Lecture 4

1

Jianguo (Jack) Ma

Department of Geographical Sciences
University of Maryland at College Park

Outline

- Point Pattern Analysis
  - Overview
  - Geometric Measurements
  - Quadrat Count Analysis (frequency-based)
  - Density Analysis
  - Nearest Neighbor Analysis
    - Distance-based clustering analysis (vs. Near Analysis)
  - Global Spatial Autocorrelation Analysis (attribute value-based)
    - Moran's I
    - General G Statistics
- Demos

Announcements

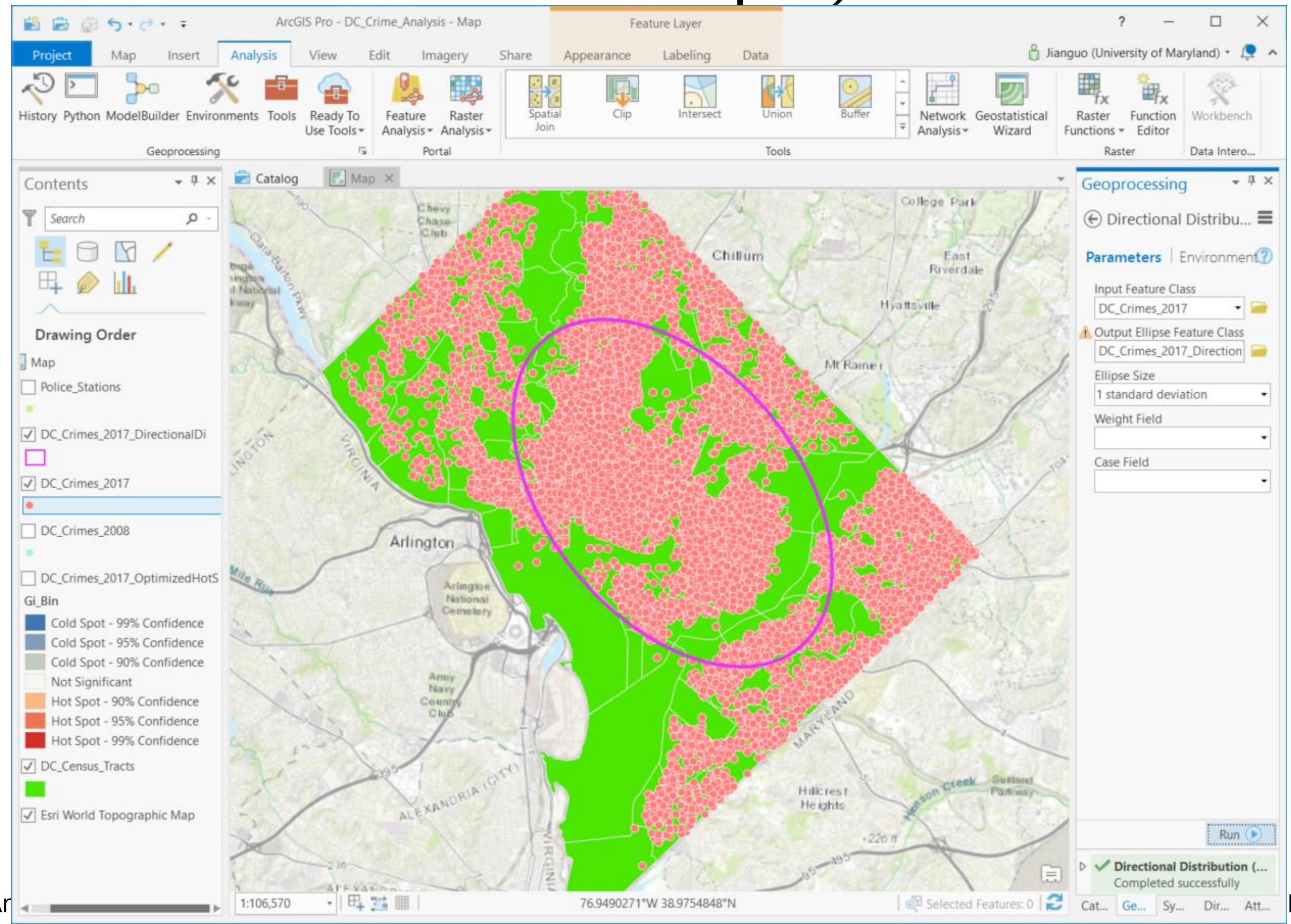
- Lab 1
  - Common questions
- Lab 2
- Reading assignment

Point Pattern Analysis

- Overview
  - Point Pattern Analysis (PPA) involves describing patterns of locations of point events and testing whether there is a significant occurrence of clustering of points in a particular area.
    - Example: crime patterns (to find out if there is any clustering or if there is any direction/orientation of incidents)
    - (Extra reading posted on ELMS.)

Point Pattern Analysis

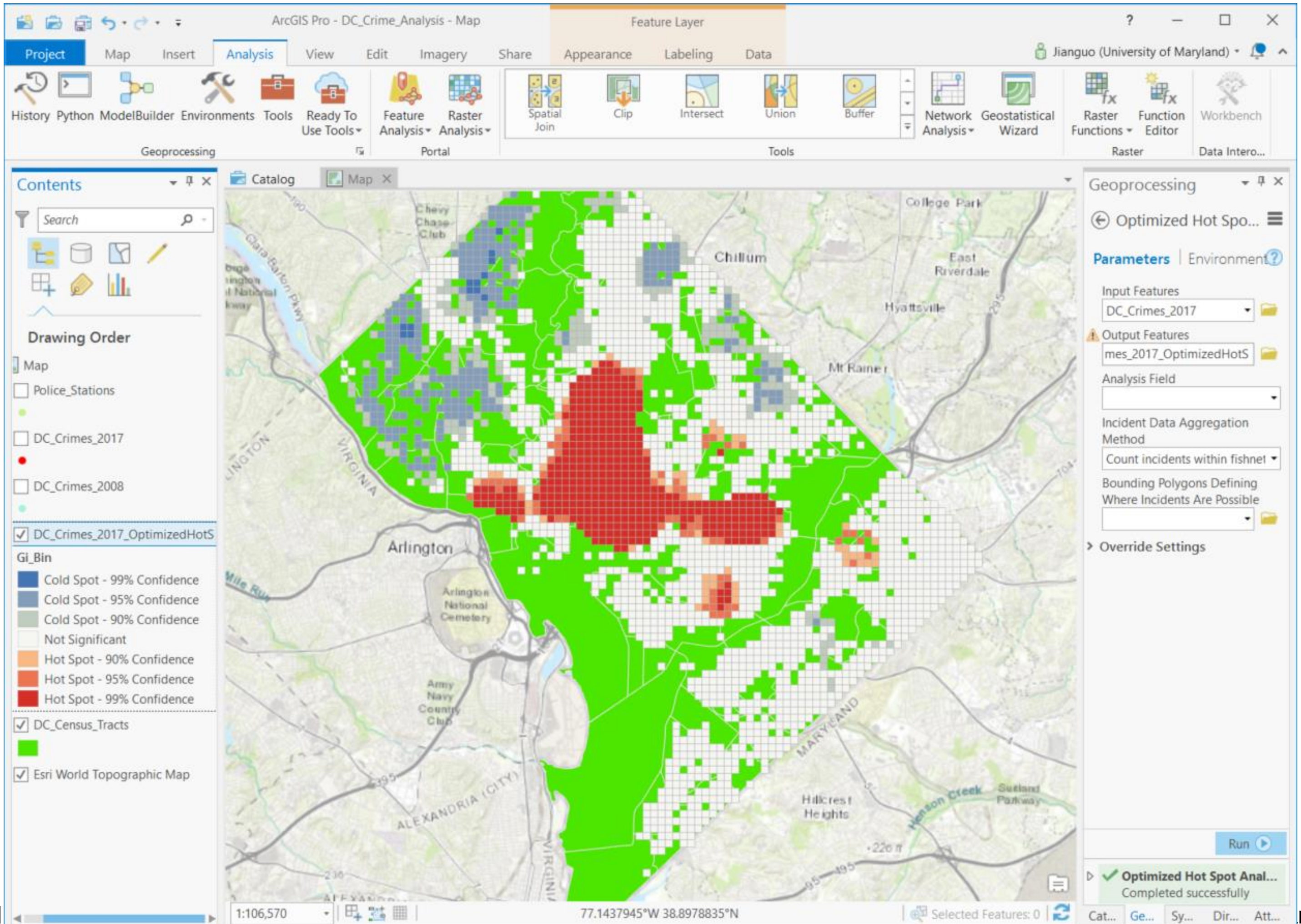
- Overview
  - Example: crime patterns (directional distribution – Standard Deviational Ellipse)



5

Point Pattern Analysis

- Overview
  - Example: crime patterns (hotspot analysis)



Point Pattern Analysis

- Overview
  - Point Pattern Analysis (PPA) is part of spatial statistics that are used to:
    - Describe and model spatial distributions, spatial patterns, spatial processes, and spatial relationships.
    - Incorporate space (area, length, proximity, orientation, and/or spatial relationships) directly into their mathematics.

Point Pattern Analysis

- Overview
  - Point Pattern Analysis (PPA) will help users investigate the point features in terms of their
    - Patterns
    - Relationships
    - Trends

- Overview
  - To conduct point pattern analysis, there are five restrictions on the data:
    - The pattern must be mapped on a plane, meaning that you will need both *x,y* coordinates.
      - Usually not appropriate to perform point pattern analysis on events scattered over a very wide geographical area due to distortions introduced by the map projection used.
    - A study area must be selected and determined objectively prior to the analysis.

Point Pattern Analysis

- Overview
  - To conduct point pattern analysis, there are five restrictions on the data: ('continued)
    - The point data should not be a selected sample, but rather the entire set of data to be analyzed.
      - Non-parametric

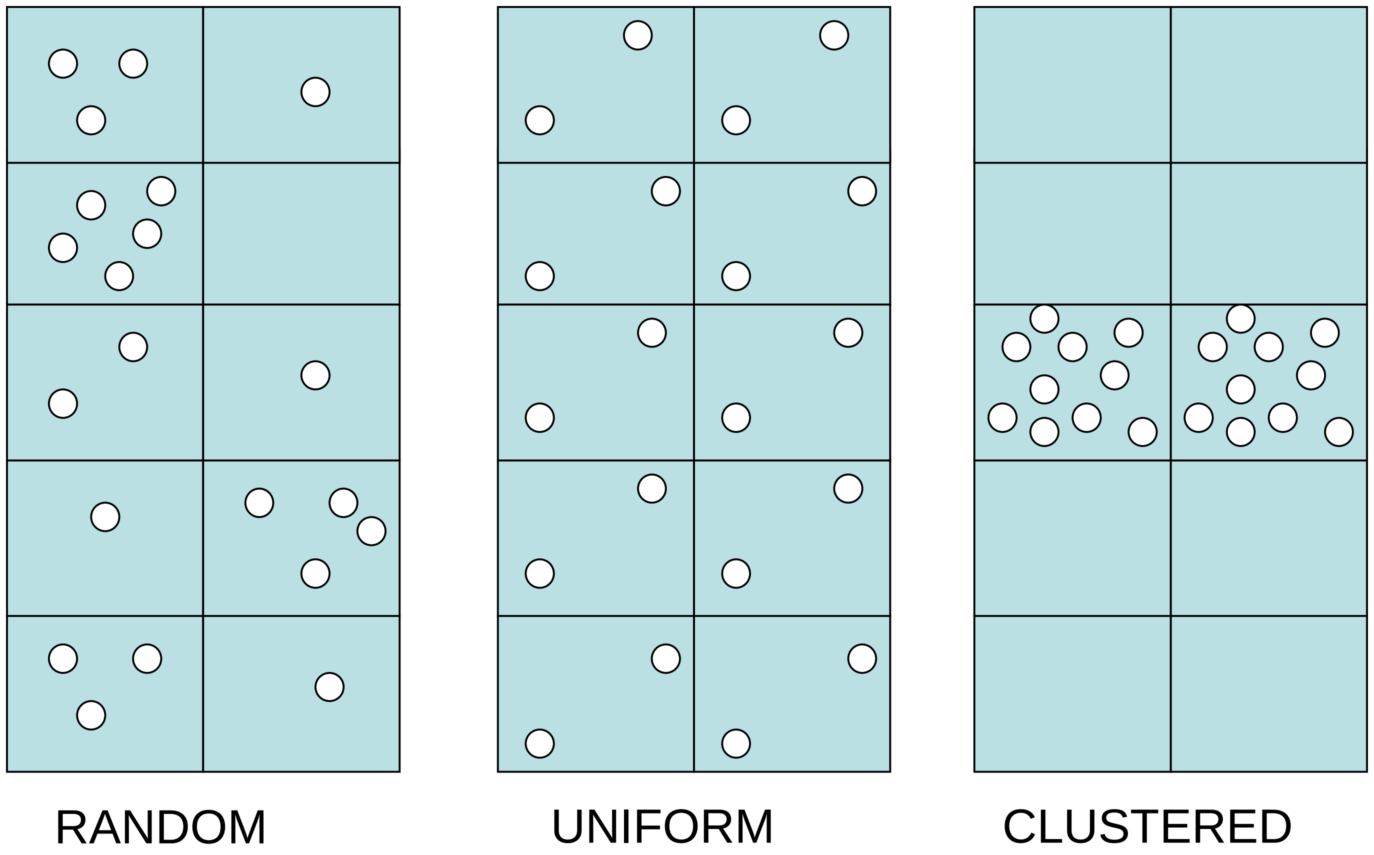
    - There should be a one-to-one correspondence between objects in the study area and events in the pattern.
      - Particularly important!
    - The points must be true incidents with real spatial coordinates.
      - For example: The points should not be the centroids of polygons.

Point Pattern Analysis

- Overview
  - The main objective of PPA is to look for specific type of pattern or clustering.
  - Typical point pattern analysis:
    - Hot spot analysis
      - Crime
      - Disease distribution

- Overview
  - Spatial patterns of points distribution:
    - Random
      - Any point is equally likely to occur at any location and the position of any point is not affected by the position of any other point. There is no apparent ordering of the distribution.
    - Uniform
      - Every point is as far from all of its neighbors as possible.
    - Clustered
      - Many points are concentrated close together, and large areas that contain very few, if any, points.

- Overview
  - Spatial patterns of points distribution:



Point Pattern Analysis

- Overview
  - In ArcGIS environment, the null hypothesis is always Complete Spatial Randomness.
  - If the result (the p-value/z score) is statistically significant, we will reject the null hypothesis. That means there is clustering.

Point Pattern Analysis

- Overview
  - You should be familiar with basic statistics.
    - Mean
    - Median
    - Standard deviation
    - Variance
    - Frequency distribution
    - Probability
    - Normal distribution
    - Confidence level
    - Hypothesis testing

Point Pattern Analysis

- Overview
  - In general, there are three types of PPA techniques:
    - Quadrat Count Analysis
    - Kernel Density Analysis Estimation
      - Also called K-Means
    - Nearest Neighbor Analysis

Point Pattern Analysis

- Overview
  - First-order spatial variation
    - An observation occur at a place due to the underlying properties of the local "environment" of the place.
      - Example: there are more crime (in terms of incidents) in places where there are more population.
    - Methods to measure first-order spatial variation:
      - Quadrat count analysis
      - kernel density analysis

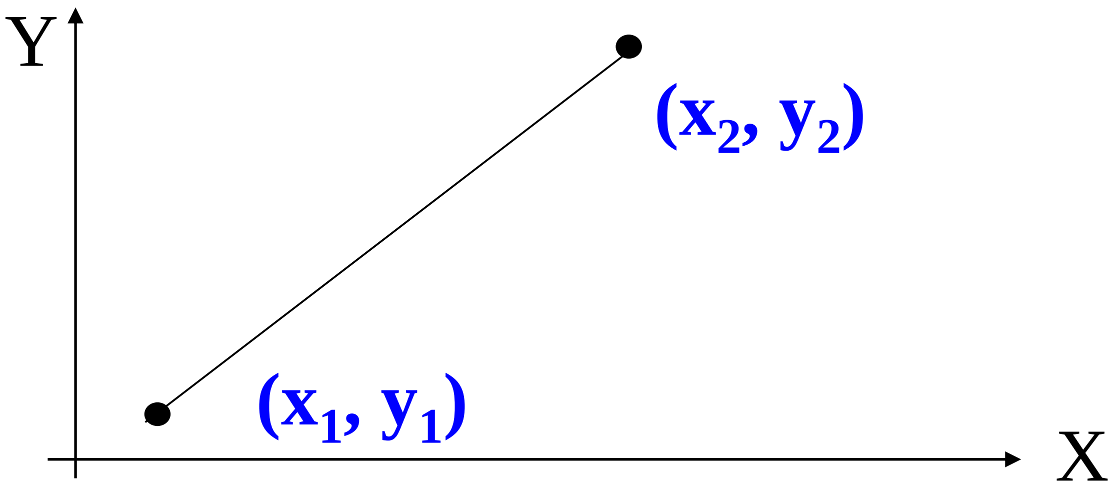
Point Pattern Analysis

- Overview
  - Second-order spatial variation
    - The existence of an observation is due to interactions with other observations.
      - Example: If there is more crime in place A, then usually there is more crime in places which are close to place A. i.e. crimes tend to be clustered.
    - Methods to measure second-order spatial variation:
      - Nearest neighbor analysis
      - K function

- Geometric Measurements
  - Distance
  - Central Tendency
  - Spatial Dispersion
  - Direction

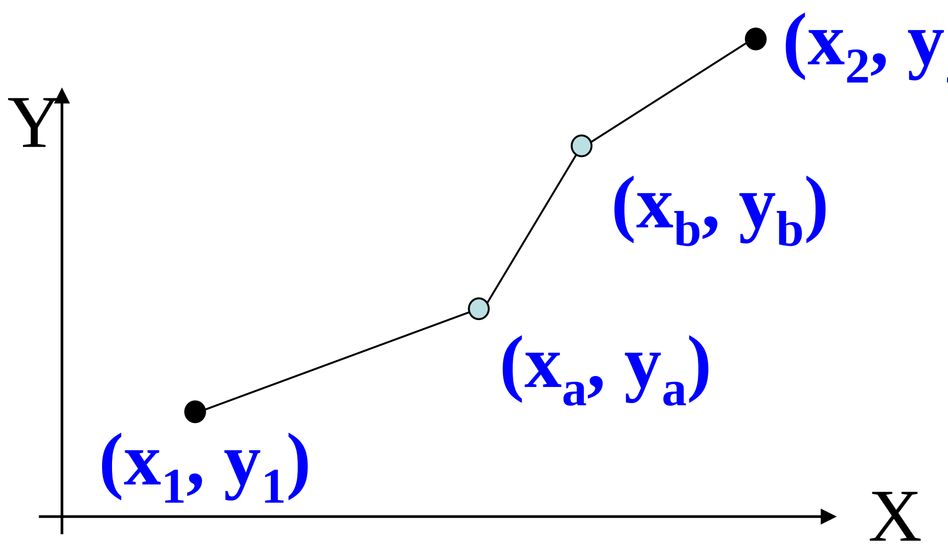
Point Pattern Analysis

- Geometric Measurements
  - Distance
    - Euclidean distance
      - The distance between two points at locations in Euclidean Cartesian space.



$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

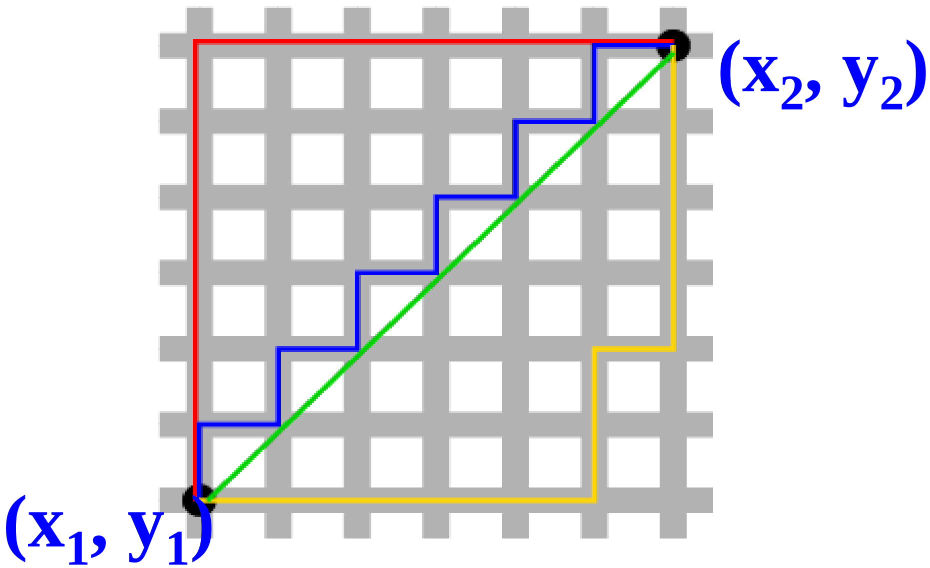
- Geometric Measurements
  - Distance
    - Distance along a path.



$$d = \sum_{i=1}^{n} d_i$$

d_i is the Euclidean distance of segment i.

- Geometric Measurements
  - Distance
    - Manhattan distance
      - The shortest distance laid out in square blocks, like in Manhattan (hence, Manhattan Distance).



$$d = |x_1 - x_2| + |y_1 - y_2|$$

Point Pattern Analysis

- Geometric Measurements
  - Central Tendency
    - Finding the center of a set of features can be very useful.
    - Examples:
      - Finding centers at different time periods can help identify the movement of certain phenomenon (e.g. epidemic disease)
      - Finding center can help identify the best location (e.g. locate the site of a new police station based on crime data)

Point Pattern Analysis

- Geometric Measurements
  - Central Tendency
    - Three kinds of center:
      - Median Center
      - Mean Center
      - Central Feature

- Geometric Measurements
  - Central Tendency
    - Median Center
      - Also called the center of minimum distance.
      - The median center (*Xe*, *Ye*) is a location that has the shortest total distance to all points.
      - It is usually a new location (not real) instead of one of the existing points (locations).

$$\min \sum_{i=1}^{N} \sqrt{(x_i - x_e)^2 + (y_i - y_e)^2}$$

- Geometric Measurements
  - Central Tendency
    - Median Center
      - The Median Center is always gravitated toward areas with most features.
      - It can be used to:
        - » Find the most accessible location. (e.g. find a new site that minimizes the sum of transport costs.)
        - » Track changes of a point distribution.
        - » Compare point distributions.
      - There is a new tool to find the Median Center in ArcGIS 10.x (didn't exist in previous versions).
        - » Spatial Statistics Tools → Measuring Geographic Distribution → Median Center.

Point Pattern Analysis

- Geometric Measurements
  - Central Tendency
    - Mean Center
      - The average x-coordinate and average y-coordinate for all points in the study area.

$$\overline{X} = \sum_{i=1}^{N} \frac{x_i}{N} \qquad \overline{Y} = \sum_{i=1}^{N} \frac{y_i}{N}$$

Point Pattern Analysis

- Geometric Measurements
  - Central Tendency
    - Mean Center
      - It can be used to:
        - » Indicate the average location.
        - » Track changes of a point distribution, for example, animal migration over time.
        - » Compare point distributions.

Point Pattern Analysis

- Geometric Measurements
  - Central Tendency
    - Weighted Centers
      - A weight indicates how important a geospatial object is.
      - The centers are biased toward locations with high weights.

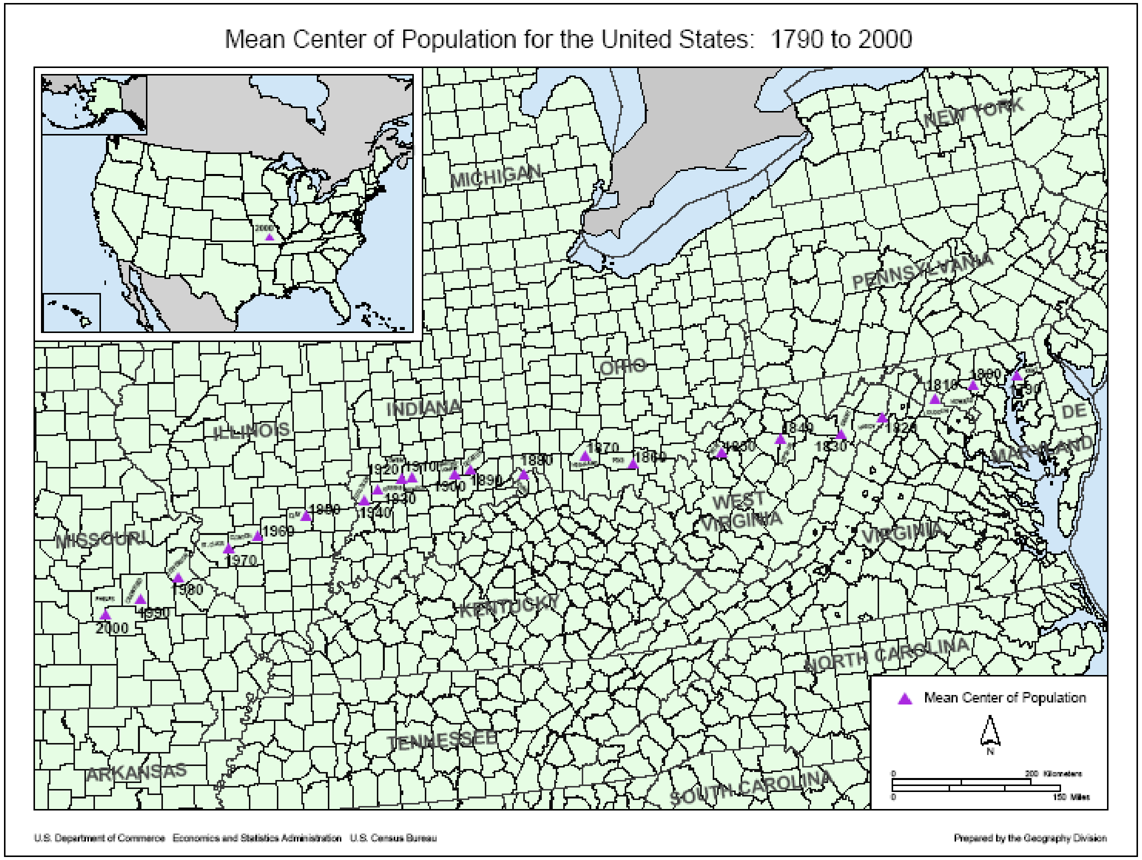
City	X coordinate (km)	Y coordinate (km)	Population (×10000)
A	1.4	2.8	3
B	2.1	0.8	20
C	0.3	1.9	13

- Geometric Measurements
  - Central Tendency
    - Weighted Median Center

If the weight of a point (x_i, y_i) is w_i, the weighted mean center is (X_{we}, Y_{we}) that minimizes:

$$\sum_{i=1}^{N} w_i \sqrt{(x_i - X_{we})^2 + (y_i - Y_{we})^2}$$

- Geometric Measurements
  - Central Tendency
    - Weighted Mean Center



- Geometric Measurements
  - Central Tendency
    - Mean Center
      - Example: (weighted) mean center of U.S. population (1790-2000)
      - Animation video here - https://archive.org/details/SVS-3163
      - ***How was it done?***

$$\bar{\phi} = \frac{\sum w_i \phi_i}{\sum w_i} \qquad \bar{\lambda} = \frac{\sum w_i \lambda_i Cos(\phi_i)}{\sum w_i Cos(\phi_i)}$$

Where *Φ, λ, w are the latitude, longitude, and population*.

- Geometric Measurements
  - Central Tendency
    - Central Feature
      - It is the point (one actual feature) at which the total distance to all other points is the shortest.

$$\min\left(\sum_{i=1}^{N}\sqrt{(x_i - x_C)^2 + (y_i - y_C)^2}\right)$$

Point Pattern Analysis

- Geometric Measurements
  - Central Tendency
    - Central Feature
      - It can be used to:
        - » Find the most accessible location.
          - » Example: find a location that minimizes the total transportation cost.
        - » Track changes of a point distribution.
        - » Compare point distributions.

- Geometric Measurements
  - Central Tendency
    - Summary

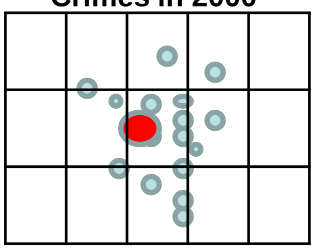
Center	What it represents	What it's good for
Mean	The average x-coordinate and average y-coordinate for all features in the study area.	Tracking changes or comparing distributions.
Median	The x,y coordinate having the shortest distance to all features in the study area.	Finding the most accessible location.
Central Feature	The feature having the shortest total distance to all other features in the study area.	Finding the most accessible feature.

Andy Mitchell, 2005

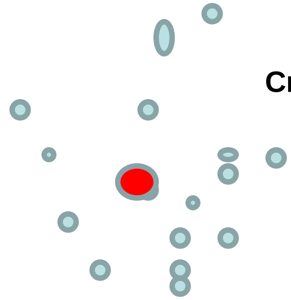
Point Pattern Analysis

- Geometric Measurements
  - Spatial Dispersion
    - It describes how a set of points scatter around a center.
      - Can be used to assess the concentration of points.

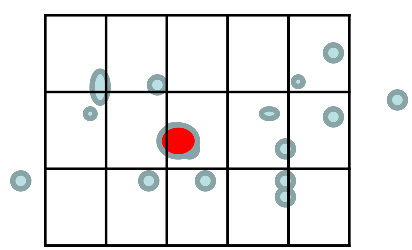
Crimes in 2000



Crimes in 2020



Crimes in 1990



- Geometric Measurements
  - Spatial Dispersion
    - Standard deviation of the *x* and *y* coordinates.

$$S_x = \sqrt{\sum_{i=1}^{N} \frac{(x_i - \overline{X})^2}{N}} \qquad S_y = \sqrt{\sum_{i=1}^{N} \frac{(y_i - \overline{Y})^2}{N}}$$

$\overline{X}, \overline{Y}$ is the mean center of the points

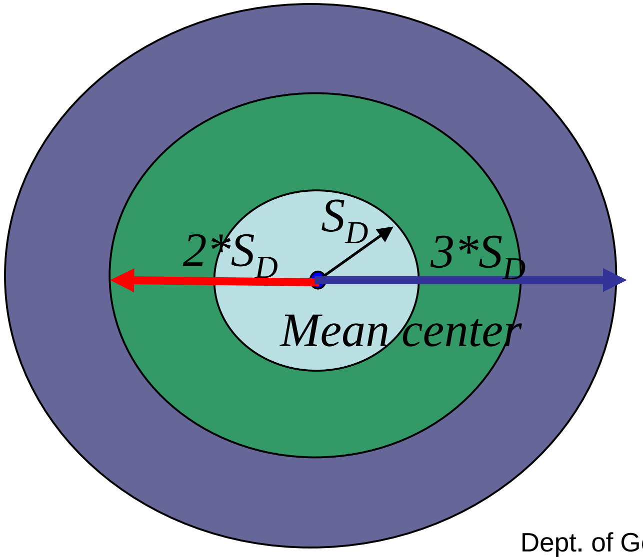
- Geometric Measurements
  - Spatial Dispersion
    - Standard Distance
      - It indicates the average distance to the mean center.
      - It provides a single value representing the dispersion of features around the center.
      - The compactness can be represented on a map by drawing a circle with the radius equal to the value.

$$S_D = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{X})^2 + \sum_{i=1}^{N}(y_i - \overline{Y})^2}{N}}$$

- Geometric Measurements
  - Spatial Dispersion
    - Standard Distance
      - One S_D covers more than 63% of the points.
      - Two S_D covers more than 98% of the points.
      - Three S_D covers more than 99% of the points.
      - Similar to the 68-95-99.7 rule for the normal distribution.

https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/standard-distance.htm

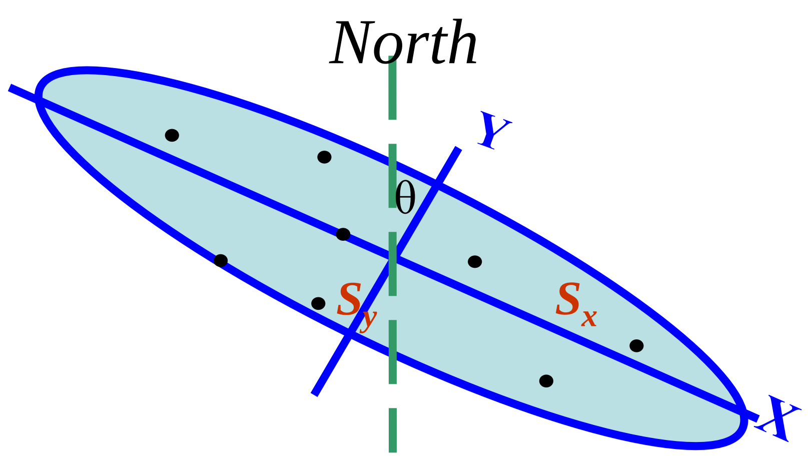
39



Point Pattern Analysis

- Geometric Measurements
  - Direction
    - Standard Distance measures the dispersion of the incidents around the mean center, but it does not capture any directional bias or the shape of the distribution. This leads to Standard Deviational Ellipse.
    - Standard Deviational Ellipse
      - It measures the trend for a set of points or areas by calculating the standard distance separately in the x and y directions.
      - The ellipse tells if the distribution of features is elongated and hence, has a particular orientation.

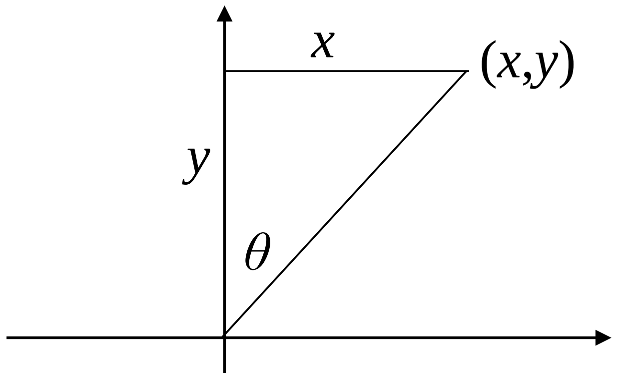
- Geometric Measurements
  - Direction
    - Standard Deviational Ellipse
      - Three components describe a standard deviational ellipse:
        - » The angle of clockwise rotation from north
        - » The standard deviation along *x*-axis
        - » The standard deviation along *y*-axis



Point Pattern Analysis

- Geometric Measurements
  - Direction
    - Standard Deviational Ellipse
      - Calculation procedures:
        1. Calculate the mean center
        2. For each point (x_i, y_i) transform the coordinate.
        3. Calculate the angle of rotation
        4. Calculate the standard deviations along the x and y axes
        5. Describe the flatness of the ellipse: Eccentricity

- Geometric Measurements
  - Direction
    - Standard Deviational Ellipse
      - Calculation procedures:
        - » Calculate the angle of rotation



$$\tan\theta = \frac{x}{y}$$

θ takes a value between [-90, 90]
If θ is positive, clockwise rotate θ from north
If θ is negative, clockwise rotation (360+θ) from north

- Geometric Measurements
  - Direction
    - Standard Deviational Ellipse
      - Calculation procedures:
        - » Calculate the standard deviations along the x and y axes

$$S_x = \sqrt{\frac{\sum_{i=1}^{N}(x_i^{'}\cos\theta - y_i^{'}\sin\theta)^2}{N}}$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{N}(x_i^{'}\sin\theta - y_i^{'}\cos\theta)^2}{N}}$$

- Geometric Measurements
  - Direction
    - Standard Deviational Ellipse
      - Calculation procedures:
        - » Describe the flatness of the ellipse: Eccentricity

$$e = \frac{\sqrt{a^2 - b^2}}{a}$$

Point Pattern Analysis

- Quadrat Count Analysis
  - This method involves simply recording and counting the number of events that occur in each quadrat.
  - It measures the density of points.
  - It can be used to compare the quadrat counts with a predefined statistic distribution in order to determine how the points are distributed statistically.

Point Pattern Analysis

- Quadrat Count Analysis
  - How does it work?
    - General procedures:
      - Divide the study area into a set of cells (quadrats).
      - Count the number of points in each quadrat.
      - Compare the counts with a known statistical distribution.

- Quadrat Count Analysis
  - How does it work?
    - Specific procedures:
      - A uniform grid network is drawn over a map of the distribution of interest.
      - The frequency count, the number of points occurring within each quadrat is recorded first.
      - These data are then used to compute a measure called the variance.
      - The variance compares the number of points in each grid cell with the average number of points over all of the cells.
      - The variance of the distribution is compared to the characteristics of a random distribution.

Point Pattern Analysis

- Quadrat Count Analysis
  - Statistics

$$N = number\_of\_quadrats$$

$$Variance = \frac{\sum x^2 - [(\sum x)^2 / N]}{N-1}$$

$$Variance - mean - ratio = \frac{\text{variance}}{\text{mean}}$$

49

Point Pattern Analysis

- Quadrat Count Analysis
  - How does it work?
    - A random distribution would indicate that that the variance and mean are the same. Therefore, we would expect a variance-mean ratio around 1.
    - Values other than 1 would indicate a non-random distribution.
      - For a uniform distribution, the variance-mean ratio is zero.
      - For a clustered distribution, the variance-mean ratio is much larger than 1.

Point Pattern Analysis

- Quadrat Count Analysis
  - Example:

3	1
5	0
2	1
1	3
3	1

Quadrat #	Number of Points Per Quadrat	x^2
1	3	9
2	1	1
3	5	25
4	0	0
5	2	4
6	1	1
7	1	1
8	3	9
9	3	9
10	1	1
	20	60
Variance	2.222	
Mean	2.000	
Var/Mean	1.111	

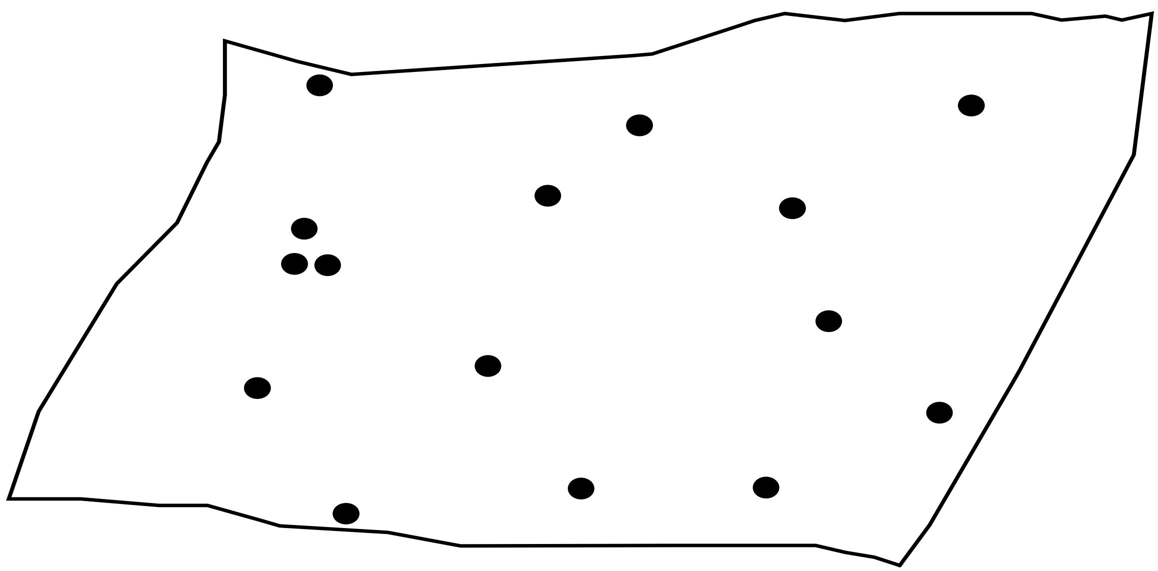
2	2
2	2
2	2
2	2
2	2

Quadrat #	Number of Points Per Quadrat	x^2
1	2	4
2	2	4
3	2	4
4	2	4
5	2	4
6	2	4
7	2	4
8	2	4
9	2	4
10	2	4
	20	40
Variance	0.000	
Mean	2.000	
Var/Mean	0.000	

0	0
0	0
10	10
0	0
0	0

Quadrat #	Number of Points Per Quadrat	x^2
1	0	0
2	0	0
3	0	0
4	0	0
5	10	100
6	10	100
7	0	0
8	0	0
9	0	0
10	0	0
	20	200
Variance	17.778	
Mean	2.000	
Var/Mean	8.889	

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.



Map extent (unit: mile): X_{min} = 19, X_{max}=37, Y_{min}=71, Y_{max} =86

Point Pattern Analysis

- Quadrat Count Analysis
  - Example:
    - Step 1: Define the quadrat
      - Parameters
        - » shape: squares, hexagons, triangle. Square is one of the most common types of quadrat
        - » size: not too large, not too small. For squares, the size can be determined as:

$$I = \sqrt{2 \bullet \frac{A}{N}} = \sqrt{2 \bullet \frac{E_x \bullet E_y}{N}}$$

I: the length of a side of a square
A: the area of study area
N: the number of points
E_x: the spatial extent on the X axis of study area
E_y: the spatial extent on the Y axis of study area
A can be $E_x * E_y$

Point Pattern Analysis

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.

E_x=18 miles, E_y=15 miles, N=15:

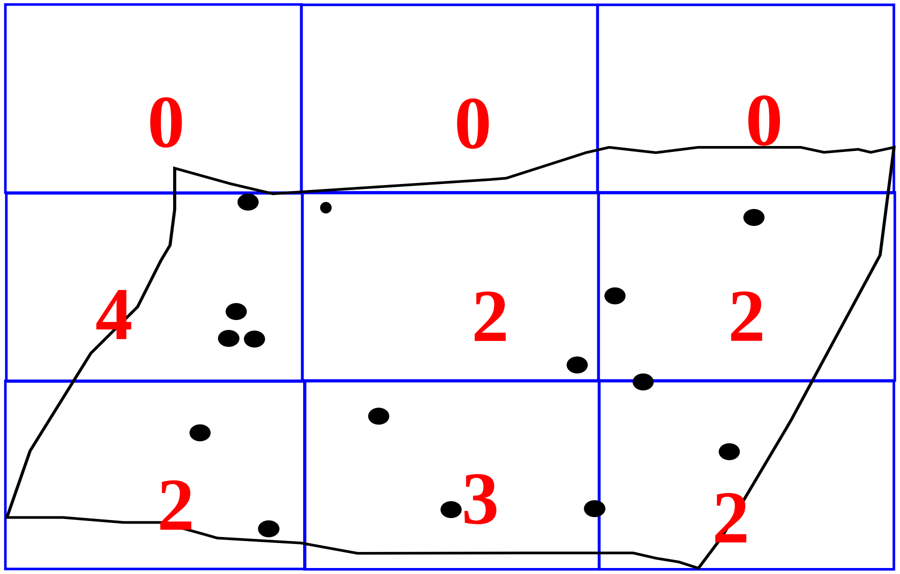
$$I = \sqrt{2 \bullet \frac{18 \bullet 15}{15}} = 6$$

6 miles

6 miles

Point Pattern Analysis

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.
    - Step 2: overlay the quadrats on the map and count the points (events) in each quadrat



Point Pattern Analysis

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.

Number of events in quadrat	Count	Proportion
0	3	0.33
1	0	0.00
2	4	0.44
3	1	0.11
4	1	0.11

Point Pattern Analysis

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.
    - Step 3: Determine the pattern
      - Compare the observed pattern of a point distribution with a predefined pattern.
      - If their difference is not statistically significant, the point distribution follows the pattern described by the predefined pattern.

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.
    - A Poisson distribution arises under the following conditions:

    1) The probability of observing a single event over a small interval (time or space) is approximately proportional to the size of that interval.
    2) The probability of two events occurring in a small enough interval is negligible.
    3) The probability of an event within a certain interval does not change over different intervals (homogenous).
    4) The probability of an event in one interval is independent of the probability of an event in any other non-overlapping interval.

Point Pattern Analysis

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.
    - A Poisson distribution arises under the following conditions:

59

$$P(x) = \frac{\lambda^x}{e^{\lambda}(x!)}$$

x: frequency of occurrence
λ: mean frequency of occurrence
e: Euler's constant, 2.72

Point Pattern Analysis

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.
    - A Poisson distribution arises under the following conditions:
      - For quadrat count analysis, $P(x)$ is the probability that x points occur in a quadrat.

λ is the average number of points per quadrat:

$$\lambda = \frac{N}{n}$$

N: the number of points
n: the number of quadrats

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.
    - A Poisson distribution arises under the following conditions:

$$\lambda = \frac{15}{9} = 1.67$$

$$P(x) = \frac{1.67^x}{2.72^{1.67}(x!)}$$

$$P(0) = \frac{1.67^0}{2.72^{1.67}(0!)} = 0.19$$

$$P(3) = \frac{1.67^3}{2.72^{1.67}(3!)} = 0.14$$

$$P(1) = \frac{1.67^1}{2.72^{1.67}(1!)} = 0.31$$

$$P(4) = \frac{1.67^4}{2.72^{1.67}(4!)} = 0.06$$

$$P(2) = \frac{1.67^2}{2.72^{1.67}(2!)} = 0.26$$

$$P(5) = \frac{1.67^5}{2.72^{1.67}(5!)} = 0.02$$

Point Pattern Analysis

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.
    - Use Kolmogorov-Smirnov test to compare the two distributions

Quadrat	Count	Proportion	Observed Cumulative Proportion	Expected Proportion	Expected Cum. Proportion	**Absolute Difference**
4	1	0.11	0.11	$P(4)$=0.06	0.06	**0.05**
1	1	0.11	0.22	$P(1)$=0.31	0.37	**0.15**
2	2	0.22	0.44	$P(2)$=0.26	0.63	**0.19**
3	2	0.22	0.66	$P(3)$=0.14	0.77	**0.11**
0	3	0.33	0.99	$P(0)$=0.19	0.96	**0.03**

- Quadrat Count Analysis
  - Example: 15 car accidents occurred in a small village during 2005.
    - Use Kolmogorov-Smirnov test to compare the two distributions
      - The maximum difference between observed cumulative proportions and expected cumulative proportions: D=0.19
      - From the KS statistics table, p-value=0.338 > 0.19 (95% confidence interval), i.e., there is no evidence against H_0: the distribution is random.
        - » K-S Table - http://www.real-statistics.com/statistics-tables/kolmogorov-smirnov-table/
      - Conclusion: the point distribution follows a Poisson distribution. The pattern is random.

Point Pattern Analysis

- Quadrat Count Analysis
  - Limitations:
    - Quadrat size and orientation
      - If the quadrats are too small, they may contain only a couple of points. If they are too large, they may contain too many points
    - Actually a measure of dispersion, and not really pattern, because it is based primarily on the density of points, and not their arrangement in relation to one another.
    - Results in a single measure for the entire distribution and thus, variations within the region are not recognized.

Point Pattern Analysis

- Density Analysis
  - The density analysis tools calculate a measured quantity of an input point layer throughout a landscape to produce a continuous surface.
    - Very much like moving a magnifier glass over the extent.
    - Useful to show where the features (points or polygons) are concentrated.

- Density Analysis
  - The density analysis tools calculate a measured quantity of an input point layer throughout a landscape to produce a continuous surface.
    - The values of the output cells always have decimals.
    - When added together, the population values of the cells equal to the sum of the population of the original point layer.
      - We can test to verify this during the demos.
    - **Discussion:**
      - How does this density surface compare to the spatial interpolation surface?
        - » Example: points representing values of chemical concentration or precipitation?

- Density Analysis
  - Key parameters:
    - Cell Size
      - Determines how fine or coarse the patterns will appear.
    - Search Radius
      - A large search radius will take longer time to process and give the surface a more generalized appearance. Small variations in the data may be missed because too many features fall within each search radius.
      - A small search radius will reflect more local variations with fewer features having an effect on each pixel.

- Density Analysis
  - Key parameters:
    - Cell Size
    - Search Radius
      - Finding the appropriate radius can be difficult.
        - » Empirical; scientific basis; expert knowledge; project specific (constraints on time or cost)
        - » Example: anaerobic digester systems on dairy farms; a deciding factor is the transportation cost which implies an “economic searching distance”.
      - Spatial distribution of the points also affect the search radius.
        - » **Discussion:** evenly distributed points across the extent comparing to unevenly distributed points.

Point Pattern Analysis

- Density Analysis
  - Key parameters:
    - Population Field
      - The interpretation of the resulting surface will depend on whether a Population Field is used.
      - If no population field is used, then the unit of cell values is a count or frequency of the points.

Point Pattern Analysis

- Density Analysis
  - Point Density
    - Spatial Analyst toolset
    - This tool calculates magnitude-per-unit area from point features that fall within a neighborhood around each cell.
  - Kernel Density
    - This tool calculates a magnitude-per-unit area from point or polyline features using a kernel function to fit a smoothly tapered surface to each point or polyline.

Point Pattern Analysis

- Density Analysis
  - Point Density
    - Only the points that fall within the neighborhood are considered when calculating the density. If no points fall within the neighborhood at a particular cell, that cell is assigned NoData.

Point Pattern Analysis

- Density Analysis
  - Kernel Density
    - This method counts the incidents in an area (a kernel), centered at the location where the estimate is made.
      - This analysis is a partitioning technique, meaning that incidents are partitioned into a number of different clusters.
      - It spreads the known quantity of the population for each point out from the point location. The resulting surfaces surrounding each point in kernel density are based on a quadratic formula with the highest value at the center of the surface (the point location) and tapering to zero at the search radius distance. The density at each output raster cell is calculated by adding the values of all the kernel surfaces where they overlay the raster cell center.

Point Pattern Analysis

- Density Analysis
  - Kernel Density
    - This method is very good for analyzing the point patterns to discover the Hot Spots.
    - Also, this method provides us with a useful link to geographical data because it is able to transform our data into a density surface.

Point Pattern Analysis

- Density Analysis
  - Kernel Density
    - General procedures:
      - Establish a region called a kernel
      - Count the number of events in the kernel
      - Calculate the density

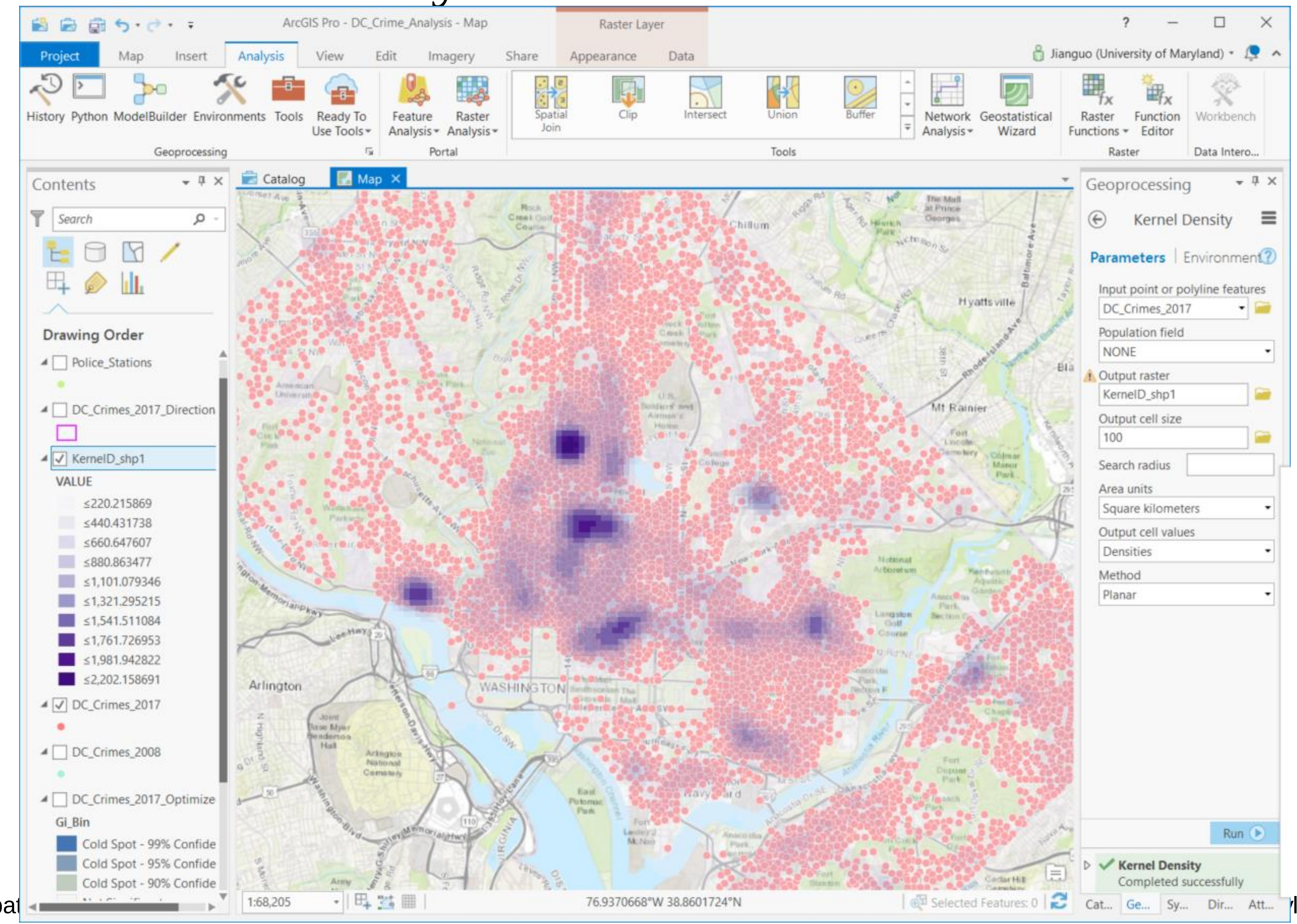
Point Pattern Analysis

- Density Analysis
  - Kernel Density
    - Using a circle as the kernel shape

$$D_i = \frac{\text{\# of points in kernel i}}{\text{kernel size} = \pi r^2}$$

Point Pattern Analysis

- Density Analysis
  - Kernel Density



Point Pattern Analysis

- Near Analysis
  - Actually a distance analysis
- Nearest Neighbor Analysis
  - Completely different from Near Analysis
  - Distance-based clustering analysis
- Global Spatial Autocorrelation Analysis
  - Moran's I
  - General G Statistics

Point Pattern Analysis

- Near Analysis
  - This tool calculates the distance and additional proximity information between the inputs features and the closest features in another layer or feature class.
    - The Input and Near Features are points in most cases, but can be polylines or polygons.
    - The Near Features can include one or more feature classes of different feature types.
    - The Input and Near Features can be the same dataset.

- Near Analysis
  - Application example:
    - Finding the distances between car accident sites and the closest police stations.
  - **Discussion**:
    - What if there are more than two nearest features for a given input feature?
    - What if there are duplicates of input or near features?
    - What if a specific input feature is coincident with a near feature?

Point Pattern Analysis

- Nearest Neighbor Analysis
  - Compares the distances between nearest points and distances that would be expected on the basis of chance
  - Ratio of two statistics
    - Nearest neighbor distance
    - Expected nearest neighbor distance based on a random distribution

Point Pattern Analysis

- Nearest Neighbor Analysis

$$d(NN) = \sum_{i=1}^{N} \left[\frac{Min(d_{ij})}{N} \right]$$

81

$Min(d_{ij})$ is the distance between each point and its nearest neighbor, and N is the number of points in the distribution

Point Pattern Analysis

- Nearest Neighbor Analysis
  - An iterative process where the nearest neighbor for each point is determined, and the total sum is divided by the number of points.

- Nearest Neighbor Analysis
  - The expected nearest neighbor distance, based on a completely random distribution is defined by

$$d(ran) = 0.5\sqrt{\frac{A}{N}}$$

Point Pattern Analysis

- Nearest Neighbor Analysis
  - The nearest neighbor index (ratio of observed nearest neighbor distance to the mean random distance) is

84

$$NNI = \frac{d(NN)}{d(ran)}$$

Point Pattern Analysis

- Nearest Neighbor Analysis
  - Clark and Evans proposed a Z-test to indicate significance

$$Z = \frac{d(NN) - d(ran)}{SE_{d(ran)}}$$

$$SE_{d(ran)} = \sqrt{\frac{(4-\pi)A}{4\pi N^2}} = \frac{0.26136}{\sqrt{\frac{N^2}{A}}}$$

Point Pattern Analysis

- Nearest Neighbor Analysis
  - NNI is only a measure of first-order randomness. It compares the average distance for the nearest neighbor to an expected random distance.
  - Sometimes we may want to evaluate the second, third, or K-*th* neighbor.

Point Pattern Analysis

- Nearest Neighbor Analysis
  - Unlike quadrat analysis uses distances between points as its basis.
  - The mean of the distance observed between each point and its nearest neighbor is compared with the expected mean distance that would occur if the distribution were random

Point Pattern Analysis

- Nearest Neighbor Analysis

$$\bar{r} = \frac{\sum r}{n}$$

$$d = \frac{n}{area}$$

$$R = \frac{r}{\bar{r}(e)}$$

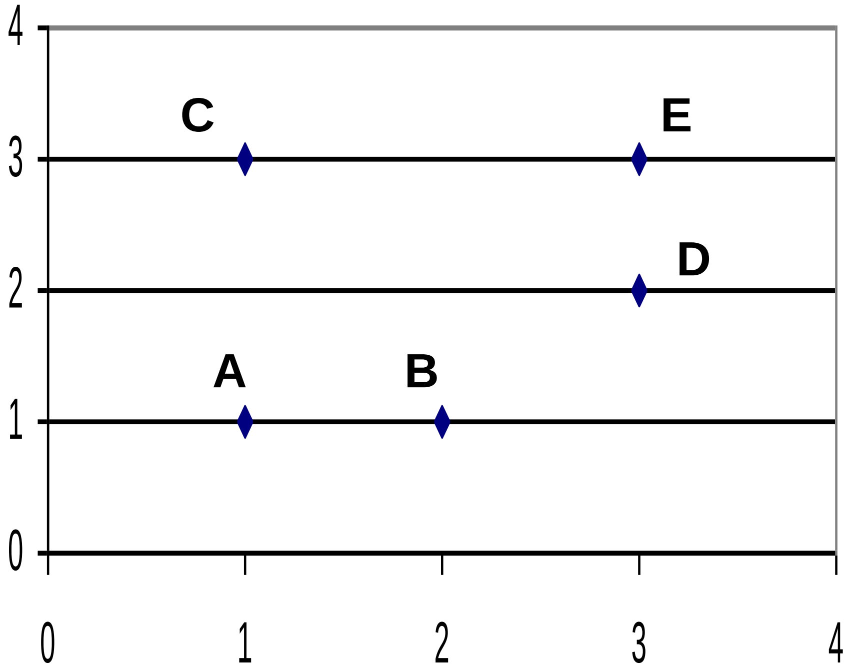
$$\bar{r}(e) = \frac{.5}{\sqrt{d}}$$

Point Pattern Analysis

- Nearest Neighbor Analysis
  - The index is an average distance from the closest neighbor to each point with a distance that would be expected on the basis of chance.
  - If the observed average distance is the same as the mean random distance, then the ratio will be 1.0 (random).
  - If the observed distance is smaller than the mean random distance, the NNI is less than 1.0 (clustered).
  - If the observed distance is higher than the mean random distance, the NNI is greater than 1.0 (dispersed).

Point Pattern Analysis

- Nearest Neighbor Analysis
  - Example:
    - Five point events: A=(1,1), B=(2,1), C=(1,3), D=(3,2), E=(3,3), unit: miles. Calculate the mean NN index



- Nearest Neighbor Analysis
  - Example:
    - Five point events: A=(1,1), B=(2,1), C=(1,3), D=(3,2), E=(3,3), unit: miles. Calculate the mean NN index
      - Step 1: calculate the distances

	A	B	C	D	E	Nearest neighbor	Nearest neighbor distance
A	0	1	2	2.2	2.8	B	1
B	1	0	2.2	1.4	2.2	A	1
C	2	2.2	0	2.2	2	A/E	2
D	2.2	1.4	2.2	0	1	E	1
E	2.8	2.2	2	1	0	D	1

Point Pattern Analysis

- Nearest Neighbor Analysis
  - Example:
    - Five point events: A=(1,1), B=(2,1), C=(1,3), D=(3,2), E=(3,3), unit: miles. Calculate the mean NN index
      - Step 2. calculate the observed mean NN distance

$$\overline{d_o} = \frac{1+2+1+1+1}{5} = 1.2$$

Point Pattern Analysis

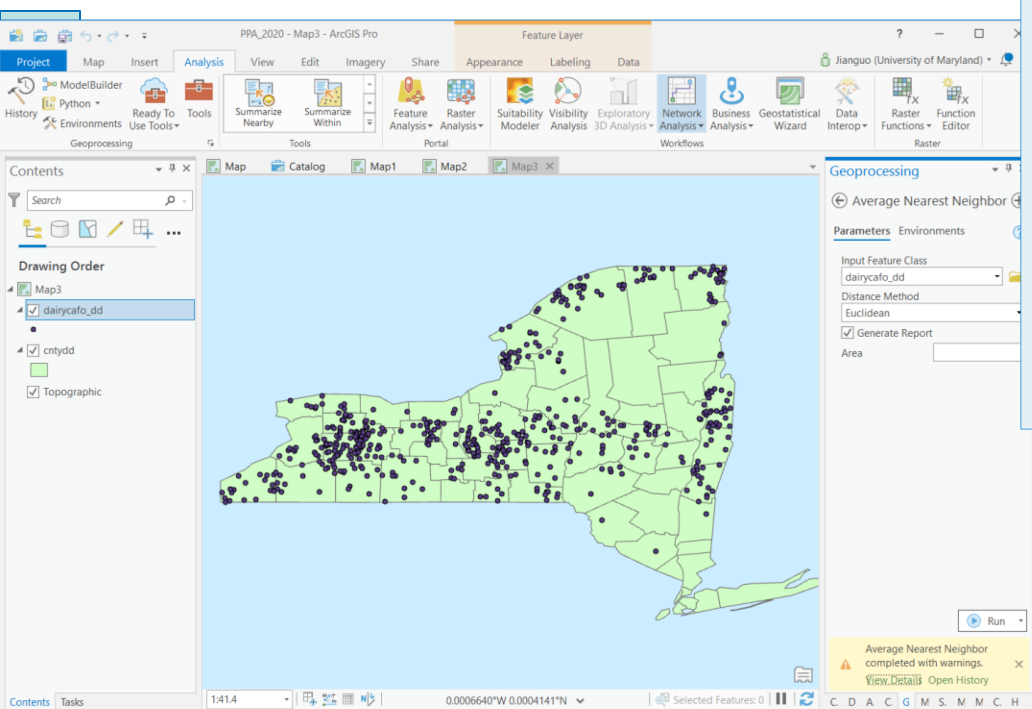
- Nearest Neighbor Analysis
  - Example:
    - Five point events: A=(1,1), B=(2,1), C=(1,3), D=(3,2), E=(3,3), unit: miles. Calculate the mean NN index
      - Step 3. calculate the expected mean NN distance of random pattern

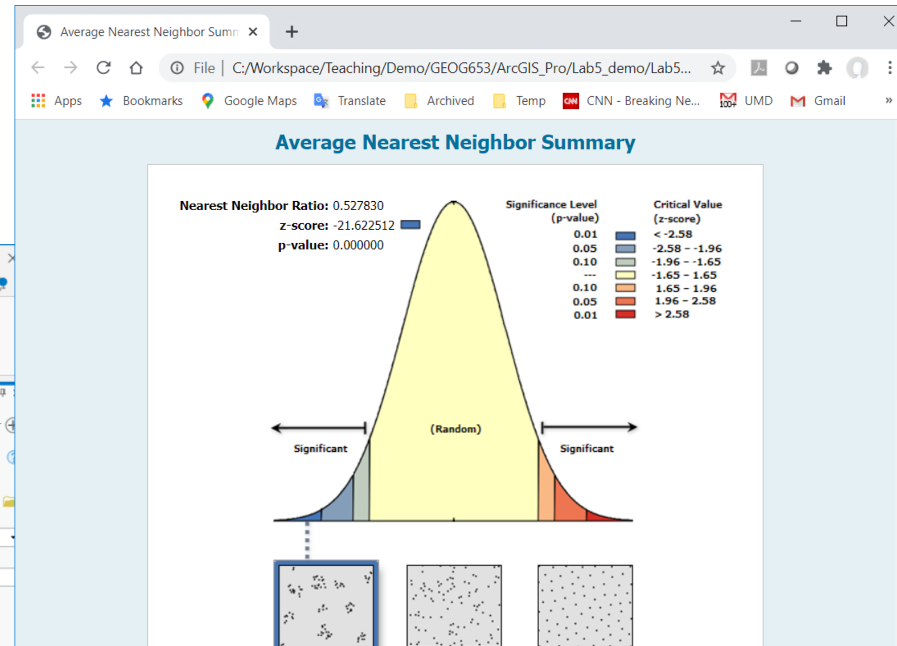
$$\overline{d_e} = \frac{1}{2} \bullet \sqrt{\frac{2 \bullet 2}{5}} = 0.45$$

Point Pattern Analysis

- Nearest Neighbor Analysis
  - Example:
    - Five point events: A=(1,1), B=(2,1), C=(1,3), D=(3,2), E=(3,3), unit: miles. Calculate the mean NN index
      - Step 4: calculate the nearest neighbor index: $r = 2.67$
      - Step 5: Conclusion: the point distribution pattern is dispersed.

- Nearest Neighbor Analysis
  - NNA tool in ArcGIS
    - Average Nearest Neighbor





- Nearest Neighbor Analysis
  - Advantages compared to Quadrat Count Analysis:
    - No quadrat size problem to be concerned with
    - Takes distance into account

Point Pattern Analysis

- Spatial Autocorrelation Analysis
  - Global Moran's I
  - General G Statistic

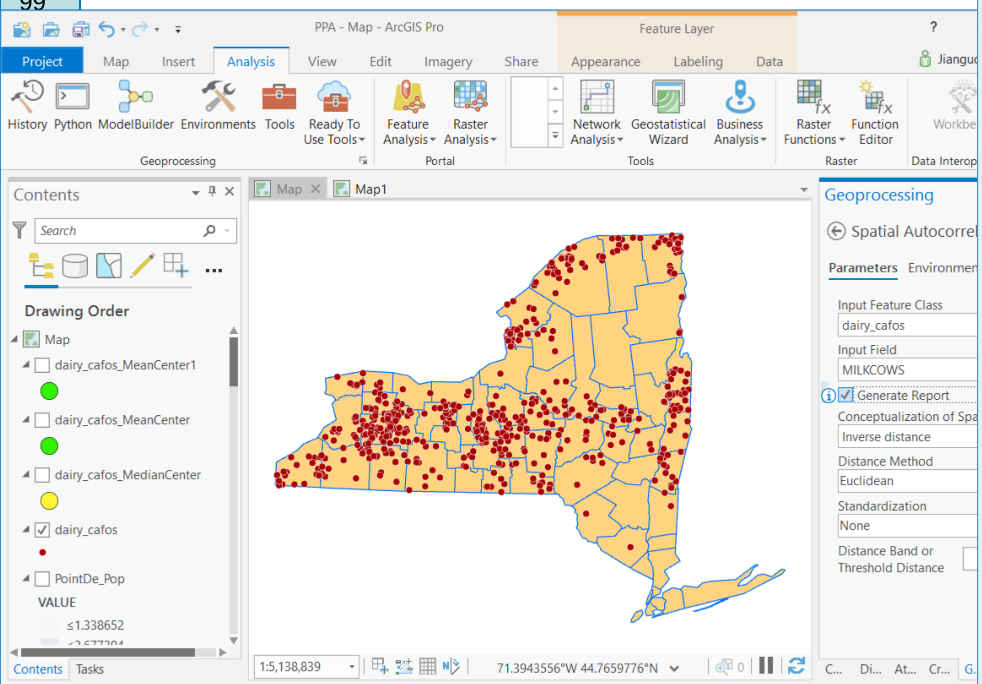
Point Pattern Analysis

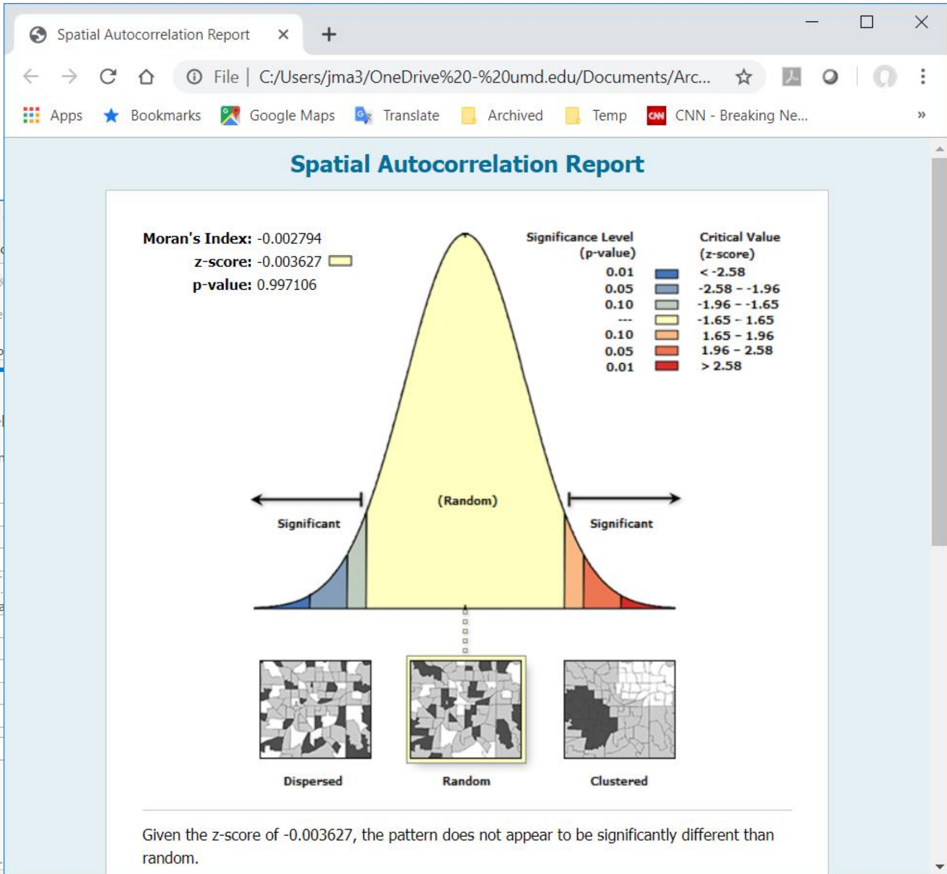
- Spatial Autocorrelation Analysis
  - Global Moran’s I
  - Tool in ArcGIS: Spatial Autocorrelation (Moran’s I)
    - Measures spatial autocorrelation based on feature locations and attribute values using the Global Moran's I statistic.
    - Example: Are those farms clustered based on the number of cows on each farm?

Point Pattern Analysis

- Spatial Autocorrelation Analysis
  - Spatial Autocorrelation (Moran's I)
    - Example: Are those farms clustered based on the number of cows on each farm?

99



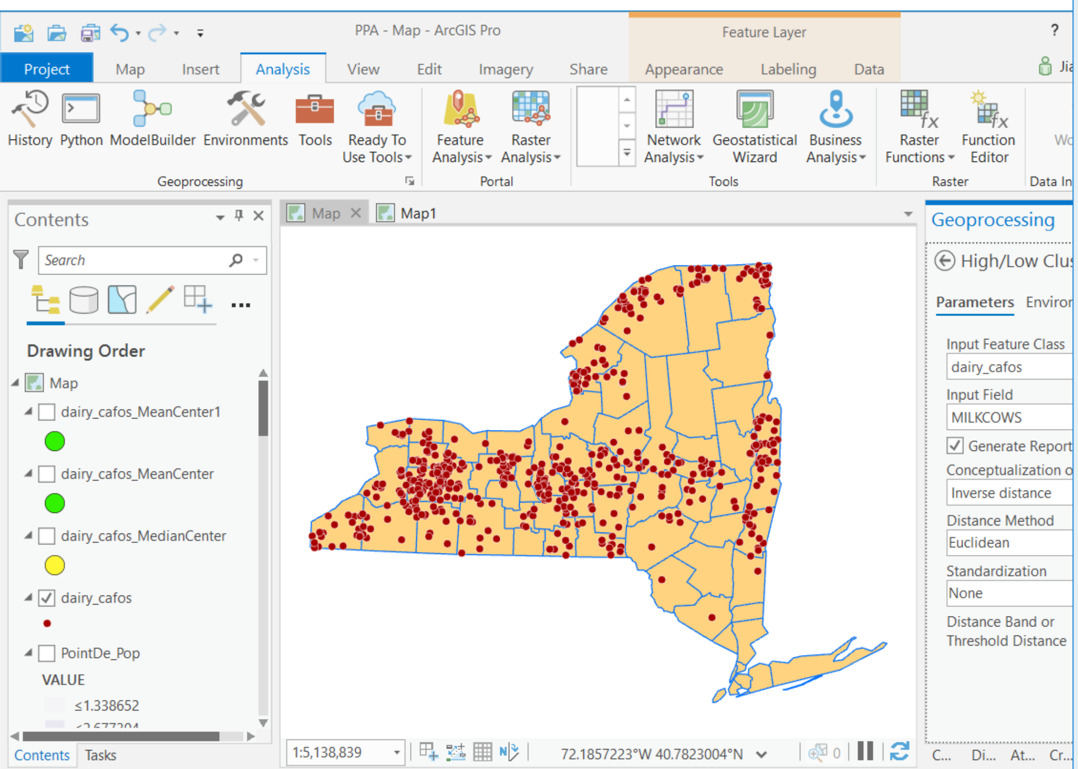


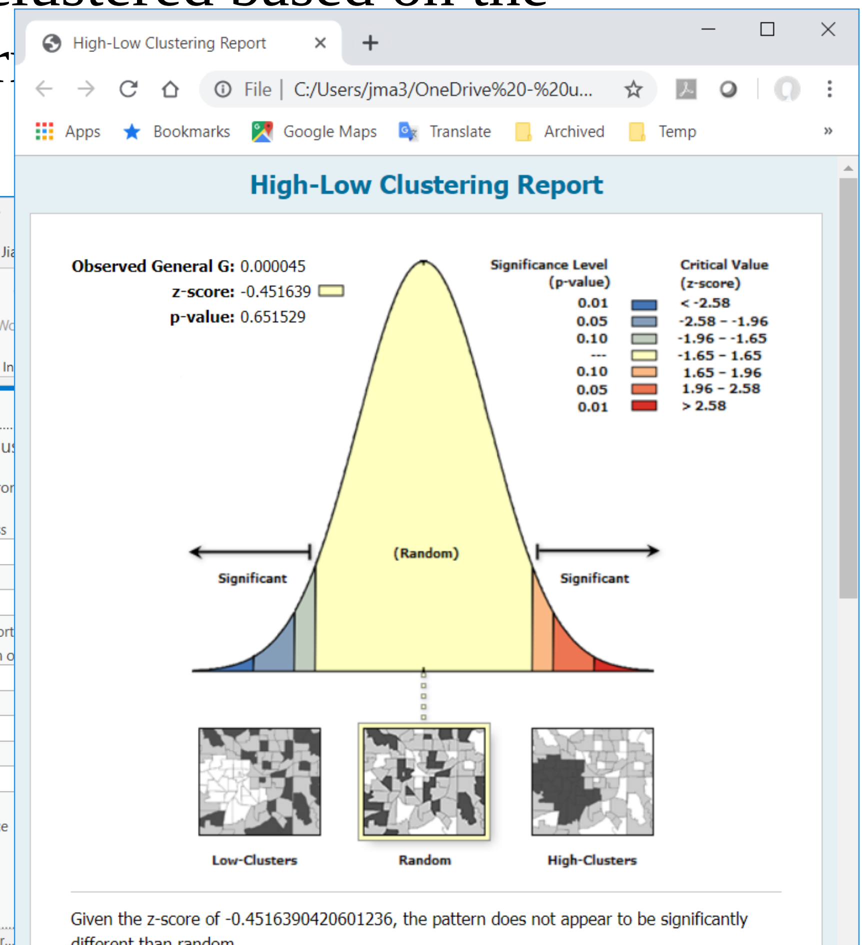
Point Pattern Analysis

- Spatial Autocorrelation Analysis
  - General G Statistic
  - Tool in ArcGIS: High/Low Clustering (Getis-Ord General G)
    - Measures the degree of clustering for either high values or low values using the Getis-Ord General G statistic.
    - Example: Are those farms clustered based on the number of cows on each farm?

Point Pattern Analysis

- Spatial Autocorrelation Analysis
  - High/Low Clustering (Getis-Ord General G)
    - Example: Are those farms clustered based on the number of cows on each far





Point Pattern Analysis

- Demos
  - Geometric Measurements of Points
  - Quadrat Count Analysis
  - Density Analysis
  - NNA
  - Clustering Analysis (Spatial Autocorrelation)
    - Moran's I
    - General G Statistics

THE END