Spatial Point Pattern Analysis: Theory and Concepts

GEOG 215 - April 13, 2020

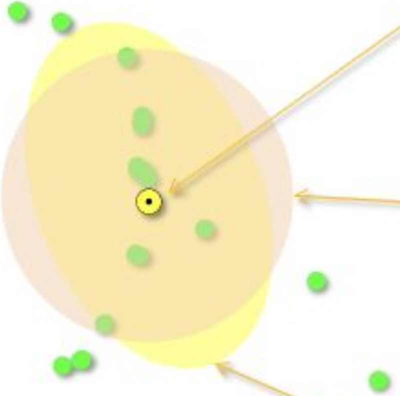
Today’s Agenda

- Spatial Point Pattern Analysis
  - Descriptive Analysis
    - Centrography
  - Density Based Analysis
    - Quadrat method
    - Kernel Density
  - Distance Based Analysis
    - Nearest Neighbor Analysis

Recall

- Tobler's first law of Geography
  - *Everything is related to everything else, but near things are more related than distant things*
    - Values at locations near each other tend to be similar, with similarity decreasing with distance
  - Implies that phenomena are not distributed randomly (throughout space)
    - Imagine how the world would appear if everything were randomly distributed

Centrography



Mean center
computed average X and Y coordinate values.

$$\bar{s} = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \right)$$

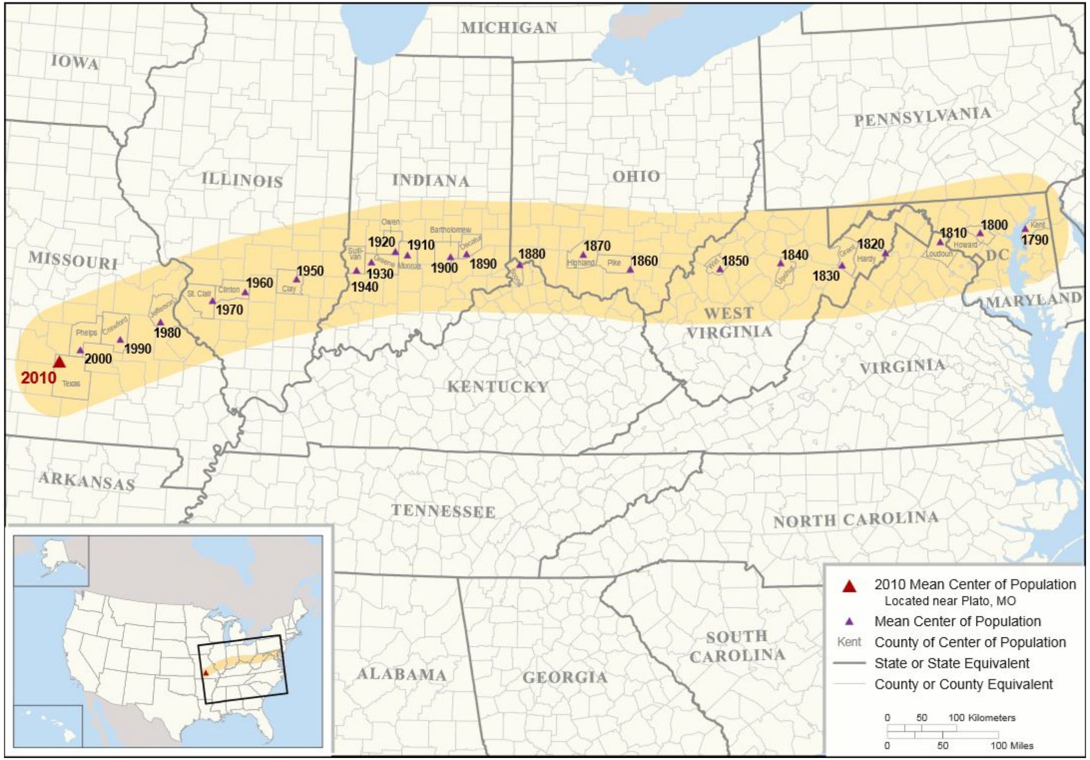
Standard distance
measure of the variance between the average distance of the features to the mean center.

$$d = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu_x)^2 + (y_i - \mu_y)^2}{n}}$$

Standard deviational ellipse
separate standard distances for each axis.

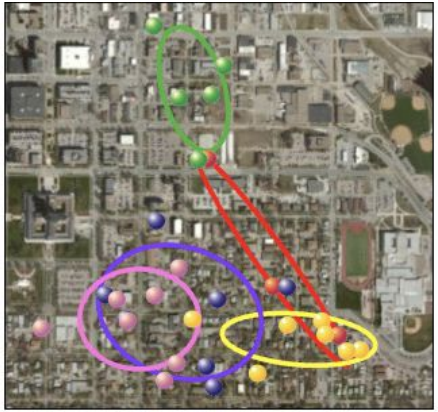
$$\begin{cases} d_x = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \mu_x)^2}{n}} \\ d_y = \sqrt{\dfrac{\sum_{i=1}^{n} (y_i - \mu_y)^2}{n}} \end{cases}$$

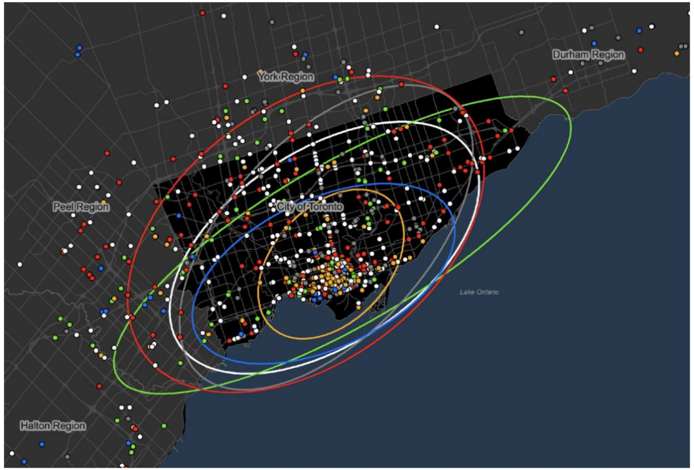
Mean Center of the United States Population

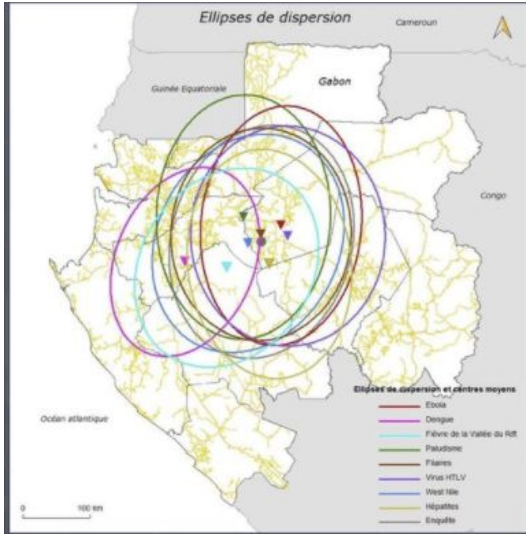


https://en.wikipedia.org/wiki/Mean_center_of_the_United_States_population

Standard Deviational Ellipses







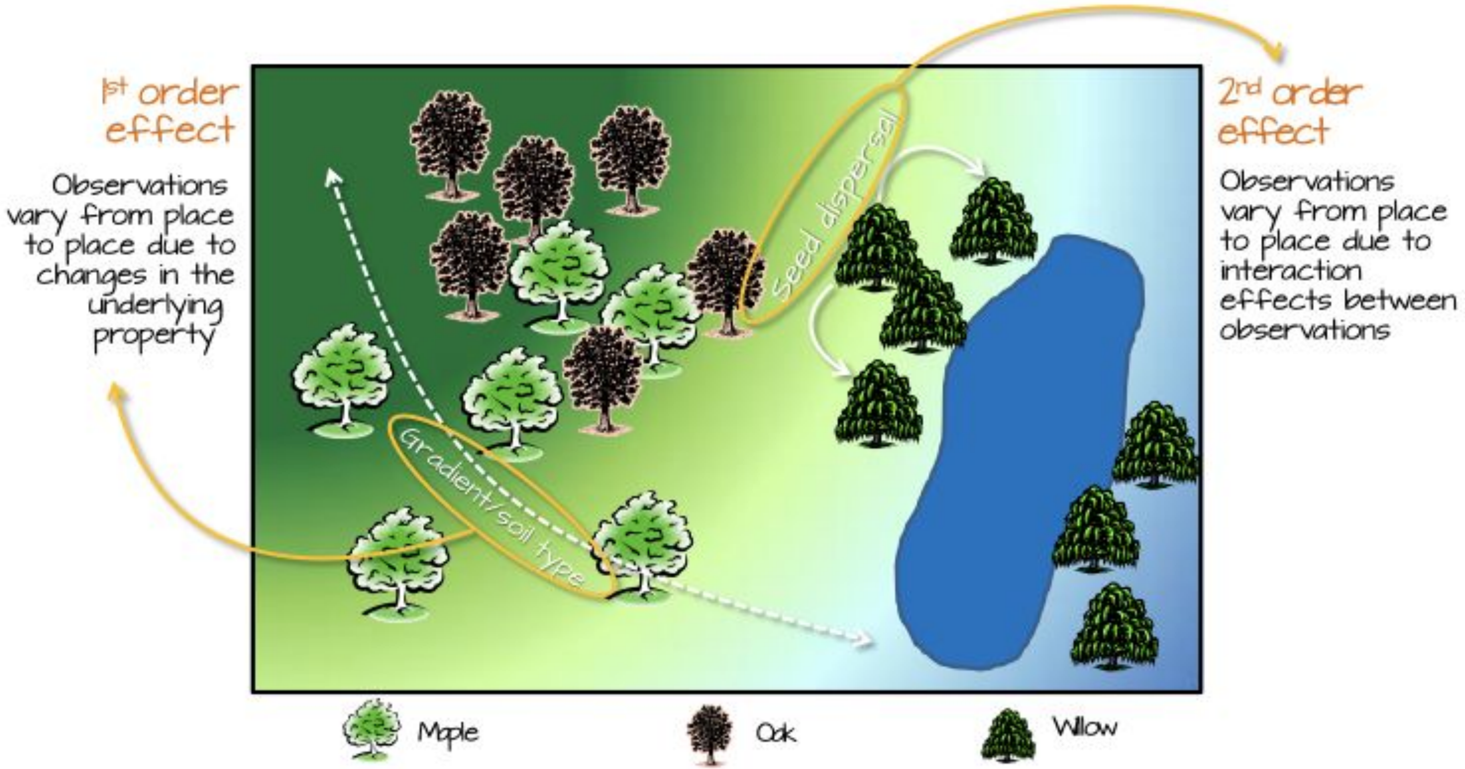
Density Based Analysis

- Characterizes the 1st order property of a spatial pattern
  - Concerns itself with variation of an observations density across the study area
    - Interest in overall trend - how are observations distributed relative to the study area
      - Mainly due to some underlying property of the study area rather than the observation itself
      - For example, the distribution of oaks will vary across a landscape based on underlying soil characteristics (resulting in areas having dense clusters of oaks and other areas not).

Distance Based Analysis

- Characterizes the 2nd order property of a spatial pattern
  - Concerns itself with an observation’s influence on one another
    - Interest in how the points are distributed relative to one another rather than relative to the study extent
    - Related to internal properties of the observation
      - For example, the distribution of oaks will be influenced by the location of parent trees–where parent oaks are present we would expect dense clusters of oaks to emerge.

1st and 2nd order effects



1st and 2nd order effects

- First order effects:
  - influence of external or environmental factors on process outcomes; e.g., abundance of plants within a sub-region could depend on soil type depend on soil type.
    - Note: first-order effects are typically assumed to influence the magnitude of process outcomes at each location, and hence are associated with the mean of all possible process outcomes at each location

- Second-order effects:
  - influence of process outcomes at one location on possible process outcomes at nearby locations; e g non outcomes at nearby locations; e.g., non-contagious contagious versus contagious diseases.
    - Note: second-order effects typically express some measure of “similarity” between possible process outcomes at different locations once the first-order effects have been removed, and are often associated with the covariance or correlation coefficient between different random variables.
- Often hard to completely distinguish

Density Based Analysis

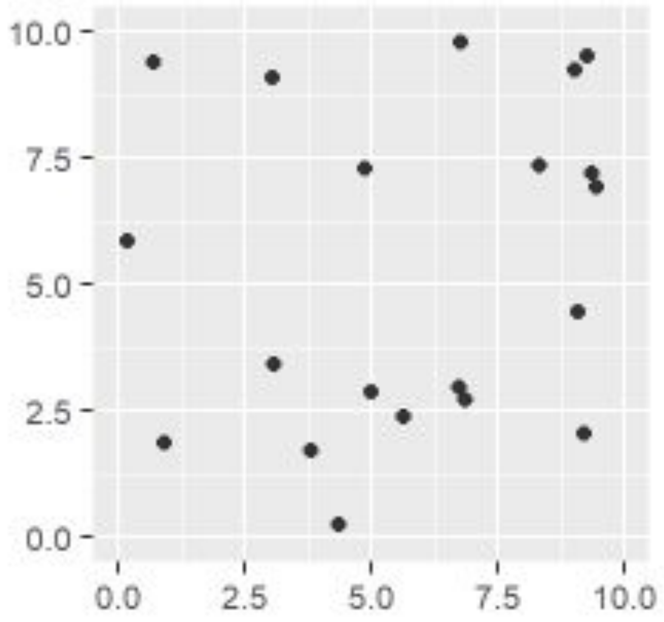
- Global measures
  - Concerns itself with variation of an observations density across the study area
- Local measures
  - Assume that variation across the study area is not uniform
    - Non-stationarity !!!!
    - There might be some 2nd order effects!

Global Density Analysis

λ = n/a

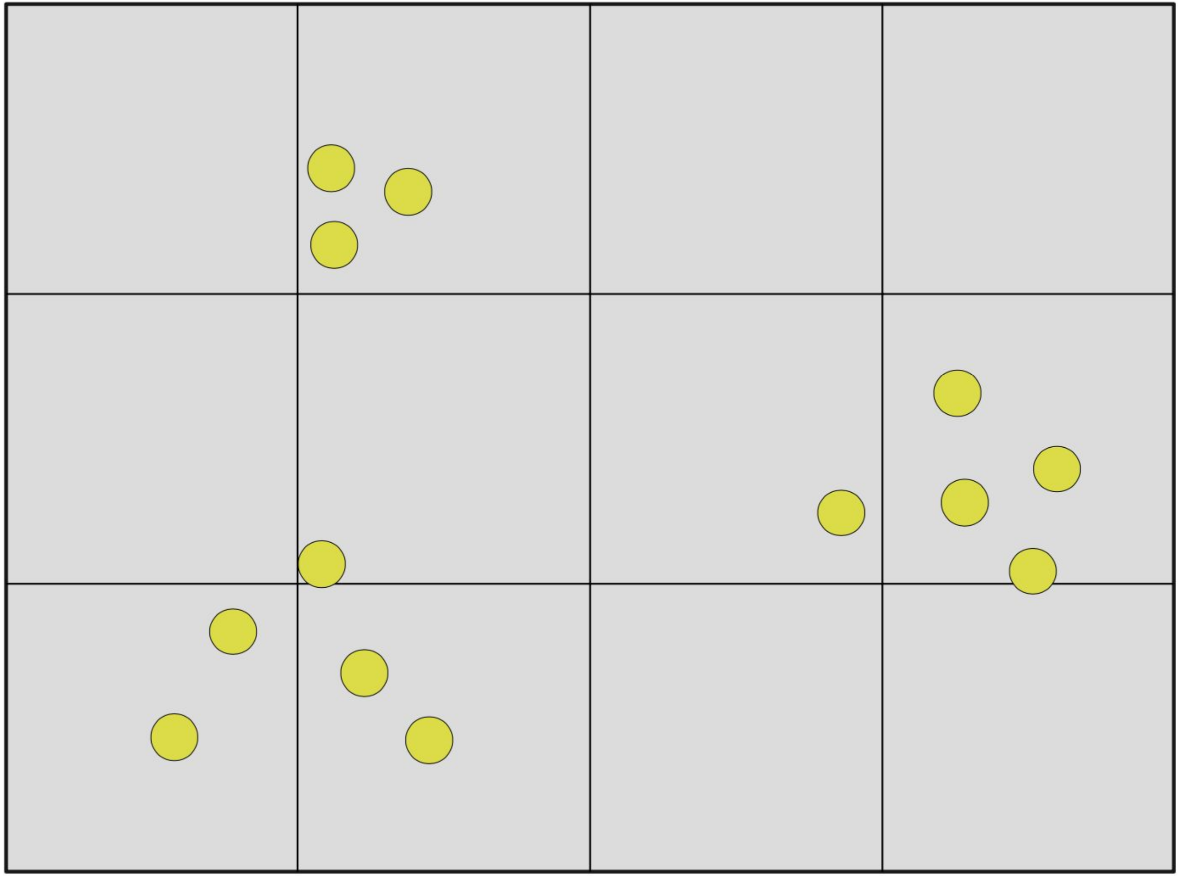
λ = estimated density
n = number of points
A = area of study area

λ is referred to as the estimated density of the observed pattern and the estimated intensity of the spatial process underlying the pattern

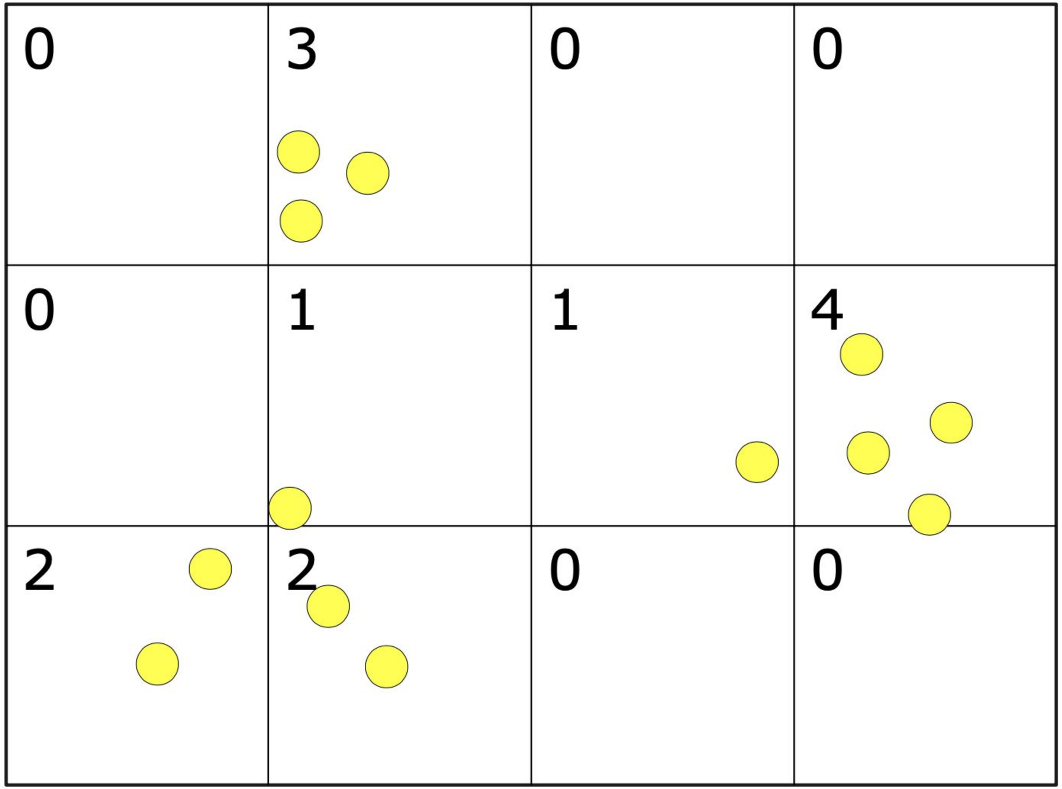


Local density analysis

- Quadrat count/density analysis
  - Examine spatial distribution of points
    - Are they clustered? Random? Dispersed?
  - Examine the frequency/density of points located throughout the regions
    - Sections the study region into quadrats
    - Considers variance among regions

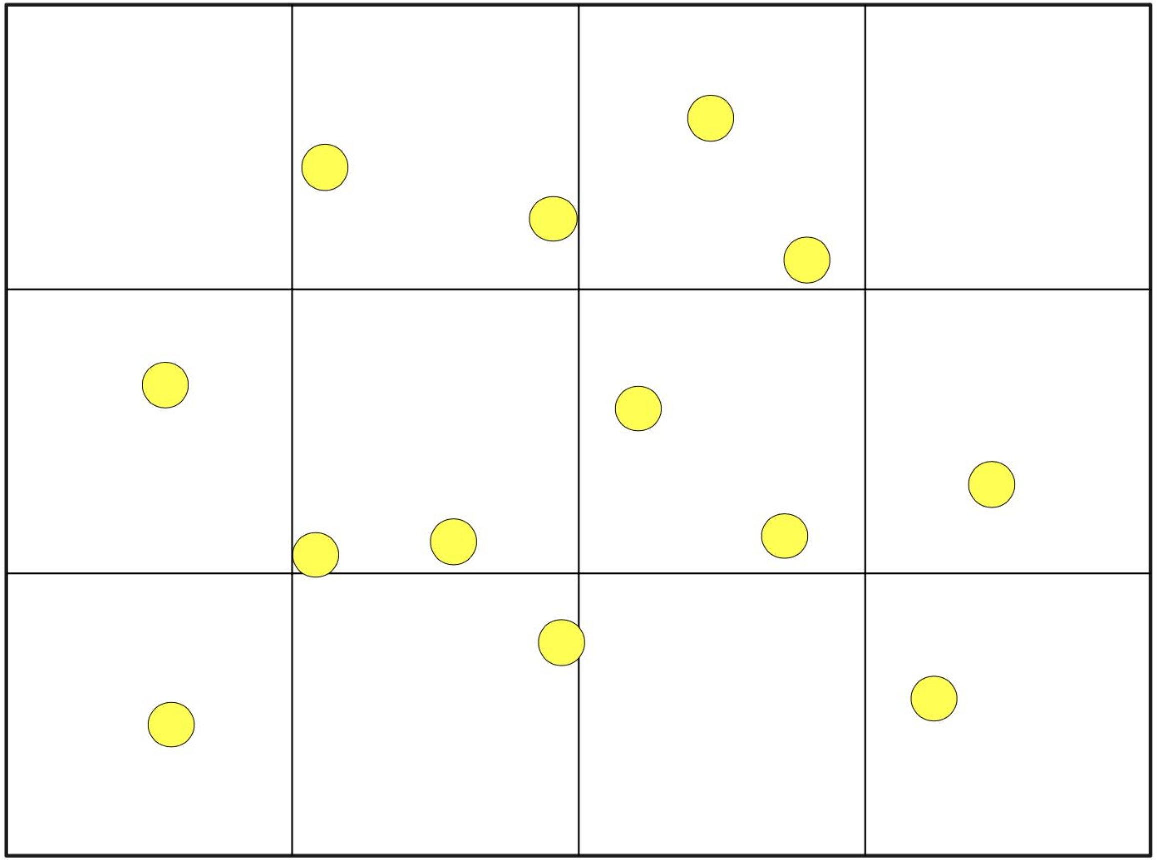


Clustered

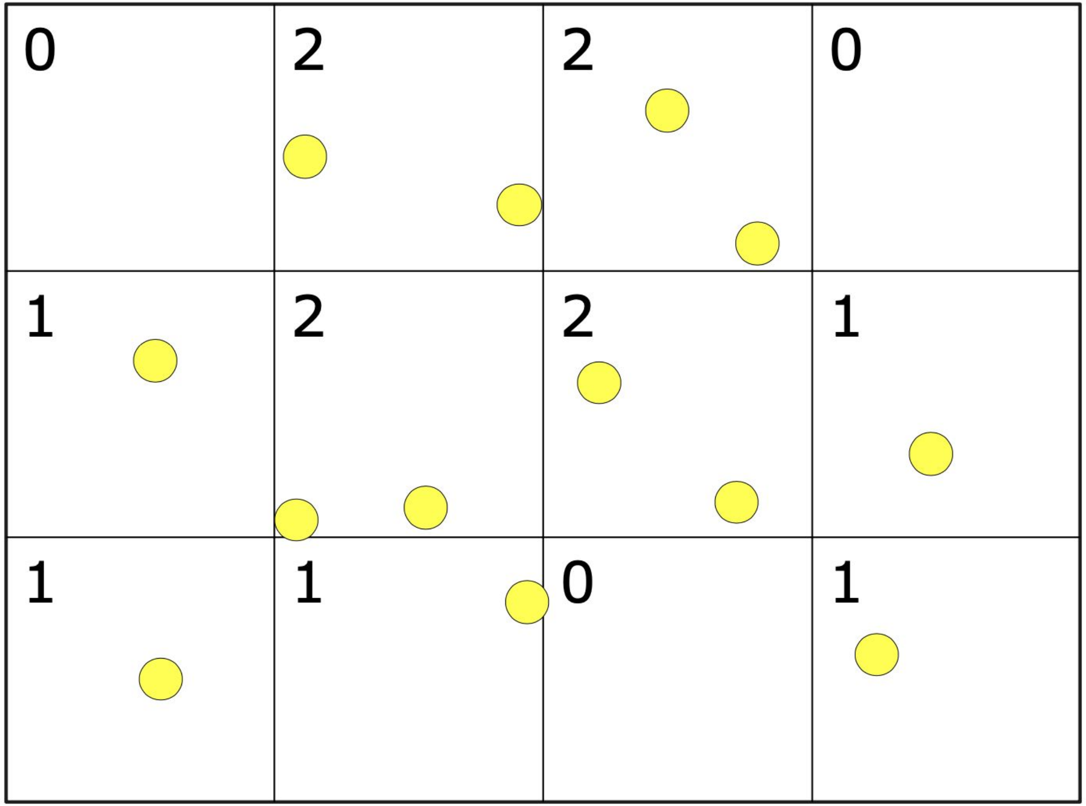


Clustered

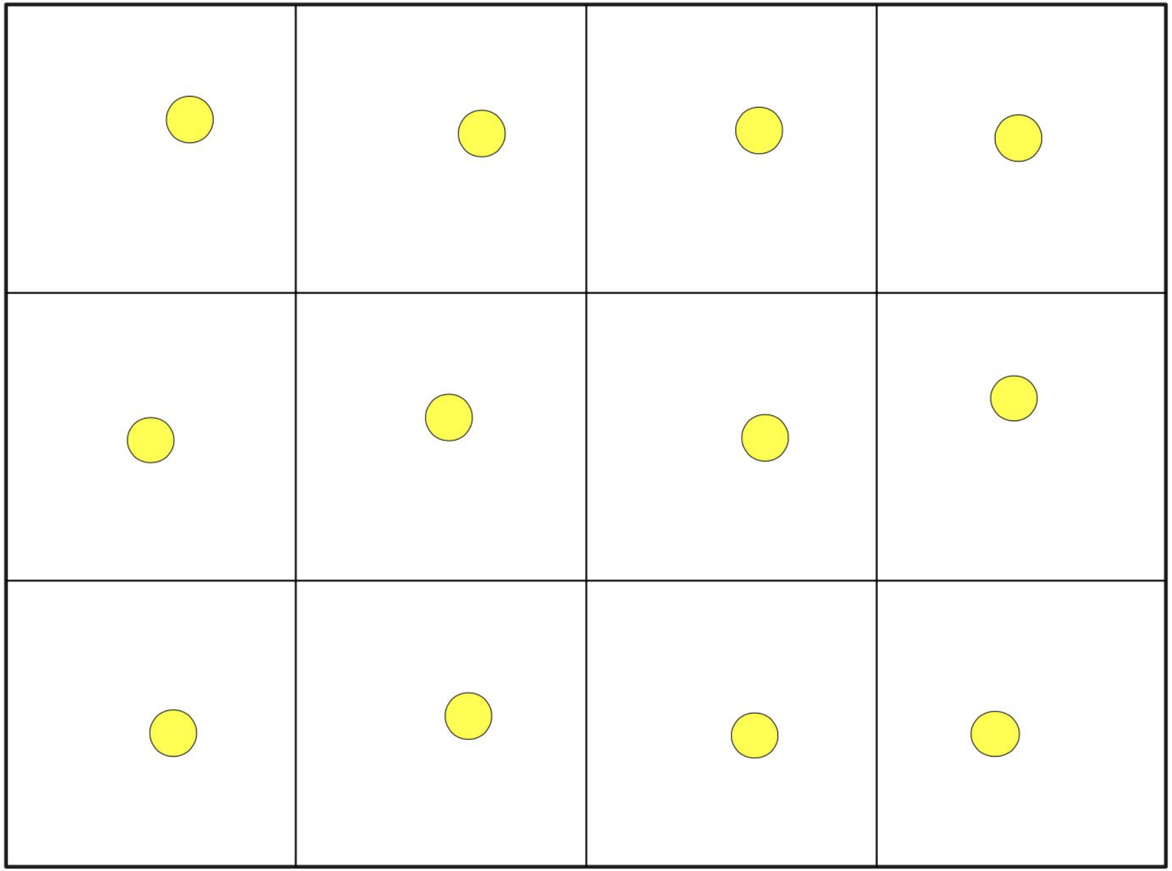
High variation



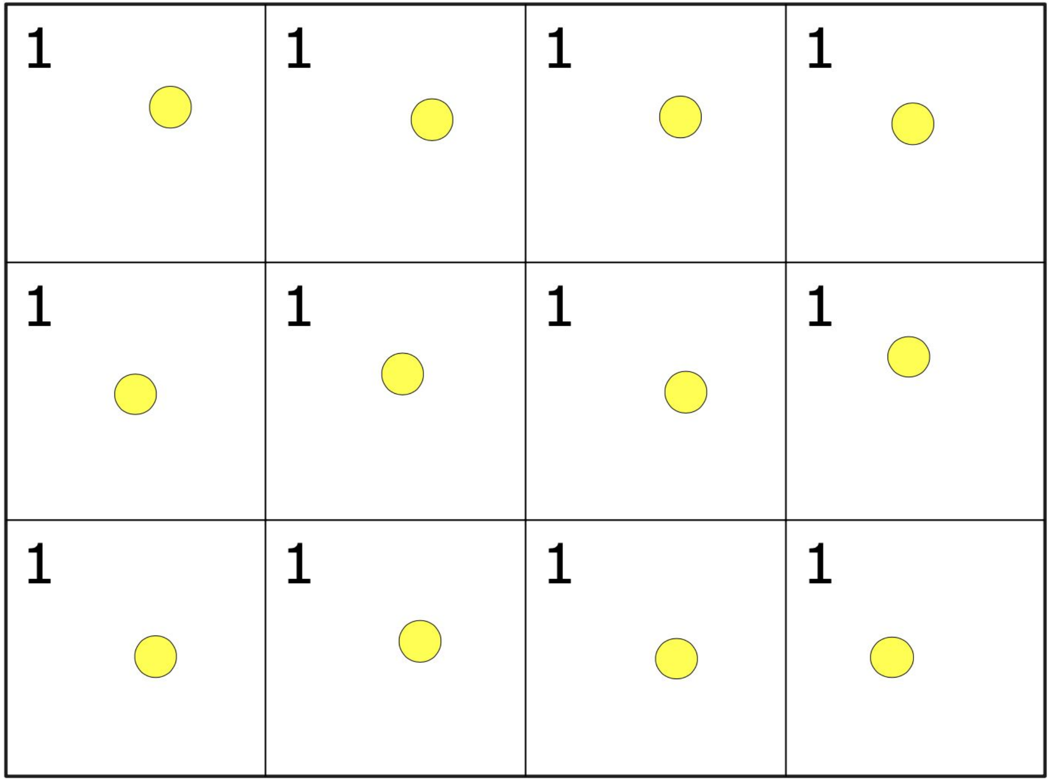
Random



Random
Moderate variation



Dispersed



Dispersed

Low variation

Quadrat Analysis

- Output
  - Variance to mean ratio (VMR)
  - p-value (based on chi-square)
  - If VMR < 1
    - Observed pattern more dispersed than random
  - If VMR > 1
    - Observed pattern more clustered than random

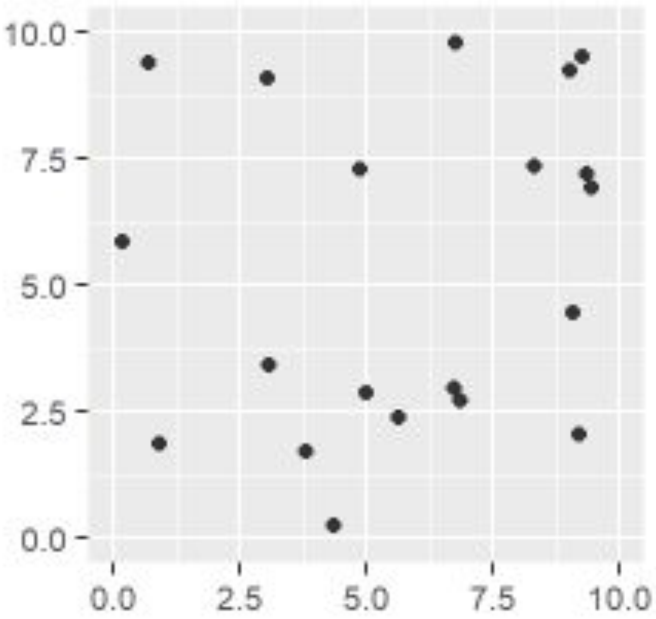
$$VMR = \frac{s^2}{\overline{X}}$$

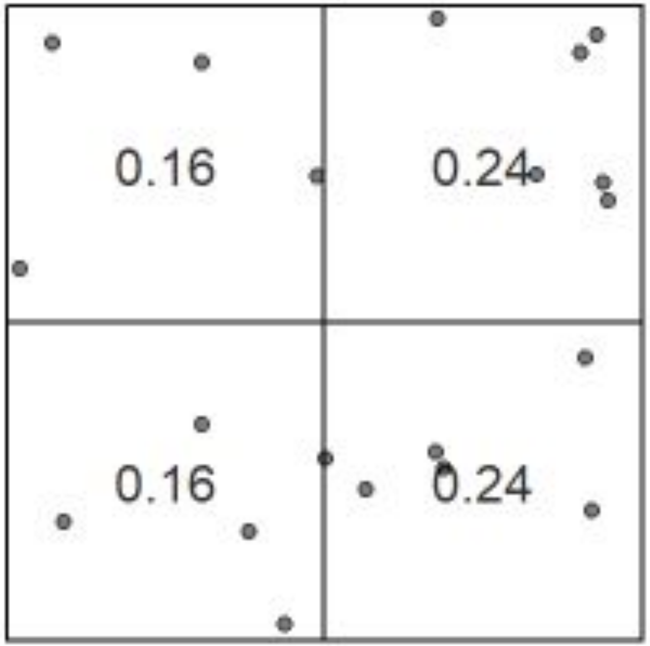
$$s^2 = \frac{\sum (X_i - \overline{X})^2}{m - 1}$$

m = the number of cells

Quadrat Analysis

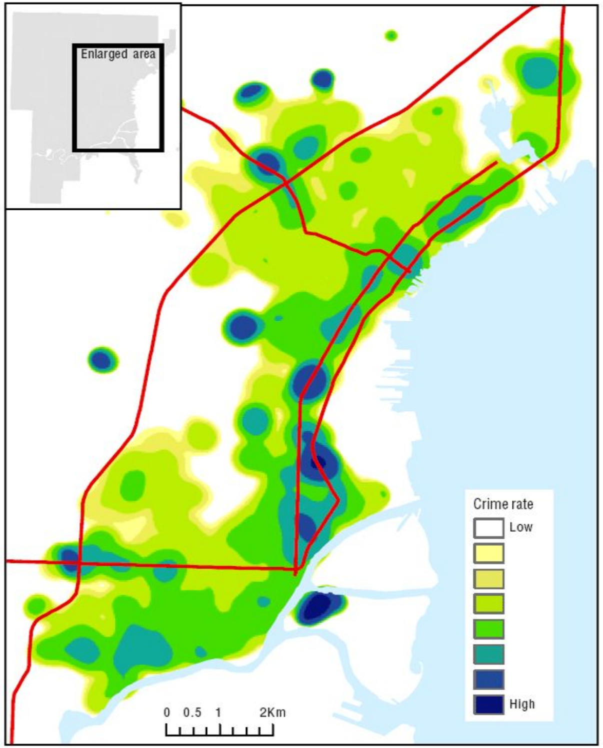
- Density based





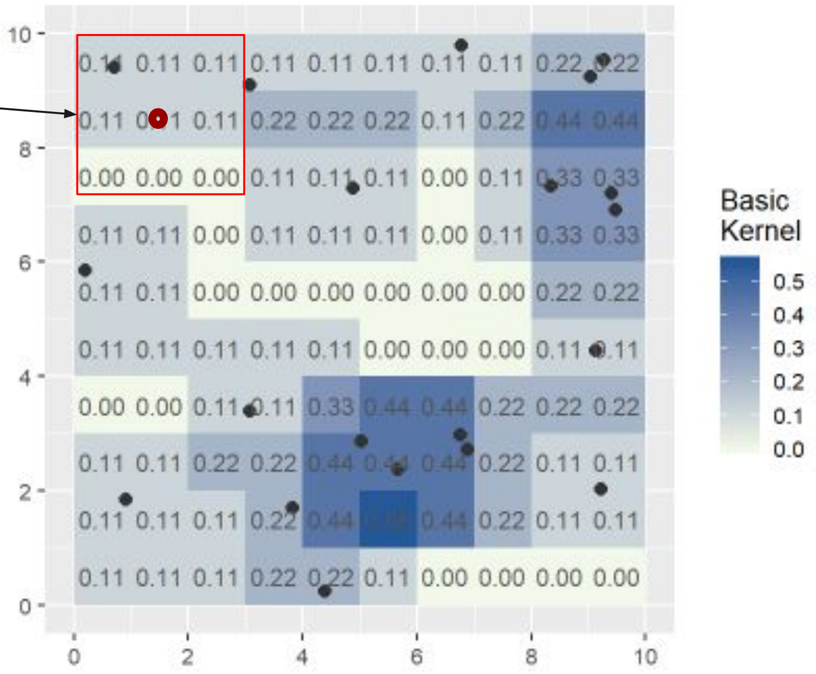
Kernal Density Analysis

- Similar to Quadrat but includes overlaps
  - Using a moving window called a kernal



Kernal Density Analysis

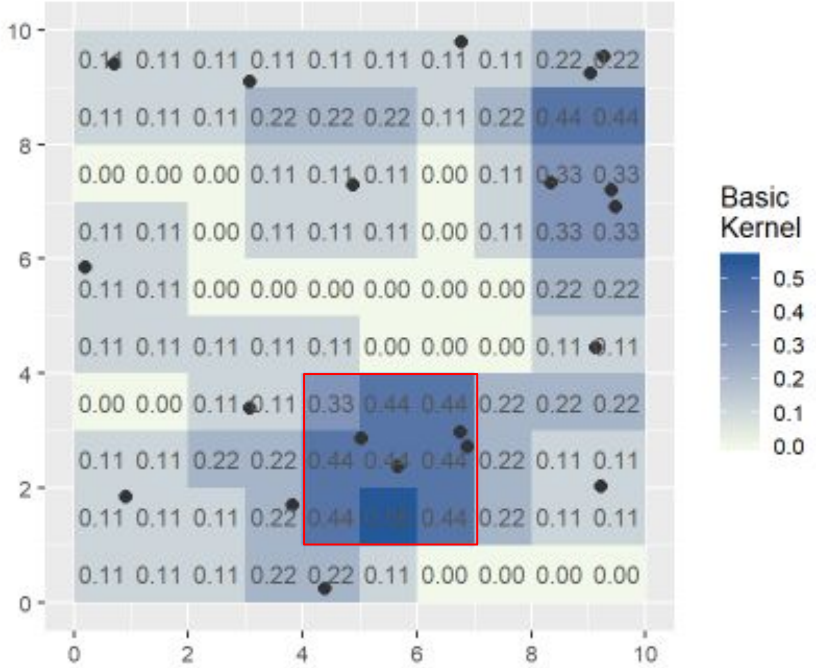
Kernal window



λ = n/a
= 1/9 = 0.11

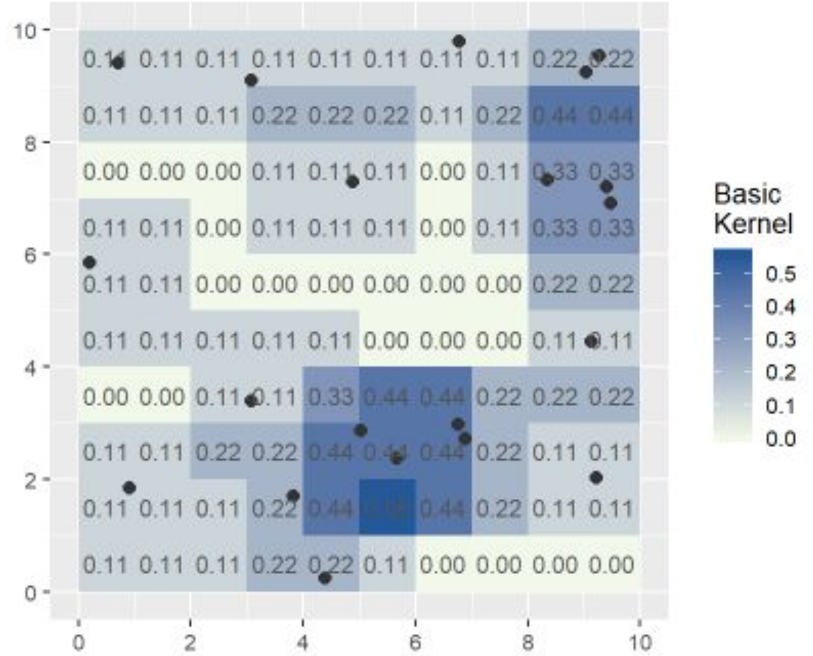
Kernal Density Analysis

λ = n/a
= ?

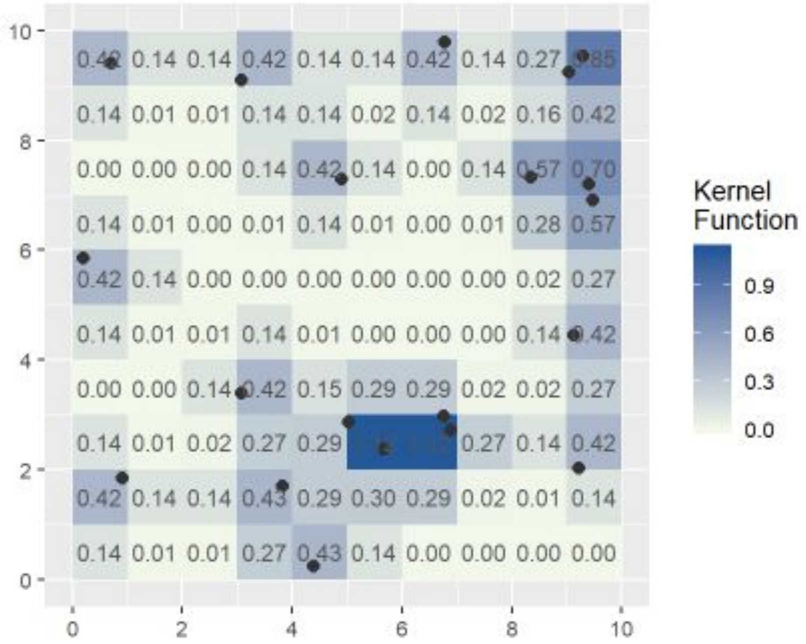


Kernal Density Analysis

Equal Weights



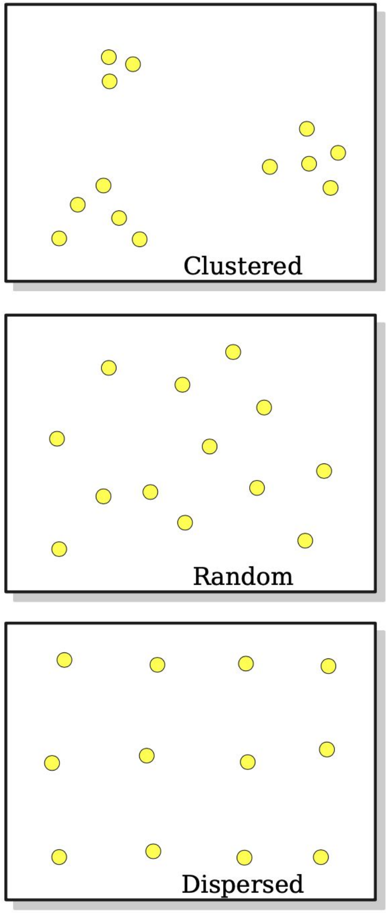
Inverse Distance Weights



Distance Based Analysis

- Average Nearest Neighbors
  - Measure distance to each point’s nearest neighbor
  - Find the mean nearest neighbor distance for the study area (NND)
    - Compare to expected NND for a random arrangements of points

Which of these 3 do you think will have the highest NND?



Nearest Neighbor Distance

- For each point, measure distance to that point's nearest neighbor
    - Find the mean nearest neighbor distance for the study area (NND)

$$\overline{NND} = \frac{\sum NND_i}{n}$$

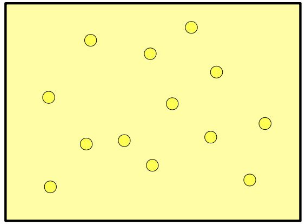
Nearest Neighbor Distance

- Perfectly clustered set of points

$$\overline{NND_C} = 0$$

How would this look like on a map?

- Randomized set of points



$$\overline{NND_R} = \frac{\sqrt{a/n}}{2} + \frac{per}{n}\left(0.0514 + \frac{0.041}{\sqrt{n}}\right)$$

$a =$ area
$per =$ perimeter of boundary

Note: Includes a “correction factor” for boundary effect

Nearest Neighbor Analysis

- Produces
  - NND
  - Z score
  - p-value

$$Z_{nnd} = \frac{\overline{NND} - \overline{NND_R}}{\sigma_{\overline{NND}}}$$

$$\overline{NND_R} = \frac{\sqrt{a/n}}{2} + \frac{per}{n}\left(0.0514 + \frac{0.041}{\sqrt{n}}\right)$$

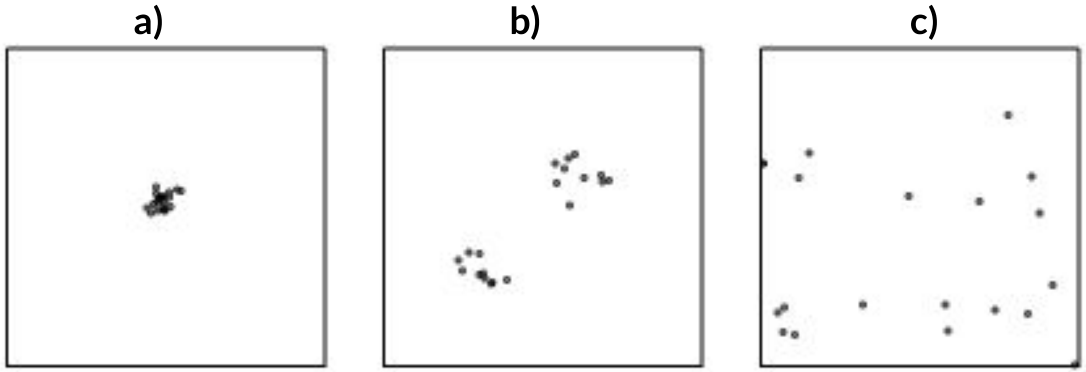
$$\sigma_{\overline{NND}} = \sqrt{\frac{0.07 * a}{n^2} + 0.037 * per * \sqrt{\frac{a}{n^5}}}$$

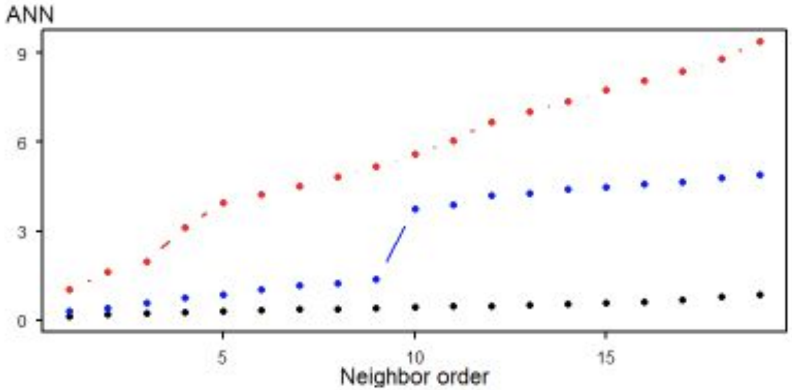
Nearest Neighbor Analysis

- Determine whether the point pattern is random or non random (clustered or dispersed)
  - If NND > NND(r) , Z score positive
    - Observed pattern more dispersed than random
  - If NND < NND(r) , Z score is negative
    - Observed pattern more clustered than random

Extending Nearest Neighbor Analysis

- To higher order neighbors





Nearest Neighbor Analysis

- Caveats
  - Sensitive to the study area
  - Assumes stationarity
    - Cannot distinguish if observed pattern is due to interaction among points or due to changes in an underlying property that changes with location