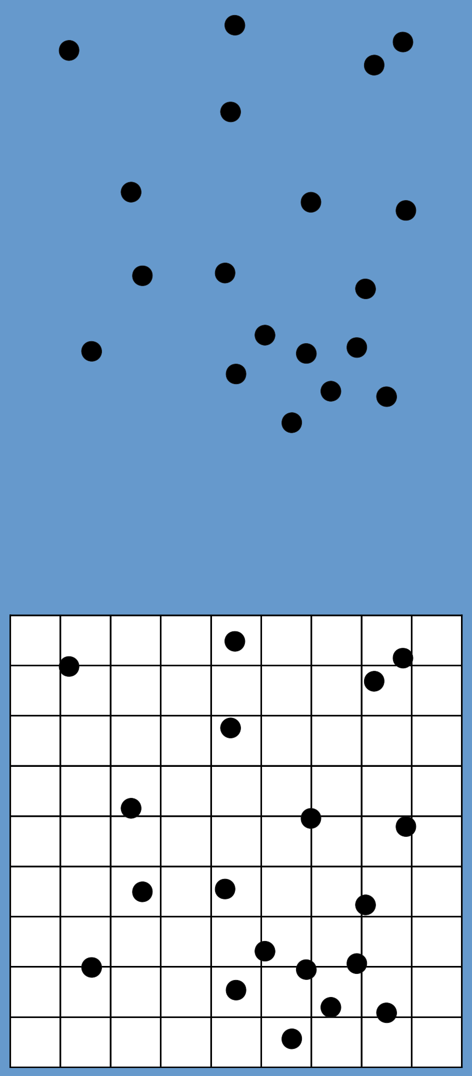
- Use extra information from neighbors to adjust rates
  - Affects observations with a small population
  - Options in GeoDa
    - Spatial Rate
    - Empirical Bayes
    - Spatial Empirical Bayes

Smoothing

- Spatial Rate Smoothing
  - Calculates rate for each areal unit based on rates of unit and neighbors
- Empirical Bayes
  - Uses global (all data) rate to adjust areal unit rates
- Spatial Empirical Bayes
  - Uses local (neighboring regions) rates to adjust areal unit rates

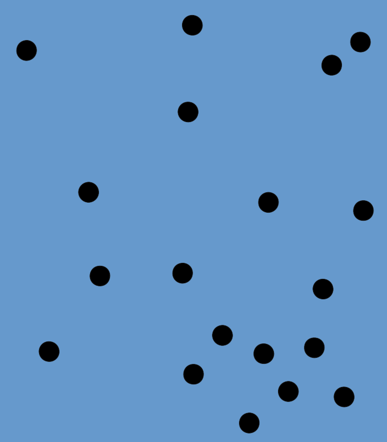
Point Density

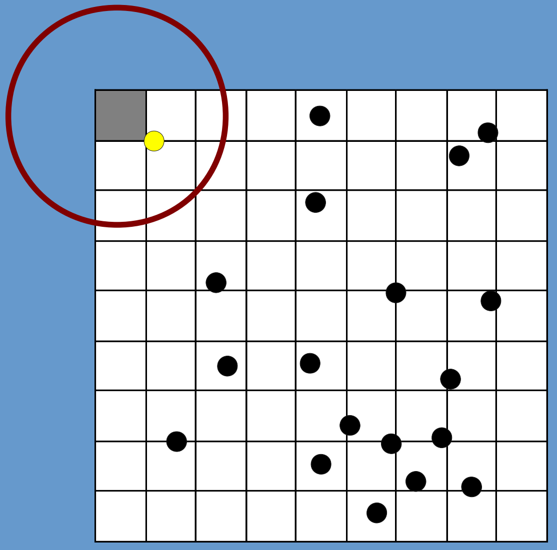
- Calculates density of points within a user-specified window
  - Inputs are "empty" grid (raster) and point locations (generally, vector), output is raster
  - Density value assigned to cell at center of window
  - Can use simple circular windows or distance-weighted kernels



Point Density

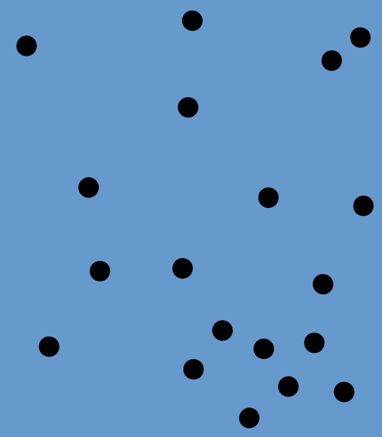
- Calculates density of points within a user-specified window
  - Inputs are “empty” grid (raster) and point locations (generally, vector), output is raster
  - Density value assigned to cell at center of window
  - Can use simple circular windows or distance-weighted kernels

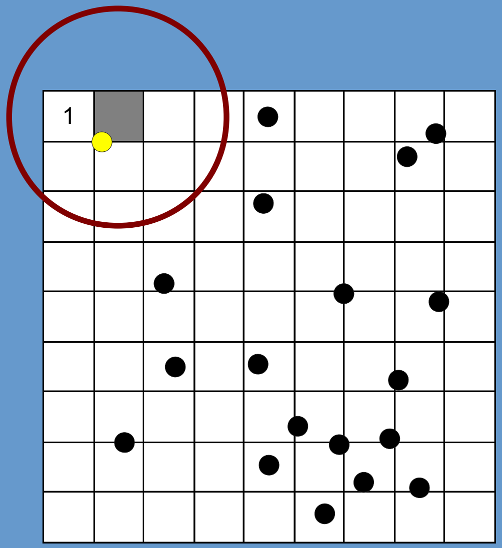




Point Density

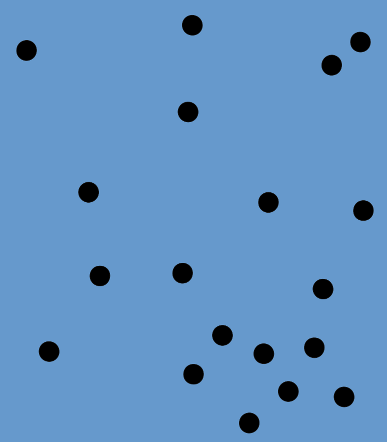
- Calculates density of points within a user-specified window
  - Inputs are "empty" grid (raster) and point locations (generally, vector), output is raster
  - Density value assigned to cell at center of window
  - Can use simple circular windows or distance-weighted kernels

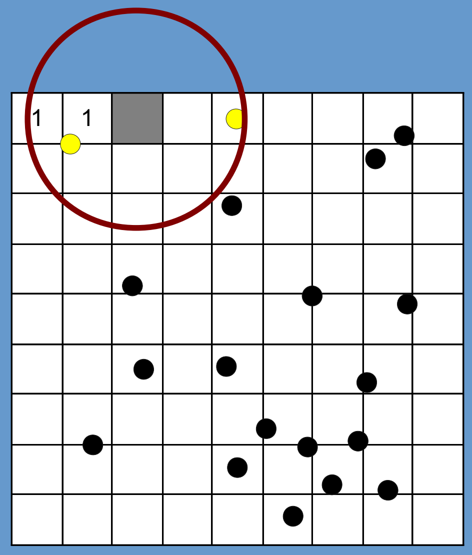




Point Density

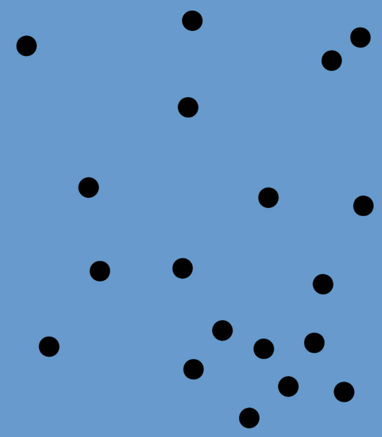
- Calculates density of points within a user-specified window
  - Inputs are "empty" grid (raster) and point locations (generally, vector), output is raster
  - Density value assigned to cell at center of window
  - Can use simple circular windows or distance-weighted kernels

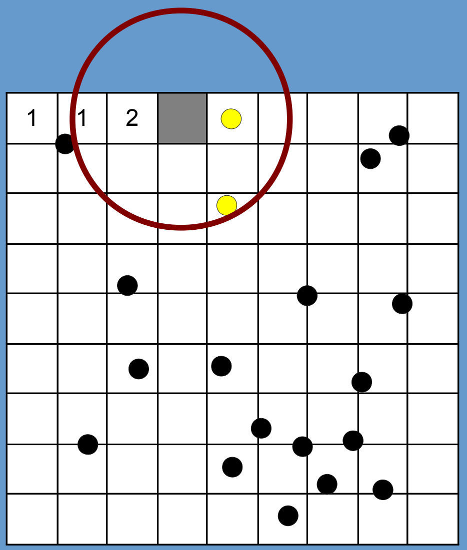




Point Density

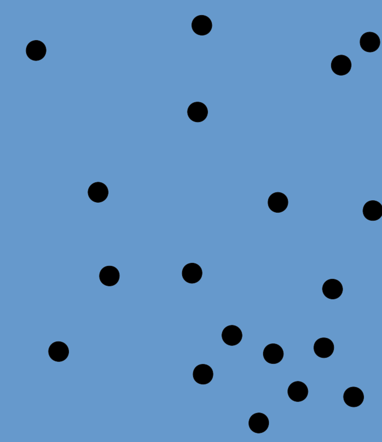
- Calculates density of points within a user-specified window
  - Inputs are "empty" grid (raster) and point locations (generally, vector), output is raster
  - Density value assigned to cell at center of window
  - Can use simple circular windows or distance-weighted kernels

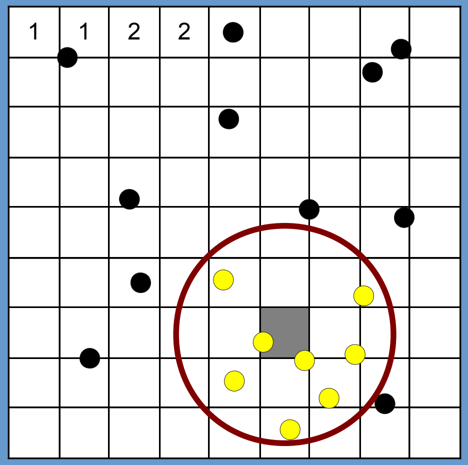




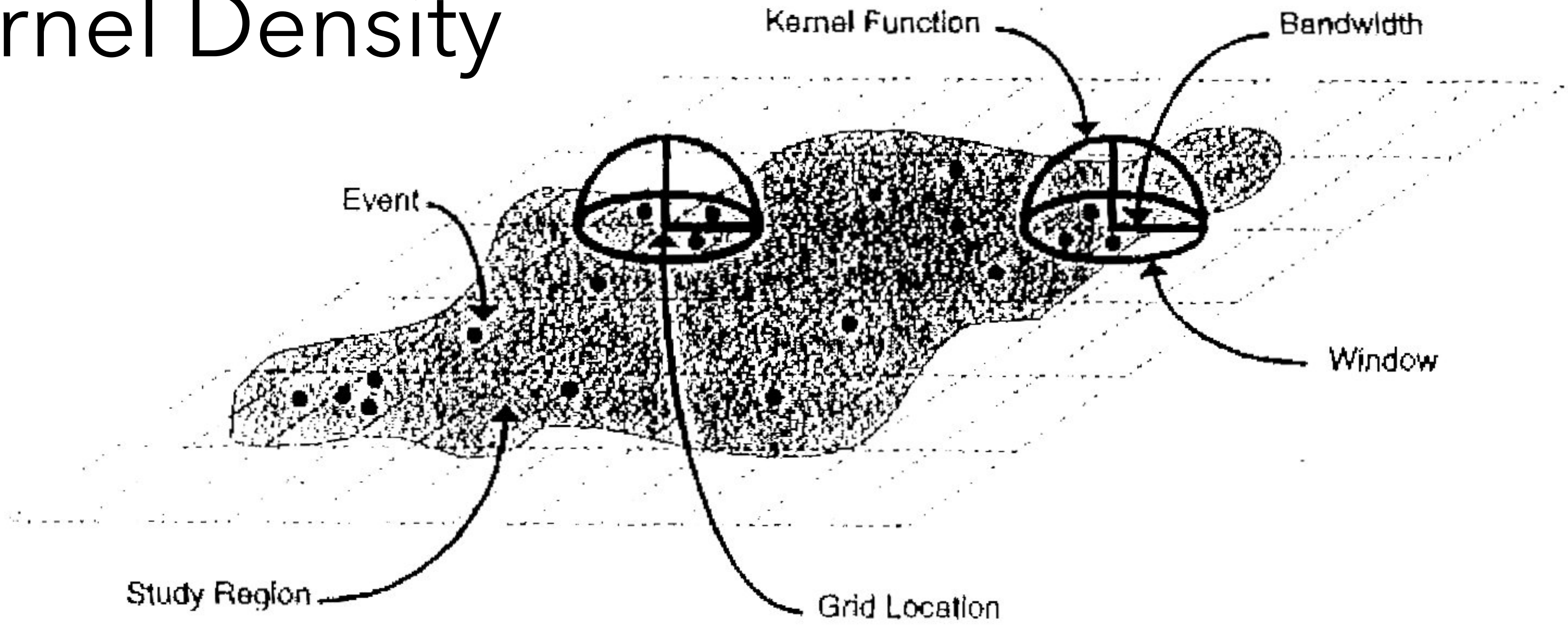
Point Density

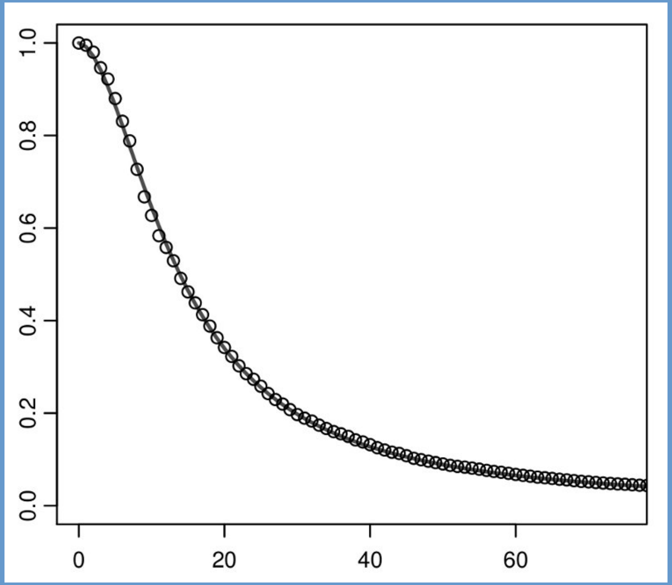
- Calculates density of points within a user-specified window
  - Inputs are "empty" grid (raster) and point locations (generally, vector), output is raster
  - Density value assigned to cell at center of window
  - Can use simple circular windows or distance-weighted kernels

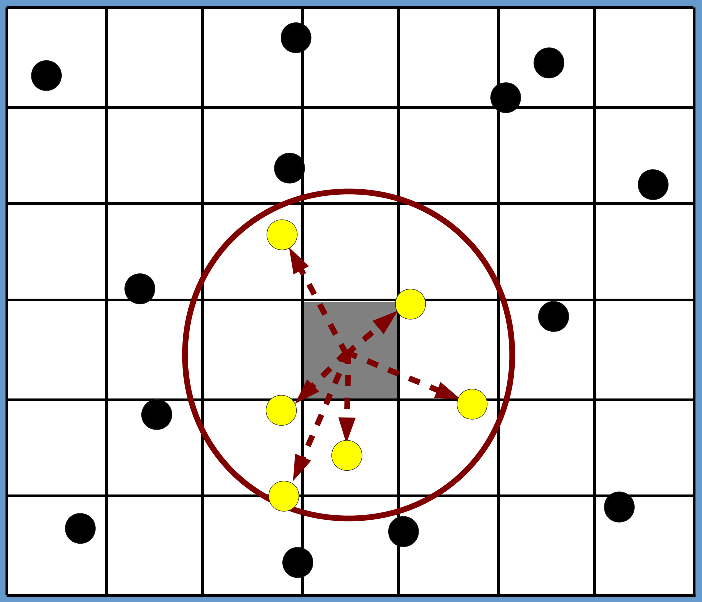




Kernel Density

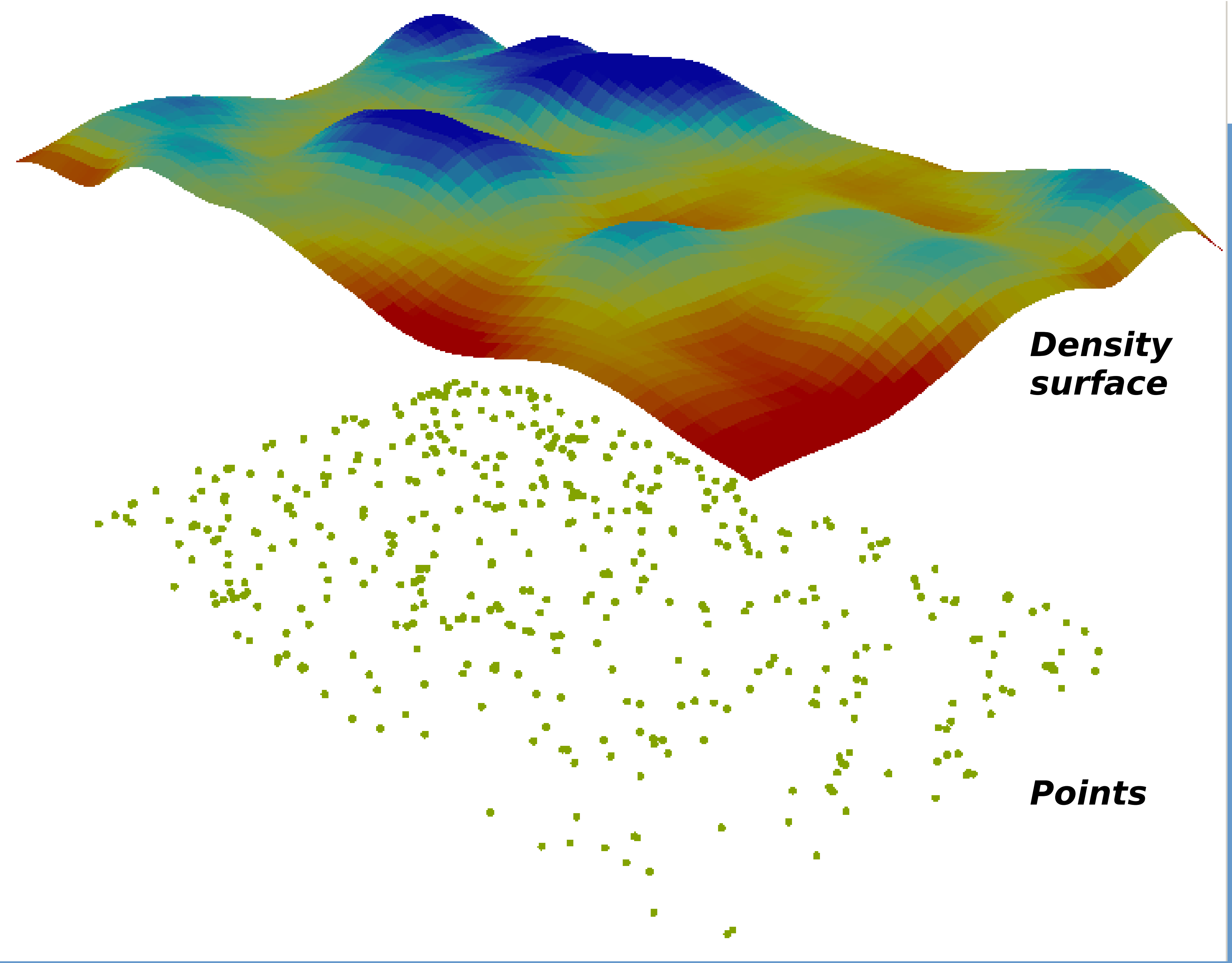


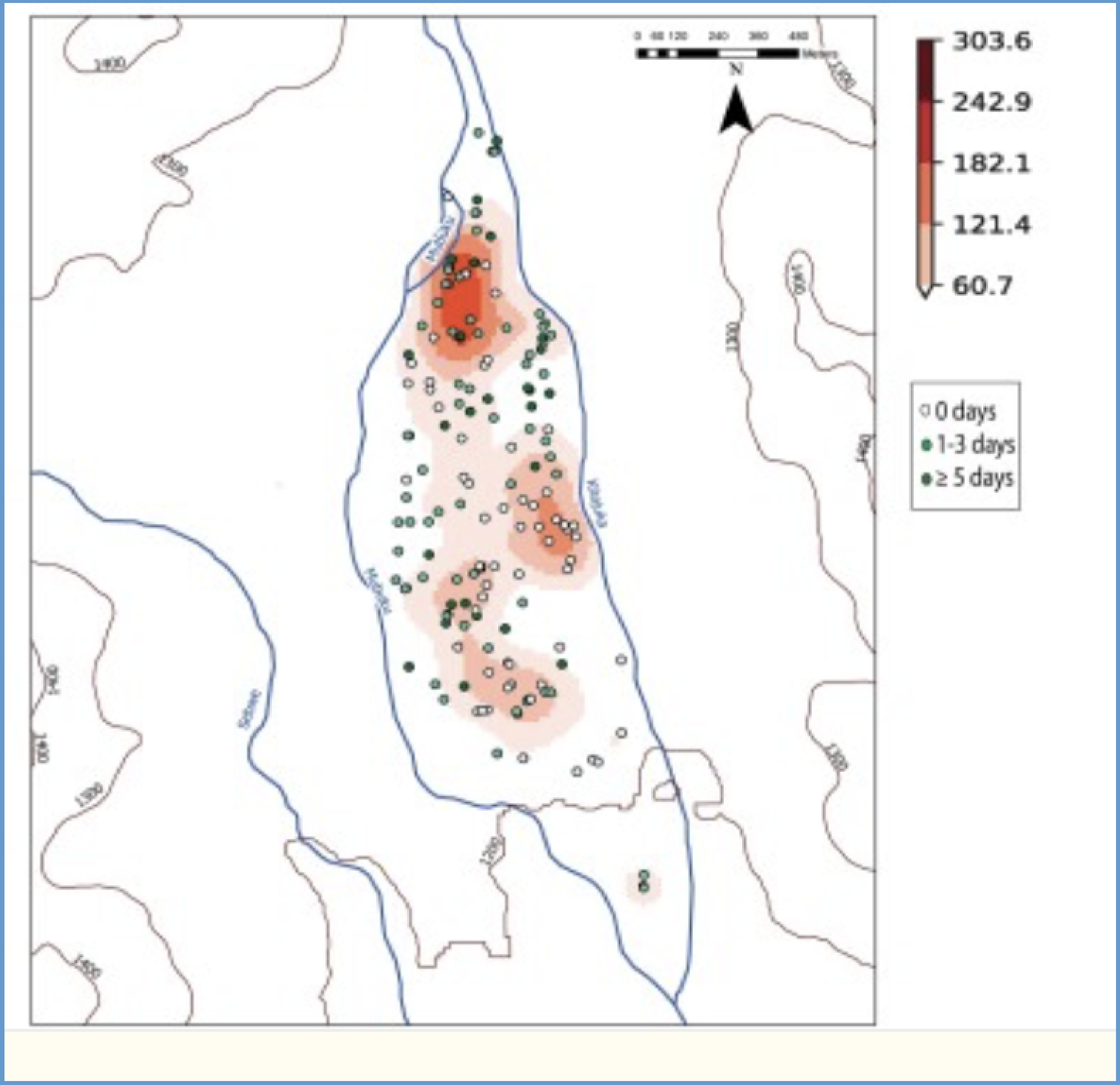


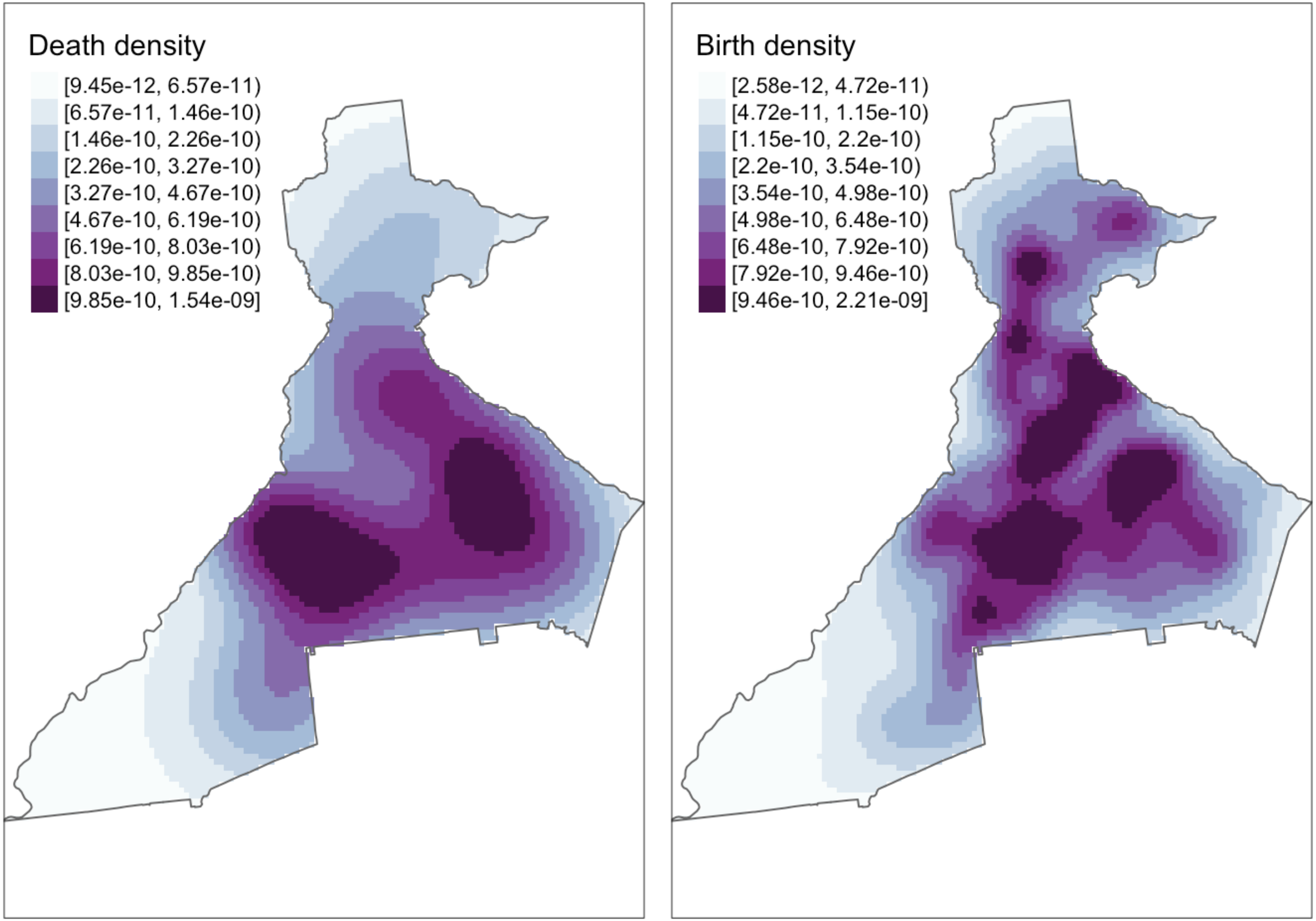


$W_1 = 0.7$
$W_2 = 0.7$
$W_3 = 0.7$
$W_4 = 0.2$
$W_5 = 0.2$
$W_6 = 0.1$

2.6


Density
surface
Points

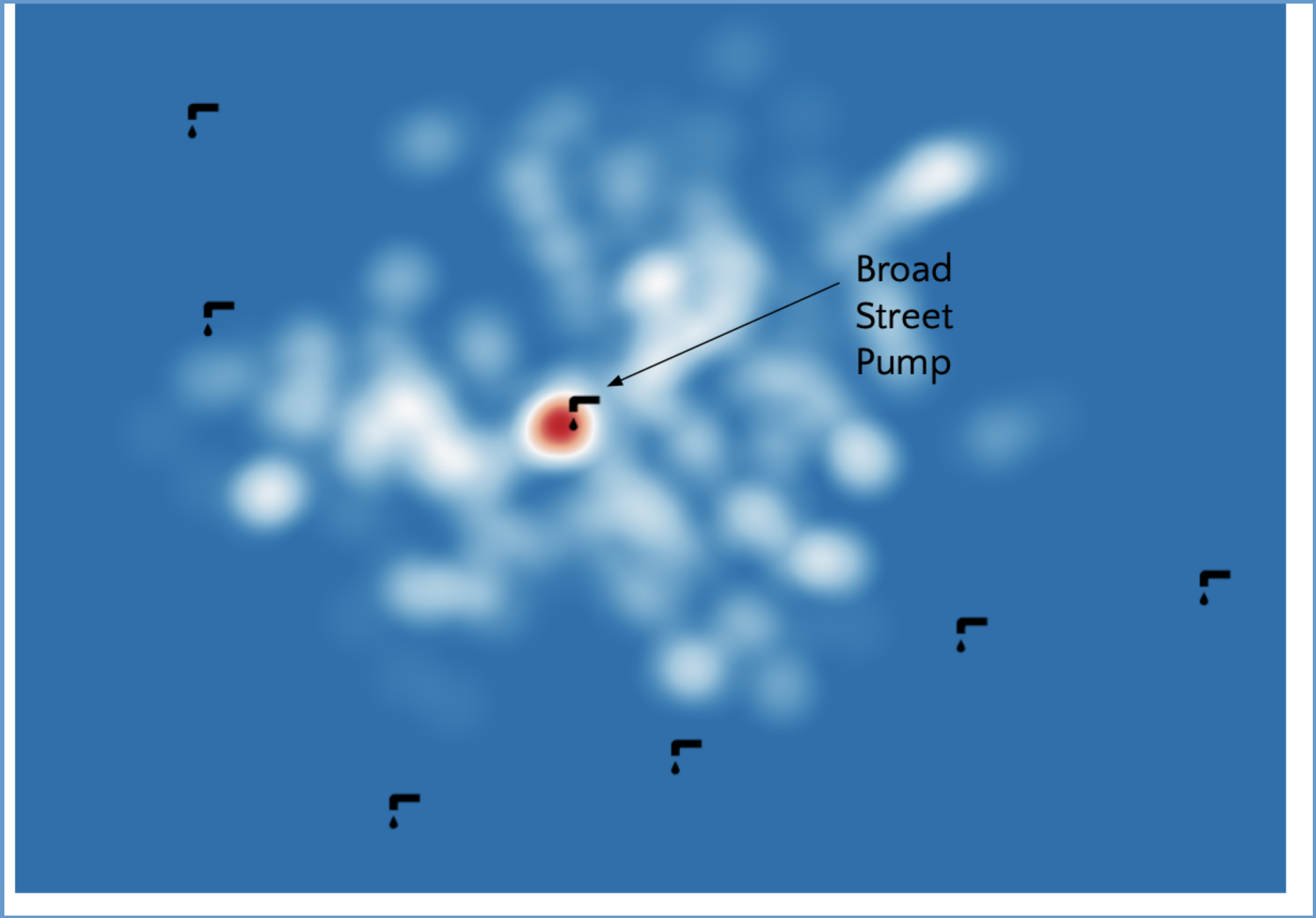

303.6
242.9
182.1
121.4
60.7
0 days
1-3 days
≥ 5 days
N


Death density
[9.45e-12, 6.57e-11)
[6.57e-11, 1.46e-10)
[1.46e-10, 2.26e-10)
[2.26e-10, 3.27e-10)
[3.27e-10, 4.67e-10)
[4.67e-10, 6.19e-10)
[6.19e-10, 8.03e-10)
[8.03e-10, 9.85e-10)
[9.85e-10, 1.54e-09]
Birth density
[2.58e-12, 4.72e-11)
[4.72e-11, 1.15e-10)
[1.15e-10, 2.2e-10)
[2.2e-10, 3.54e-10)
[3.54e-10, 4.98e-10)
[4.98e-10, 6.48e-10)
[6.48e-10, 7.92e-10)
[7.92e-10, 9.46e-10)
[9.46e-10, 2.21e-09]

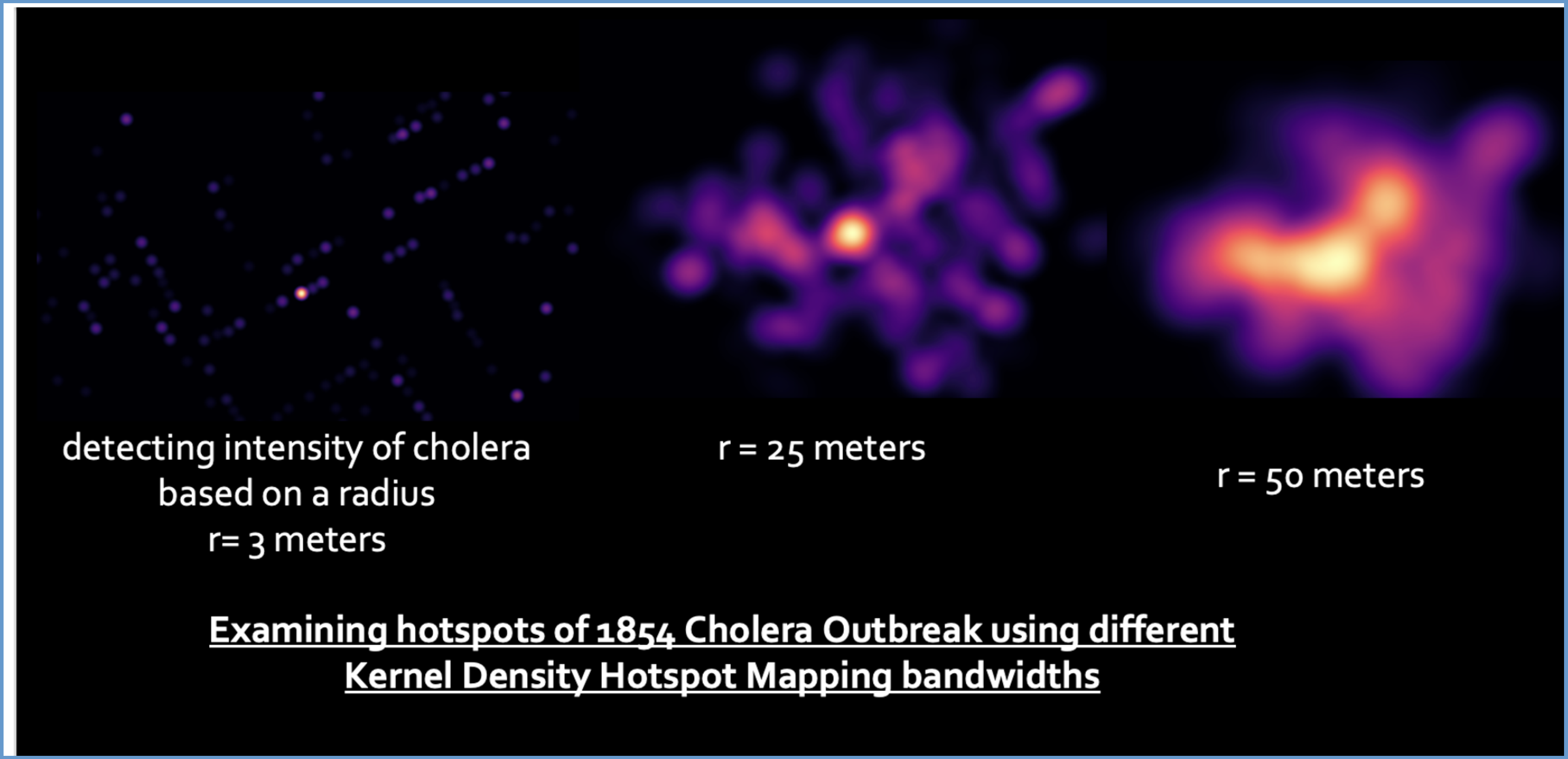
Kernel Density

- For display only?
  - Output is EXTREMELY sensitive to size and shape of the kernel
    - The distance and the function
  - Beware of using the output values in analysis

Kernel Density



Kernel Density



Keywords

- Population size
- At risk population
- Population composition
- Age standardize
- Poisson probabilities
- Smoothing
- Point and Kernel density