Spatial Data Analysis

Spatial Data Selection

- Selection operations identifies features that meet one or several conditions.
- Selection on a table is performed by 'Queries'
- Algebraic notations in queries

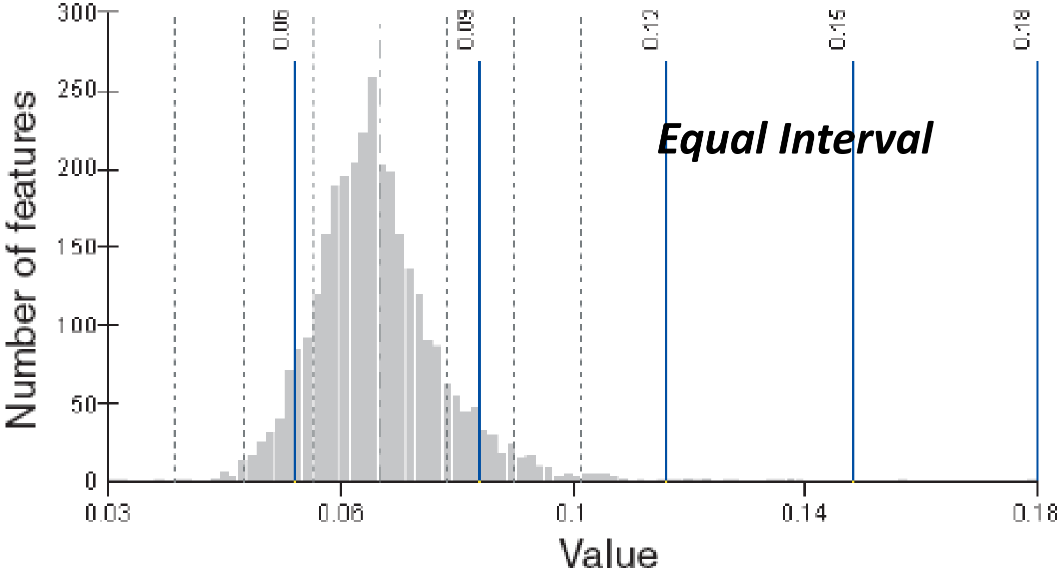
Set Algebra	Boolean Algebra
• Less than • Greater than • Equal to • Not equal to	• AND • OR • NOT

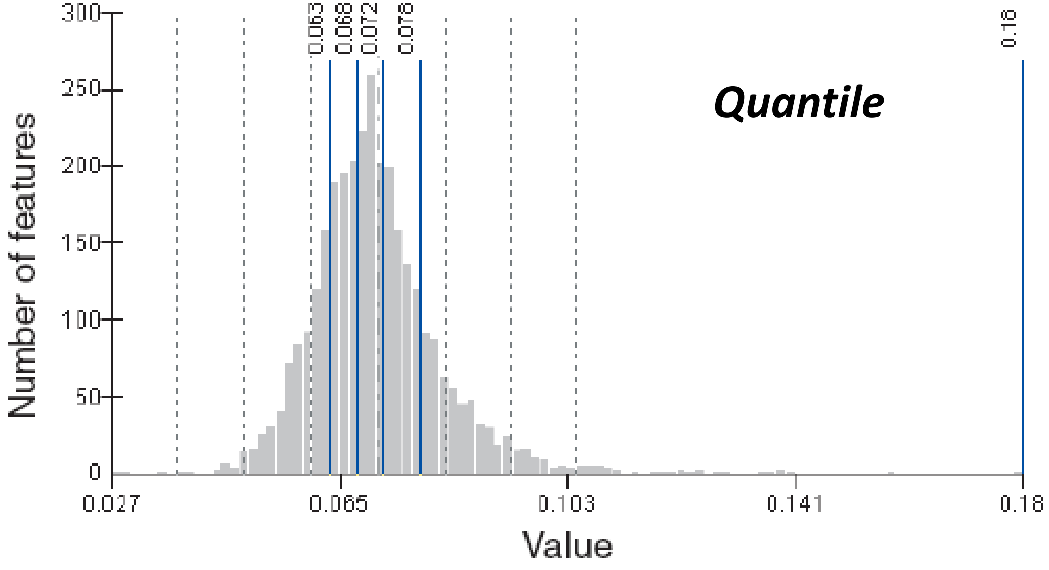
Spatial data Classification

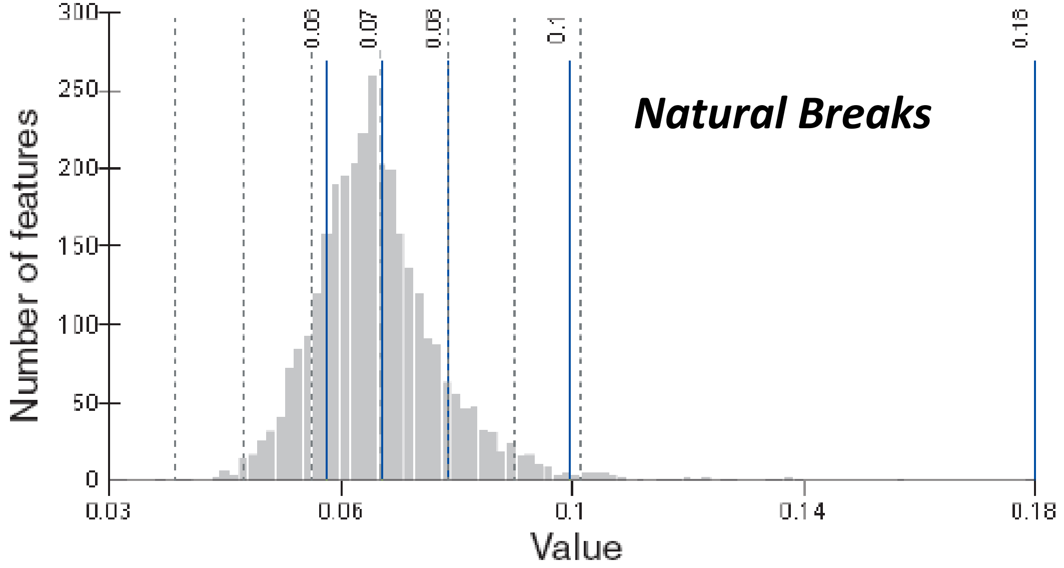
- Categorize geographic objects based on a set of conditions.
- Equal Interval
  - Divides the range of attribute values into equal-sized subranges.
  - This allows you to specify the number of intervals, and the class breaks based on the value range are automatically determined.
- Natural Breaks (Jenks)
  - Classes are based on natural groupings inherent in the data.
  - groups similar values together and maximizes the differences between classes.
  - boundaries are set where there are relatively big differences in the data values.

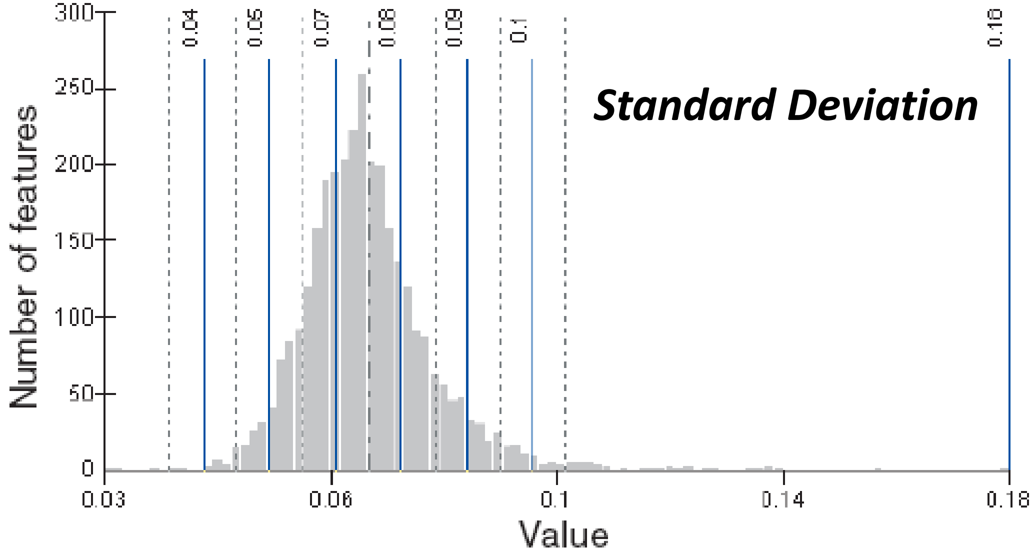
- **Quantile**
  - Each class contains an equal number of features.
  - Quantile assigns the same number of data values to each class.
  - There are no empty classes or classes with too few or too many values.
- **Standard Deviation**
  - SD shows you how much a feature's attribute value varies from the mean.
  - The mean and standard deviation are calculated automatically.
  - Class breaks are created with equal value ranges that are a proportion of the standard deviation usually at intervals of one, one-half, one-third, or one-fourth—using mean values and the standard deviations from the mean.

- **Geometrical Interval**
  - Geometric Interval creates class breaks based on class intervals that have a geometric series.
  - The geometric coefficient in this classifier can change once to optimize the class ranges.
  - The algorithm creates geometric intervals by minimizing the sum of squares of the number of elements in each class.
  - This ensures that each class range has approximately the same number of values in each class and that the change between intervals is fairly consistent.









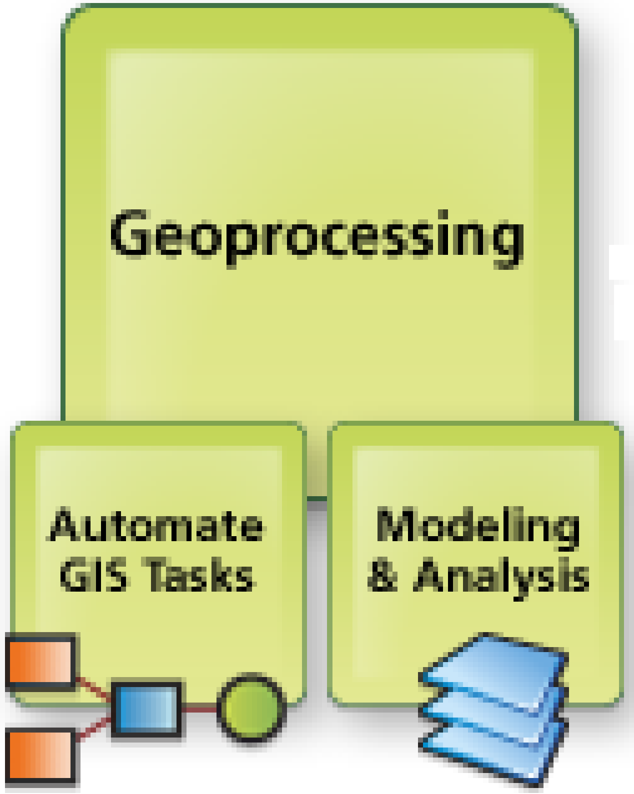
http://individual.utoronto.ca/lackner/ggr272/DataClassificationMethods.pdf

Spatial Data Analysis

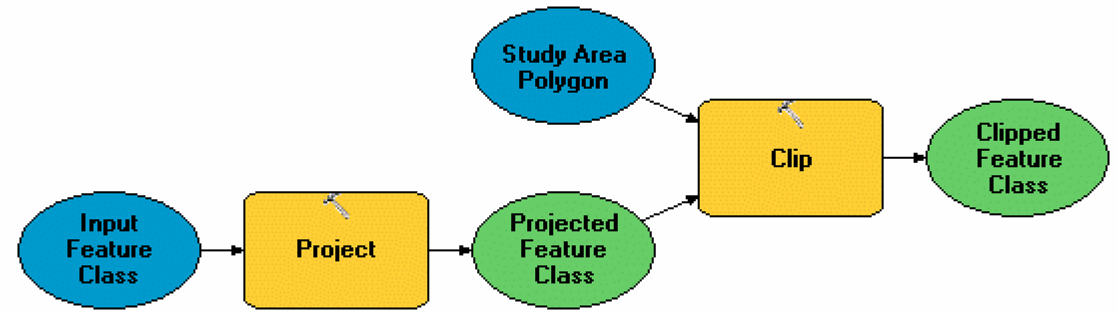
Spatial Data Analysis

- The fundamental purpose of SDA is to provide **tools and framework** for performing analysis and managing geographic data.
- Spatial analysis is any of the formal techniques which studies entities using their **topological, geometric, or geographic properties.**
- Known as '**Geoprocessing**' in ArcGIS
- The modeling and analysis capabilities of geoprocessing make GIS complete.
- Tasks can be quite creative, using a sequence of operations to model and analyze complex spatial relationships

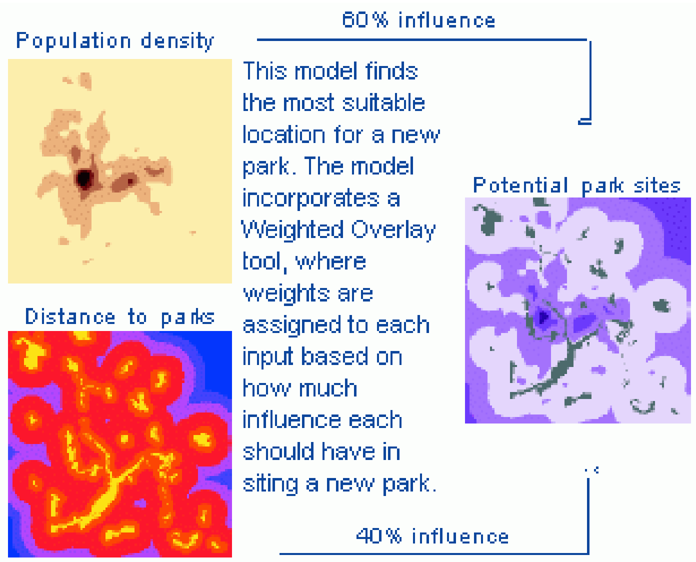
What SDA offers



Task Automation



Modelling and Analysis



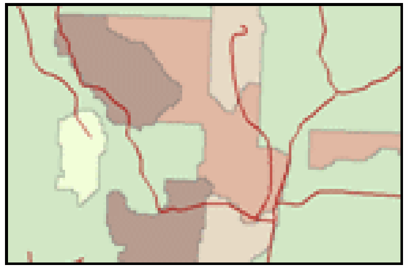
SDA Essentials

- Tool
  - A tool performs a small, essential operation on GIS data. Tools can be used in ModelBuilder, and you can call them from software programs. Tools are two categories, system and custom.
- Toolbox
  - Toolbox contain tools and toolsets. Toolsets are simple organization folders, just like folders on your system.
- Environments
  - Environment settings can be thought of as additional parameters that affect the execution of a tool. They are values you set once using a separate dialog and are used by tools when they are executed.
- Models and ModelBuilders
  - A ModelBuilder is how you quickly and easily turn your ideas into software by chaining together tools into a sequence, feeding the output of one tool to the input of another.

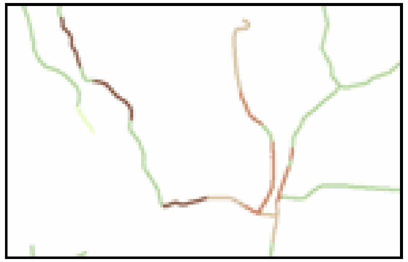
Overlay Analysis

I. Overlay Analysis

- One of the most basic questions asked of a GIS is **"What is on what?"** For example:
  - What land use is on top of what soil type?
  - What roads are within what counties?
- Overlay analysis can **combine** the characteristics of several datasets into one. Find specific locations or areas that have a certain set of attribute values that match the criteria specified.



=



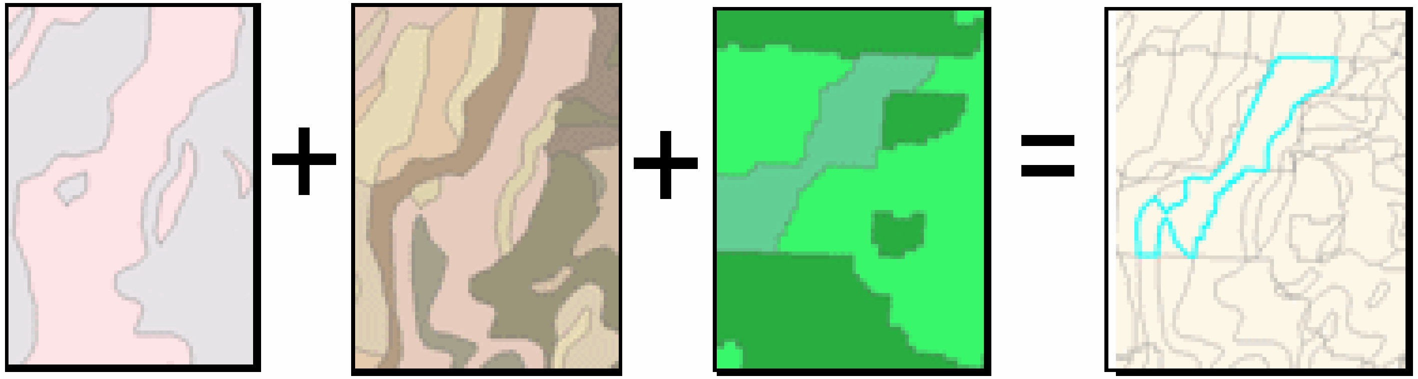
FID	Shape*	LOCALID	RD_TYPE
6	Polyline	222192	1500
7	Polyline	220893	1500

+

FID	Shape*	VEG_TYPE
6	Polygon	CC
7	Polygon	SO
8	Polygon	SS

FID	Shape*	LOCALID	RD_TYPE	VEG_TYPE
6	Polyline	219378	1500	FC
7	Polyline	219384	1500	FC
8	Polyline	219380	1500	FC
9	Polyline	219380	1500	SO
10	Polyline	224631	1500	FC

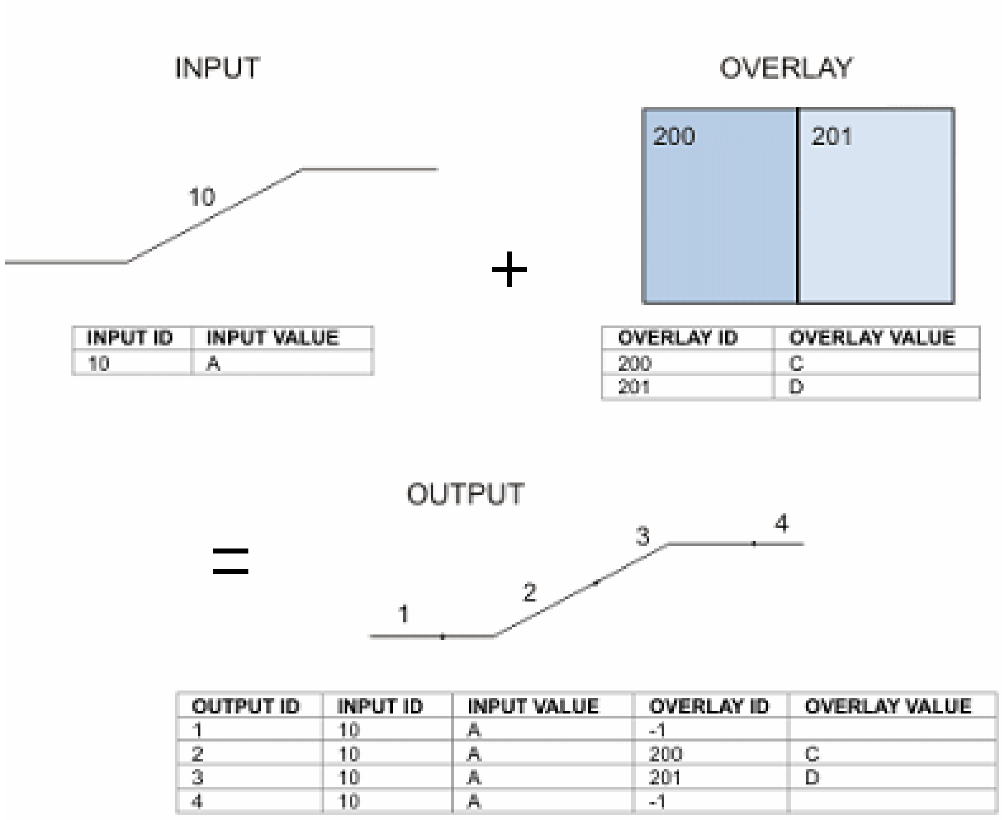
- Below is an example of an overlay of steep slopes, soils, and vegetation. New polygons are created by the intersection of the input polygon boundaries. The resulting polygons have all the attributes of the original polygons.



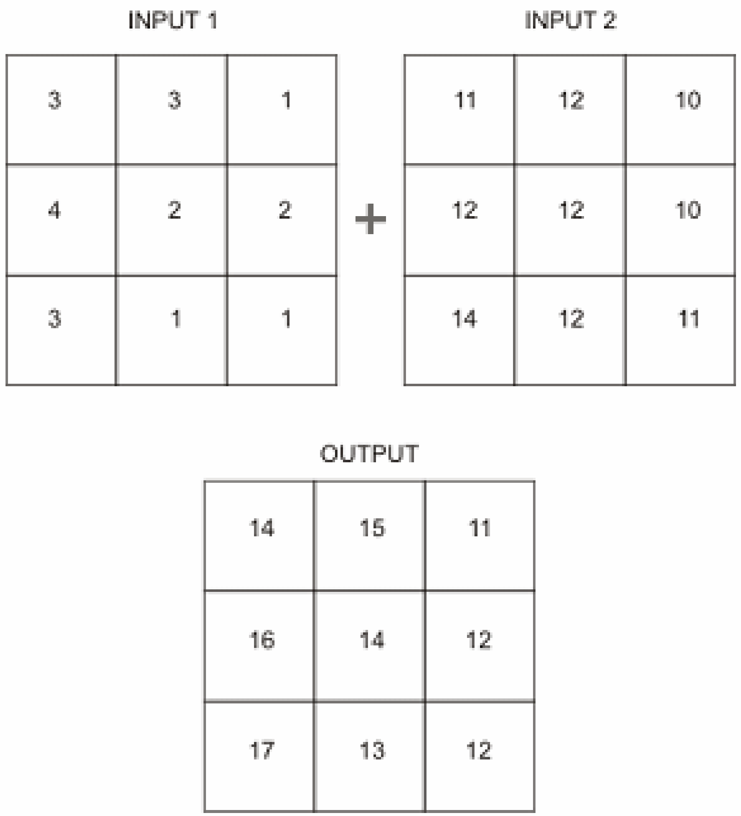
	FID	Shape*	FID_soils	CODE	CLASS	FID_sl	SLOPE	FID_veg	DET_TYPE
	3039	Polygon	508	38F	6	0	60	117	A
	3040	Polygon	508	38F	6	0	60	119	SS
	3041	Polygon	508	38F	6	0	60	157	U
	3042	Polygon	508	38F	6	0	60	158	A
	3043	Polygon	508	38F	6	0	60	160	FC

Overlay Types

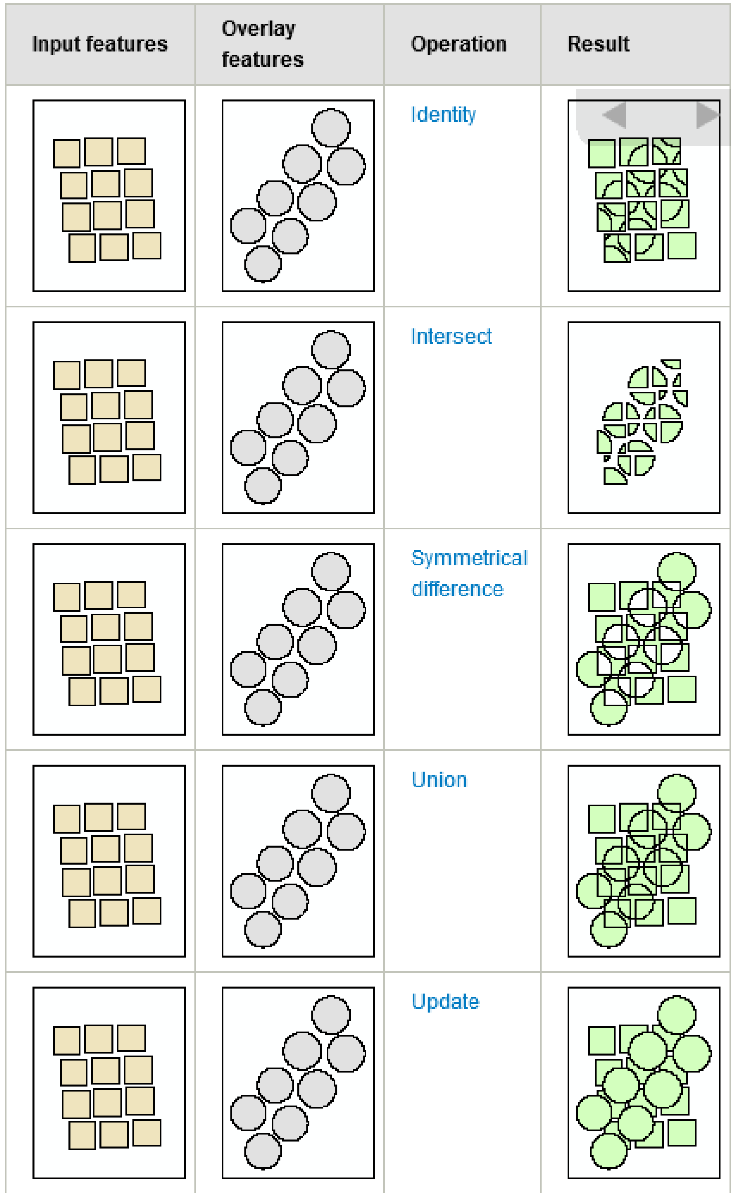
Feature Overlay



Raster Overlay



Vector Overlay



Raster Overlay

Tool	Location	What it does
Zonal Statistics	Zonal toolset	Summarizes values in a raster layer by zones (categories) in another layer—for example, calculate the mean elevation for each vegetation category.
Combine	Local toolset	Assigns a value to each cell in the output layer based on unique combinations of values from several input layers.
Weighted Overlay	Overlay toolset	Automates the raster overlay process and lets you assign weights to each layer before adding (you can also specify equal influence to create an unweighted overlay).
Weighted Sum	Overlay toolset	Overlays several rasters, multiplying each by their given weight and summing them together.

Proximity Analysis

II. Proximity Analysis

- The most basic questions asked of a GIS is "what's near what?" For example:
  - What is the distance between two locations?
  - Does any road pass within 1,000 meters from a stream?
- The Proximity toolset contains tools that are used to determine the proximity of features within one or more feature classes or between two feature classes. These tools can identify features that are closest to one another or calculate the distances between or around them.

Rules of Distance Measurement

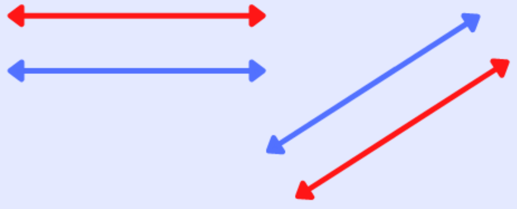
Basic operations for finding distance

- Calculating distance is dependent on the geometry type as well as other factors such as coordinate system.
- However, there are three basic rules, described in detail below, that determine how distance is calculated.
    - The distance between two points is the straight line connecting the points.
    - Distance from a point to a line is either the perpendicular or the closest vertex.
    - Distance between polylines is determined by segment vertices.

https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/how-near-analysis-works.htm

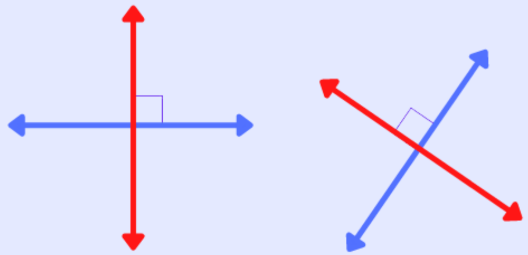
Parallel and Perpendicular Lines

Parallel



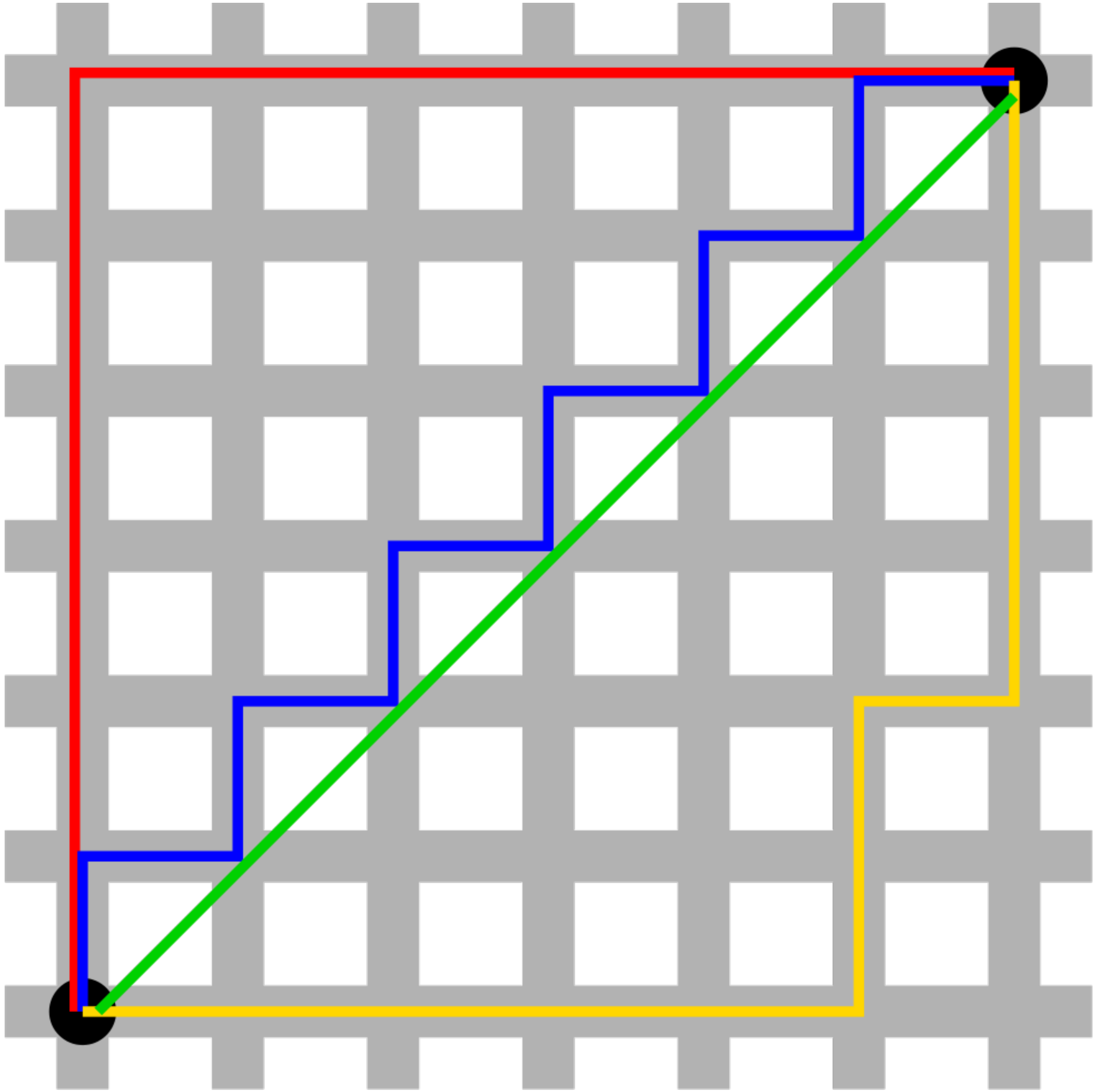
- In the same plane
- Never intersect
- Always the same distance apart
- Have the same slope
- Symbol is ||
- Railroad tracks are an example

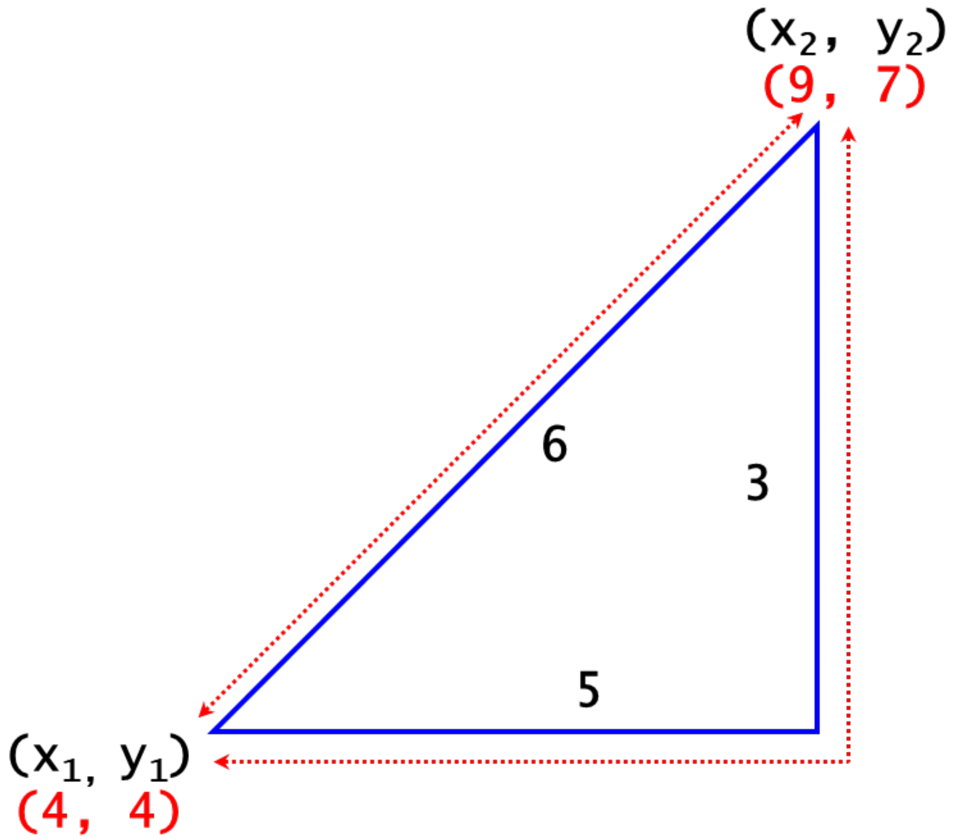
Perpendicular



- In the same plane
- Always intersect
- Intersect at a 90° angle
- One line has slope m, other has slope -1/m
- Symbol is ⊥
- The letter "T" is an example

- In taxicab/manhattan geometry, the lengths of the red, blue, green, and yellow paths all equal 12.
- The taxicab distance between the opposite corners, and all three paths are shortest paths.
- In Euclidean geometry, the green line has length $6 \times \sqrt{2} \approx 8.48$, and is the unique shortest path.





Euclidean Distance:

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

$= \sqrt{(9-4)^2 + (7-4)^2}$

$= \sqrt{(5)^2 + (3)^2}$

$= \sqrt{25+9}$

$= \sqrt{34}$

<u>= 5.83</u>

Manhattan Distance:

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

$= |9-4| + |7-4|$

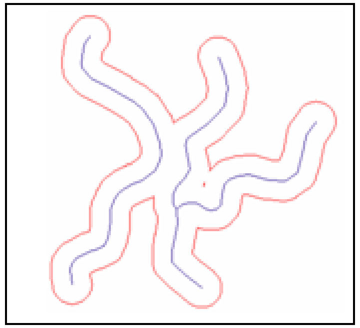
$= 5 + 3$

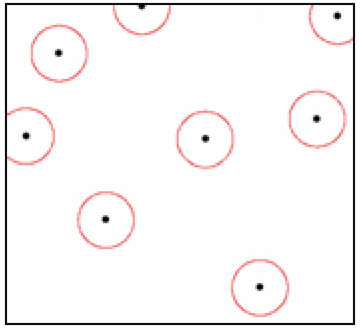
<u>= 8</u>

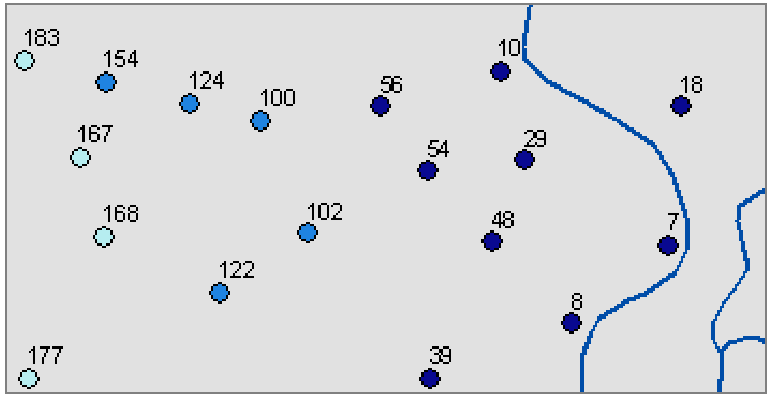

(x1,y1)
(x2,y2)
Manhattan
Euclidean

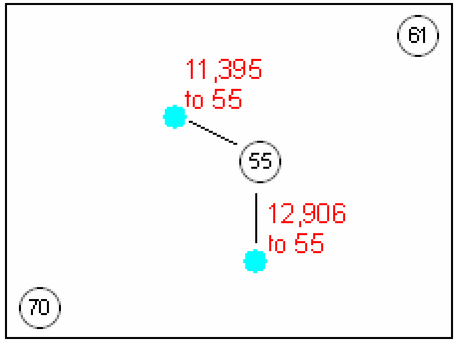
Proximity Tools: Vector

Buffer	Creates new feature data with feature boundaries at a specified distance from input features
Near	Adds attribute fields to a point feature class containing distance, feature identifier, angle, and coordinates of the nearest point or line feature
Point Distance	Creates a new table with distance and feature identifier attributes showing the distance from each point in the input feature class to all points in the Near feature class, within a given search radius
Select By Location	Selects features from a target feature class within a given distance of (or using other spatial relationships) the input features
Create Thiessen Polygons	Creates polygons of the areas closest to each feature for a set of input features



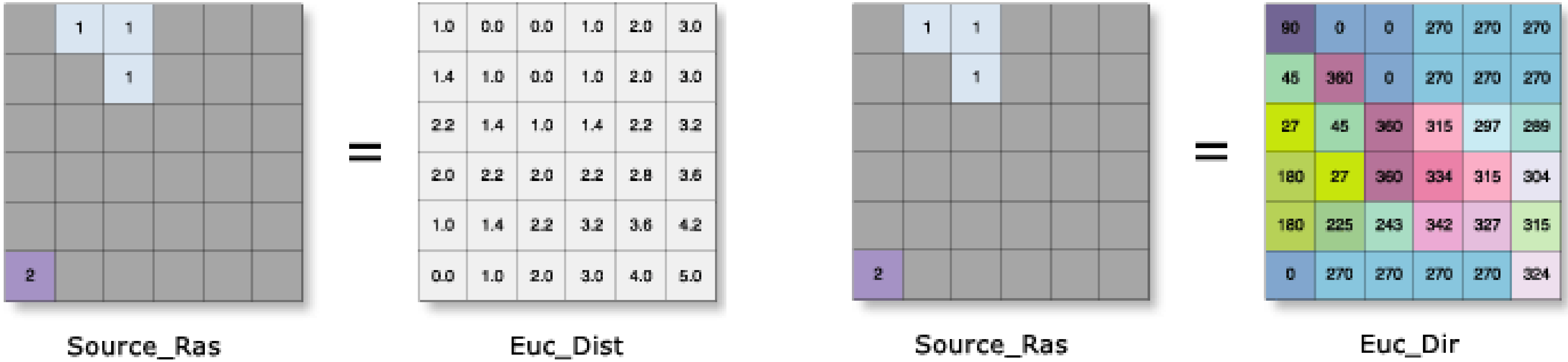






Proximity Tools: Raster

Euclidean Distance	Calculates the distance to the nearest source for each cell.
Euclidean Direction	Calculates the direction to the nearest source for each cell.
Cost Distance	Calculates the distance to the nearest source for each cell, minimizing cost specified in a cost surface.
Cost Allocation	Gives each cell the identifier of the closest source, minimizing cost specified in a cost surface.
Cost Path	Calculates the least-cost path from a source to a destination, minimizing cost specified in a cost surface.
Path Distance	Calculates the distance to the nearest source for each cell, minimizing horizontal cost specified in a cost surface, as well as the terrain-based costs of surface distance and vertical travel difficulty specified by a terrain raster and vertical cost parameters.



Network Analysis

III. Network Analysis

- What is the quickest way to get from point A to point B?
- What is the 'market area' does a business cover?

- What is Network
  - A network is a system of interconnected elements, such as edges (lines) and connecting junctions (points), that represent possible routes from one location to another.
- People, resources, and goods tend to travel along networks
- Importance and Applications in Various Industries (Transportation, Urban Planning, Utilities…)
- ArcGIS groups networks into two categories:
  - Geometric Networks (Utility and River networks)
  - Transportation Networks (Road, rail, street, and pedestrian networks)

Fundamentals of Networks

Overview of Network Components

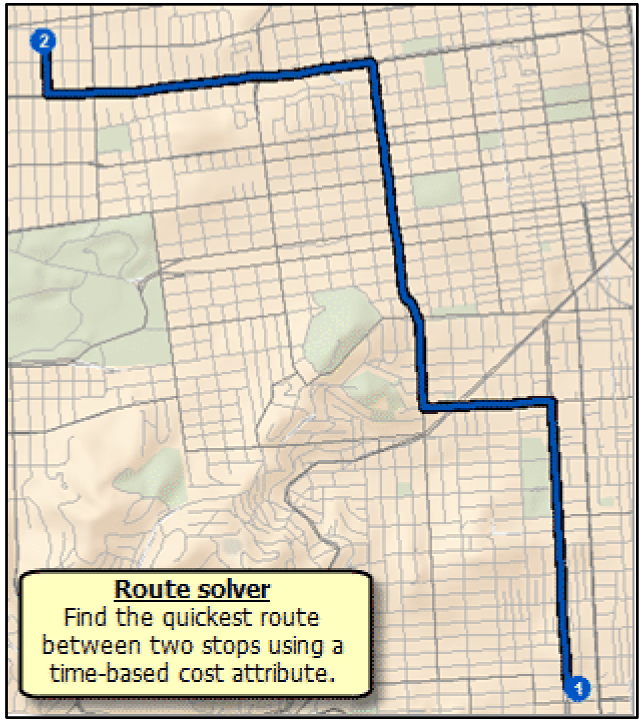
- **Edges** - Connect to other elements and are the links over which agents travel
- **Junctions** - Connect edges and facilitate navigation from one edge to another
- **Turns** - Store information that can affect movement between two or more edges

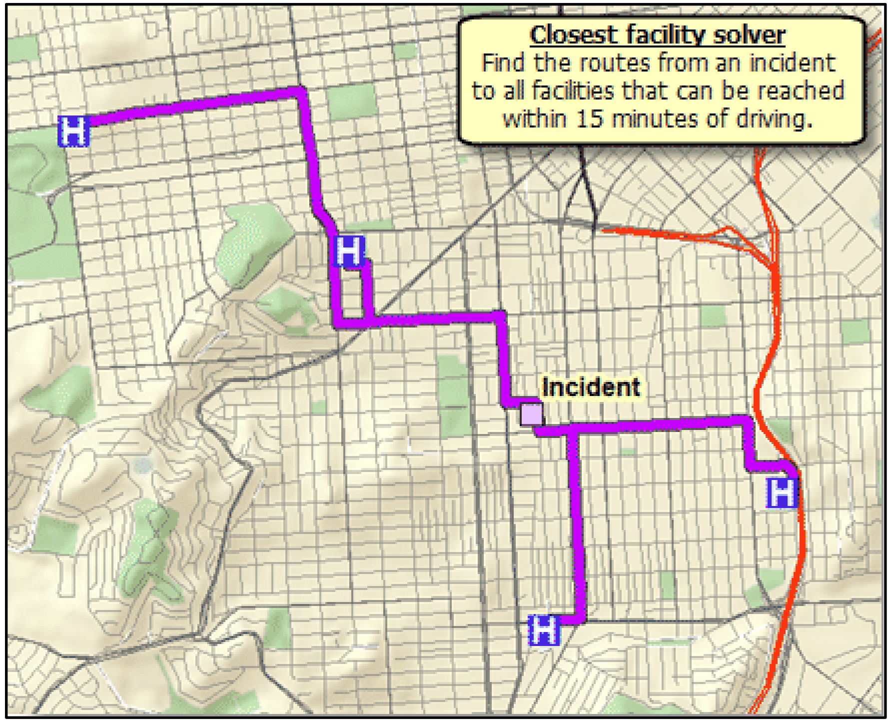
You must know…

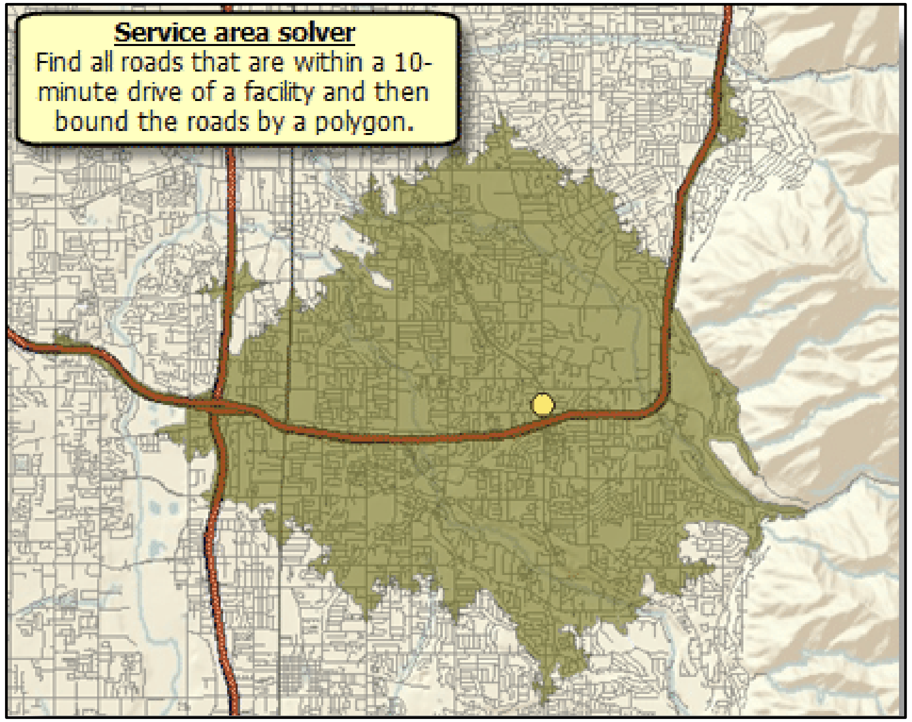
- Types of Networks (Transportation, Utility, Telecommunication)
- Data Structures and formats
- Attributes Associated with Network Elements (Length, Speed, Capacity)

Types of Network Analysis

- Route Analysis (Shortest Path, Closest Facility, Origin-Destination)
- Service Area Analysis (Isochrones, Drive-Time Analysis)
- Network Allocation (Optimal Facility Location, Resource Allocation)







Advanced Concepts

Network Optimization

(Vehicle Routing, Facility Location, Traffic Management)

Spatial Analysis with Multimodal Networks

(Integration of Multiple Transportation Modes)

Dynamic Network Analysis

(Real-time Data, Live Traffic Updates)

3D Network Models

(Mining, Floor mapping)

Case Studies and Applications

- Urban Planning and Traffic Management
- Emergency Response and Routing
- Utility Network Management
- Telecommunication Network Planning

Surface Analysis

IV. Surface Analysis

- Basics of Surface/Terrain Data in GIS
- Triangulated Irregular Network (TIN)
- Digital Elevation Models (DEM)
- Digital Terrain models (DTM)
- Digital Surface Models (DSM)
- Understanding Elevation, Slope, and Aspect
- Different Types of Surface/Terrain Data Sources (LIDAR, Satellite Imagery, Contour Maps)

DEM

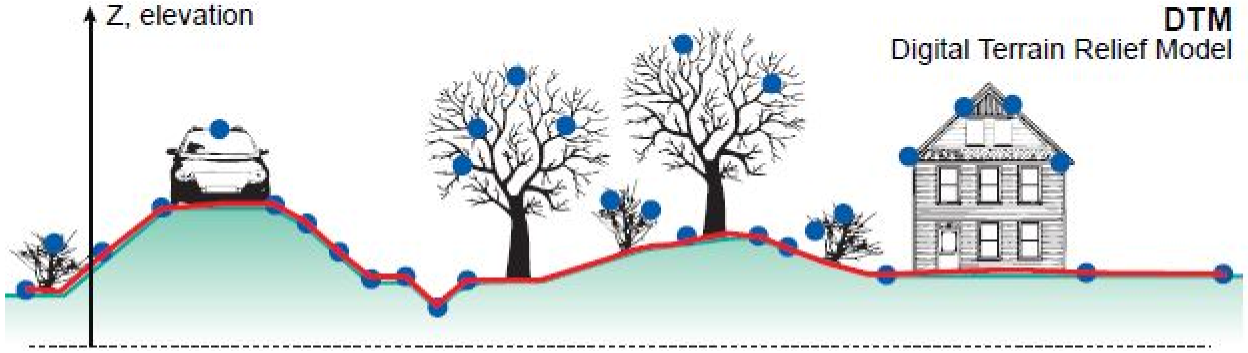
linked to a common vertical datum, which filters out and excludes all ground objects, both built and natural

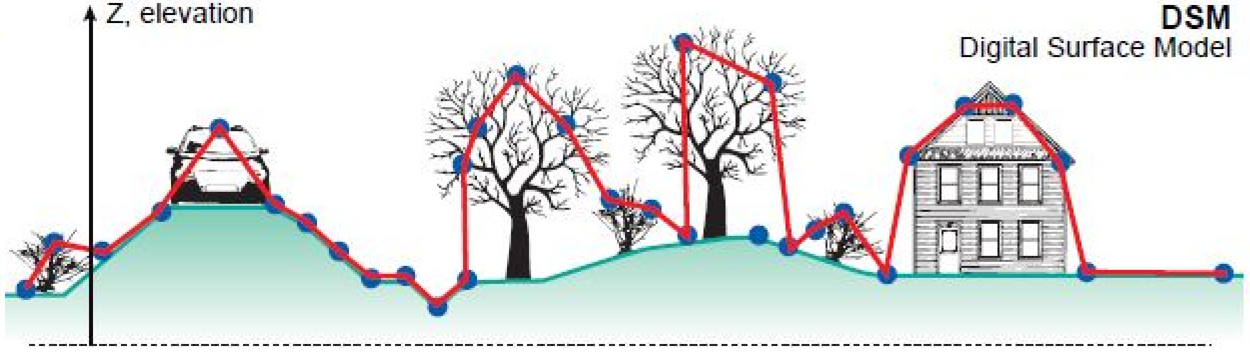
DTM

includes not only heights and elevations, but also other geographical elements such as rivers, ridge lines, etc.

DSM

includes a terrain model, buildings, vegetation, and any other features above the earth surface





Data Processing and Preparation

- Data Acquisition and Preprocessing Techniques
- DEM Resolution, Accuracy, and Quality Assessment
- DEM Data Integration and Conversion

Core Surface Analysis

Tool	Description
Aspect	The aspect identifies the downslope direction of the maximum rate of change in value from each cell to its neighbors.
Contour	Creates a line feature class of contours (isolines) from a raster surface.
Contour List	Creates a feature class of selected contour values from a raster surface.
Contour with Barriers	Creates contours from a raster surface. The inclusion of barrier features will allow you to independently generate contours on either side of a barrier.
Curvature	Calculates the curvature of a raster surface, optionally including profile and plan curvature.
Cut Fill	Calculates the volume change between two surfaces. This is typically used for cut and fill operations.
Hillshade	Creates a shaded relief from a surface raster by considering the illumination source angle and shadows.
Slope	Identifies the slope (gradient, or rate of maximum change in z-value) from each cell of a raster surface..
Observer Point	Identifies which observer points are visible from each raster surface location.
Viewshed	Determines the raster surface locations visible to a set of observer features.
Visibility	Determines the raster surface locations visible to a set of observer features, or identifies which observer points are visible from each raster surface location.

Associated Surface Analysis

- Hydrological Analysis
  - Watershed Delineation and Basin Analysis
  - Flow Accumulation and Drainage Network Extraction
  - Stream Network Generation and Channel Identification

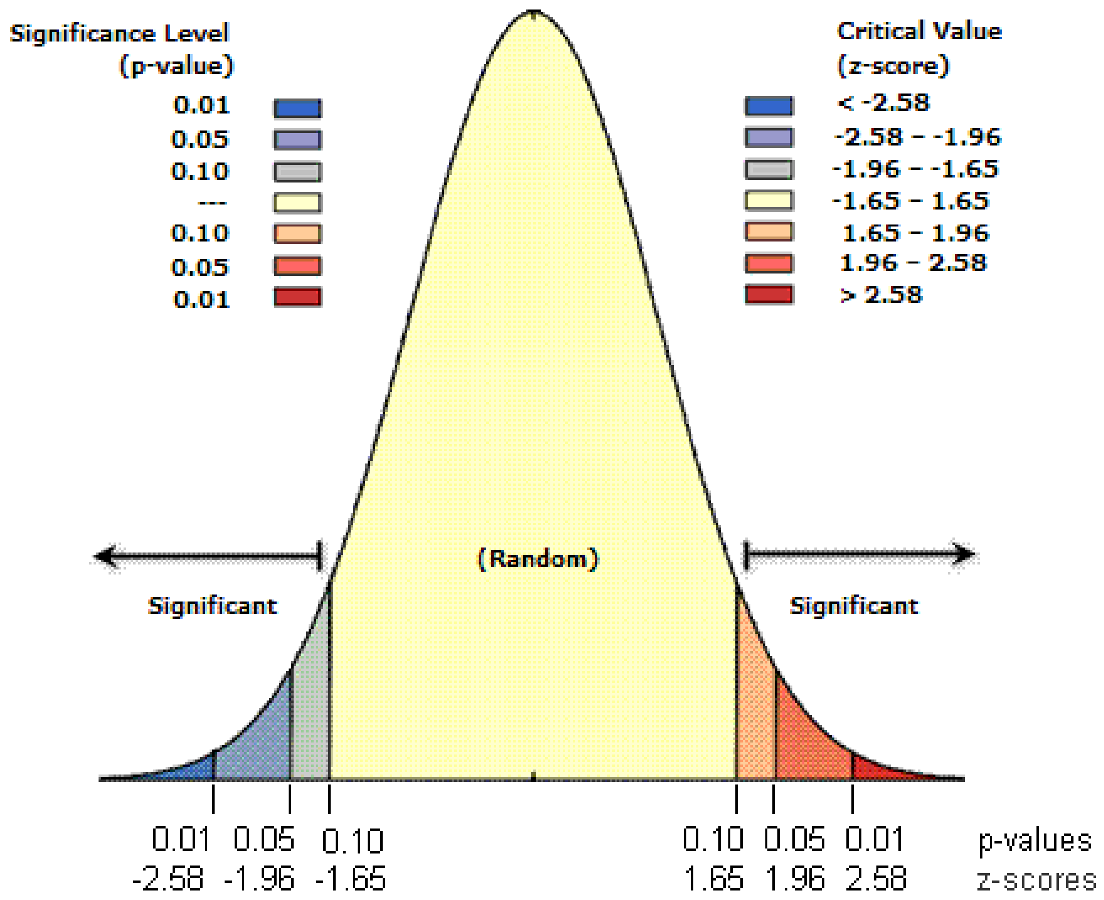
- Geomorphological Analysis
  - Landform Classification and Morphometry
  - Geomorphological Mapping Techniques
  - Identification of Geomorphic Features (Valleys, Ridges, etc.)

Spatial Statistical Analysis

V. Spatial Statistical Analysis

- Statistical analysis for spatial data
- The tools in the Spatial Statistics allow you to summarize:
  - Spatial distribution (determine the mean center or overarching directional trend, for example)
  - Identify statistically significant spatial clusters (hot spots/cold spots) or spatial outliers
  - Assess overall patterns of clustering or dispersion
  - Group features based on attribute similarities
  - Identify an appropriate scale of analysis, and
  - Explore spatial relationships.

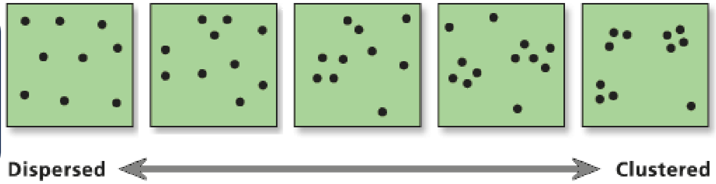
- **Mapping Patterns:**
  - These tools evaluate if features, or the values associated with features, form a clustered, dispersed, or random spatial pattern.
- **Mapping Clusters:**
  - These tools may be used to identify statistically significant hot spots, cold spots, or spatial outliers. There are also tools to identify or group features with similar characteristics.
- **Measuring Geographic Distributions:**
  - These tools address questions such as Where's the centre? What's the shape and orientation? How dispersed are the features?
- **Modelling Spatial Relationships:**
  - These tools model data relationships using regression analyses or construct spatial weights matrices.



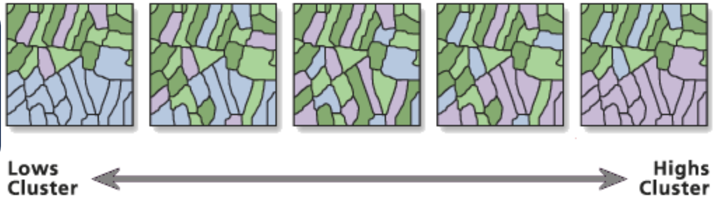
z-score (Standard Deviations)	p-value (Probability)	Confidence level
< -1.65 or > +1.65	< 0.10	90%
< -1.96 or > +1.96	< 0.05	95%
< -2.58 or > +2.58	< 0.01	99%

Mapping Patterns

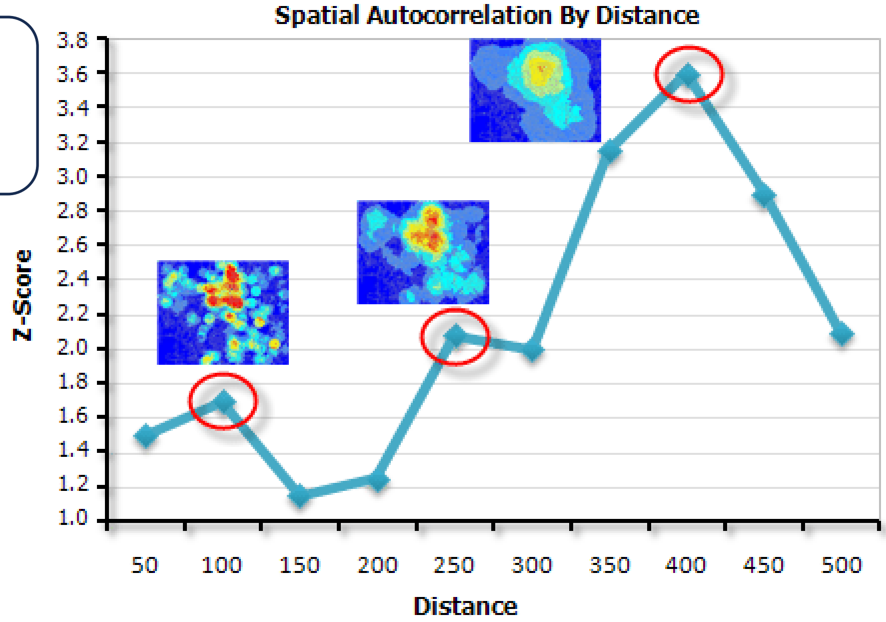
1. Average Nearest Neighbour



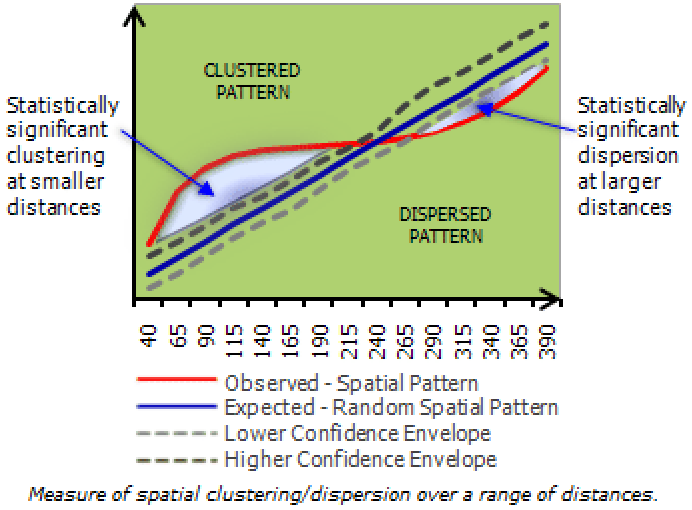
2. High/Low Clustering (Getis-Ord General G)



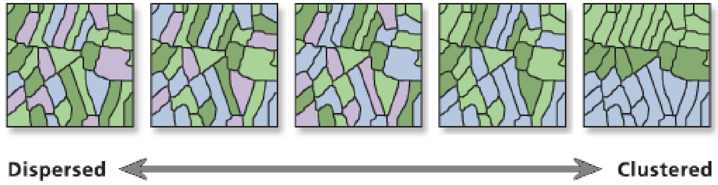
3. Incremental Spatial Autocorrelation



4. Multi-Distance Spatial Cluster Analysis (Ripley's K Function)

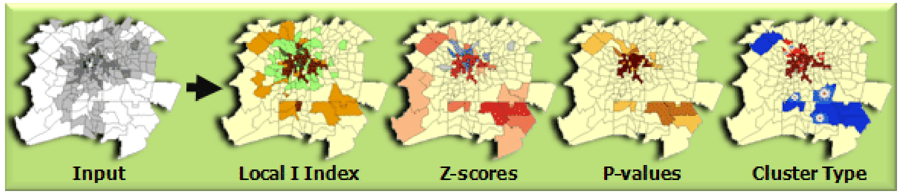


5. Spatial Autocorrelation (Global Moran's I)

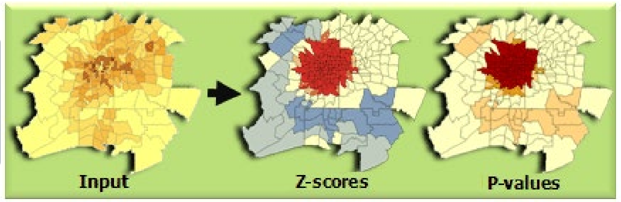


Mapping Clusters

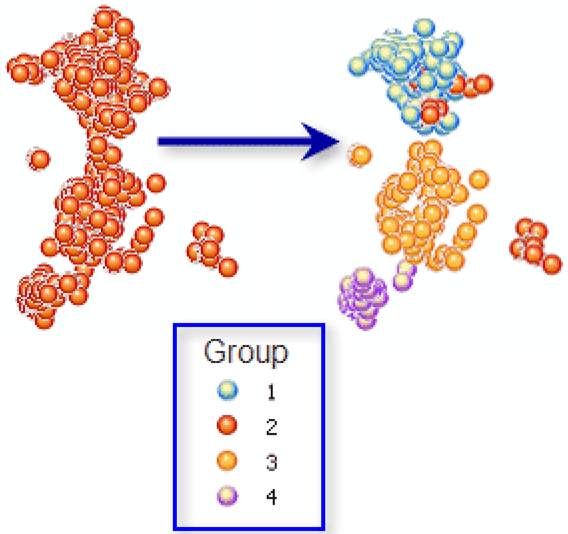
1. Cluster and Outlier Analysis (Anselin Local Moran's I)



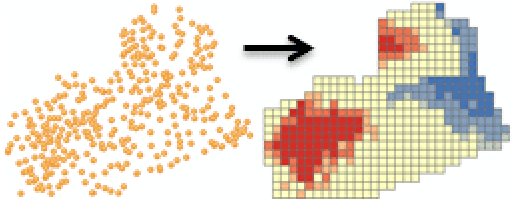
2. Hot Spot Analysis (Getis-Ord Gi*)



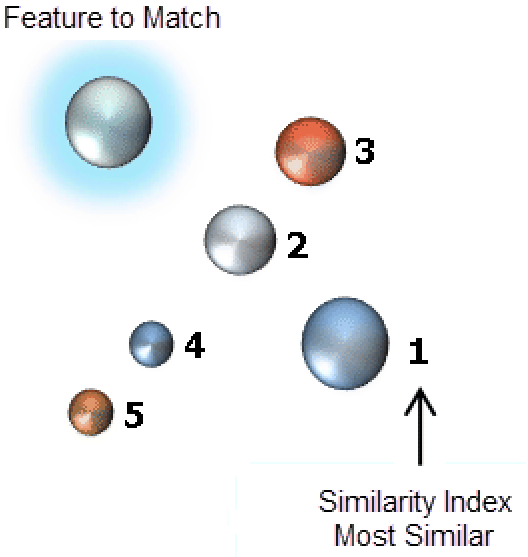
3. Grouping Analysis



4. Optimized Hot Spot Analysis

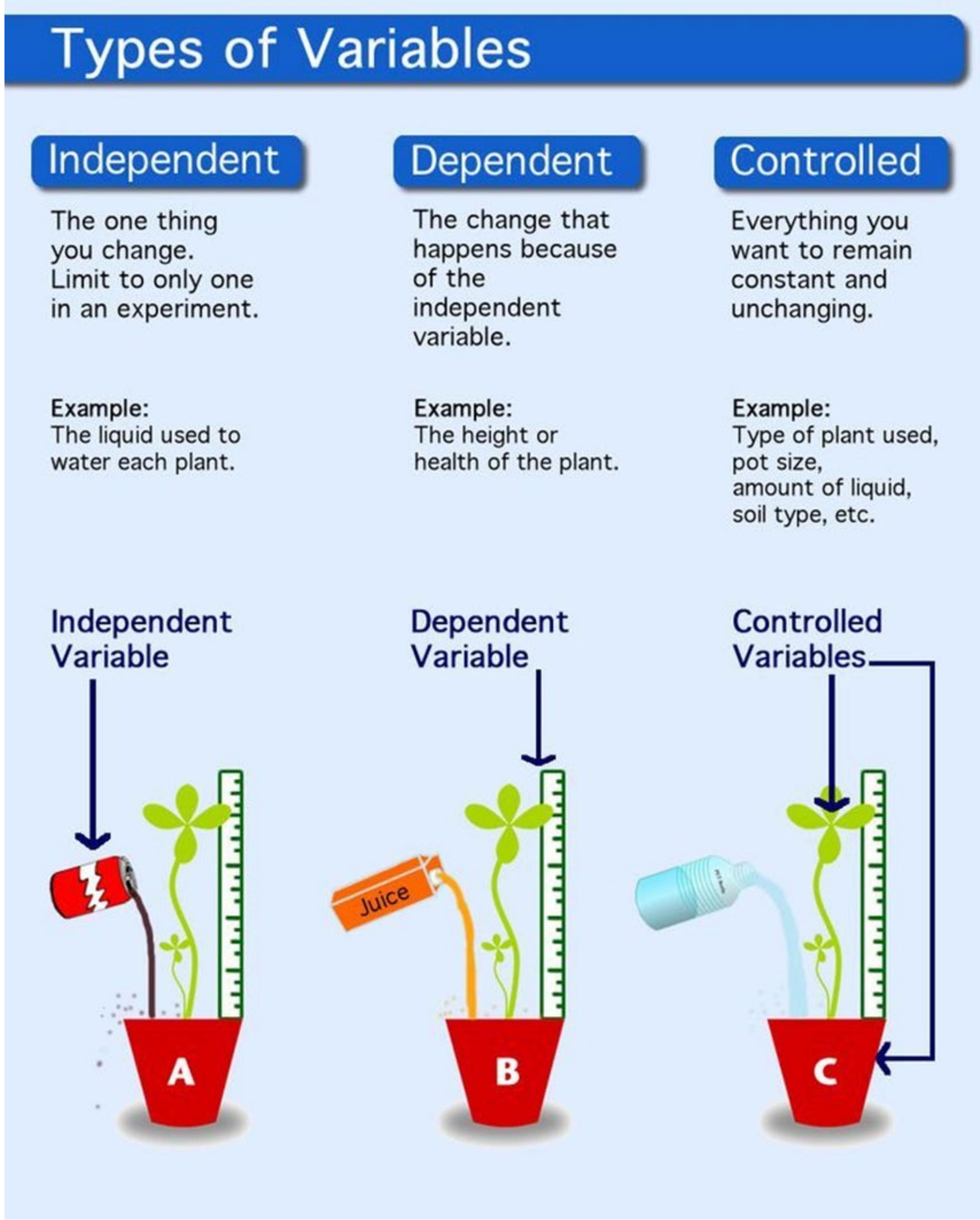


5. Similarity Search



Modelling Spatial Relationships

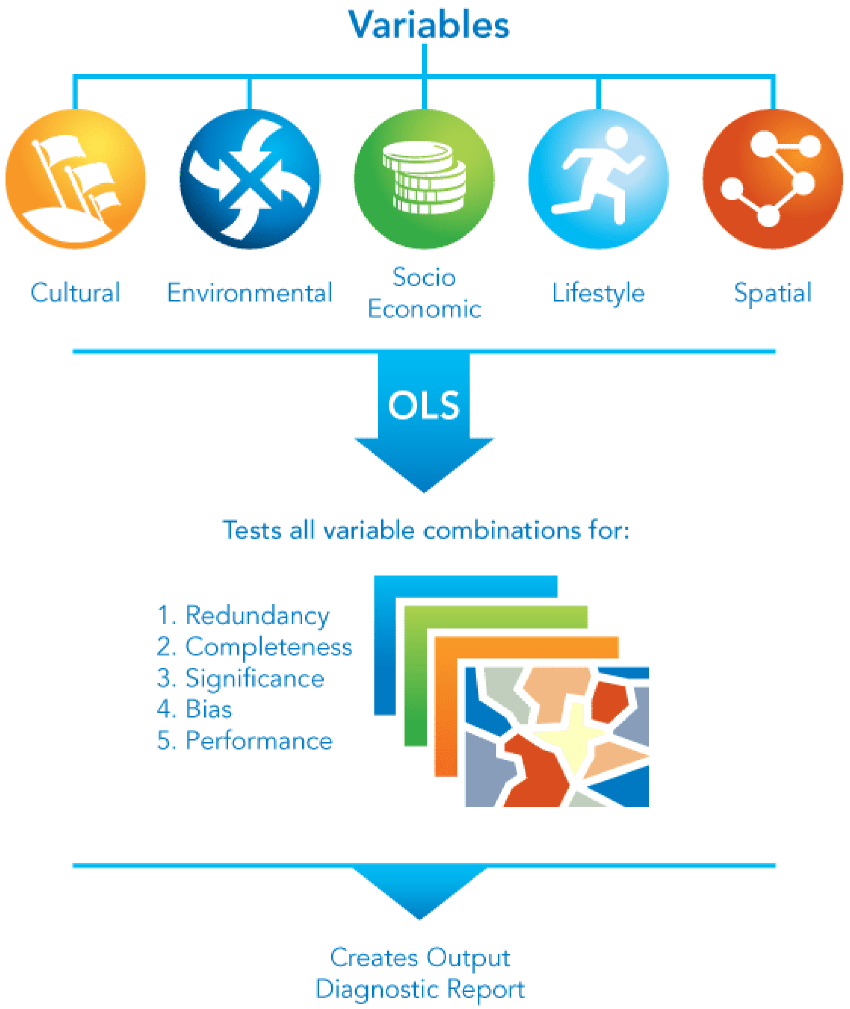
- The regression tools provided in the Spatial Statistics Toolbox model relationships among data variables associated with geographic features, allowing you to make predictions for unknown values or to better understand key factors influencing a variable you are trying to model.



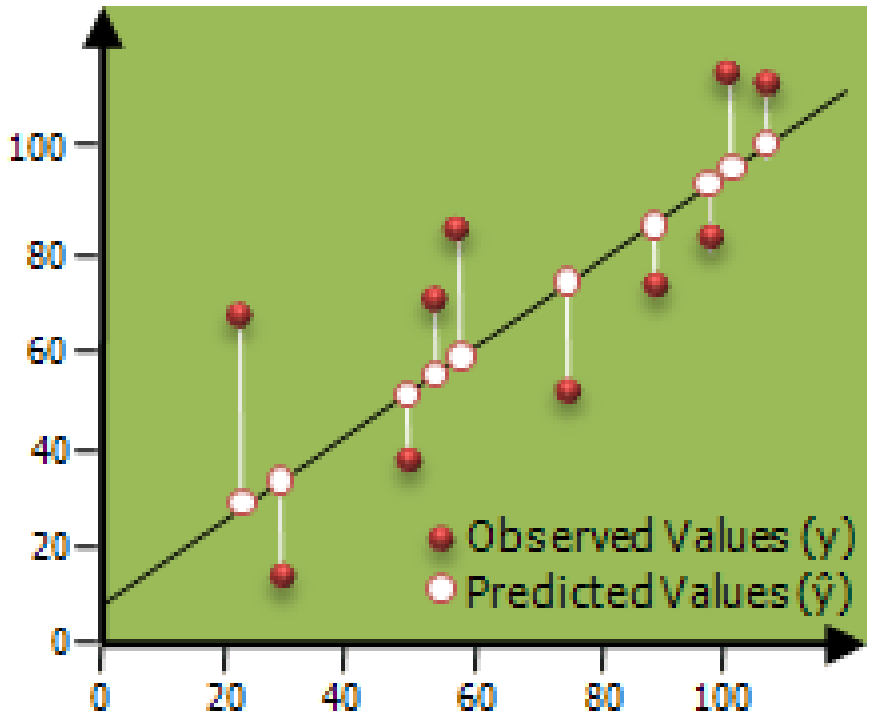
	D	I	I	I	I
	Happy Old People	Live with Spouse	Live with Kids	Events organised for them	Close to Nature
Place 1	xx	xx	xx	xx	xx
Place 2	xx	xx	xx	xx	xx
Place 3	xx	xx	xx	xx	xx
Place 4	xx	xx	xx	xx	xx
Place 5	xx	xx	xx	xx	xx

A sample data collected from 100 old people across five different places about their happy life.

1. Exploratory Regression

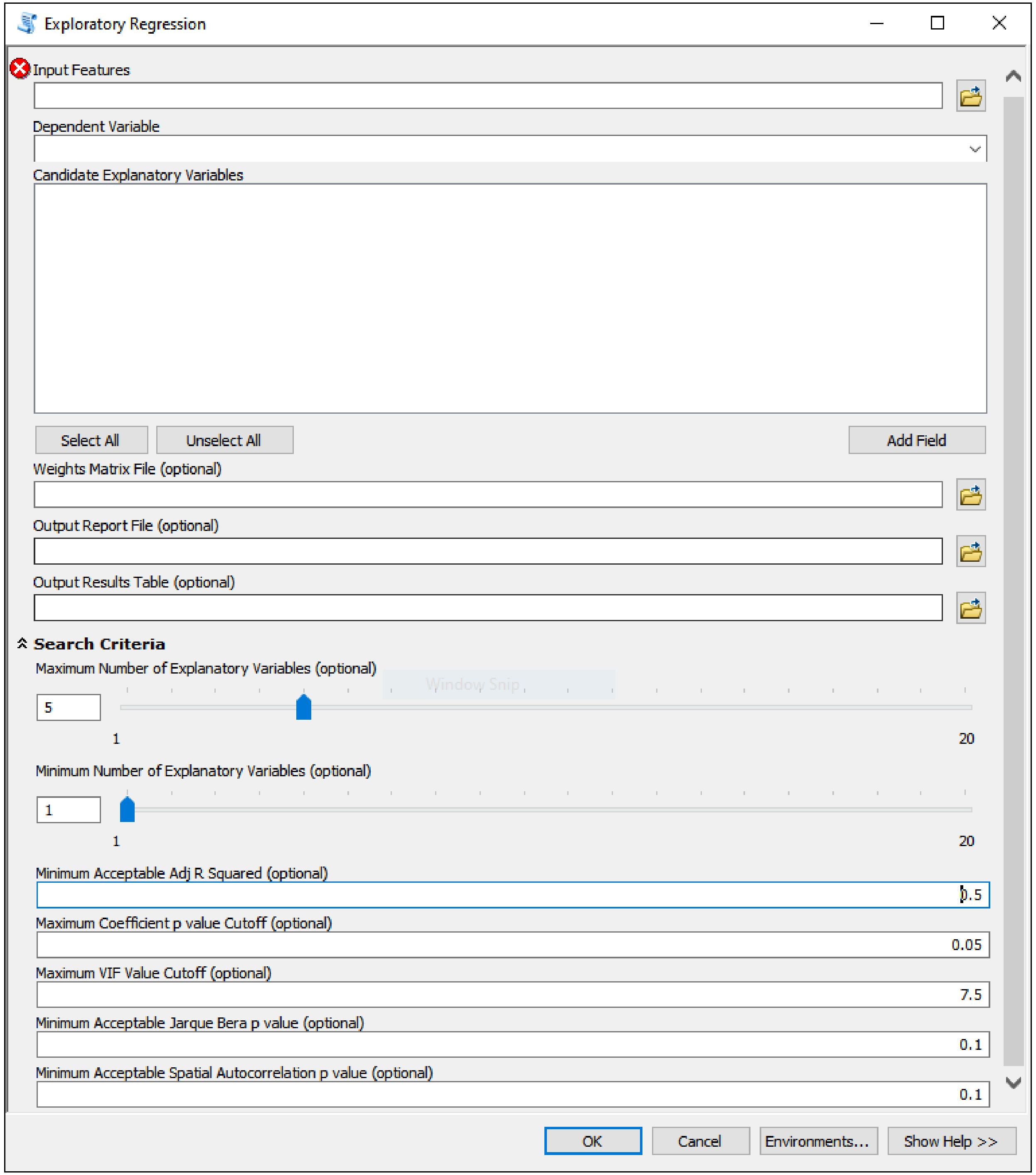


2. Ordinary Least Squares (OLS)



Ordinary Least Squares Regression: predicted values in relation to observed values

- Finding a properly specified OLS model can be difficult, especially when there are lots of potential explanatory variables you think might be important contributing factors to the variable you are trying to model (your dependent variable).
- The Exploratory Regression tool can help.
- It is a data mining tool that will try all possible combinations of explanatory variables to see which models pass all of the necessary OLS diagnostics.
- By evaluating all possible combinations of the candidate explanatory variables, you greatly increase your chances of finding the best model to solve your problem or answer your question.



https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-exploratory-regression-works.htm

Spatial Autocorrelation - Moran's I

- Given a set of features and an associated attribute, this tool evaluates whether the pattern expressed is clustered, dispersed, or random.
- When the **z-score** or **p-value** indicates statistical significance,
  - a positive Moran's I index value indicates tendency toward **clustering**
  - a negative Moran's I index value indicates tendency toward **dispersion**.
- The **z-score** and **p-value** are measures of statistical significance which tell you whether or not to reject the null hypothesis.
- For this tool, the null hypothesis states that the values associated with features are randomly distributed.

The p-value is not statistically significant.

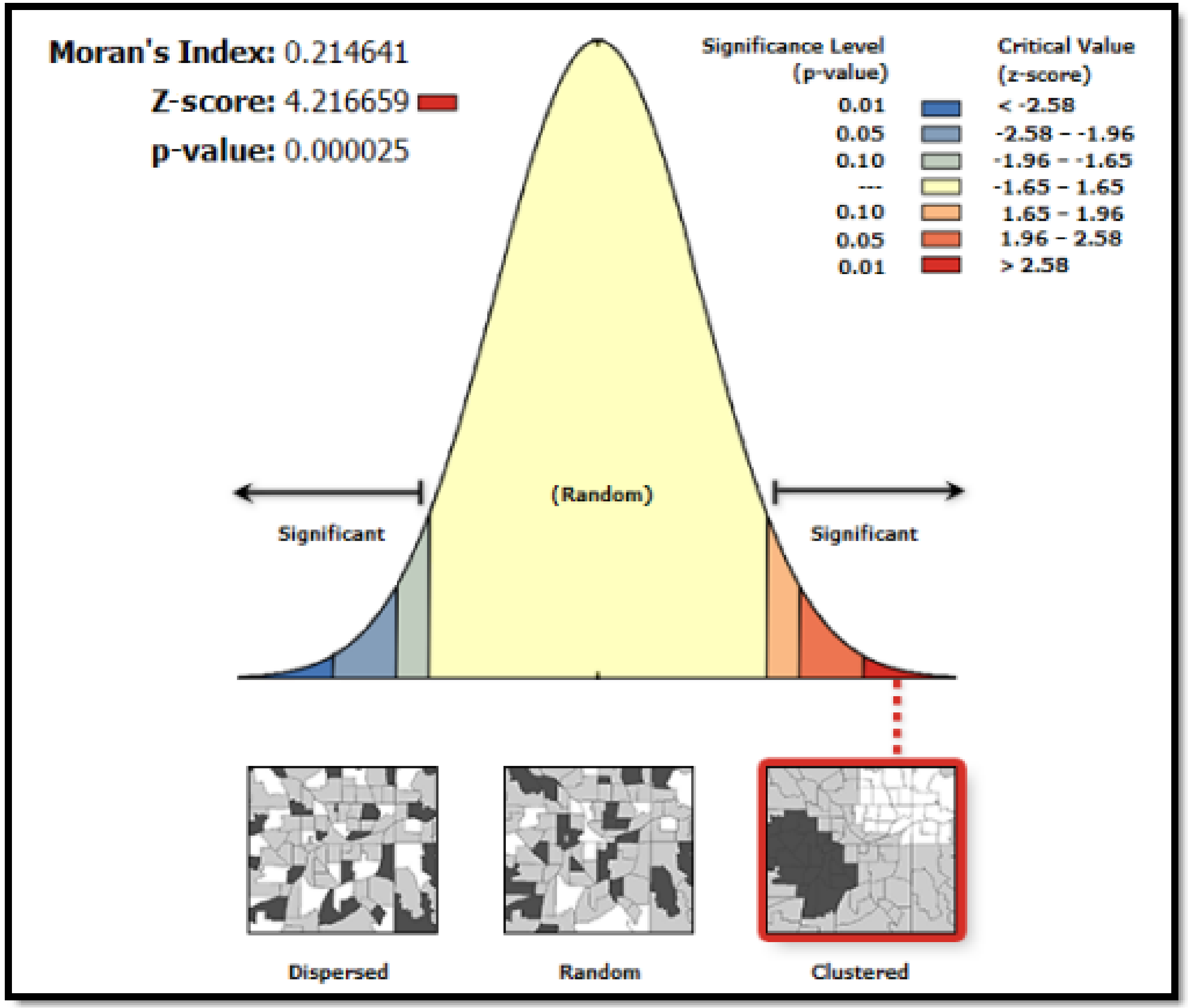
- You cannot reject the null hypothesis.
- It is quite possible that the spatial distribution of feature values is the result of random spatial processes.
- The observed spatial pattern of feature values could very well be one of many, many possible versions of complete spatial randomness (CSR).

The p-value is statistically significant, and the z-score is positive.

- You may reject the null hypothesis.
- The spatial distribution of high values and/or low values in the dataset is more spatially clustered than would be expected if underlying spatial processes were random.

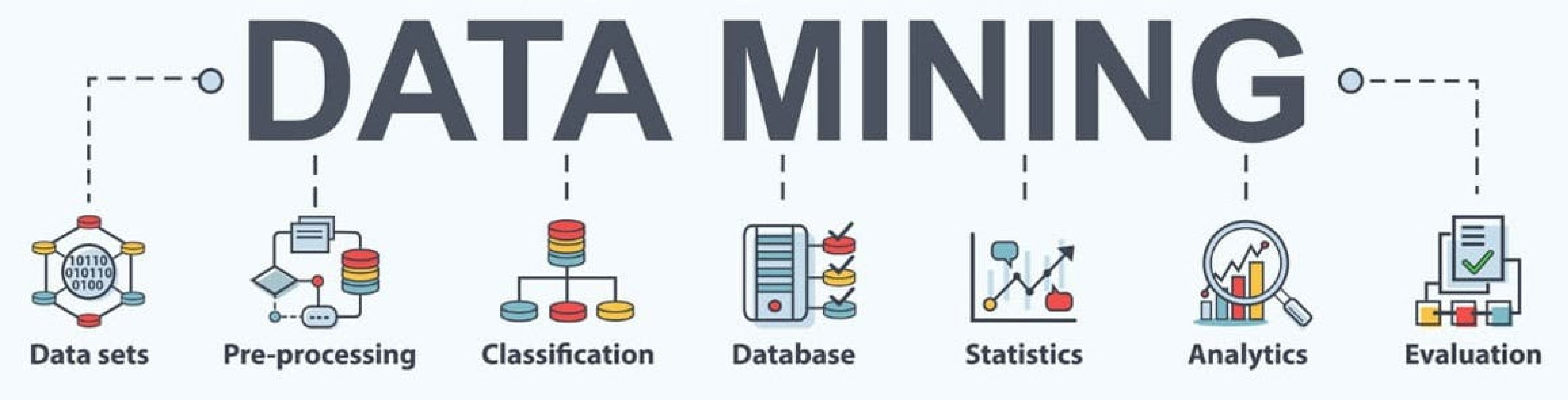
The p-value is statistically significant, and the z-score is negative.

- You may reject the null hypothesis.
- The spatial distribution of high values and low values in the dataset is more spatially dispersed than would be expected if underlying spatial processes were random.

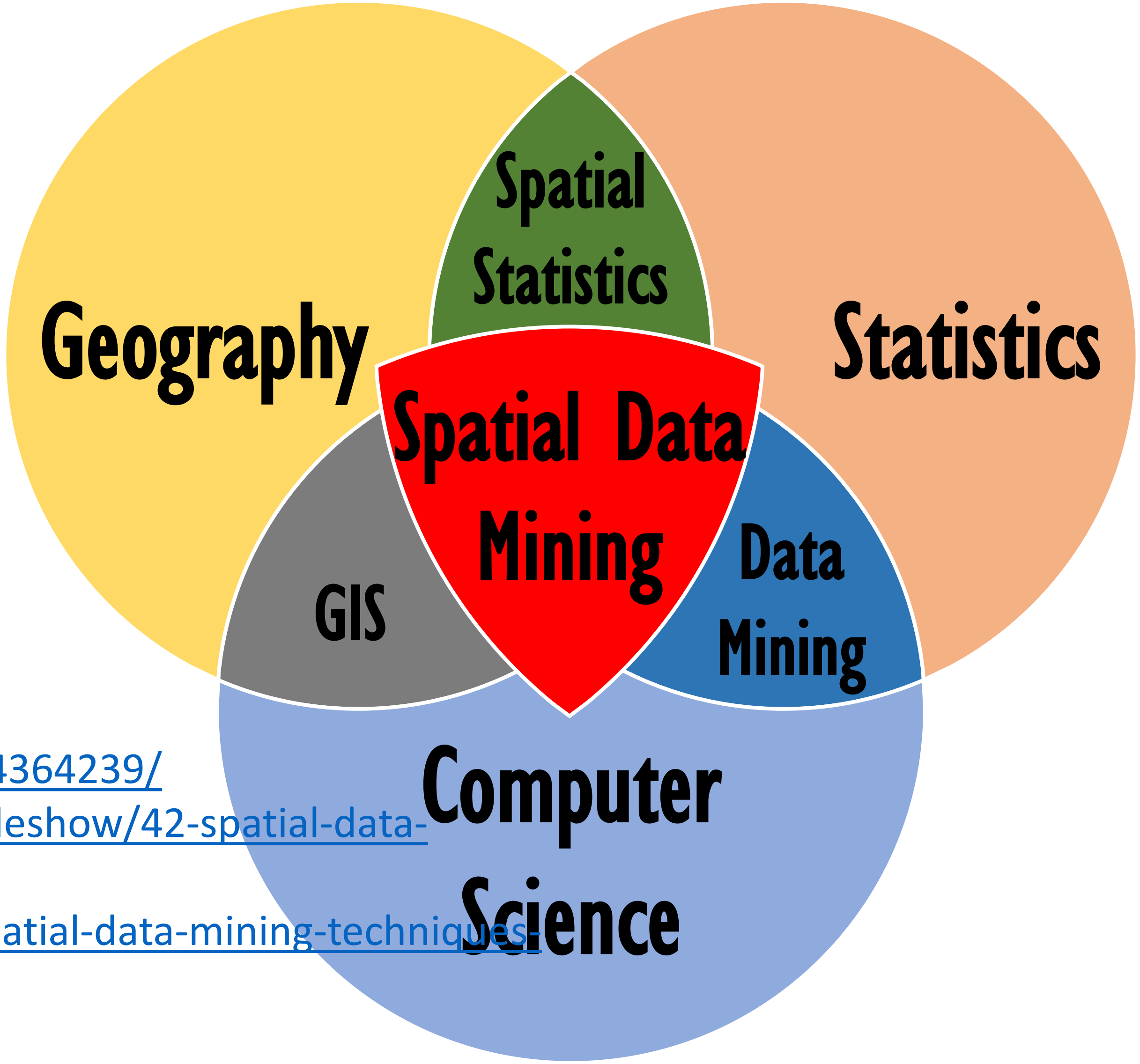

Moran's Index: 0.214641
Z-score: 4.216659
p-value: 0.000025
Significance Level (p-value)
Critical Value (z-score)
0.01 < -2.58
0.05 -2.58 - -1.96
0.10 -1.96 - -1.65
--- -1.65 - 1.65
0.10 1.65 - 1.96
0.05 1.96 - 2.58
0.01 > 2.58
(Random)
Significant
Significant
Dispersed
Random
Clustered

Data Mining

- Data mining is the process of sorting, usually on a large data to:
  - identify patterns and relationships that can help solve business problems through data analysis.



Spatial Data Mining



https://slideplayer.com/slide/14364239/
https://www.slideshare.net/slideshow/42-spatial-data-mining/47853756
https://dynamicdatalabs.net/spatial-data-mining-techniques-in-under-5-minutes