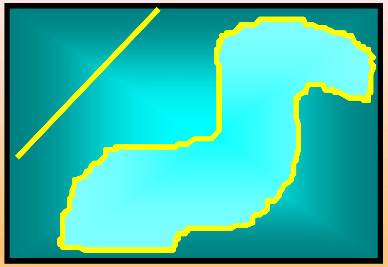
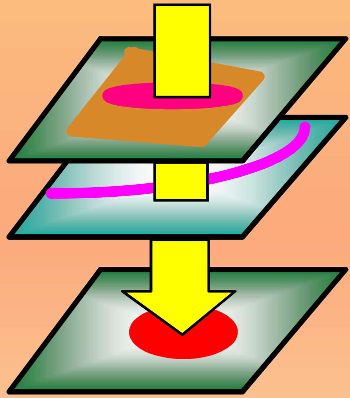
PATTERN ANALYSIS

Prof. Seema Jalan

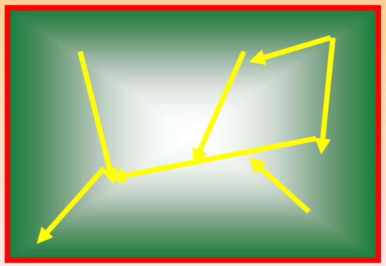
GIS As a Scientific Technology



Patterns



Relationships



Processes

Components of spatial analysis

Spatial analysis is more than asking questions ?????

WHERE?

↓

WHAT?

↓

WHY?

↓

WHAT IF?

- Spatial data analysis
  Finding interesting patterns
- Visualization
  Showing the patterns
- Spatial data modelling
  Explaining the patterns and predicting

?

- What is the state or condition of a particular geographic space
- What are the probable causes of the state

  (cause & effect relationships)
- Possible interrelationships among spatial features or phenomenon
- What the state will possibly be like in future?

PATTERN ANALYSIS

Analyzing spatial distribution of phenomenon

" Everything Is Related To Everything Else, But Near Things Are More Related Than Distant Things"

- Tobler's First Law Of Geography (Tobler, 1970)

PATTERN ANALYSIS…….

Pattern analysis refers to the use of quantitative methods for describing and analysing <u>the distribution pattern of spatial features</u>

GENERAL

Reveals if a distribution pattern is random, dispersed or clustered

Focus on whether or not the features form a pattern in the study area..identification & quantification

LOCAL

Detects the presence of local clusters of high or low values

Focus on individual features and their relationship with nearby features

PATTERN ANALYSIS…….

Knowing a pattern is useful if you need to…

- *Better understand a geographic phenomenon*
- *Monitor conditions on the ground*
- *Compare patterns of different phenomenon, features or areas*
- *Track changes*

PATTERN ANALYSIS…….

Two ways of identifying patterns:

1. *Display features or values on the map*

2. *Use statistics to measure the extent to which features or values are clustered, dispersed or random – allows comparison across space or time*

   Use of statistics to measure patterns is more accurate than mere looking at a map, because…

   On a thematic map the number of classes, the class ranges, classification method…all affect whether there appears a pattern or not

PATTERN ANALYSIS…….

Two ways of identifying patterns:

1. *Display features or values on the map*

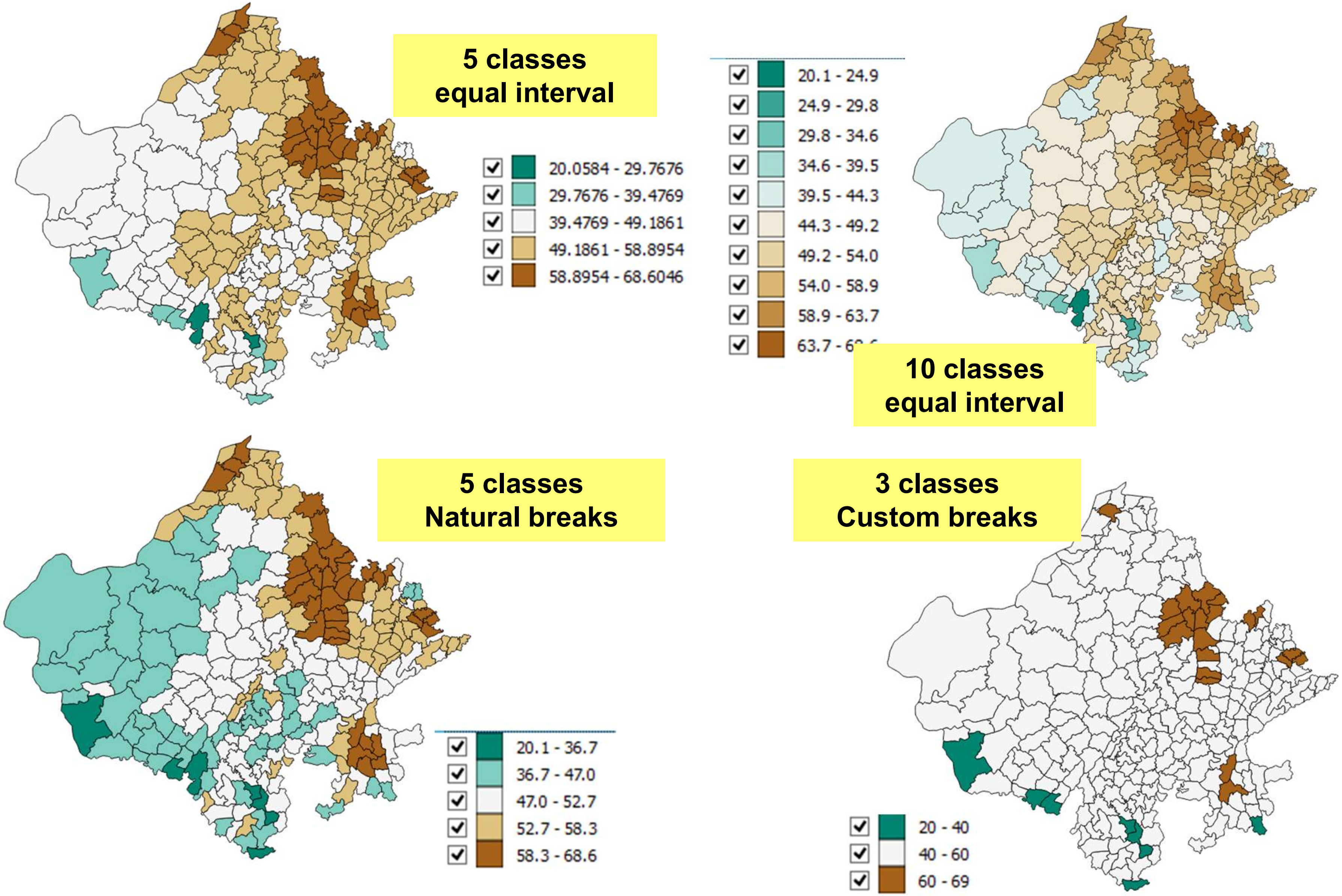
2. *Use statistics to measure the extent to which features or values are clustered, dispersed or random – allows comparison across space or time*

   Use of statistics to measure patterns is more accurate than mere looking at a map, because…

   On a thematic map the number of classes, the class ranges, classification method…all affect whether there appears a pattern or not

IMPACT OF CLASS DEFINITIONS ON SPATIAL PATTERNS IN THEMATIC MAP

RURAL LITERACY IN RAJASTHAN -2011



STATISTICAL TOOLS FOR PATTERN ANALYSIS…….

Global statistical tools fall into two categories

Feature Based

Measures the pattern formed by discrete features – points, lines and non-contiguous polygons

Feature location based

1. Quadrat analysis
2. Nearest neighbor index
3. K- function

Attribute based

Used for analyzing contiguous areas such as census tracts.

Measure the pattern formed by attribute values associated with features.

1. Joint count statistic (Categorical data)
2. Geary's c (Continuous data)
3. Moran's I (Continuous data)

STATISTICAL PATTERN ANALYSIS……. THE PROCESS

- *Compute relevant statistical parameter for observed distribution – index*
- *Compare actual distribution to a hypothetical random distribution of the same number of features over the same area*
- *Analyse the extent to which observed distribution deviates from the random distribution.This indicates the extent to which the pattern is more clustered or more dispersed than the random distribution*
- *Use statistical inference to confirm the significance of results and strength of the pattern (clustering or dispersal)*
- *Calculate the probability that pattern is not simply due to chance*

TESTING STATISTICAL SIGNIFICANCE

Am I certain that my conclusions about the pattern or relationship are correct????

1. ***Significance tests give us the probability that what a statistic is telling you is true***

2. ***Probability is a measure of chance..what is the role of chance on the outcome of the analysis?***

HYPOTHESIS TESTING

- What do you understand by " Hypothesis" ?

THE OBSERVATIONS OR BELIEFS ARE STATED AS A HYPOTHESIS

A proposition whose truth and falsity is capable of being tested

- **Hypothesis testing is a fundamental way in which inferences about a population are made from a sample**
- **Assessing - do the sample characteristics provide a precise estimate of the population characteristics????**

HYPOTHESIS TESTING

- Its human to favour any patterns or relationships one sees or expects to see
- To maintain impartiality, we set out to prove the opposite – the so called NULL HYPOTHESIS
- Our initial hypothesis is called the ALTERNATIVE HYPOTHESIS
- Significance tests help us to decide **whether we should or should not reject the null hypothesis**

- **In this course there is a possibility of two kinds of errors**

Type 1 Error

Likelihood of rejecting a true hypothesis

Type 2 Error

Likelihood of accepting a false hypothesis

Likelihood of making Type 1 Error is denoted by ‘α’ and is referred to as significance level

HYPOTHESIS TESTING

- In order to decide whether to reject a Null Hypothesis, we first decide the risk we are willing to accept for being wrong..i.e. erroneously rejecting the null hypothesis
- This degree of risk..often referred to as the confidence level (or significance level) is expressed as a probability ranging from 0 to 1.
- The desired level of confidence is compared with an observed level of confidence to decide whether or not to reject the null hypothesis
- The observed level of confidence…known as the p-value, is calculated using the sample data
- Thus characteristics of the sample (size and randomness) affect the observed level of confidence

CONFIDENCE LEVELS

- **Most common confidence levels for statistical tests are 0.10, 0.05 & 0.01**
- **If study could be repeated 100 times, each time with a different sample, probabilities indicate that at 0.05 confidence level 95 out of 100 studies would yield the same result.**
- **In other words 5 out of 100 would likely to yield erroneous results due to sampling error**

TESTING STATISTICAL SIGNIFICANCE

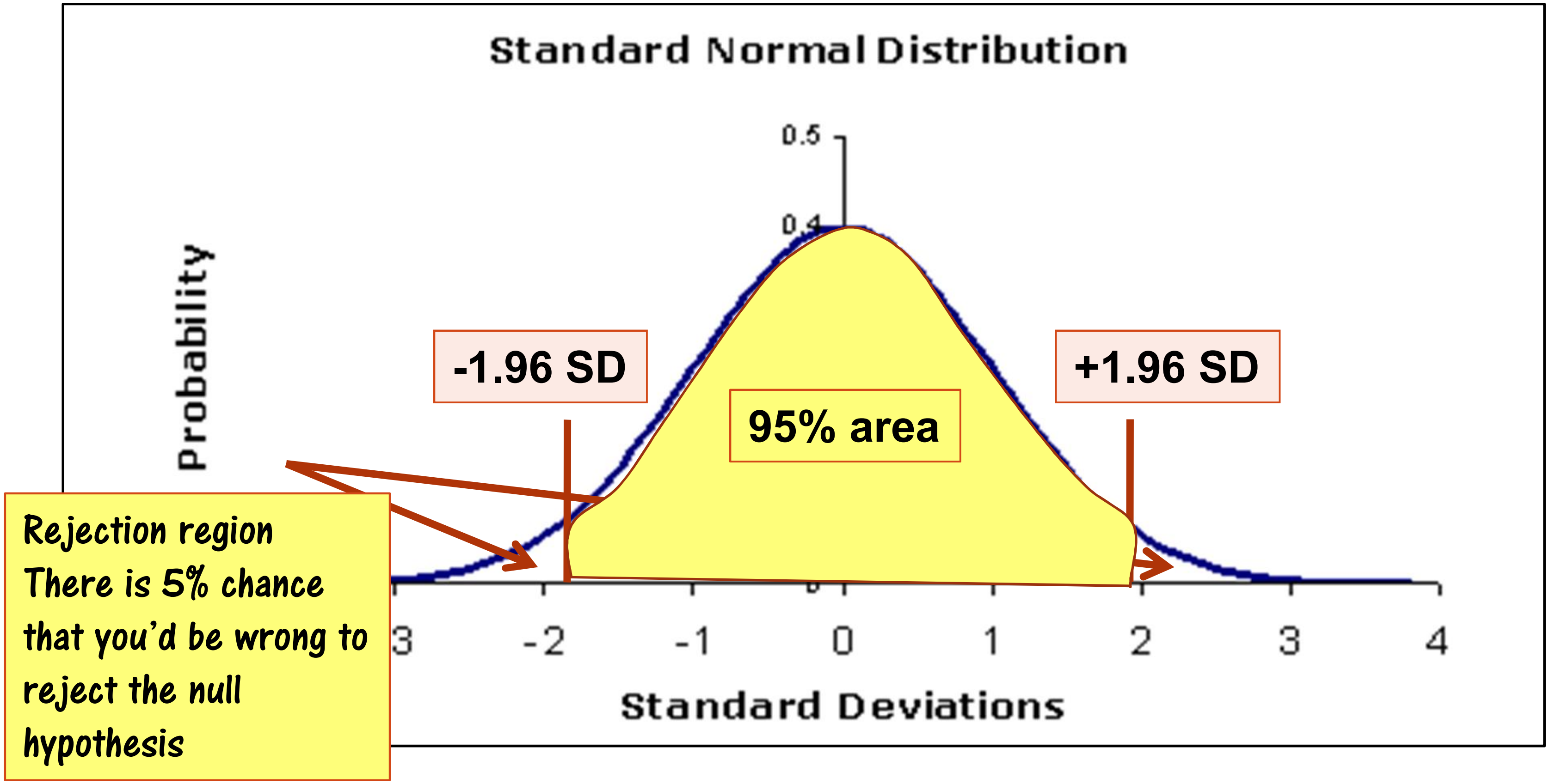
- **Each statistical tool has an appropriate significance test**
- **The test provides a statistic that represents the p-value**
- **The desired confidence level has a corresponding critical value (which depends on specific test)**
- **If the value of test statistic exceeds the critical value, you reject the null hypothesis**
- **THE RESULTS OF YOUR ANALYSIS ARE SAID TO BE STATISTICALLY SIGNIFICANT AT THE SPECIFIED CONFIDENCE LEVEL**

TESTING STATISTICAL SIGNIFICANCE

- **Most spatial statistics tools calculate a test statistics at the same time they calculate the initial statistic and report both.**
- **Many of the tools calculate a Z-score**

Confidence level	Z-score critical values	Area under the curve
0.01	± 2.58	99 %
0.05	± 1.96	95 %
0.10	± 1.65	68 %
0.20	± 1.28	

The critical value for the Z-score at the confidence level of 0.05 is 1.96. If z-score is within the range -1.96 and +1.96, the null hypothesis cannot be rejected. If it falls outside the range, you can reject the null hypothesis



Very high or a very low Z scores are found in the tails of the normal distribution. From the graph above, it is evident that the probabilities in the tails of the distribution are very low. When you perform a feature pattern analysis and **it yields either a very high or a very low Z Score, this indicates it is very UNLIKELY that the observed pattern is some version of the theoretical spatial pattern represented by your null hypothesis**.

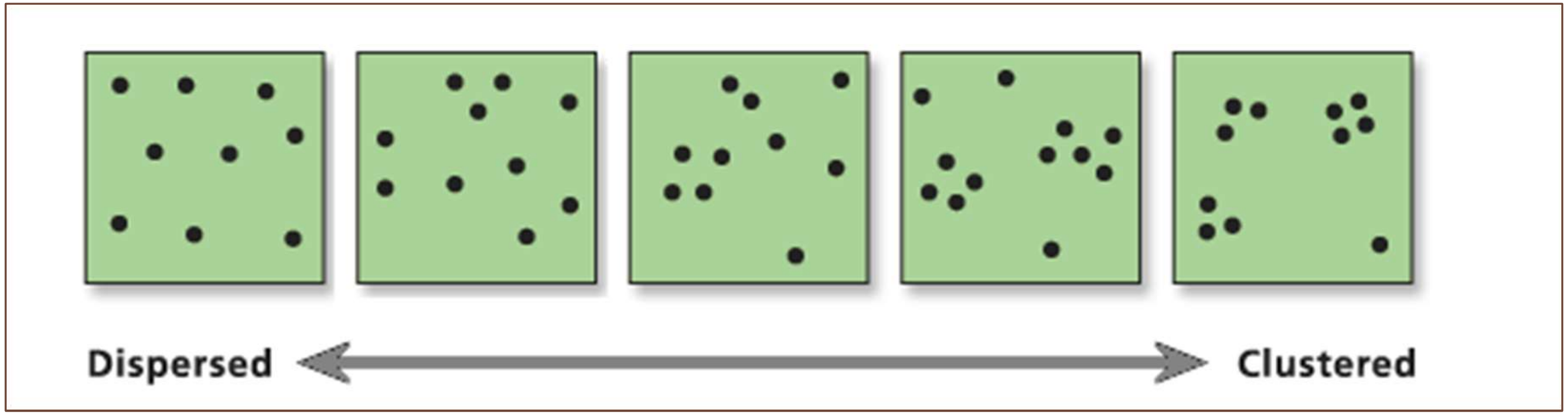
Using significance tests with spatial data

- **Each statistical tool has an appropriate significance test**
- **The test provides a statistic that represents the p-value**
- **The desired confidence level has a corresponding critical value (which depends on specific test)**
- **If the value of test statistic exceeds the critical value, you reject the null hypothesis**
- **THE RESULTS OF YOUR ANALYSIS ARE SAID TO BE STATISTICALLY SIGNIFICANT AT THE SPECIFIED CONFIDENCE LEVEL**

NEAREST NEIGHBOUR ANALYSIS – POINT PATTERN ANALYSIS

- **Spatial patterns are indicative of causal processes which may be important indicators of physical and socio-economic processes at work.**

- **Spatial patterns may be *Regular* or *Irregular.* Each type may be further divided as *Clustered, Random and Anticlustered (Dispersed).***



THE MAIN CONCERN IS WITH THE LOCATIONAL CHARACTERISTICS OF THE POINTS RATHER THAN THEIR ATTRIBUTES

Nearest Neighbor Analysis.....

- Makes use of the distance measurement between each point and its closest neighboring point in a layer in determining if the point pattern is random, regular or clustered.

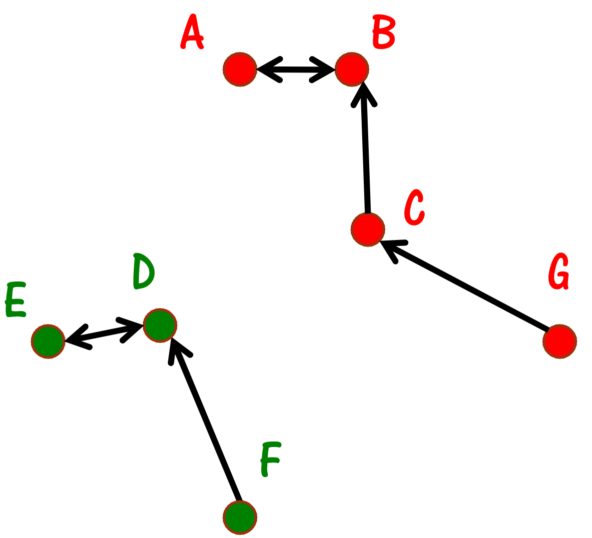
THE MECHANISM

- The distance between each feature centroid and its nearest neighbor's centroid location is measured.
- All nearest neighbor distances are averaged- **Observed distance** -**d_{obs}**
- Expected mean distance is calculated which is the mean distance for a hypothetical random distribution (with the same number of features covering the same total area) – **Expected distance** – **d_{exp}**
- The nearest neighbor statistic is the **ratio (R) of d_{obs} and d_{exp}**

$$\mathbf{R = d_{obs} / d_{exp}}$$

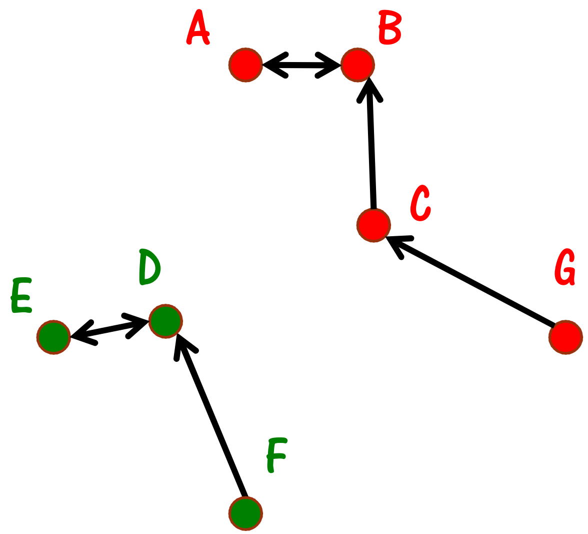
Calculation of Observed Mean Distance

	A	B	C	D	E	F	G
A		998	2117	2494	3538	3858	4267
B	988		1725	3348	4308	4004	3601
C	2117	1725		2804	4034	2567	2309
D	2494	3348	2804		1196	2510	4897
E	3538	4308	4034	1196		3277	6034
F	3858	4004	2567	2510	3277		3433
G	4267	3601	2309	4897	6034	3433	



Calculation of Observed Mean Distance

- GIS calculates the distance from each feature to all other features in the set
- Then it finds the shortest distance – nearest neighbor
- Adds the distances between each pair of nearest neighbor
- Divides by number of features in the set to get the mean distance



Observed mean distance

$$d_o = \frac{\sum_i c_i}{n}$$

Measure the distance to each feature's nearest neighbor, and sum the distances…

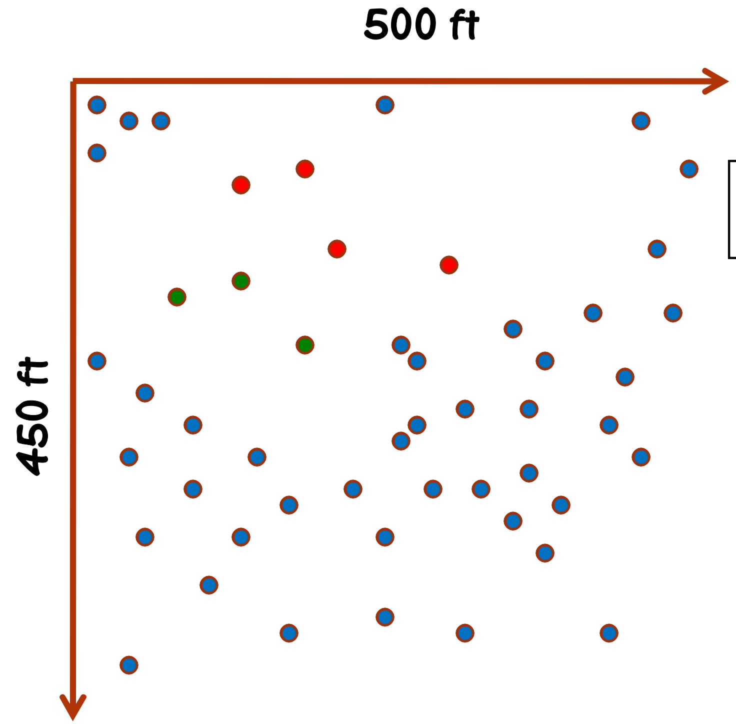
…then divide by total number of features

Calculation of Expected Mean Distance

Expected mean distance
For a random distribution

$$d_e = \frac{0.5}{\sqrt{\frac{n}{A}}}$$

… divide 0.5 by square root of total number of features divided by the area



N = 50

$$d_e = \frac{0.5}{\sqrt{\frac{50}{225000}}}$$

$= 33.56$ ft

Calculation of INDEX

- ✓ GIS subtracts the expected mean distance from the observed mean distance

$$d = do - de$$

- ✓ If the observed and expected means are equal, the difference is zero

- ✓ If expected mean is greater than the observed mean, **difference will be negative, observed distribution is clustered**

- ✓ If expected mean is less than the observed mean, the difference will be greater than zero **(positive), then the observed distribution is dispersed**

Nearest Neighbor Analysis..... INTERPRETATION

✓ Alternatively, ratio between the two mean distances is calculated

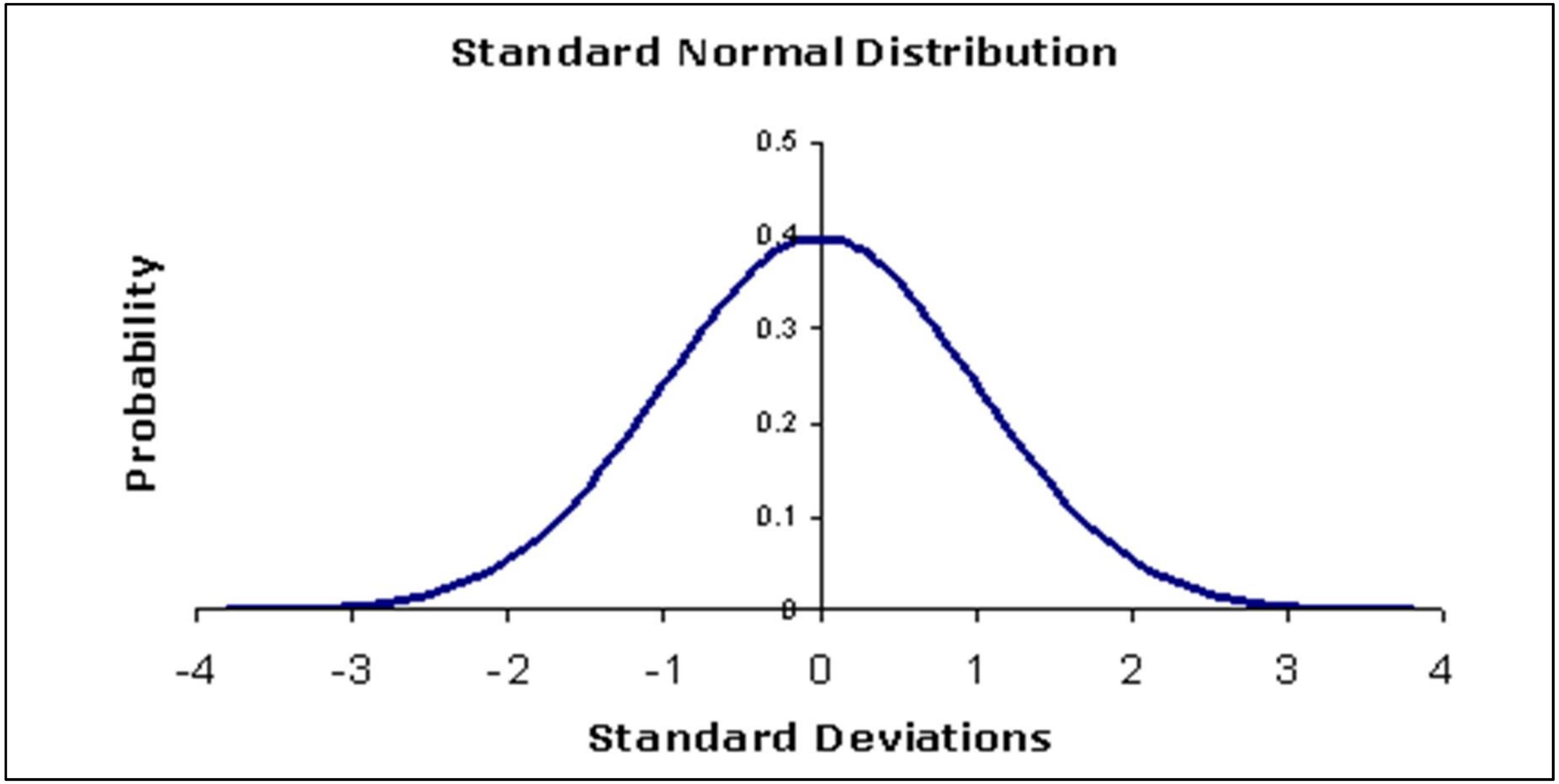
$$r = do / de$$

- If the means are the same, the ratio is 1 and observed distribution is **RANDOM**
- If expected mean is greater than observed mean, $R < 1$, **pattern exhibits clustering. The closer the index to zero, the more clustered the pattern**
- If $R > 1$, i.e., expected mean is less than the observed mean, **pattern indicates dispersion**

DIFFERENCE	RATIO	PATTERN
d<0	r<1	Clustered
0	1	Random
d>0	r>1	Dispersed

Nearest Neighbor Analysis….. INTERPRETATION

- The analysis also produces a Z-Score which indicates the likelihood that the pattern could be the result of a random chance
- Null hypothesis states that there is no pattern…. points are randomly distributed
- The critical Z score values when using a 95% confidence level are -1.96 and +1.96 standard deviations.
- If your Z score is between -1.96 and +1.96 you cannot reject your null hypothesis; the pattern exhibited is a pattern that could very likely be one version of a random pattern.
- If the Z score falls outside that range(for example -2.5 or +5.4), it is possible to reject the null hypothesis and proceed with figuring out what might be causing either the statistically significant clustered or statistically significant dispersed pattern.

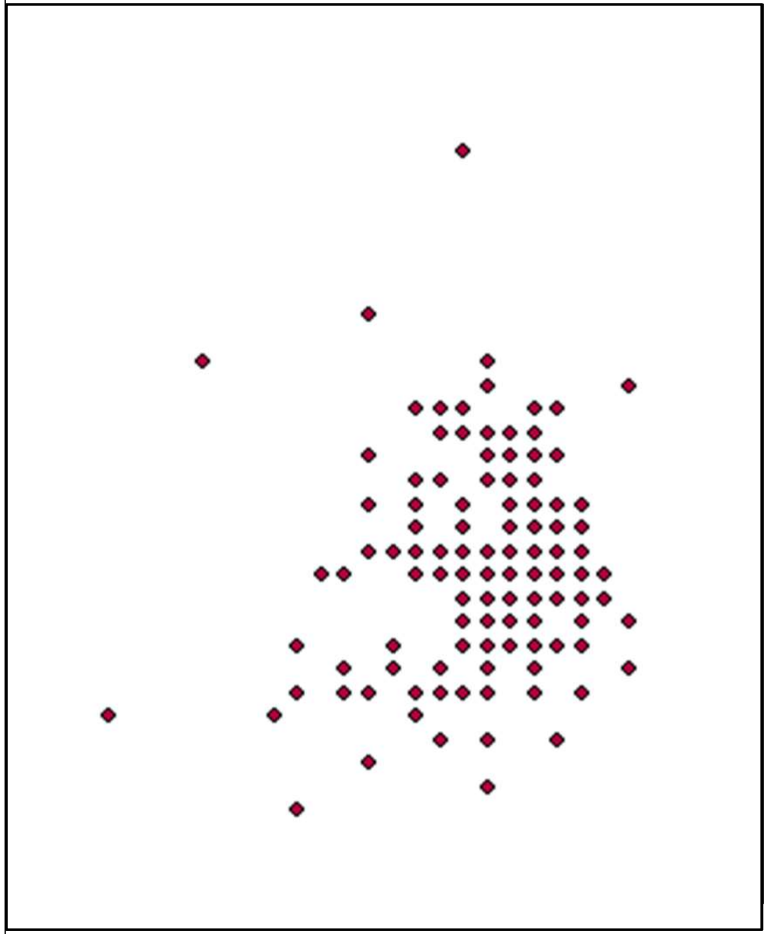


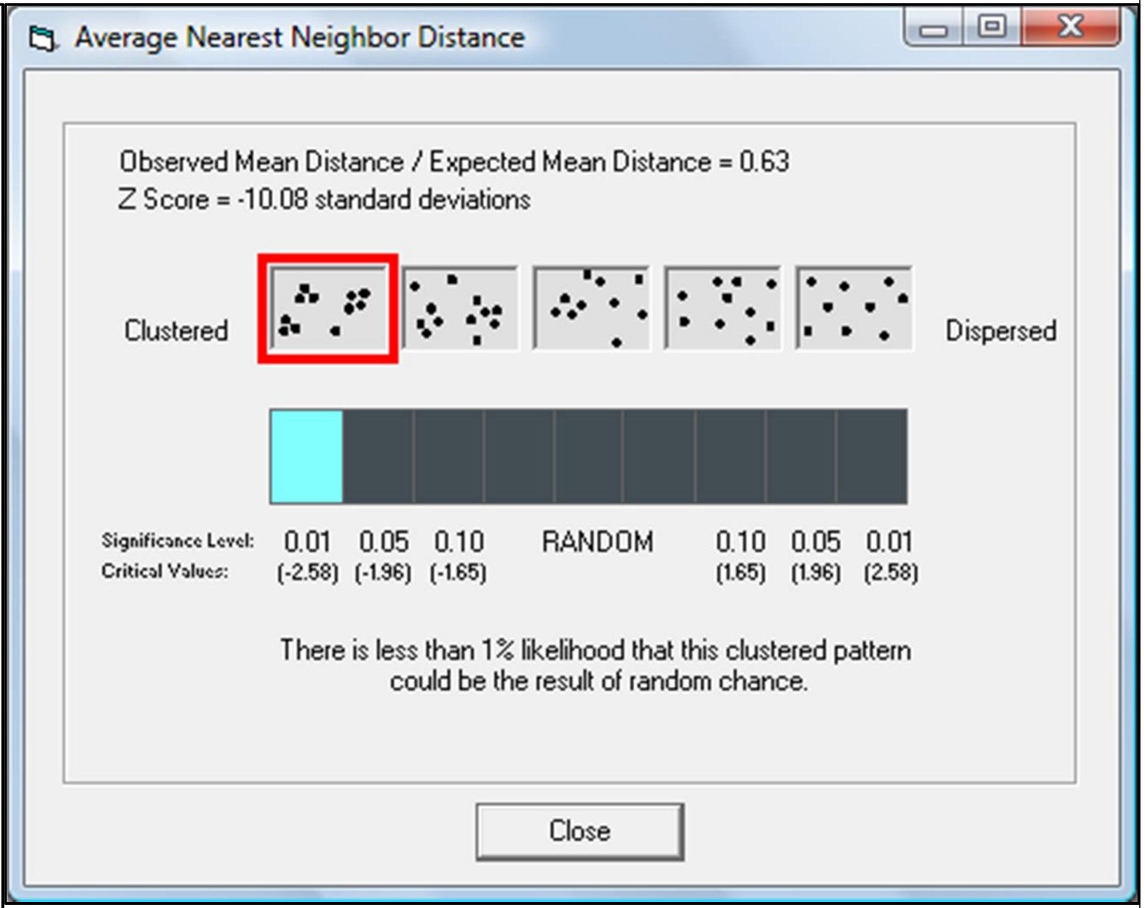
Very high or a very low Z scores are found in the tails of the normal distribution. From the graph above, it is evident that the probabilities in the tails of the distribution are very low. When you perform a feature pattern analysis and **it yields either a very high or a very low Z Score, this indicates it is very UNLIKELY that the observed pattern is some version of the theoretical spatial pattern represented by your null hypothesis**.

Nearest Neighbor Analysis..... CONSIDERATIONS

- Although this tool will work with polygon or line data, it is really only appropriate for event, incident, or other fixed-point feature data. For line and polygon features, feature centroids are used in the computations.
- The equations used to calculate the Average Nearest Neighbor Distance Index and Z score are based on the assumption that the points being measured are free to locate anywhere within the study area (for example, there are no barriers, and all cases or features are located independently of one another).
- The index and Z score for this statistic are sensitive to changes in the study area
- The nearest neighbor function is very sensitive to the area value (small changes in the area can result in considerable changes in the results).
- If an area value is not specified, then the area of the minimum enclosing rectangle around the features is used.
- The units of the area parameter are the input feature class' coordinate system's units squared.
- Distance calculations are based on either Euclidean or Manhattan distance and require projected data to accurately measure distances

Example: Nearest Neighbor Analysis of Deer locations





Nearest Neighbor Analysis..... APPLICATIONS

- Analysis of settlement patterns
- Evaluate competition or territory
- Quantify and compare patterns in distributions for a variety of plant or animal species

This statistic is most appropriate when the study area is fixed: comparing average nearest neighbor distances for different types of retail stores within a particular county or comparing a single type of retail for a fixed study area over time.

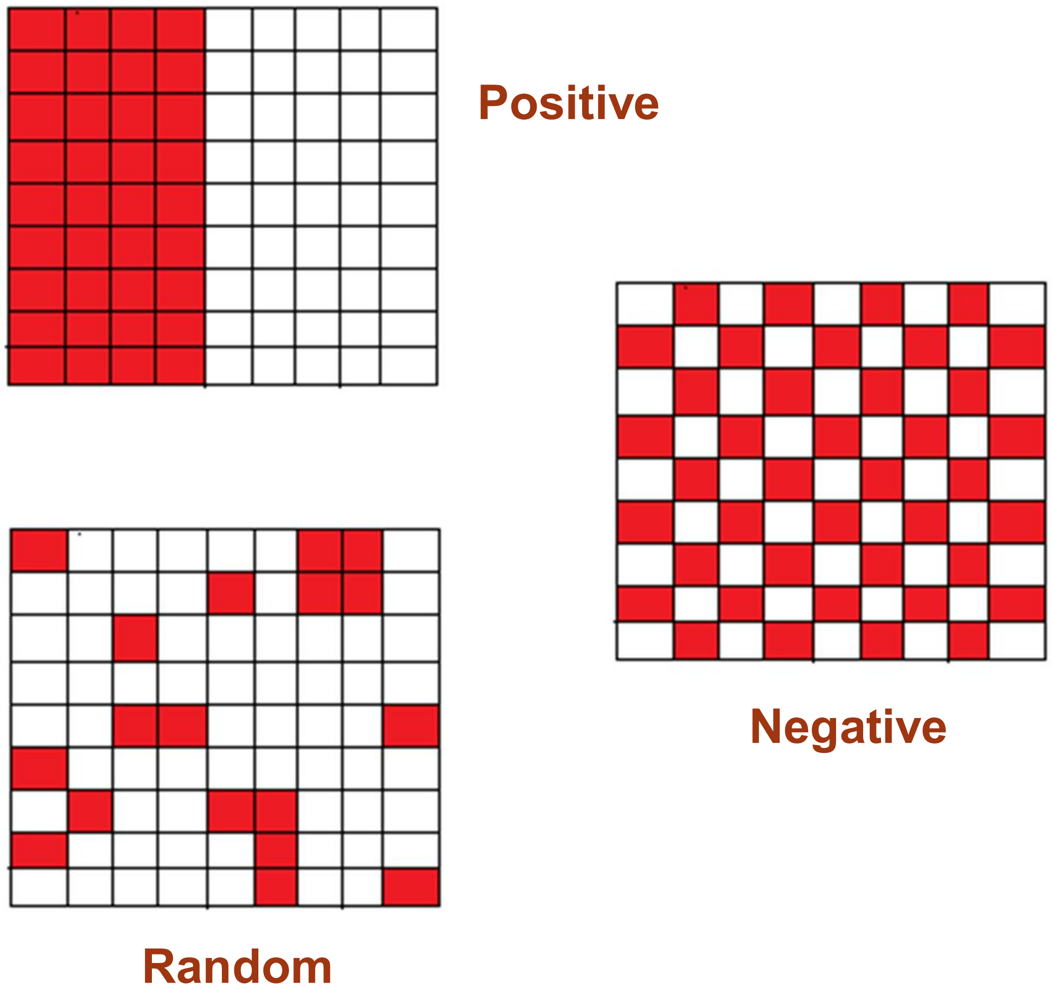
SPATIAL AUTOCORRELATION

SPATIAL AUTOCORRELATION refers to the fact that data from locations near one another in space are more likely to be similar than data from locations remote from one another

- Deals with **both** the ***attributes and location*** of spatial features
- Measures the relationship among spatial objects and their neighbors (Cliff and Ord, 1973)
- Indicates whether adjacent or neighboring values in geo-spatial data vary together? If so. HOW???
- The derived statistic makes it possible to measure interdependence in a spatial distribution and to use formal statistical methods to test hypothesis about spatial interdependence

Same spatial pattern at different scales may produce different spatial auto-correlation results

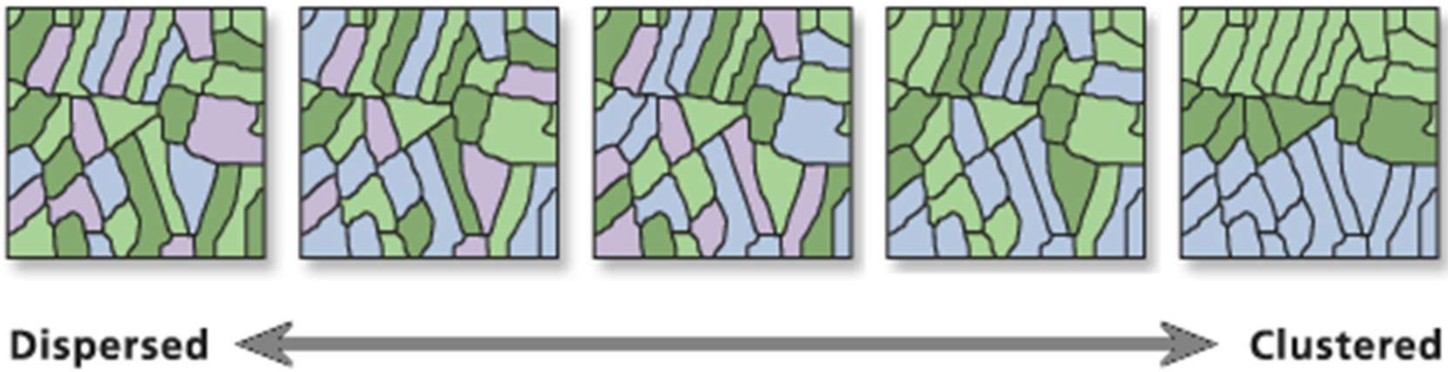
TYPES OF SPATIAL AUTOCORRELATION



<u>POSITIVE</u>: When spatial objects with similar values vary/ cluster together

<u>RANDOM</u>: No pattern of clustering

<u>NEGATIVE</u>: Even distribution over a large geo-spatial space



Relationship is described as

- highly correlated (Clustered) when like values are spatially close to each other
- Random or independent (Dispersed) when no pattern can be discerned from the arrangement of values

METHODS OF MEASURING SPATIAL AUTOCORRELATION

- JOINT COUNT STATISTICS –NOMINAL DATA
- GEARY’S INDEX (c) – GEARY (1968)
- MORAN’S INDEX (I) – MORAN (1948)

FINDING PATTERNS FOR FEATURES HAVING CONTINUOUS VALUES

- Geary's *c*
- Moran's *I*
- *Getis Ord General G*

Use the magnitude of feature values to identify and measure the strength of spatial patterns

Measuring similarity of nearby features (Ratio/ Interval data)

- **Geary's contiguity ratio or Geary's *c*** Developed by Economist and Statistician Robert Geary in early 1950s
- **Moran's Index or Moran's *I*** developed by Australian Statistician Patrick Moran in late 1940s

- The methods are used for features having interval or ratio values
- The analysis compares the attribute values between neighbouring features to the distribution of values for the dataset as a whole.
- Geary's C uses difference in values between any two neighbouring features
- Moran's I compares the value of each feature in a pair to the mean value of each dataset
- For both methods, if the difference in values of nearby features is less than the difference in values among all features…like values are clustered

THE METHODS ONLY INDICATE WHETHER SIMILAR VALUES OCCUR TOGETHER

NOT THAT THESE CLUSTERS ARE COMPOSED OF HIGH OR LOW VALUES

GEARY'S INDEX…How does GIS do it?

- Calculates the difference in values between the target feature and each of its neighbours
- Neighbourhood is defined based on adjacency, a set distance, or the distance of all features in the dataset
- GIS starts with one feature and subtracts the value of a neighbouring feature from the value of the original, or target, feature
- GIS only wants to find out how large the difference is.. Not whether it is positive or negative..hence it squares the difference to make sure its positive
- It then multiplies the difference by the weight value (as per neighbourhood we defined)
- The process is repeated for the target feature with all other features in the study area

GEARY'S INDEX...How does GIS do it?

- It then moves to the next feature and does the same thing
- The results are summed as it goes on
- It then sums the weights for each pair of features, multiplies by 2, and multiplies the variance of distribution by it
- Finally it divides this value into the initial value it calculated (sum of weighted difference in values) to get the C-ratio.

What have we done?

We have compared the sum of differences in attribute values between any two features within the neighbourhood to the sum of the differences for the dataset as a whole

GEARY'S INDEX…

- **Developed as a measure of spatial autocorrelation for area objects with interval attributes**
- **Suitable measure for use in the analysis of data aggregated by statistical reporting zones (e.g. census tracts)**

$$c = \frac{\sum_i \sum_j w_{ij} c_{ij}}{2 \sum_i \sum_j wij \sigma^2}$$

where $c_{ij} = (z_i - z_j)^2$ (A measure of similarity in attributes)

z_i is the value of attribute of interest for object *i*

z_j is the value of attribute of interest for object *j*

w_{ij} (A measure of locational similarity/ spatial proximity/ degree of adjacency of locations of ***i*** & ***j***)

w_{ij} = 1 if i & j share a common boundary

w_{ij} = 0 otherwise

$\sigma^2 =$ Variance of the attribute z values

$$\sigma^2 = \frac{\sum_i (z_i - \bar{z})^2}{(n-1)}$$

where

$$\bar{z} = \frac{\sum_i^n z_i}{n}$$

- Value of **c** will be largest when large values of ***wij*** coincide with large values of ***cij***
- ***C=1 indicates*** *no spatial autocorrelation*
- ***C<1 indicates*** *positive spatial autocorrelation (similar attributes coincide with similar locations*
- ***C>1 indicates*** *negative spatial autocorrelation, attributes & locations are dissimilar*

MORAN'S index (I)

- More logical than Geary's Index: Positive values imply positive spatial autocorrelation, Negative values implying dissimilarity and a zero value indicating uncorrelated, random arrangement of attribute values
- Given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed, or random.
- The tool calculates the Moran's I Index value and a Z score evaluating the significance of the index value.

$$I = \frac{\sum_i \sum_j w_{ij} c_{ij}}{s^2 \sum_i \sum_j w_{ij}}$$

$$s = \frac{\sum_i (z_i - z_m)^2}{n}$$

c_{ij} is measure of attribute similarity

$\boxed{w_{ij}}$ is measure of spatial proximity

$c_{ij} = (z_i - z_m)(z_j - z_m)$

$z_i =$ value of the attribute of interest for object i

$z_j =$ value of the attribute of interest for object j

$z_m =$ mean of attribute of interest

$s^2 =$ sample variance

n = number of points

W_{ij}

- **For point objects** : compute distance between pairs of points and use inverse distance weighting to compute similarity

$$w_{ij} = 1/\,d_{ij} \quad \text{OR} \quad 1/\,d_{ij}^{\,2}$$

- **For line objects**: If the lines represent links between nodes that have attributes c_{ij} will measure the attributes of each pair of nodes while w_{ij} will measure the links between them. If links carry attributes w_{ij} will measure proximity between two links (feature centroid) and c_{ij} will measure attribute similarity between the links.

- **For Area objects**: w_{ij} measures **adjacency between two locations.**

$$w_{ij} = 1 \text{ if } i \,\&\, j \text{ share a common boundary}$$

$$w_{ij} = 0 \text{ if otherwise.}$$

MORAN'S I index……INTERPRETATION

Values are anchored at the expected value E(I) for a random pattern

$$E\,(I) = -1/(N-1)$$

E(I) approaches zero when no. of values is large

Moran's I is **close to E (I)** if pattern is **random**

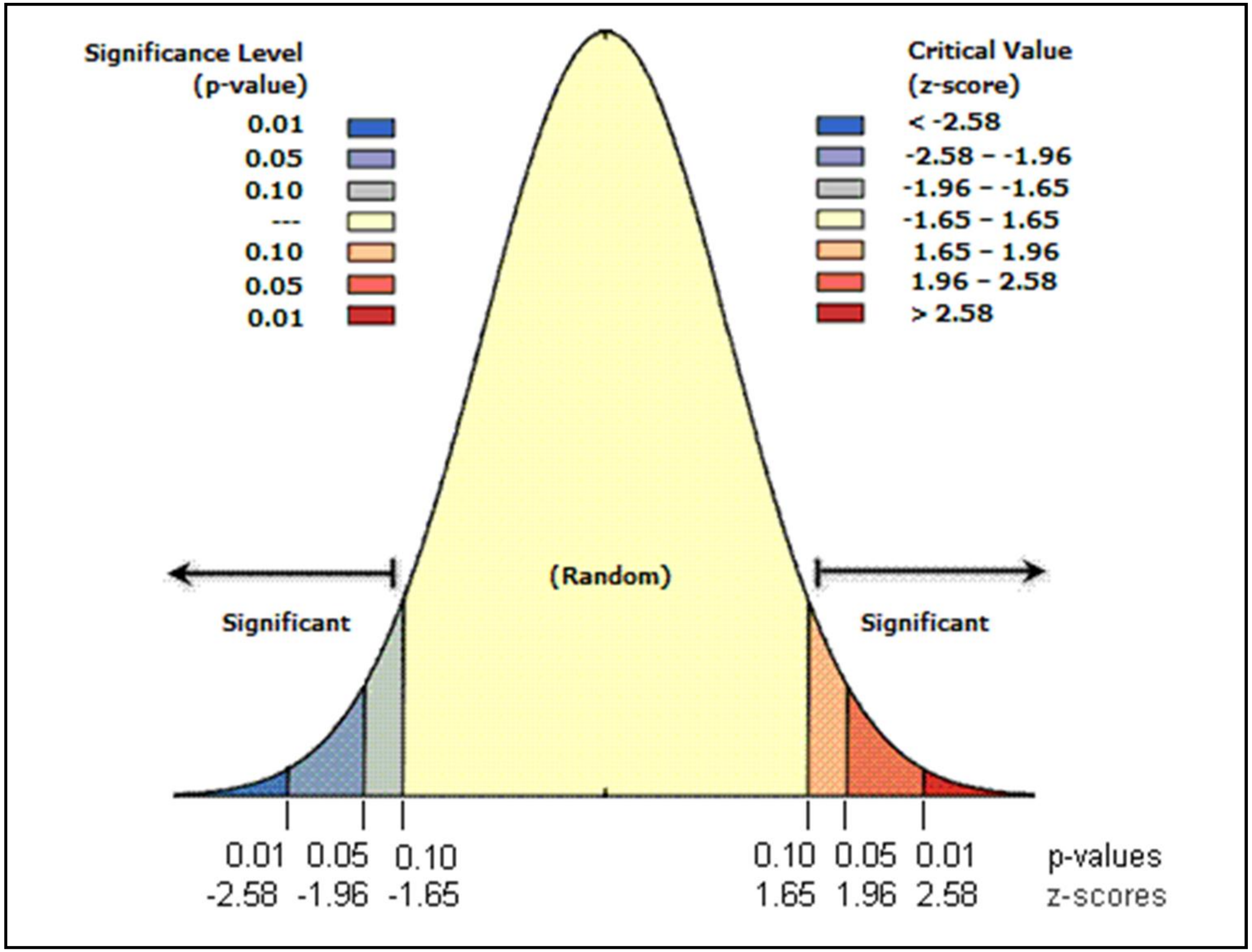
Moran's I **is greater than E (I)** if adjacent points tend to have similar values (i.e. **spatially correlated**)

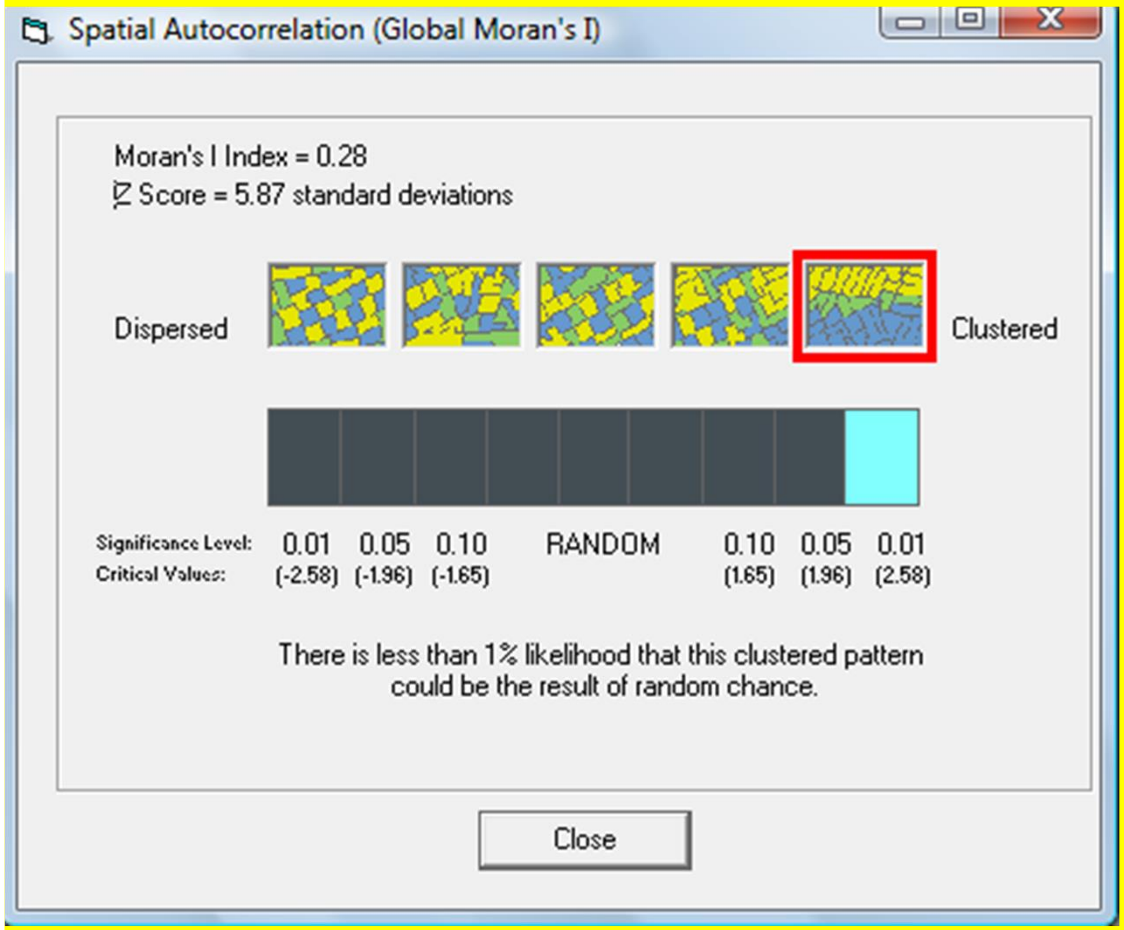
Moran's I is **less than E (I)** if adjacent points tend to have different values (i.e. **are not spatially correlated**)

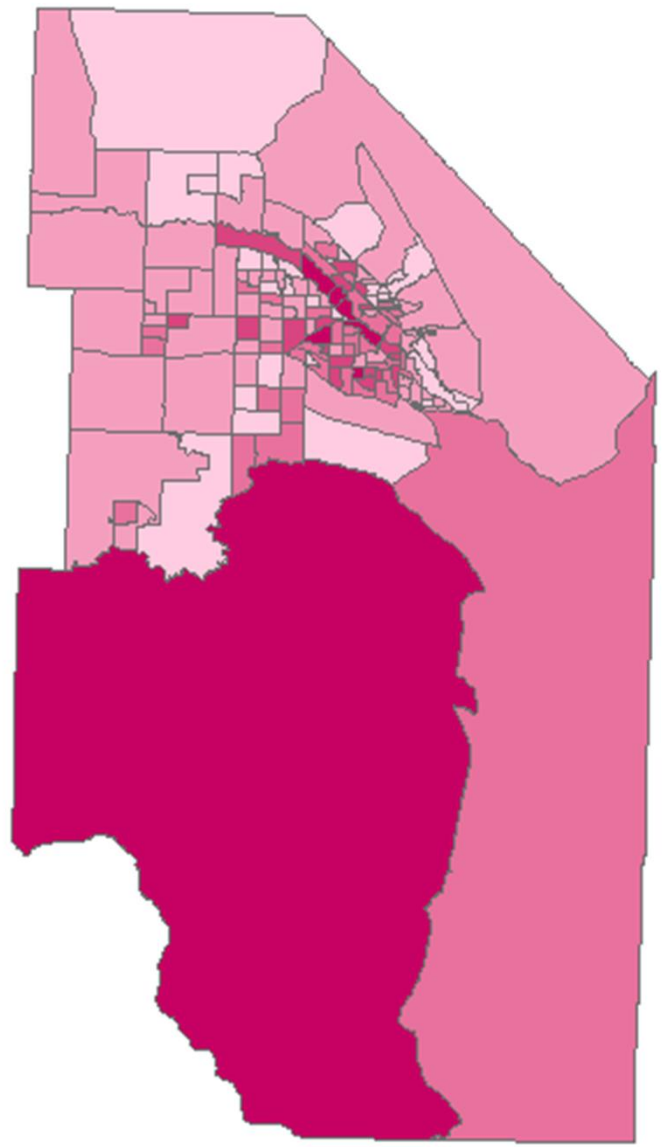
MORAN'S I index……INTERPRETATION

- A positive Moran's Index value indicates clustering
- An negative index value indicates dispersion.
- The Global Moran's I function also calculate a Z score value that indicates whether or not we can reject the null hypothesis. Z score indicates statistical significance at the specified level of confidence.
- The null hypothesis states "there is no spatial clustering of the values".
- To determine if the Z score is statistically significant, compare it to the range of values for a particular confidence level. For example, at a significance level of 0.05, a z score would have to be less than –1.96 or greater than 1.96 to be statistically significant.
- When the Z score is large (or small) enough to such that it falls outside of the desired significance, the null hypothesis can be rejected.
- When the null hypothesis is rejected, the next step is to inspect the value of the Moran's I Index. If the value is greater than 0, the set of features exhibits a clustered pattern. If the value is less than 0, the set of features exhibits a dispersed pattern.

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%




Spatial Autocorrelation (Global Moran's I)
Moran's I Index = 0.28
Z Score = 5.87 standard deviations
Dispersed
Clustered
Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01
Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)
There is less than 1% likelihood that this clustered pattern could be the result of random chance.
Close



MORAN'S I index……CONSIDERATIONS

- Spatial autocorrelation is affected by the scale of spatial pattern. Same pattern at different scales may produce different results (Modifiable Areal Unit Problem).

- The input field you select should only contain positive numeric values. Negative weights will be converted to zero for the calculations.

- The values in the input field should have at least some variation. The statistic will not compute if the values have no variation (if they are all one value).

- Calculations are based on either Euclidean or Manhattan distance and require projected data to accurately measure distances.

- The units of the "Distance Band or Threshold Distance" parameter are the units of the input feature class' coordinate system.

MORAN'S I index……APPLICATIONS

- **Determine the feasibility of using a particular statistical method (for example, linear regression analysis and many other statistical techniques require independent observations)**
- **Help identify an appropriate neighborhood distance for a variety of spatial analysis methods. For example, find the distance where spatial autocorrelation is strongest.**
- **Measure broad trends in ethnic or racial segregation over time—is segregation increasing or decreasing.**
- **Summarize the diffusion of an idea, disease or trend over space and time—is the idea, disease, or trend remaining isolated and concentrated or spreading and becoming more diffuse.**

MORAN'S index (I)

- More logical than Geary's Index: Positive values imply positive spatial autocorrelation, Negative values implying dissimilarity and a zero value indicating uncorrelated, random arrangement of attribute values
- Given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed, or random.
- The tool calculates the Moran's I Index value and a Z score evaluating the significance of the index value.

MORAN'S (I)

1. **For each pair of features** GIS subtracts the value of each feature from the mean value for all the features in the study area. Then multiplies them to get the **Cross Product (Cij).**
2. **Cij is multiplied by the weight (Wij)** for that pair, and added to the sum for all features
3. The process is repeated for all the features of the study area and results are summed

$$I = \frac{\sum_i \sum_j w_{ij} c_{ij}}{s^2 \sum_i \sum_j w_{ij}}$$

$$s = \frac{\sum_i (z_i - z_m)^2}{n}$$

c_{ij} is measure of attribute similarity

w_{ij} is measure of spatial proximity

$c_{ij} = (z_i - z_m)(z_j - z_m)$

z_i = value of the attribute of interest for object i

z_j = value of the attribute of interest for object j

z_m = mean of attribute of interest

s^2 = sample variance

n = number of points

MORAN'S (I)

4. GIS then calculates **the variance from the mean value for all the features** in the study area, sums the weights for each pair of features, and multiplies the variance by this sum.
5. The sum of weighted cross products is divided by the product of sum of weights and variance to get the **Ratio (I)**

$$I = \frac{\sum_i \sum_j w_{ij} c_{ij}}{s^2 \sum_i \sum_j w_{ij}}$$

$$s = \frac{\sum_i (z_i - z_m)^2}{n}$$

c_{ij} is measure of attribute similarity

w_{ij} is measure of spatial proximity

$c_{ij} = (z_i - z_m)(z_j - z_m)$

z_i = value of the attribute of interest for object i

z_j = value of the attribute of interest for object j

z_m = mean of attribute of interest

s^2 = sample variance

n = number of points

$$W_{ij}$$

- **For point objects** : compute distance between pairs of points and use inverse distance weighting to compute similarity

$$w_{ij} = 1/\, d_{ij} \quad \text{OR} \quad 1/\, d_{ij}{}^{2}$$

- **For line objects**: If the lines represent links between nodes that have attributes c_{ij} will measure the attributes of each pair of nodes while w_{ij} will measure the links between them. If links carry attributes w_{ij} will measure proximity between two links (feature centroid) and c_{ij} will measure attribute similarity between the links.

- **For Area objects**: w_{ij} measures **adjacency between two locations.**

$$w_{ij} = 1 \text{ if } i \,\&\, j \text{ share a common boundary}$$

$$w_{ij} = 0 \text{ if otherwise.}$$

MORAN'S I index……INTERPRETATION

Values are anchored at the expected value E(I) for a random pattern

$$E(I) = -1/(N-1)$$

E(I) approaches zero when no. of values is large

Moran's I is **close to E (I)** if pattern is **random**

Moran's I **is greater than E (I)** if adjacent points tend to have similar values (i.e. **spatially correlated**)

Moran's I is **less than E (I)** if adjacent points tend to have different values (i.e. **are not spatially correlated**)

W H A T D O E S I T M E A N ?

$$I = \frac{\sum_i \sum_j w_{ij} c_{ij}}{s^2 \sum_i \sum_j w_{ij}}$$

1. A **High C_{ij} (+VE)** indicates nearby features have **similar values (I > 0)**
2. A **Low C_{ij} (- VE)** indicates nearby features have **dissimilar values (I < 0)**

- *If both neighbouring values are higher than mean, the cross product will be positive and high*
- *If both neighbouring values are lower than mean, the cross product will be positive and high*

THUS OVERALL SUM IS POSITIVE – CLUSTERED (SIMILAR VALUES FOUND TOGETHER)

If one value in the pair is higher and the other is lower than the mean, the product is negative.

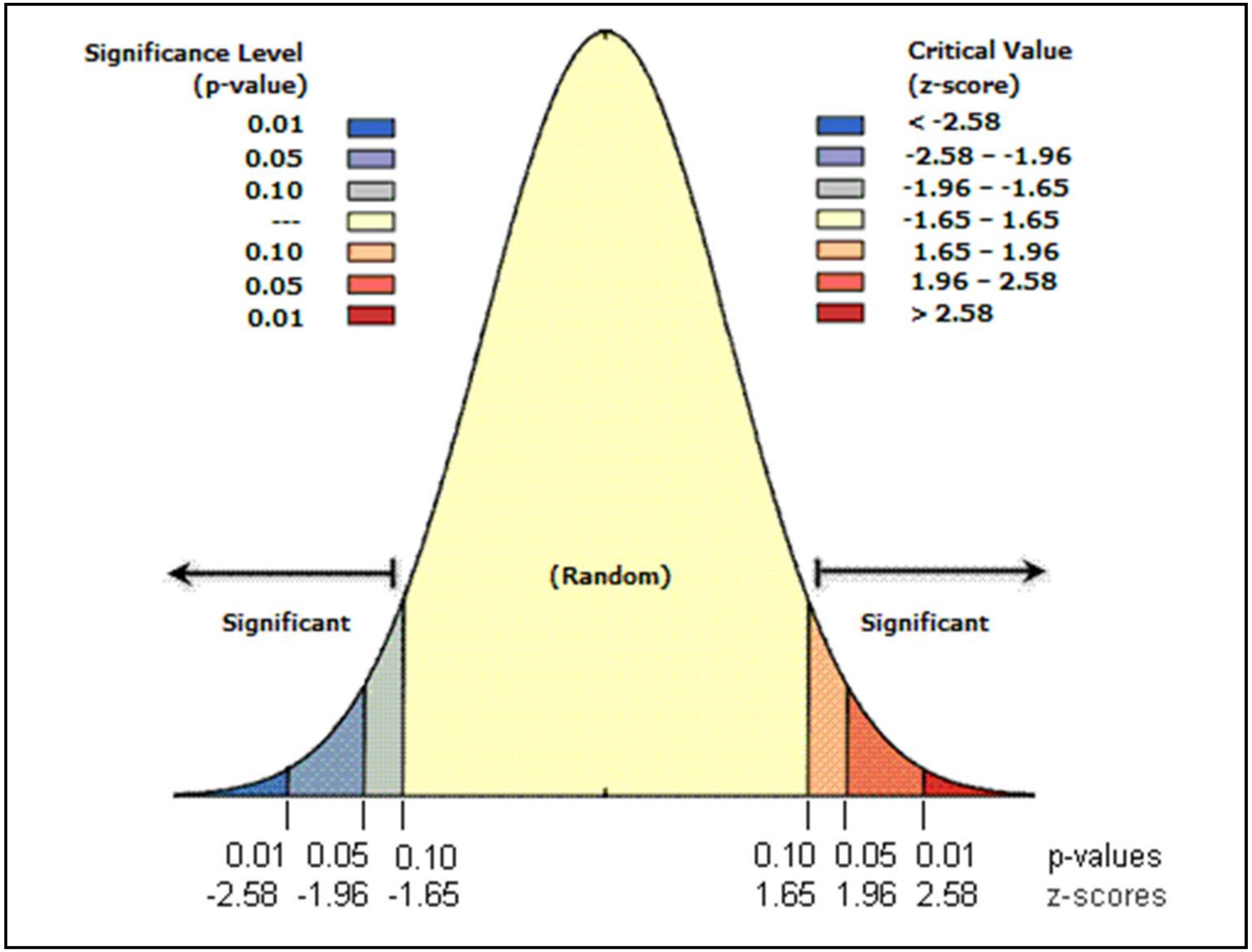
LARGER NUMBER OF DISSIMILAR PAIRS - OVERALL SUM IS NEGATIVE – DISPERSED (High and Low values are interspersed)

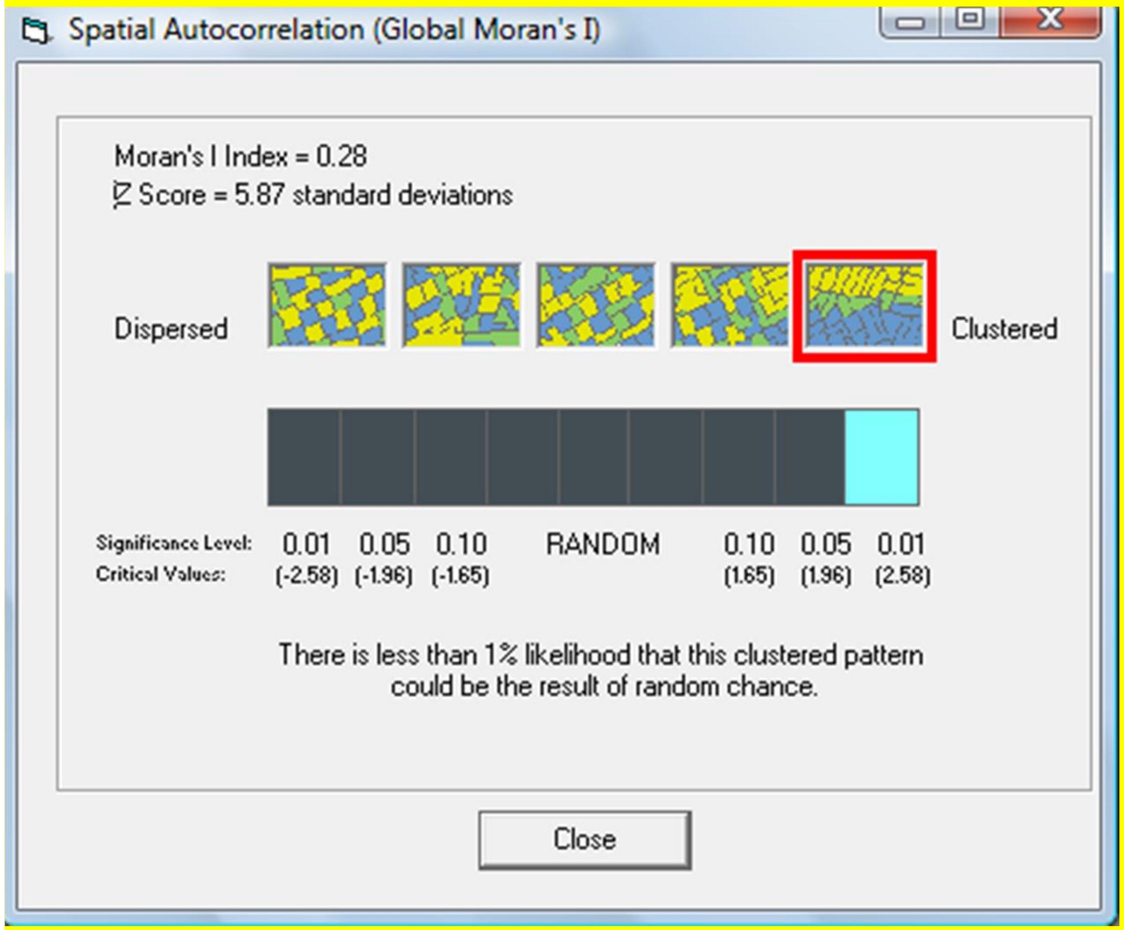
- *If there are roughly as many pairs with positive CPs as there are with negative CPs, the summation result will be close to zero -* **RANDOM**

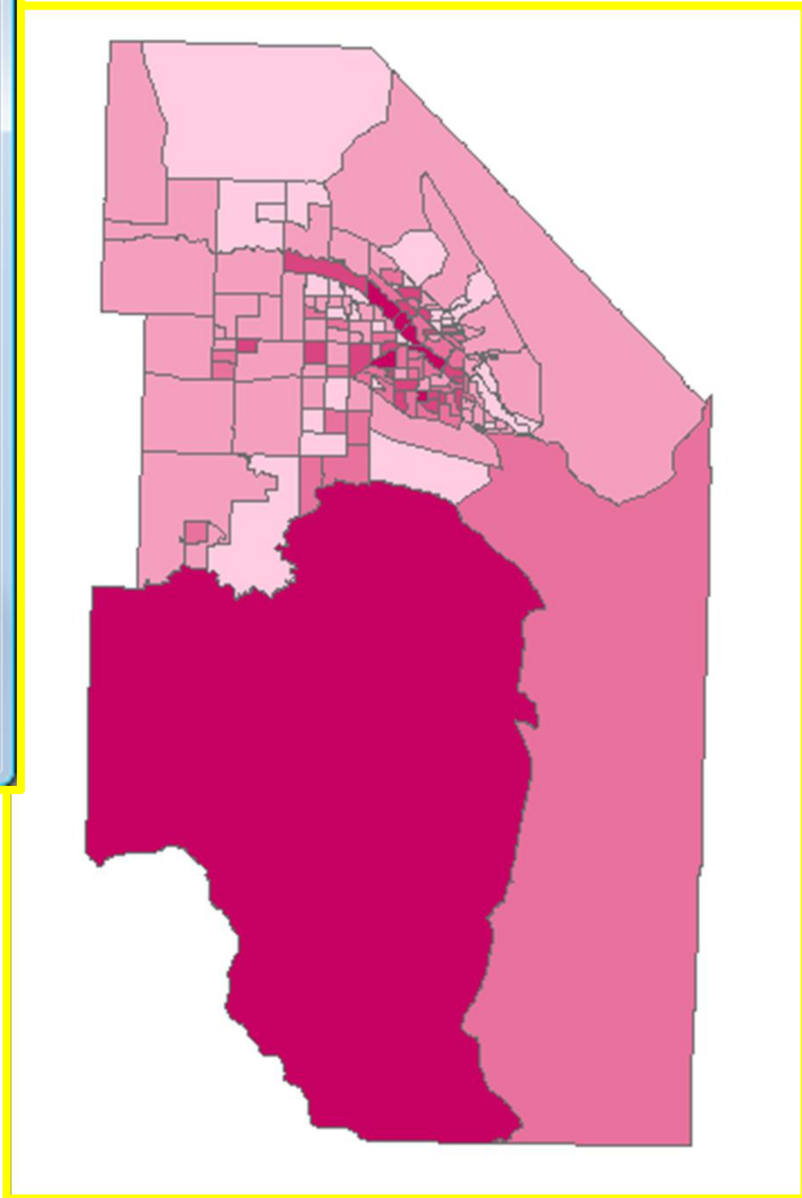
MORAN'S I index……HOW TO INTERPRET ??

- A positive Moran's Index value indicates clustering
- An negative index value indicates dispersion.
- The Global Moran's I function also calculates a Z score value which indicates statistical significance at the specified level of confidence.
- The null hypothesis states "there is no spatial clustering of the values".
- To determine if the Z score is statistically significant, compare it to the range of values for a particular confidence level. For example, at a significance level of 0.05, a z score would have to be less than -1.96 or greater than 1.96 to be statistically significant.
- When the Z score is large (or small) enough to such that it falls outside of the desired significance, the null hypothesis can be rejected.
- When the null hypothesis is rejected, the next step is to inspect the value of the Moran's I Index. If the value is greater than 0, the set of features exhibits a clustered pattern. If the value is less than 0, the set of features exhibits a dispersed pattern.

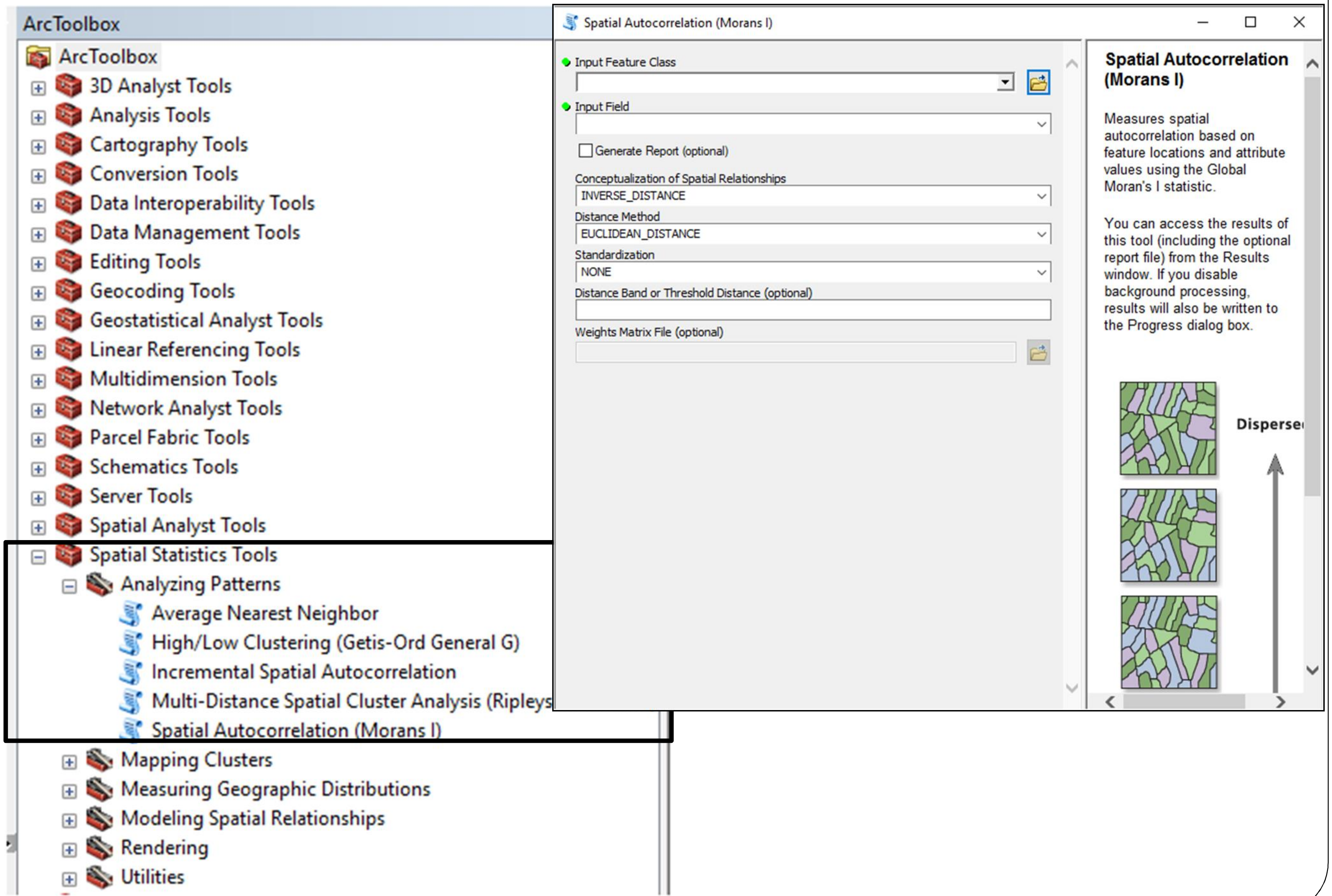
z-score (Standard Deviations)	p-value (Probability)	Confidence level
< -1.65 or > +1.65	< 0.10	90%
< -1.96 or > +1.96	< 0.05	95%
< -2.58 or > +2.58	< 0.01	99%




Spatial Autocorrelation (Global Moran's I)
Moran's I Index = 0.28
Z Score = 5.87 standard deviations
Dispersed
Clustered
Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01
Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)
There is less than 1% likelihood that this clustered pattern could be the result of random chance.
Close

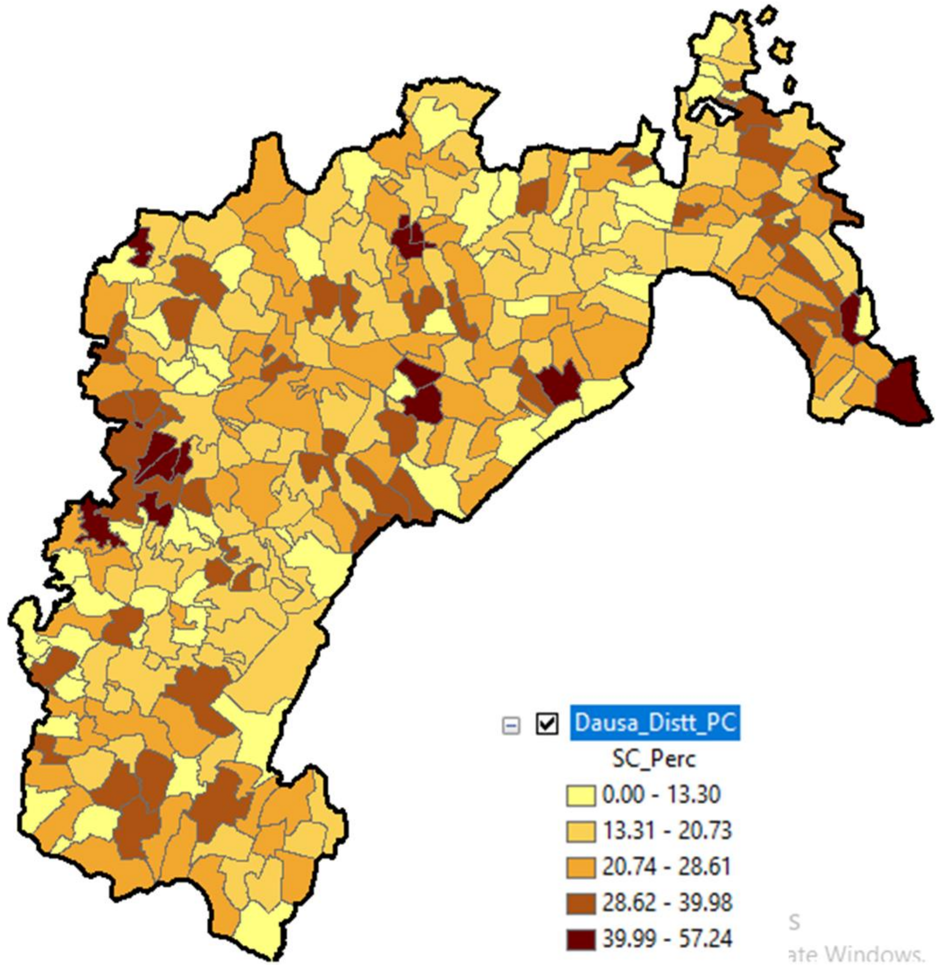


The ArcGIS Spatial Statistics Tool Box



EXAMPLE

Is the SC Population in Dausa district clustered?



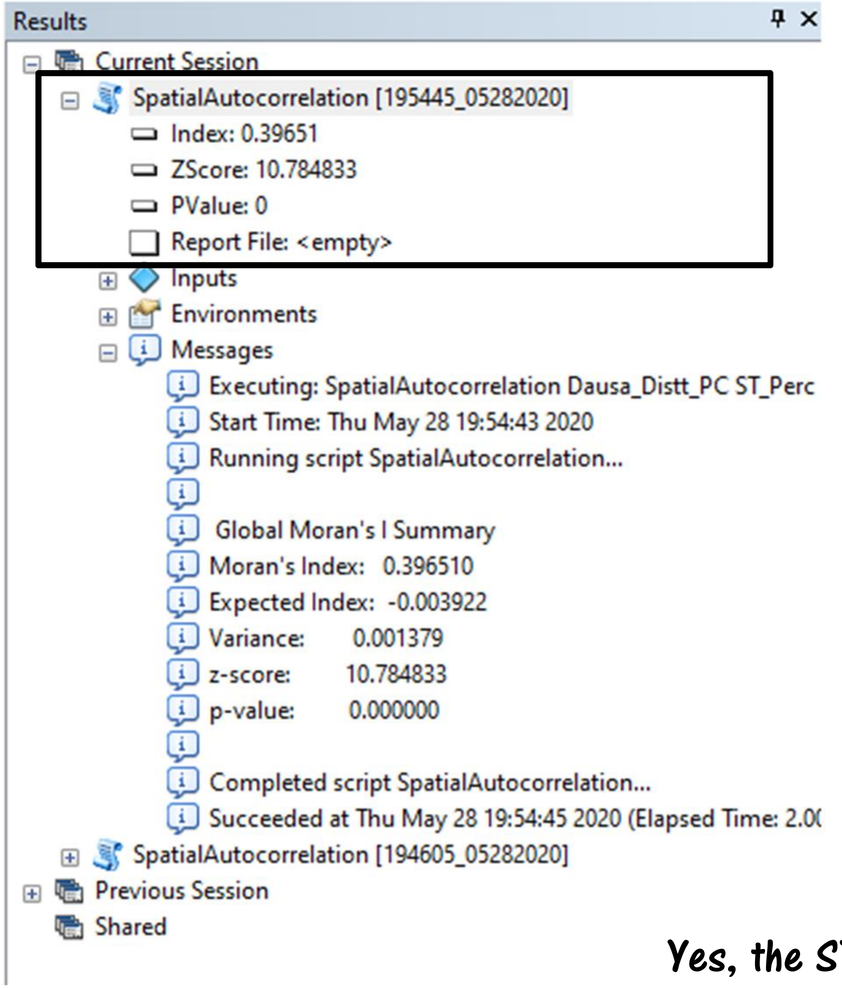
Results

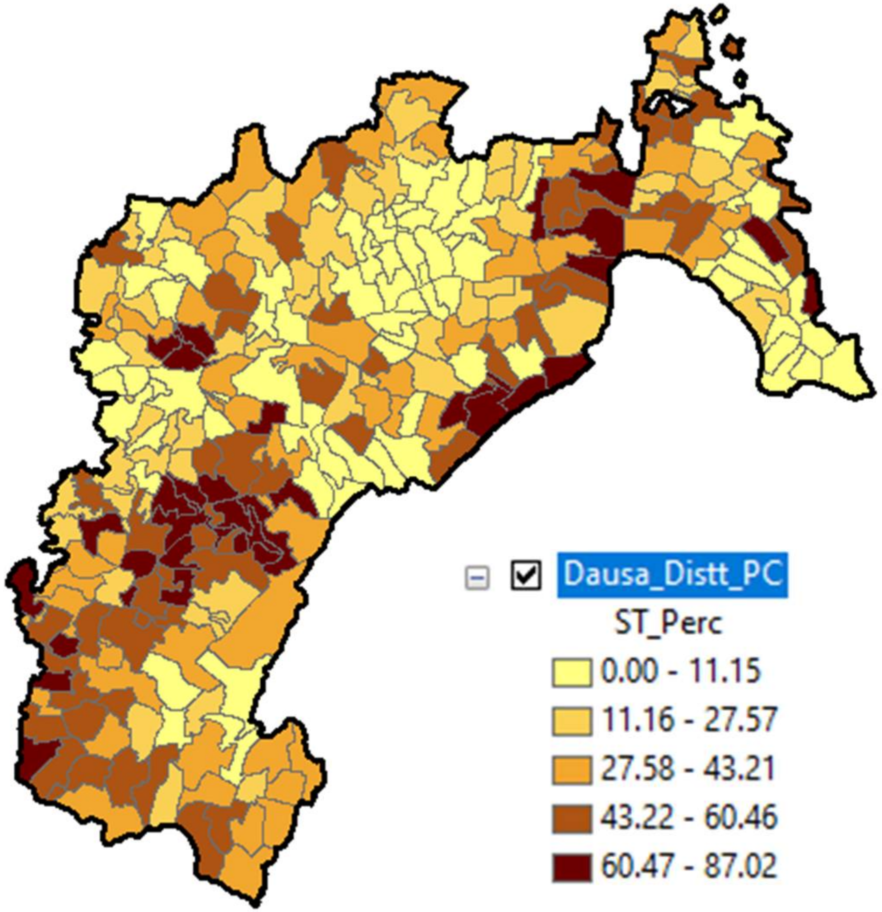
- Current Session
  - SpatialAutocorrelation [194605_05282020]
    - Index: 0.079608
    - ZScore: 2.259192
    - PValue: 0.023871
    - Report File: MoransI_Result.html
    - Inputs
      - Input Feature Class: Dausa_Distt_PC
      - Input Field: SC_Perc
      - Generate Report: GENERATE_REPORT
      - Conceptualization of Spatial Relationships: CONTIGUITY_EDGES_CORNERS
      - Distance Method: EUCLIDEAN_DISTANCE
      - Standardization: NONE
      - Distance Band or Threshold Distance:
      - Weights Matrix File: <empty>
    - Environments
    - Messages
- Previous Session
- Shared

Yes, the SC population is clustered at 0.05 level of significance.
There is 5 % likelihood that this result may be due to random chance

EXAMPLE

Is the ST Population in Dausa district more clustered than SC Population?





Yes, the ST population is clustered at 0.001 level of significance. The clustering is higher than that of SC population

MORAN'S I index……CONSIDERATIONS

Moran's I only measures whether similar/dissimilar values occur together – NOT whether the clusters are composed of high or low values

- Spatial autocorrelation is affected by the scale of spatial pattern. Same pattern at different scales may produce different results (Modifiable Areal Unit Problem).
- The index is sensitive to the number of features and extent of study area
- The input field you select should only contain positive numeric values. Negative weights will be converted to zero for the calculations.
- The values in the input field should have at least some variation. The statistic will not compute if the values have no variation (if they are all one value).
- Calculations are based on either Euclidean or Manhattan distance and require projected data to accurately measure distances.
- The units of the "Distance Band or Threshold Distance" parameter are the units of the input feature class' coordinate system.

MORAN'S I index......APPLICATIONS

- Determine the feasibility of using a particular statistical method (for example, linear regression analysis and many other statistical techniques require independent observations)
- Help identify an appropriate neighborhood distance for a variety of spatial analysis methods. For example, find the distance where spatial autocorrelation is strongest.
- Measure broad trends in ethnic or racial segregation over time—is segregation increasing or decreasing.
- Summarize the diffusion of an idea, disease or trend over space and time—is the idea, disease, or trend remaining isolated and concentrated or spreading and becoming more diffuse.

G-STATISTIC FOR MEASURING HIGH / LOW CLUSTERING

SEPERATES CLUSTERS OF HIGH VALUES FROM CLUSTERS OF LOW VALUES (Getis and Ord 1992)

- The G-statistic indicates whether clusters of high values (Hot Spots) or clusters of low values (Cold Spots) exist in the study area.
- However it does not show where the concentration of values is
- Makes it possible to assess the spatial association of a variable within a particular distance of each observation – based on neighborhood you specify
- Feature pairs for which the neighbouring feature is within the distance of the target feature are assigned a weight of 1, all other pairs are assigned a weight of zero
- It is a **multiplicative measure**

THE MECHANISM

- The **General G tool** calculates the value of the General G statistic and associated Z score for a given input feature class.

$$G(d) = \frac{\sum \sum w_{ij}(d)\, x_i x_j}{\sum \sum x_i x_j} \qquad i \neq j$$

$G(d)$ =	General G statistic based on a specified distance
x_i =	Value at location **i**
x_j =	Value at location *j* if *j* is within **d** of *i*
$w_{ij}(d)$=	Spatial weight based on some weighted distance (e.g. inverse distance)

THE MECHANISM

$$G(d) = \frac{\sum \sum w_{ij}(d)\, x_i x_j}{\sum \sum x_i x_j} \qquad i \neq j$$

- GIS multiplies the attribute values for the first feature pair, then multiplies this product by the weight (1 if features are neighbors; 0 if they are not)
- It does the same for all other pairs of features in the dataset and sums the results
- For pairs where the distance is greater than the specified distance, the value ends up being zero
- Finally sum is divided by unweighted sum of the products of all feature pairs in the dataset

Interpretation

$$G(d)obs = \frac{\sum\sum w_{ij}(d)\, x_i x_j}{\sum\sum x_i x_j} \qquad i \neq j$$

- If the pairs within the distance have relatively high values, the numerator will be larger, hence G will be larger
- If the pairs within the distance have relatively low values, the numerator, hence value of G – will be smaller
- G is a relative value. You don't really know what a large or small value means unless you compare it to the expected G- Statistic for a random distribution at the distance you specified
- You have to check whether the distribution of values is significantly different from a random distribution?

The expected value of G(d) is

$$G(d)e = \frac{\sum \sum w_{ij}(d)}{n(n-1)}$$

E (G) is a typically small value when n is large.

The expected value of G-Statistic at the given distance is what the value of G would be were there no particular concentration of high or low values

$$Z_{G(d)} = \frac{G(d)_o - G(d)_e}{SD_{G(d)}}$$

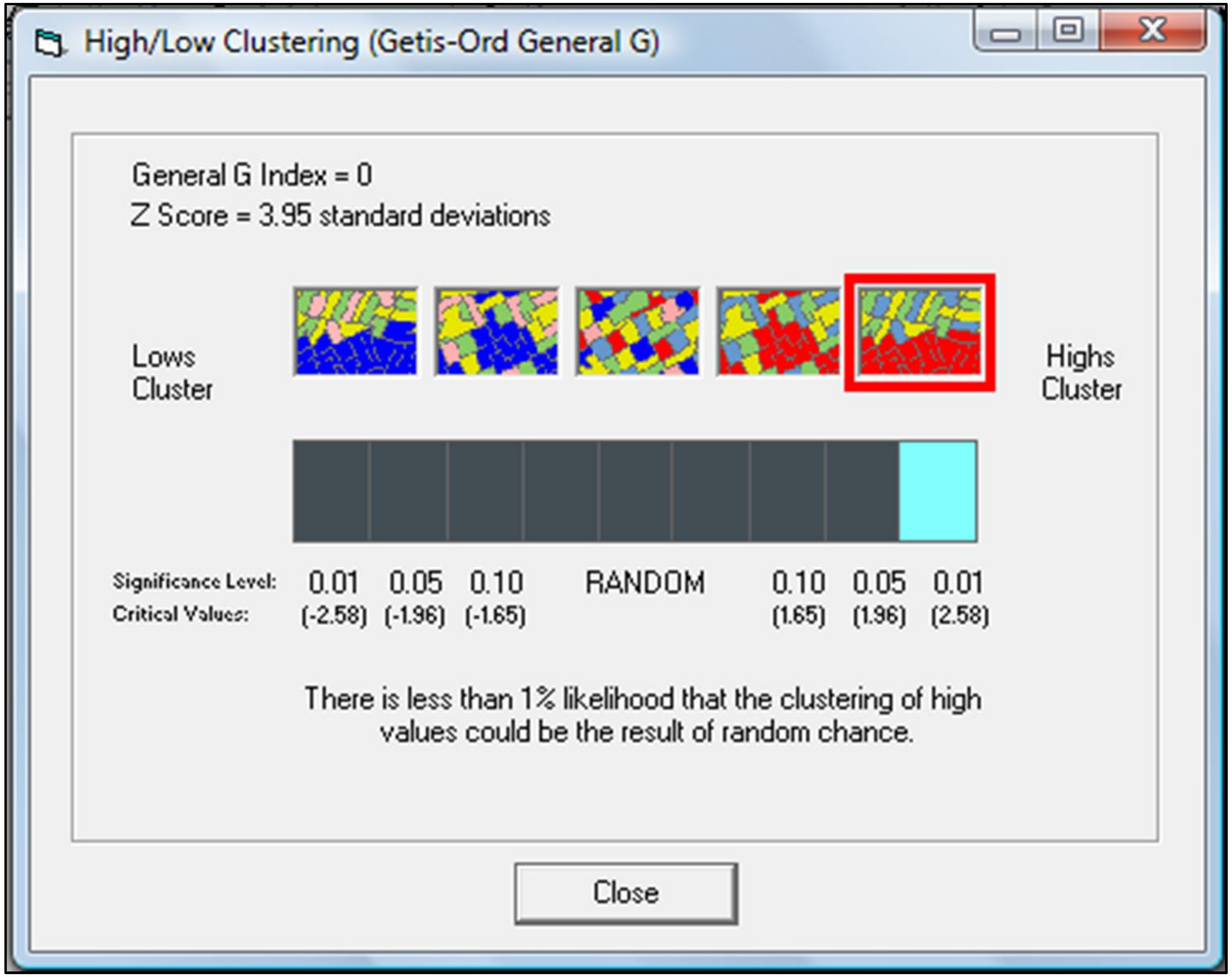
SD of G(d)e

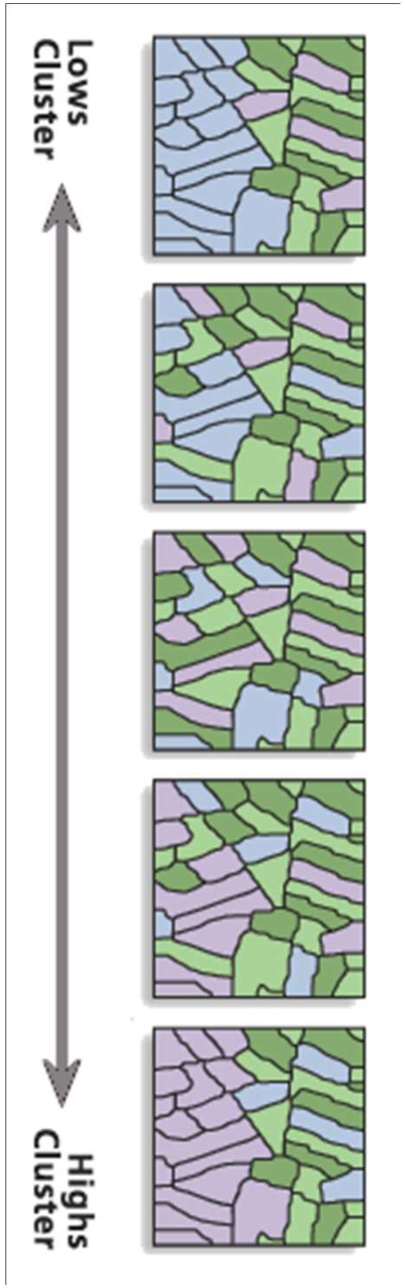
GIS tests whether the observed value of G is significantly different than expected G at a given confidence level

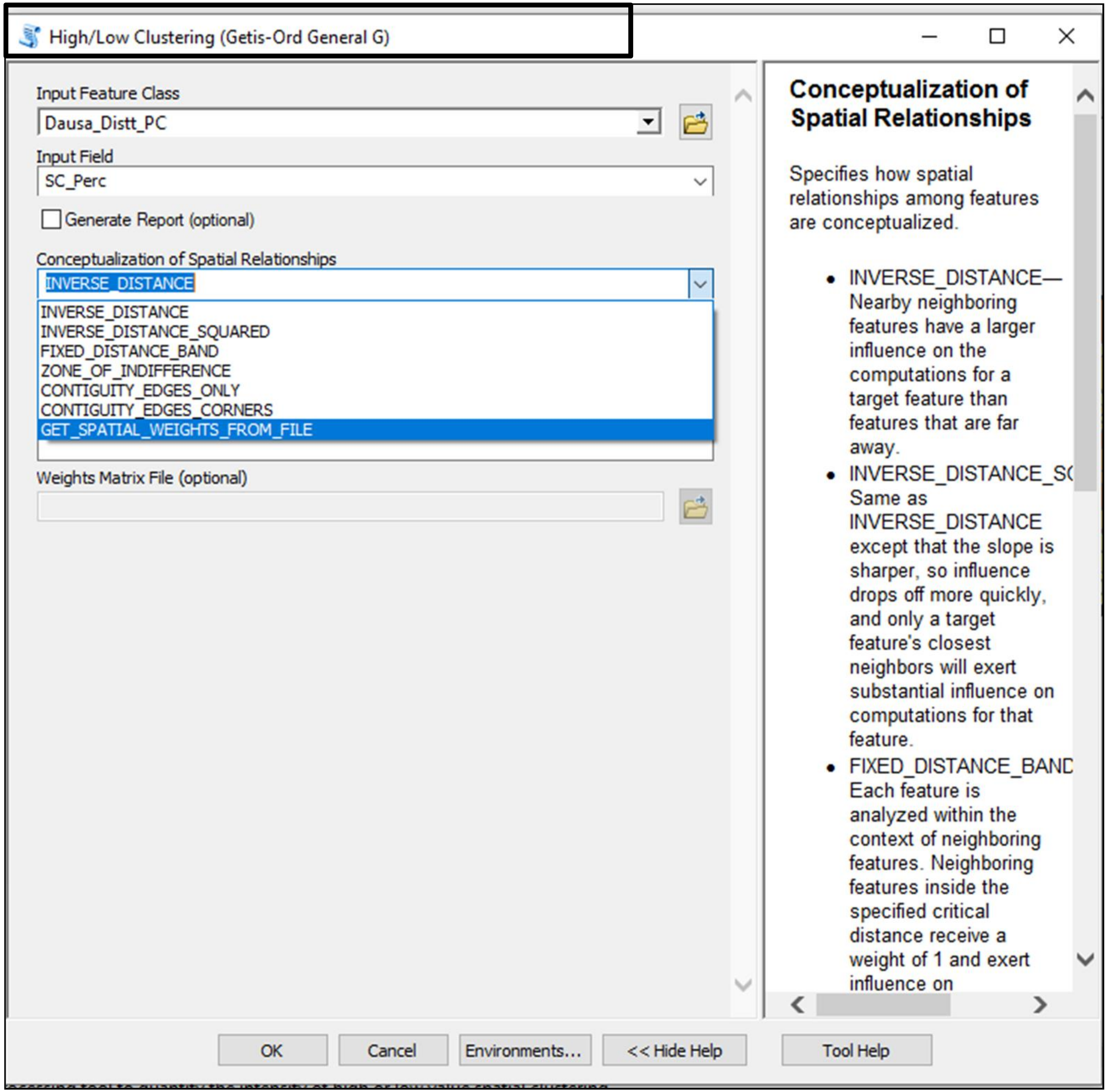
- **A high G (*d*) value suggests a clustering of high values**
- **A low G (*d*) value suggests a clustering of low values**
- **A Z score is computed for a G(*d*) to evaluate its statistical significance.**
- **THE NULL HYPOTHESIS STATES "THERE IS NO SPATIAL CLUSTERING".**
- **A Z score near zero indicates no apparent clustering within the study area.**
- **A positive Z score indicates clustering of high values [G(d) o > G(d)e]**
- **A negative Z score indicates clustering of low values [G(d)o < G(d)e]**
- **The higher (or lower) the Z score, the stronger the intensity of the clustering.**

G-statistic …CONSIDERATIONS

- The input field you select **should only contain positive numeric values**. Negative weights will be converted to zero for the calculations.
- The values in the input field **should have at least some variation**. The statistic will not compute if the values have no variation (if they are all one value).
- For line and polygon features, feature centroids are used in computations
- Calculations are based on either **Euclidean** or **Manhattan** distance and require projected data to accurately measure distances.
- Definition of distance/ neighbourhood influences the results
- Results also depend on **range of values of features**. If there are one or few very high values – relative to the mean value of the dataset – the G statistic may show concentration of high values- even if there are more features of low values near each other

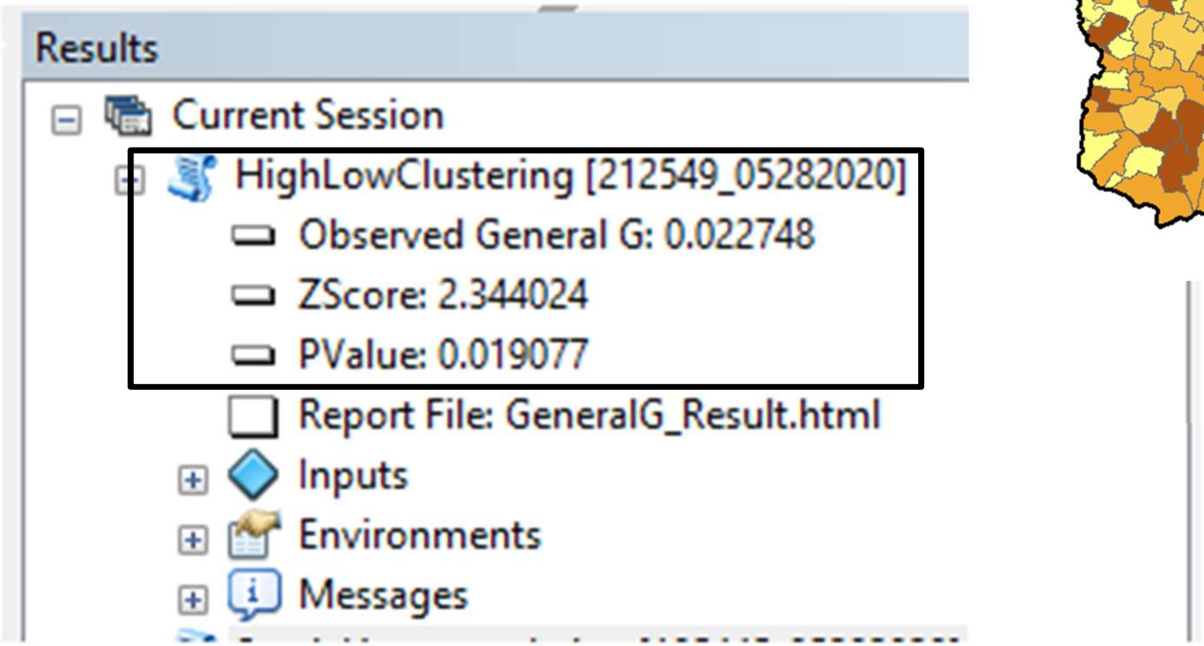

High/Low Clustering (Getis-Ord General G)
General G Index = 0
Z Score = 3.95 standard deviations
Lows Cluster
Highs Cluster
Significance Level: 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01
Critical Values: (-2.58) (-1.96) (-1.65) (1.65) (1.96) (2.58)
There is less than 1% likelihood that the clustering of high values could be the result of random chance.
Close

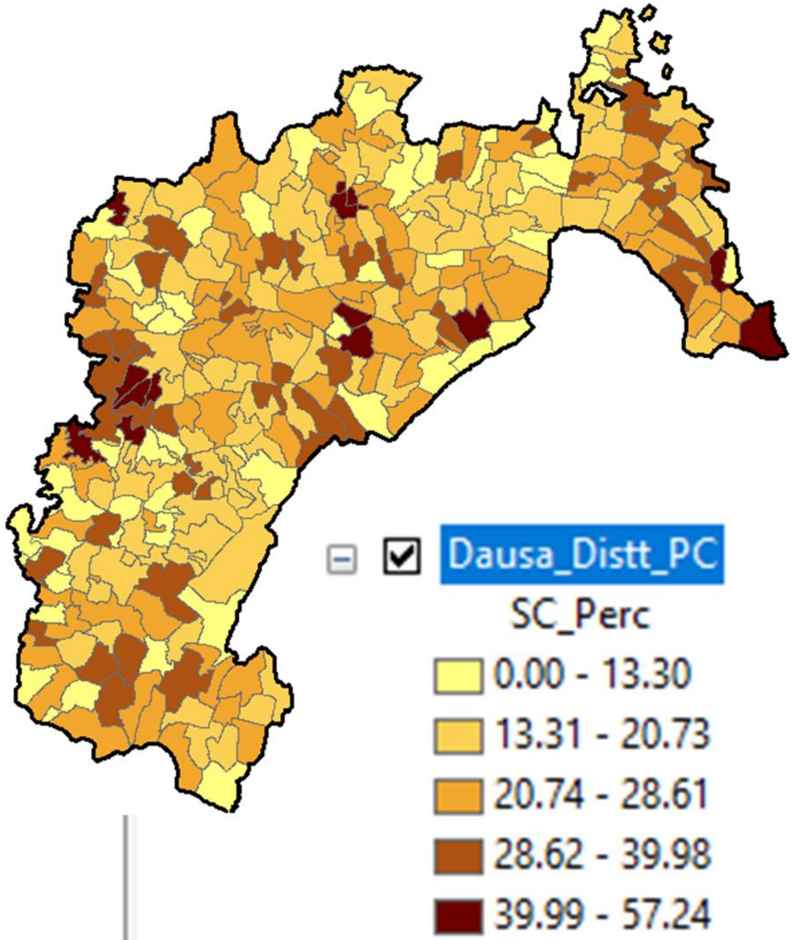

Lows Cluster
Highs Cluster


High/Low Clustering (Getis-Ord General G)
Input Feature Class
Dausa_Distt_PC
Input Field
SC_Perc
Generate Report (optional)
Conceptualization of Spatial Relationships
INVERSE_DISTANCE
INVERSE_DISTANCE
INVERSE_DISTANCE_SQUARED
FIXED_DISTANCE_BAND
ZONE_OF_INDIFFERENCE
CONTIGUITY_EDGES_ONLY
CONTIGUITY_EDGES_CORNERS
GET_SPATIAL_WEIGHTS_FROM_FILE
Weights Matrix File (optional)
Conceptualization of Spatial Relationships
Specifies how spatial relationships among features are conceptualized.
INVERSE_DISTANCE— Nearby neighboring features have a larger influence on the computations for a target feature than features that are far away.
Same as INVERSE_DISTANCE except that the slope is sharper, so influence drops off more quickly, and only a target feature's closest neighbors will exert substantial influence on computations for that feature.
Each feature is analyzed within the context of neighboring features. Neighboring features inside the specified critical distance receive a weight of 1 and exert influence on
OK
Cancel
Environments...
<< Hide Help
Tool Help

G -Statistic

Is the SC Population in Dausa district exist in clusters of high and low values?

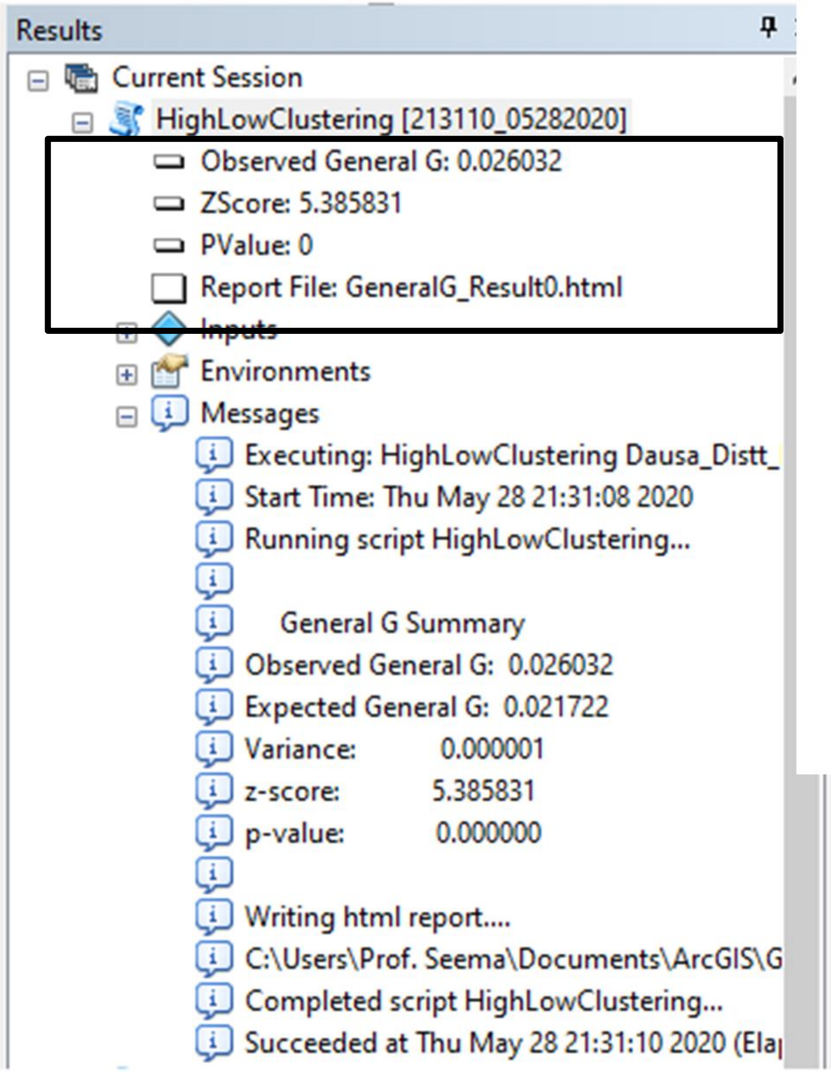


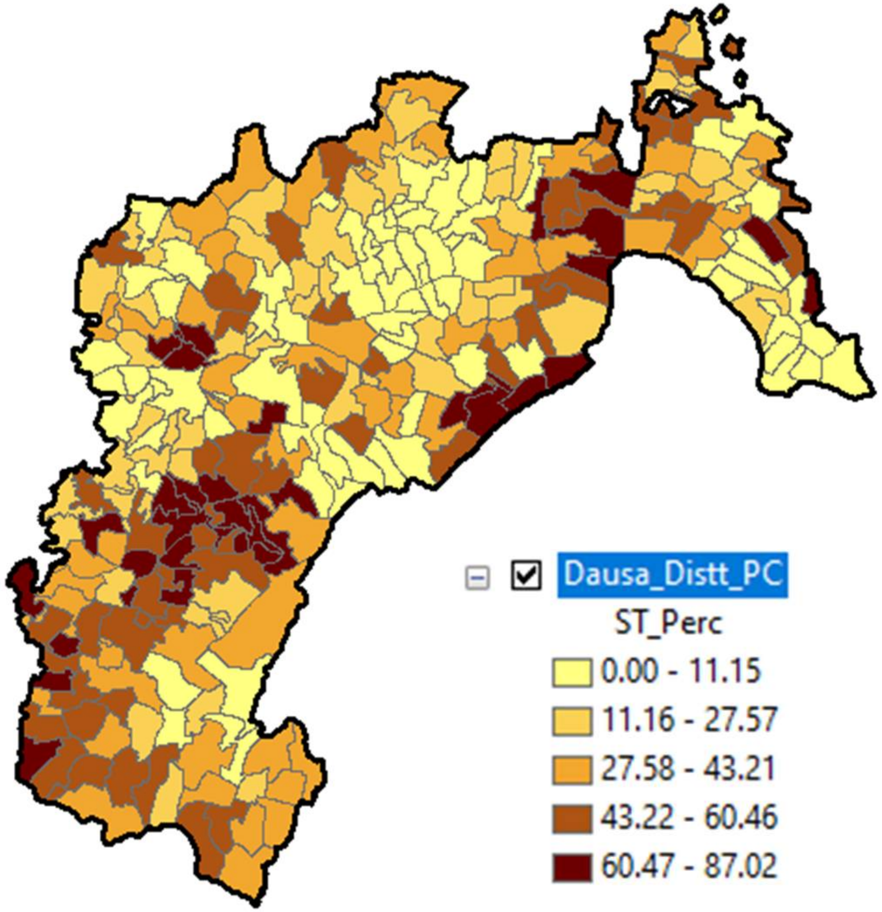


High values of SC population are clustered . The results are significant at 0.05 level of significance. There is 5 % likelihood that this result may be due to random chance

G - STATISTIC

Is the ST Population in Dausa district more clustered than SC Population?





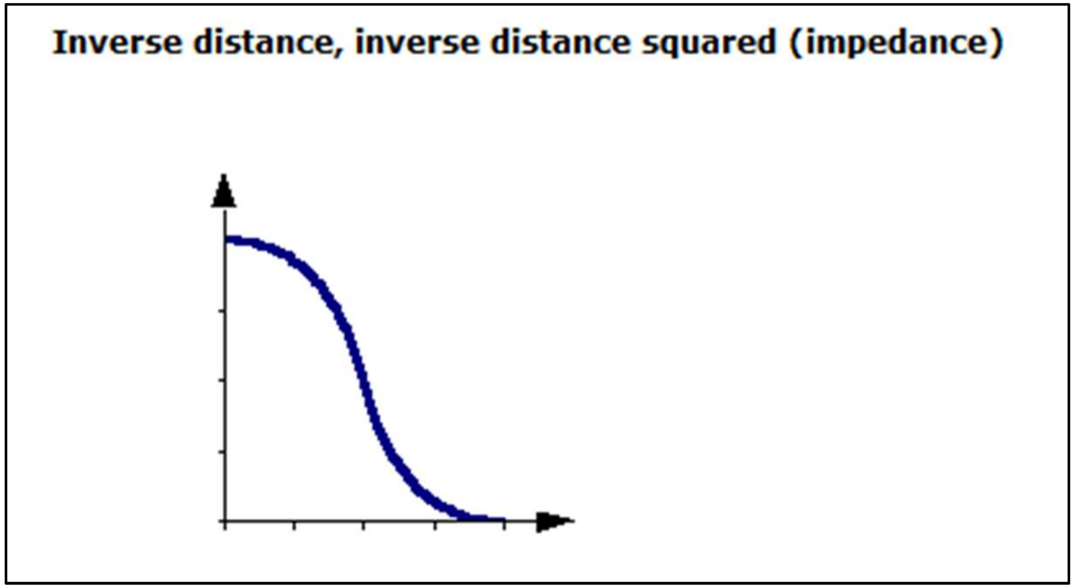
Yes, the ST population is more clustered at 0.001 level of significance. The clustering is higher than that of SC population. Hot spots of ST population exist in the district.

DEFINING THE NEIGHBOURHOOD

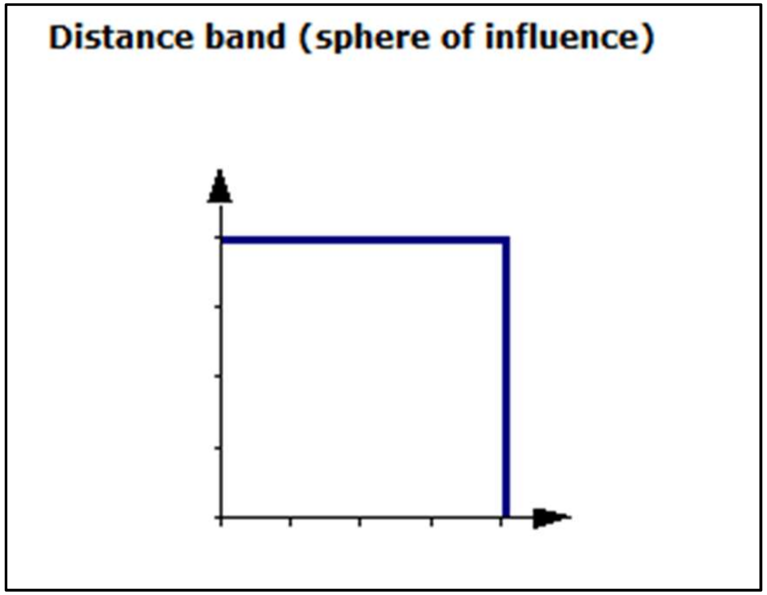
The distance used for analysis should be based on your understanding of spatial interaction among the features being analyzed. Spatial relationships between features may be conceptualized in the following ways

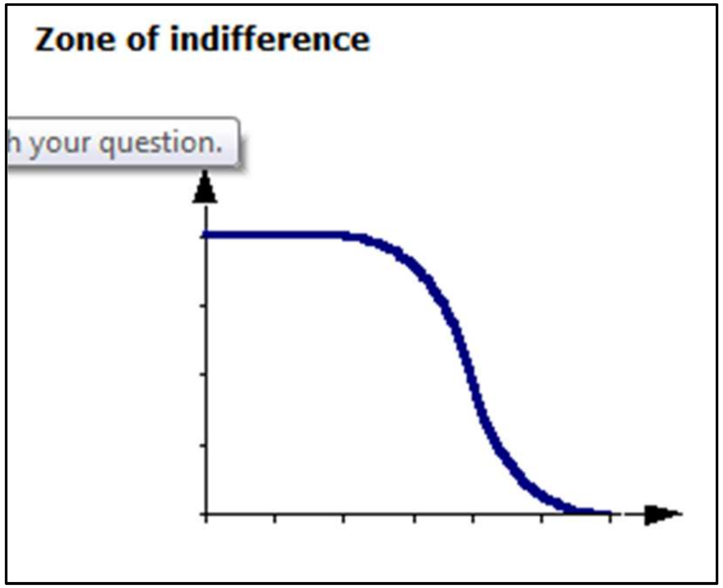
- **Inverse Distance**—The impact of one feature on another feature decreases with distance. All features impact or influence all other features, but the farther away something is, the smaller the impact it has.
- **Inverse Distance Squared**—Same as Inverse Distance, but the impact decreases more sharply over distance. Only the very closest features exert an influence.
- **Fixed Distance Band**—Everything within a specified critical distance is included in the analysis; everything outside the critical distance is excluded.

CONCEPTUALIZATION OF SPATIAL RELATIONSHIPS



Your conceptualization of the spatial relationship will depend on what you are measuring. If you are measuring clustering of a particular species of tree in a forest, for example, inverse distance is probably most appropriate. However, if you are assessing the geographic distribution of a region's commuters, travel time or travel cost might be better choices for describing spatial relationships.





- **Zone of Indifference**—A combination of Inverse Distance and Fixed Distance Band. Anything up to a critical distance has an impact on your analysis. Once that critical distance is exceeded, the level of impact quickly drops off.
- **Polygon Contiguity (First Order)**—The neighbors of each feature are only those with which the feature shares a boundary. Polygons that share an edge (those with coincident boundaries) influence each other. All other features have no influence.
- **Get Spatial Weights From File**—Spatial relationships are defined in a spatial weights file. Spatial weights are numbers that reflect the distance, time, or other cost between each feature and every other feature in the dataset. You can provide a pathname to an ASCII text file that represents your conceptualization of spatial relationships (travel time, travel costs, spatial interaction, or more abstract relationships, such as familiarity).

- The units of the "Distance Band or Threshold Distance" parameter are the units of the input feature class' coordinate system.

- Any value entered for the "Distance Band or Threshold Distance" parameter is not considered when "Inverse Distance", "Inverse Distance Squared", "Polygon Contiguity" or "Get Spatial Weights from File" are selected for the "Conceptualization of Spatial Relationships" parameter.

LOCAL INDICES

Answer WHERE there are areas with similar values; and where similar and dissimilar values are interspersed.

Similarity of neighbouring values

High/Low values

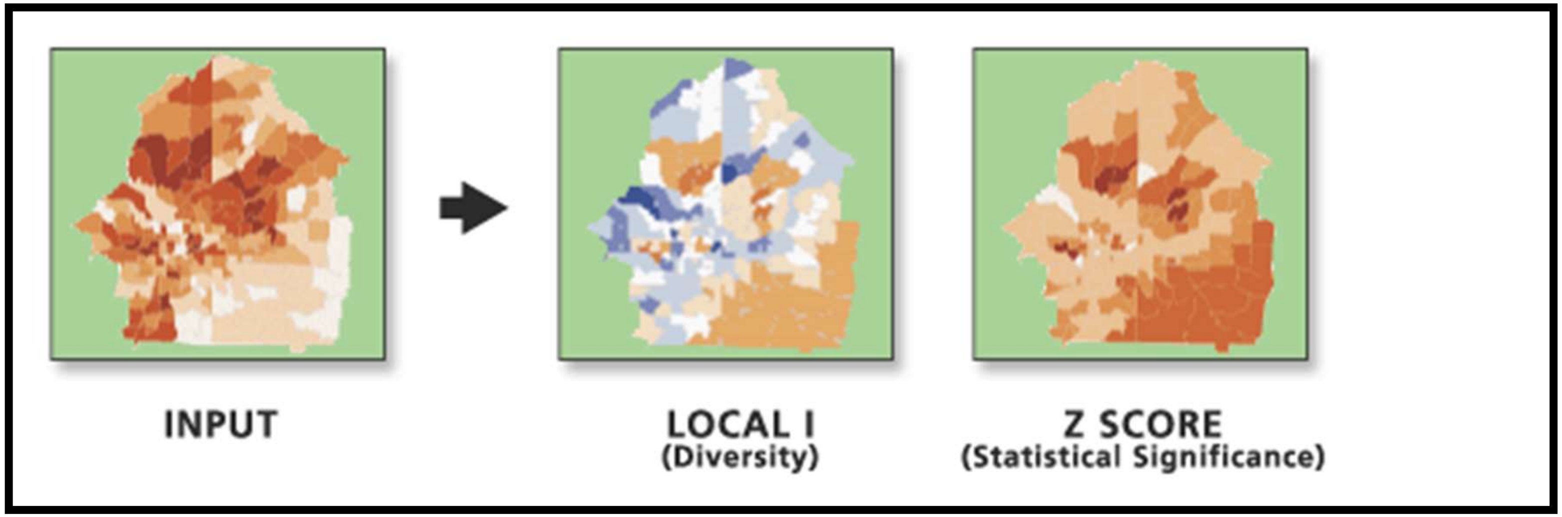
Moran's Ii

Getis-Ord Gi*

The i subscript indicates that a statistic is calculated for each feature

Anselins Local MORAN’S I (1995)/ LISA

- **Identifies clusters of features with values similar in magnitude.**
- **Calculates for each feature (point or polygon) an index value and a Z-Score.**
- **The Local Moran's index can only be interpreted within the context of the computed Z score**



Anselins Local MORAN'S I (1995)/ LISA

- **A neighbourhood based on adjacency is more appropriate.**
- **Use either binary weighting or row standardized weighting**

A	B	C
D	E	F

Divide each number by the row sum

Total number of neighbors
--some have more than others

	A	B	C	D	E	F	Row Sum
A	0	1	0	1	0	0	2
B	1	0	1	0	1	0	***3***
C	0	1	0	0	0	1	2
D	1	0	0	0	1	0	2
E	0	1	0	1	0	1	***3***
F	0	0	1	0	1	0	2

Row standardized
--usually use this

	A	B	C	D	E	F	Row Sum
A	0.0	0.5	0.0	0.5	0.0	0.0	1
B	0.3	0.0	0.3	0.0	0.3	0.0	1
C	0.0	0.5	0.0	0.0	0.0	0.5	1
D	0.5	0.0	0.0	0.0	0.5	0.0	1
E	0.0	0.3	0.0	0.3	0.0	0.3	1
F	0.0	0.0	0.5	0.0	0.5	0.0	1

Briggs Henan University 2010

Anselins Local MORAN'S I (1995)/ LISA

- **Compares each value in the pair (target and neighbour) to the mean value for all the features in the study area**
- **Emphasizes how features differ from the values in the STUDY AREA AS A WHOLE**

Local Moran's I, calculated for each feature (i)

The mean value ($\bar{x}$) is subtracted from the value of the neighbor (x_j) and the difference multiplied by the weight (w_{ij}) for the target-neighbor pair; the results for all neighbors are summed....

$$I_i = \frac{(x_i - \bar{x})}{s^2} \cdot \sum_j w_{ij} (x_j - \bar{x})$$

....then the sum is multiplied by: the difference between the mean value ($\bar{x}$) and the target feature value (x_i), divided by the variance (s^2)

Anselins Local MORAN'S I (1995)/ LISA

- A **large +ve value** of Moran's I*i* indicates that the feature is surrounded by features with similar values
- Several adjacent features with high values of I*i* define cluster of similar values
- A **-ve value** of Moran's I*i* indicates that the feature is surrounded by features with dissimilar values
- The statistic doesn't indicate if the attribute values themselves are high or low. **Create a map by classifying the I*i* values into ranges**
- Attribute values themselves may be mapped to see whether the cluster is comprised of high value or low value

Anselins Local MORAN'S I (1995)/ LISA…..

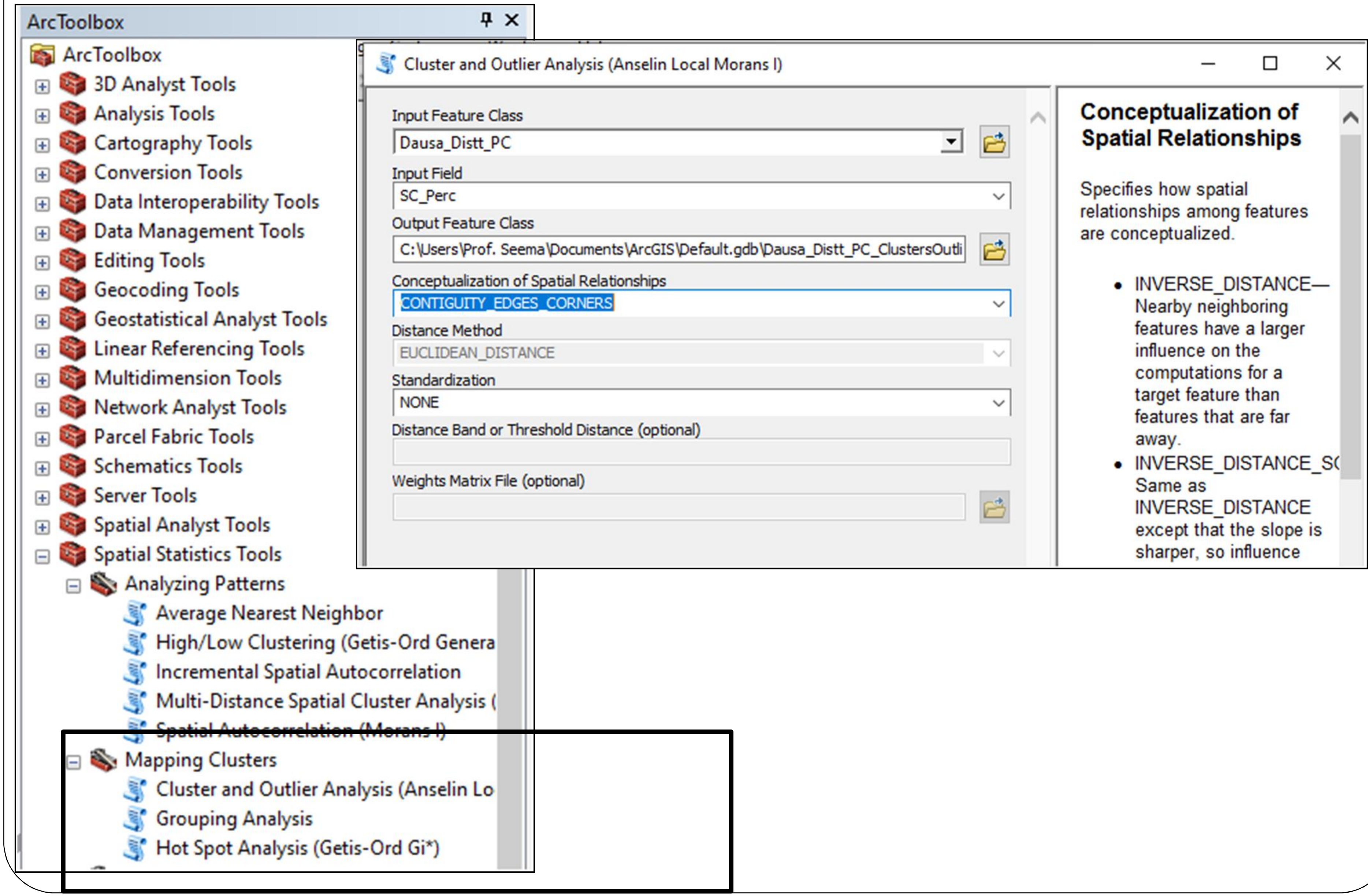
- **The Z score represents the statistical significance of the index value.** It, in effect, indicates whether the apparent similarity (or dissimilarity) in values between the feature and its neighbors is greater than one would expect simply by chance.

- A high **+ve Z-score** suggests that the feature is **adjacent to features of similar values (high or low)** ***Such a feature is part of a cluster.***

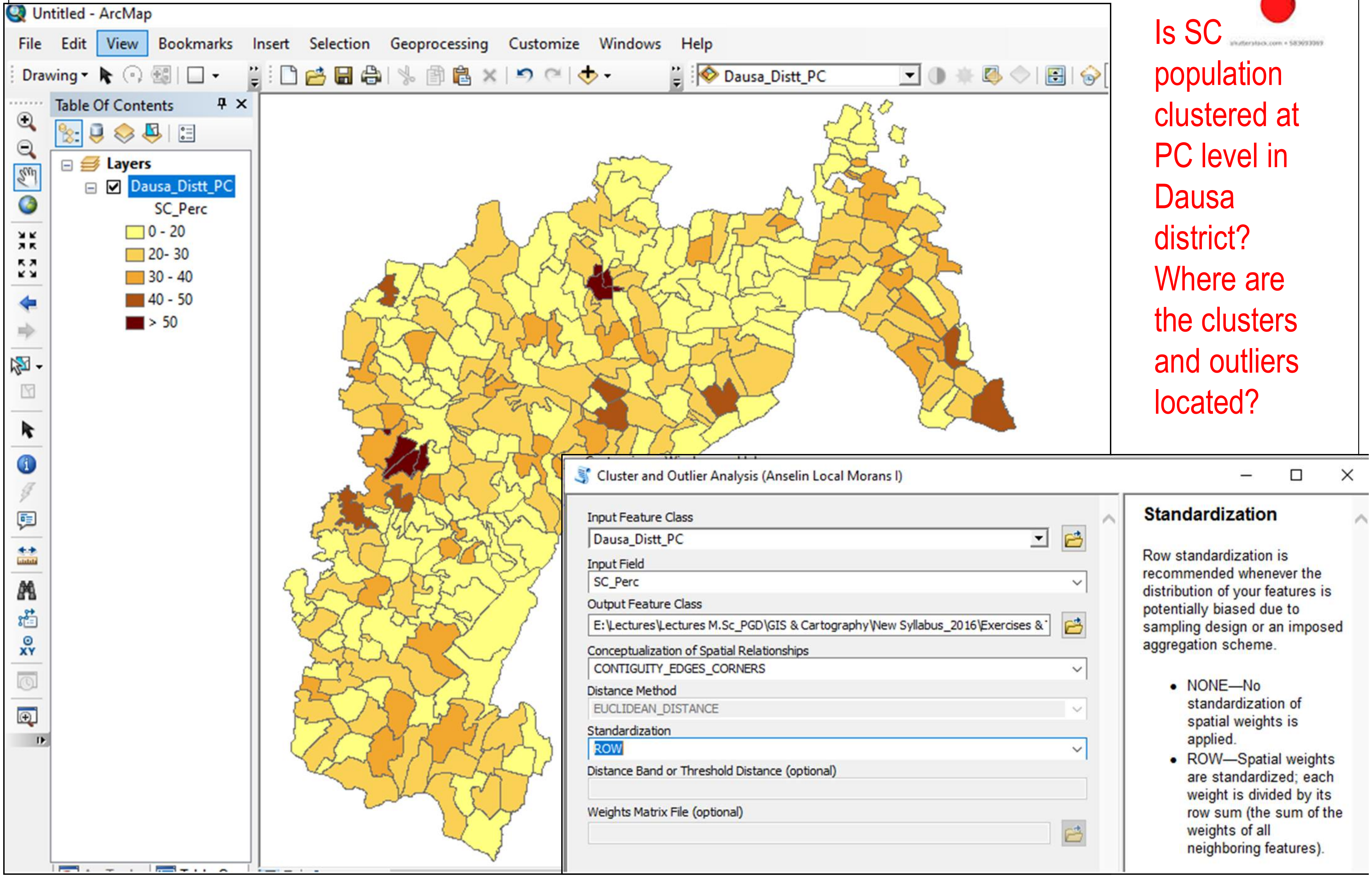
- A high **-ve Z-score** indicates that the feature is **adjacent to features of dissimilar values** (a high value relative to a neighborhood that has low values or a low value relative to a neighborhood that has high values). ***Such a feature is an outlier.***

- ***Mapping the Z-scores shows which clusters are statistically significant***

The ArcGIS Spatial Statistics Tool Box



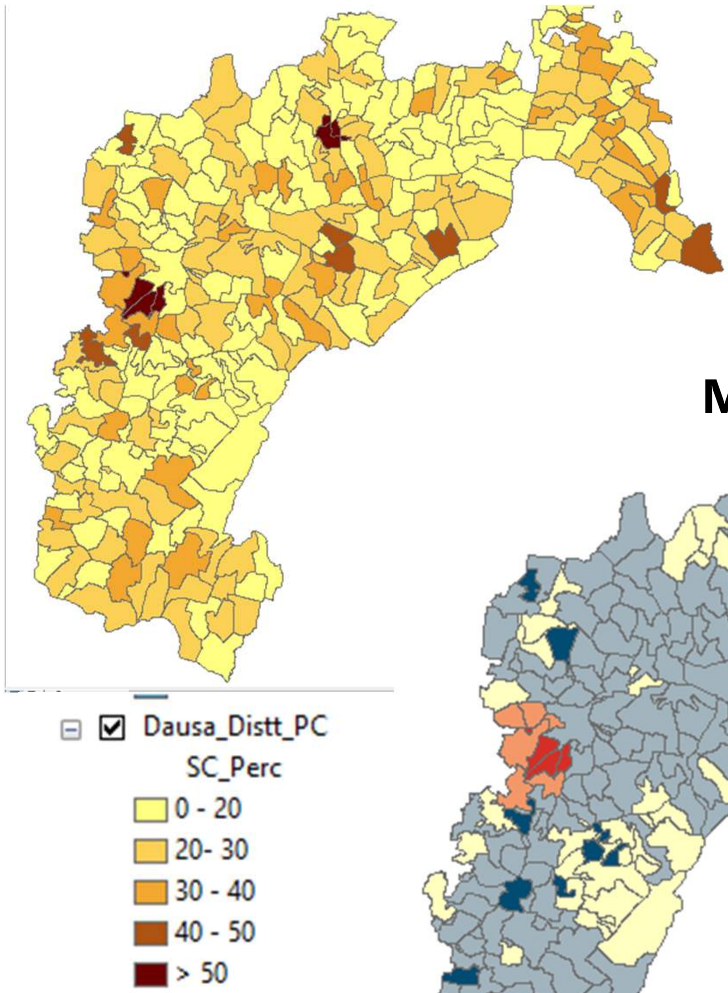
Anselin Local Moran's I



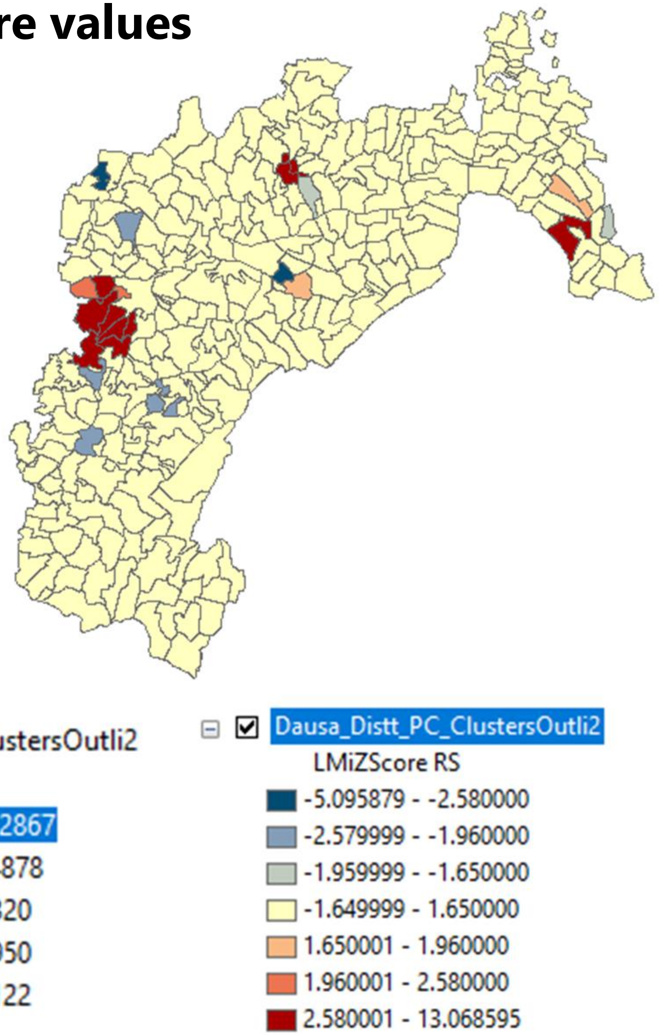
Is SC population clustered at PC level in Dausa district? Where are the clusters and outliers located?

Anselin Local Moran's I

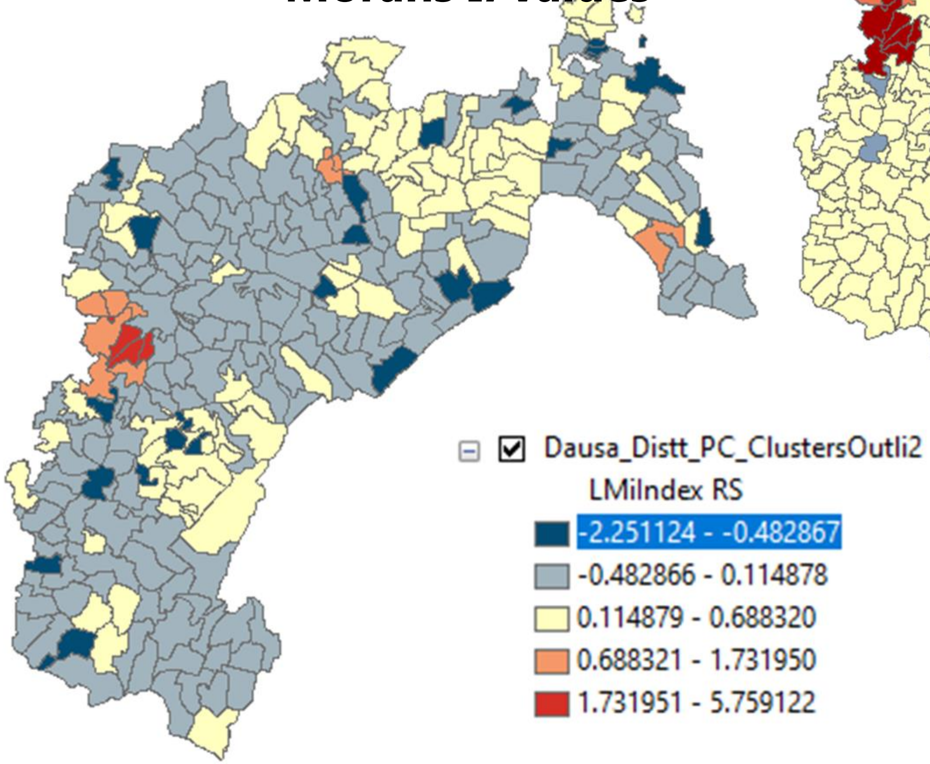
Compare the results with simple mapping of proportions



Z- Score values

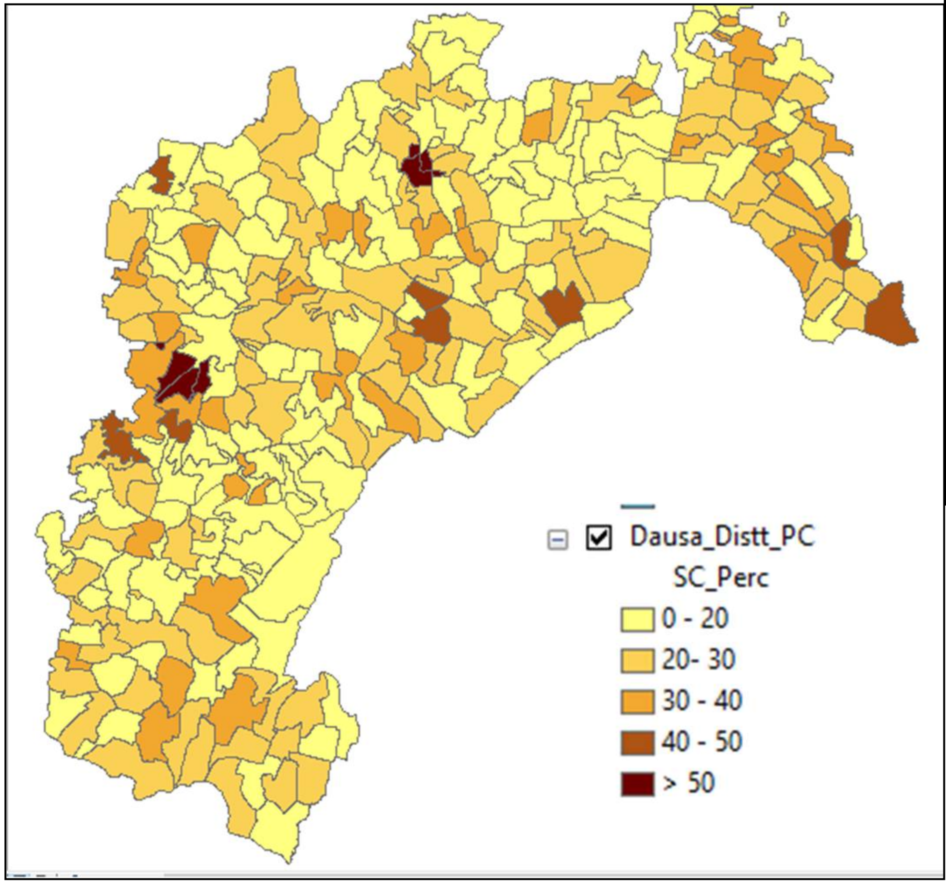


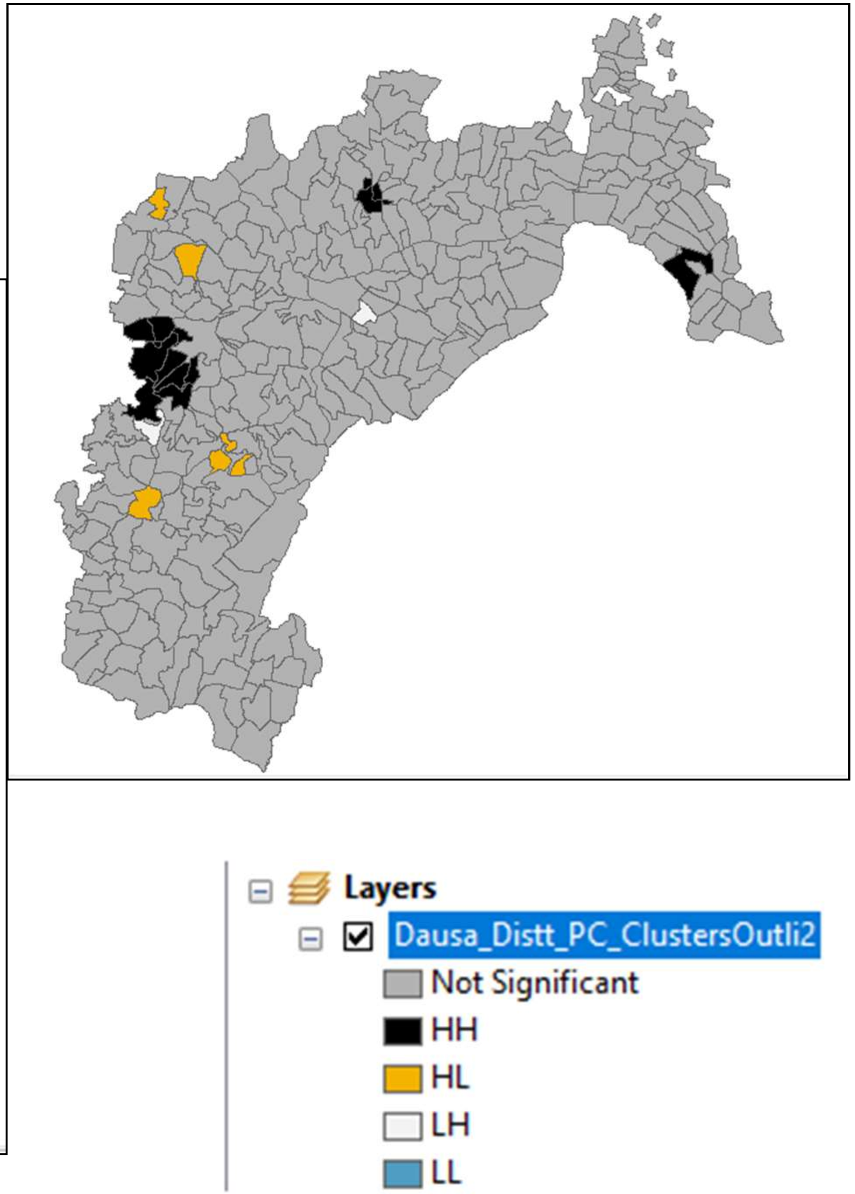
Morans I*i* values



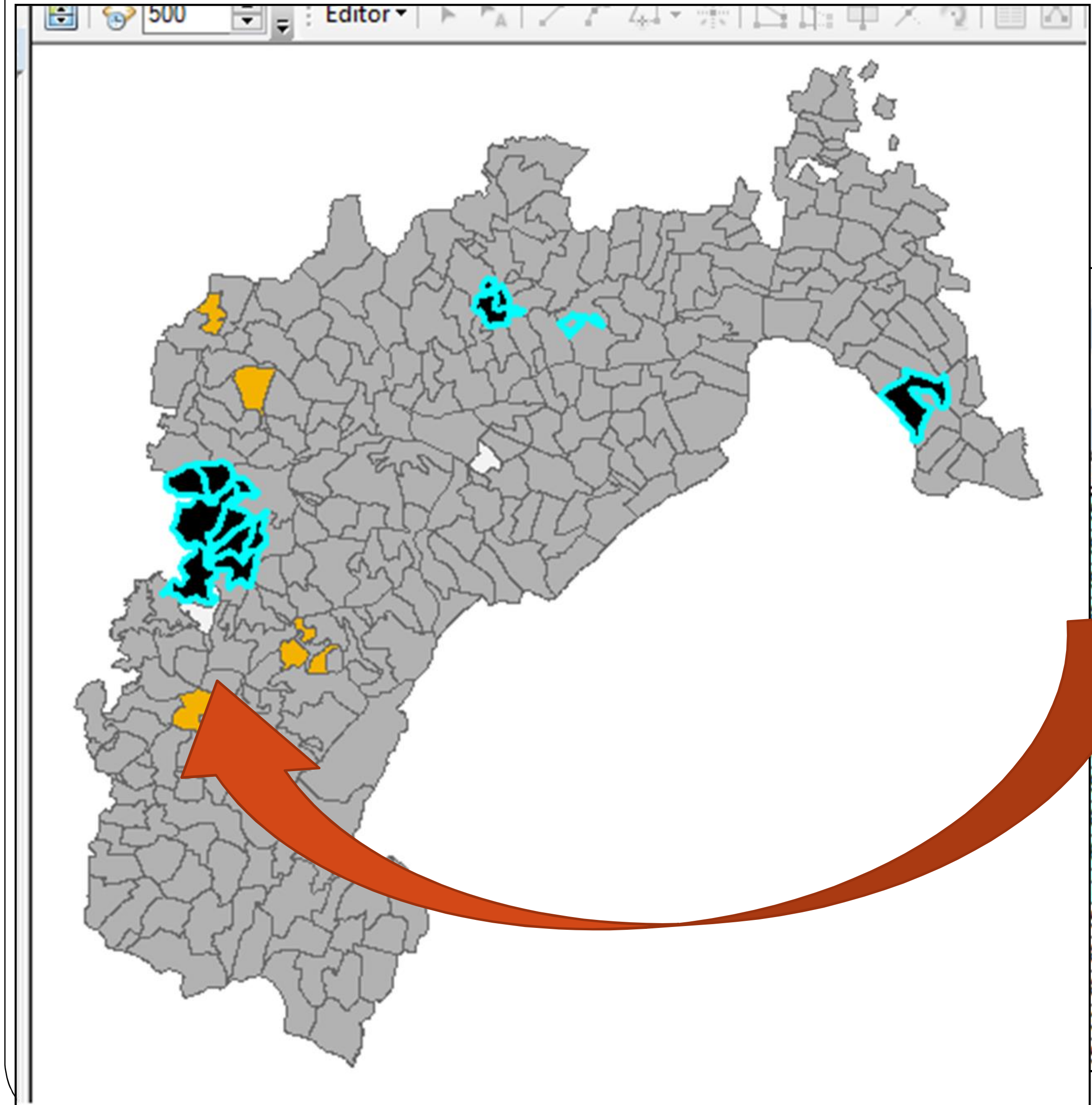
Anselin Local Moran's I

Cluster and Outlier Analysis combines the I*i* values and z-scores to map significant clusters of high and low values in one map





Anselin Local Moran's I



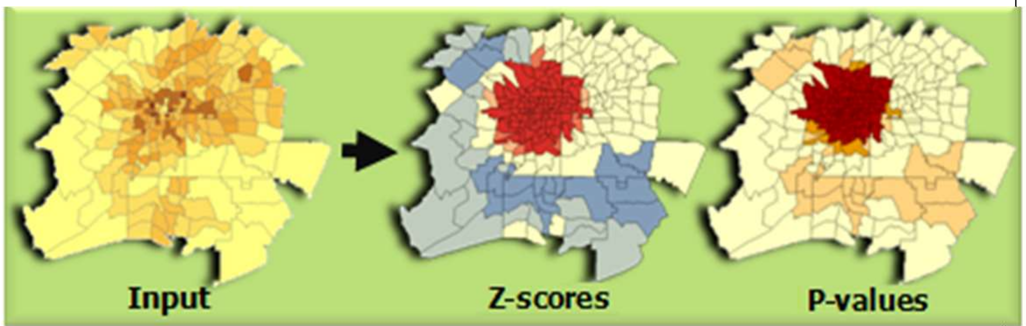
Understand the output

1. Local Moran's I
2. Z-Score statistic
3. P value

_Area	LMiIndex RS	LMiZScore RS	LMiPValue RS
5.817182	5.759122	13.068595	0
97.06636	3.834022	9.552686	0
9.555838	3.284854	6.657334	0
8.703453	1.73195	5.008904	0.000001
6626	1.578528	4.566203	0.000005
0282	1.486575	3.709863	0.000207
89701	1.460908	3.645979	0.000266
.831668	1.330186	3.025297	0.002484
9.513415	1.222991	2.782215	0.005399
8.727465	0.980292	2.839975	0.004512
6.575578	0.894032	2.03625	0.041725
1.576899	0.68832	1.211162	0.225833
5.557337	0.682118	1.848043	0.064596
8.335455	0.671119	1.818414	0.069001
8.533813	0.626524	1.569187	0.116604
3.369812	0.623612	1.27029	0.203982
5.604807	0.590685	0.847771	0.396565

Hot Spot Analysis (Getis- Ord Gi*)

- Indicates the extent to which each feature is surrounded by similarly high or low values
- Identifies statistically significant hot spots and cold spots
- Hot Spots refer to clusters of high values ..a feature with high value surrounded by other features with high values as well
- Cold spots refer to clusters of low values
- Calculates Getis Ord Gi* statistic (pronounced G-i-star) for each feature in the dataset
- Creates a new output feature class with a z-score and p-value for each feature in the input feature class



Hot Spot Analysis (Getis- Ord Gi*)....how GIS does it??

- Works by looking at each feature within the context of neighbouring features.
- The local sum for a feature and its neighbours is compared proportionally to the sum of all features
- When the local sum is very different from the expected local sum, and the difference is too large to be the result of random chance, a statistically significant z-score results
- The recent version of the Gi* statistic combines the Gi* and the Z-Score in a single measure and returns a z-score for each feature in the dataset.

The value of each neighbor (x_j) is multiplied by the weight for the target-neighbor pair (w_{ij}), and the results summed....

Gi for a feature (i), at a distance (d)*

$$G_i^*(d) = \frac{\sum_j w_{ij}(d)\,x_j}{\sum_j x_j}$$

....then the sum is divided by the sum of the values of all neighbors (x_j), that is, all features in the data set

Hot Spot Analysis (Getis- Ord Gi*)....Input Parameters

- Uses a neighbourhood based either on adjacent features or on a set distance
- When using a distance -based neighbourhood, the specified distance is based on knowledge of features and their behavior
- With a larger distance, there may be few large clusters, and with a smaller distance there may be more smaller clusters
- Distance can be defined as Euclidean distance, or travel time etc..

Hot Spot Analysis (Getis- Ord Gi*)....understanding the output

The z-scores and p-values are measures of statistical significance which tell you

- whether or not to reject the null hypothesis, feature by feature
- Whether observed spatial clustering of high and low values is more pronounced than one would expect in a random distribution of the same values

A high z-score and small p-value indicates spatial clustering of high values

A low negative z-score and small p-value indicates spatial clustering of low values

The higher (or lower) the z-score, the more intense the clustering

A z-score near zero indicates no apparent clustering – No concentration of high or low values surrounding the target feature. This occurs when the surrounding values are near the mean, or when the target feature is surrounded by a mix of high or low values

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

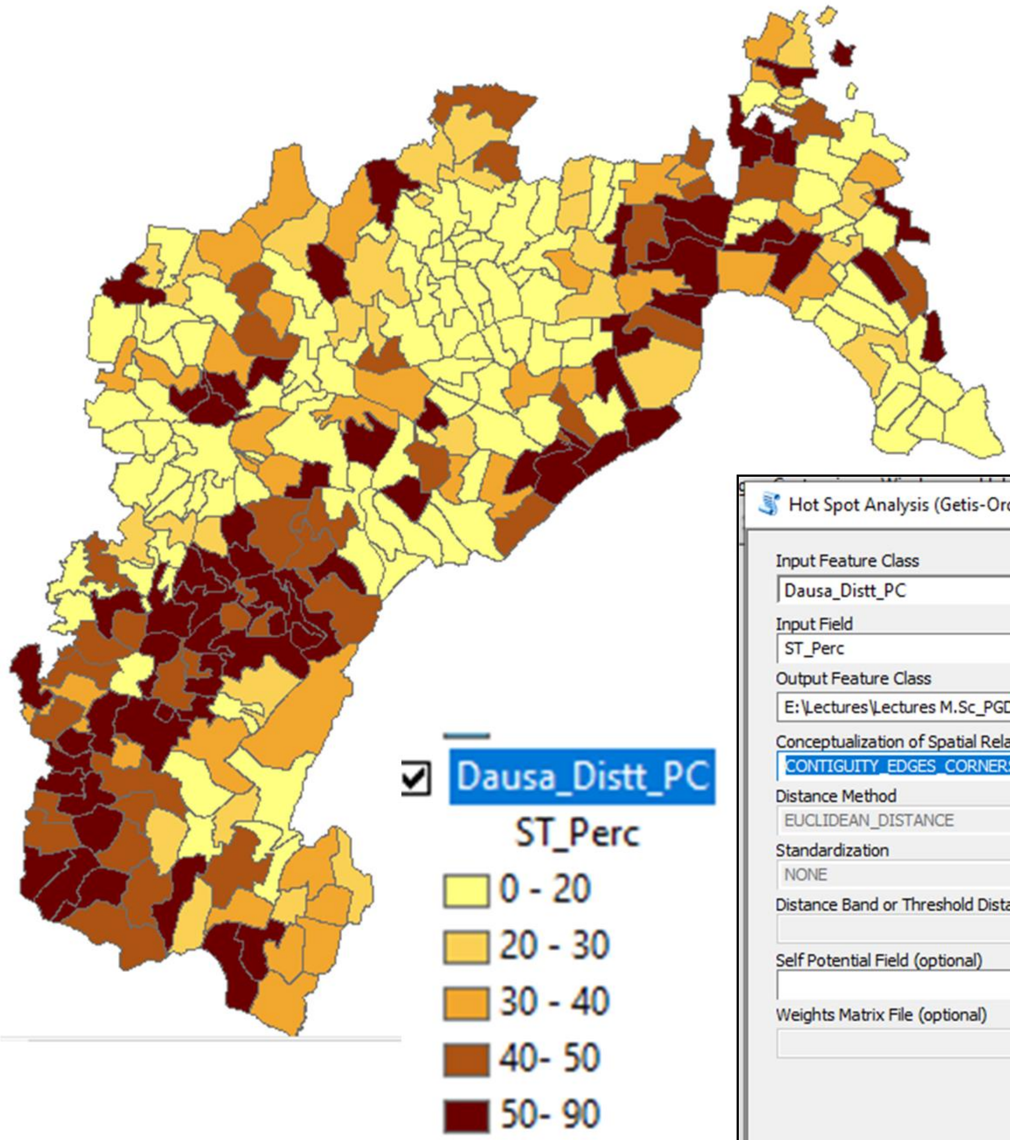
- Z-score is based on randomization null hypothesis computation
- Calculations based on either Euclidean / Manhattan distance require projected data to accurately measure distances
- Input field should contain a variety of values..i.e. some variation in the variable being analysed
- Input feature class should contain at least 30 features..if less results are not reliable
- Fixed distance band method for defining neighborhood is recommended
- Ensure all features should have at least one neighbor and no feature should have all other features as neighbors (8 neighbors optimum)

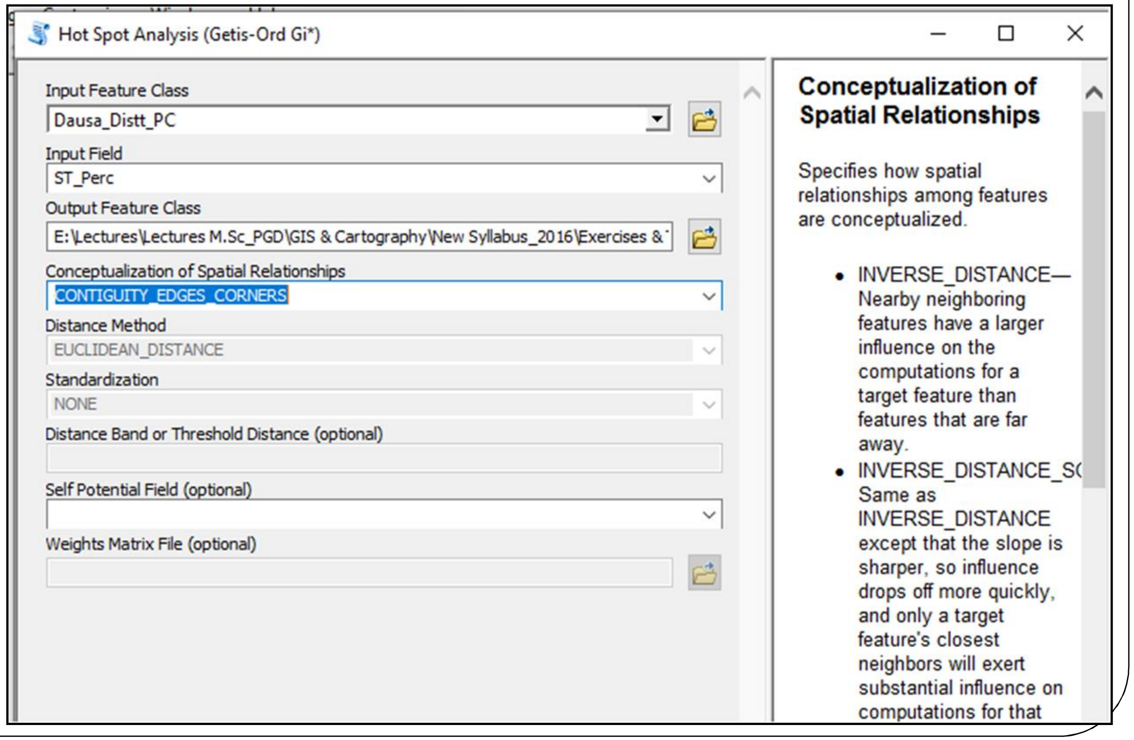
Hot Spot Analysis (Getis- Ord Gi*)….answers questions like

- Where is the disease outbreak concentrated
- Where should evacuation sites be located
- Where do peak intensities occur
- Crime analysis
- Voting pattern analysis
- Retail analysis
- Traffic incident analysis
- Demographics…..

Hot Spot Analysis – Gi *

Is ST population clustered at PC level in Dausa district? Where are the hot spots and cold spots located?

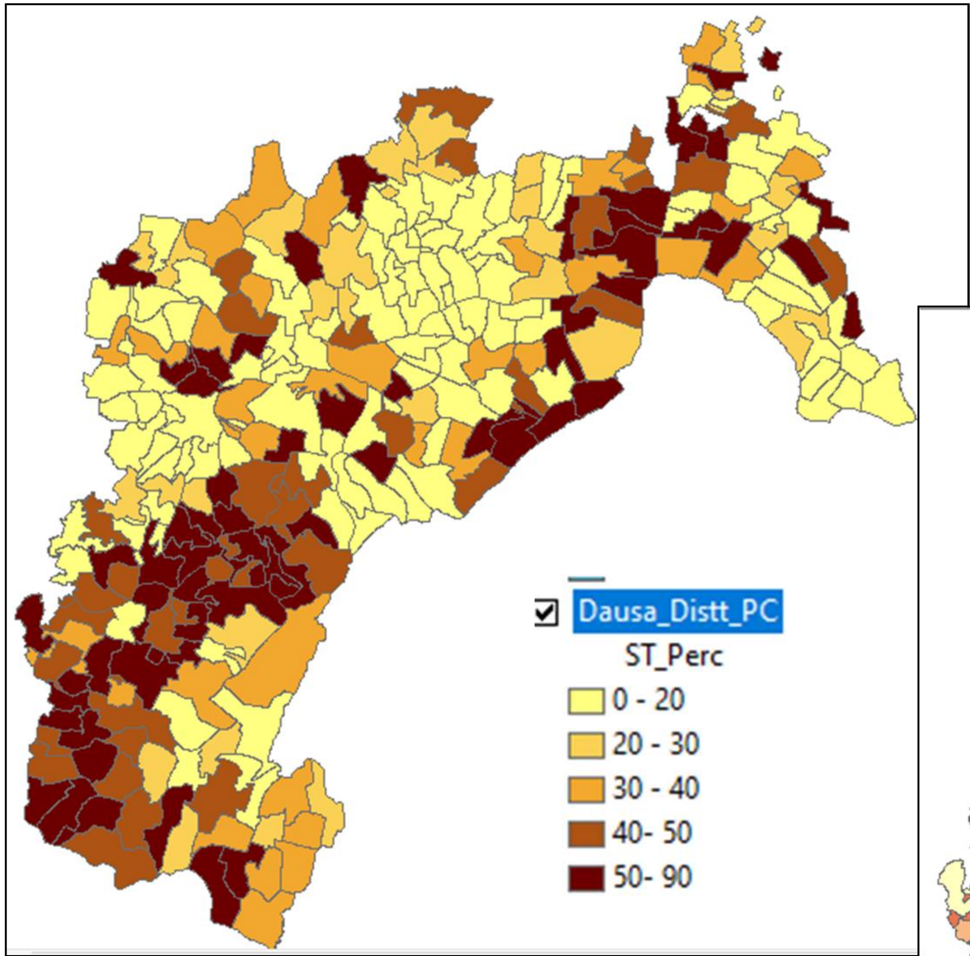


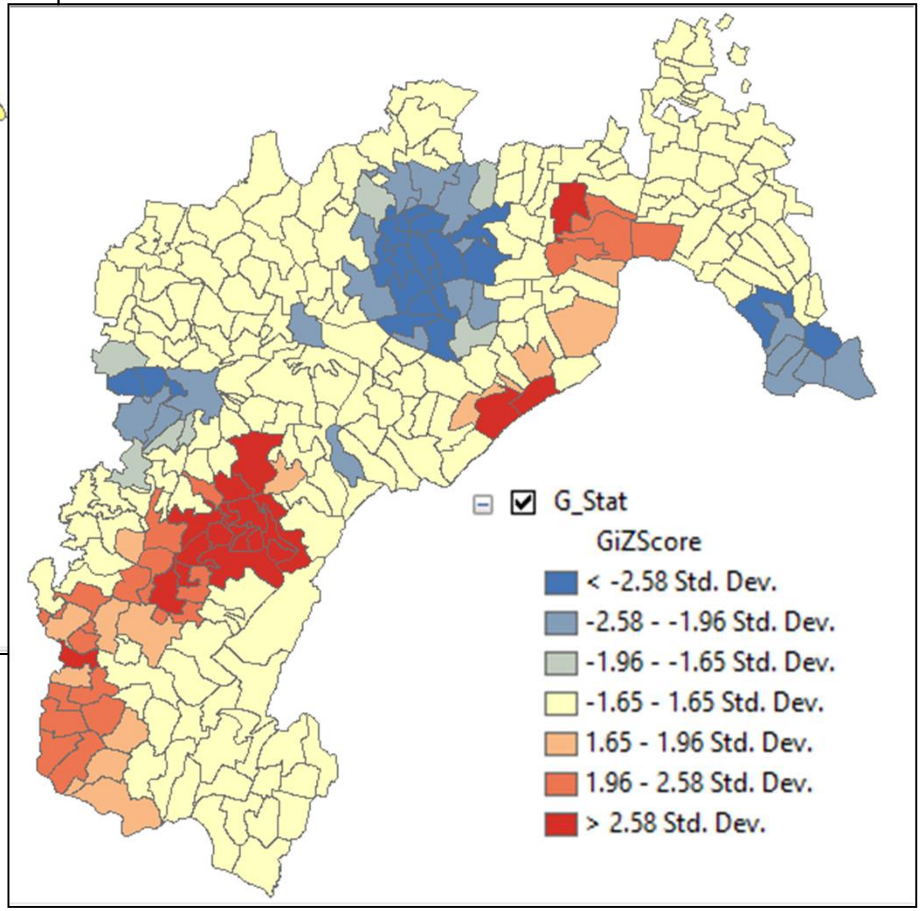


Hot Spot Analysis – Gi *

Correlate the output

1. Z-Score statistic
2. Proportions





Hot Spot Analysis – Gi *

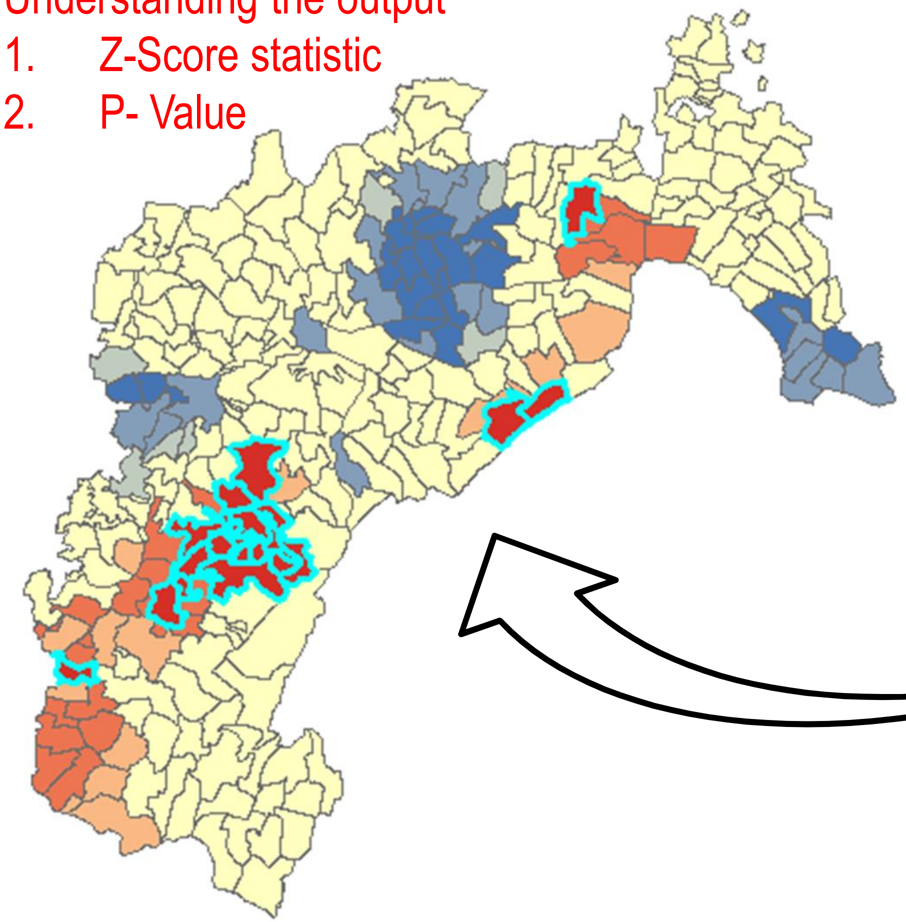
HOT SPOTS

Table

G_Stat

Understanding the output

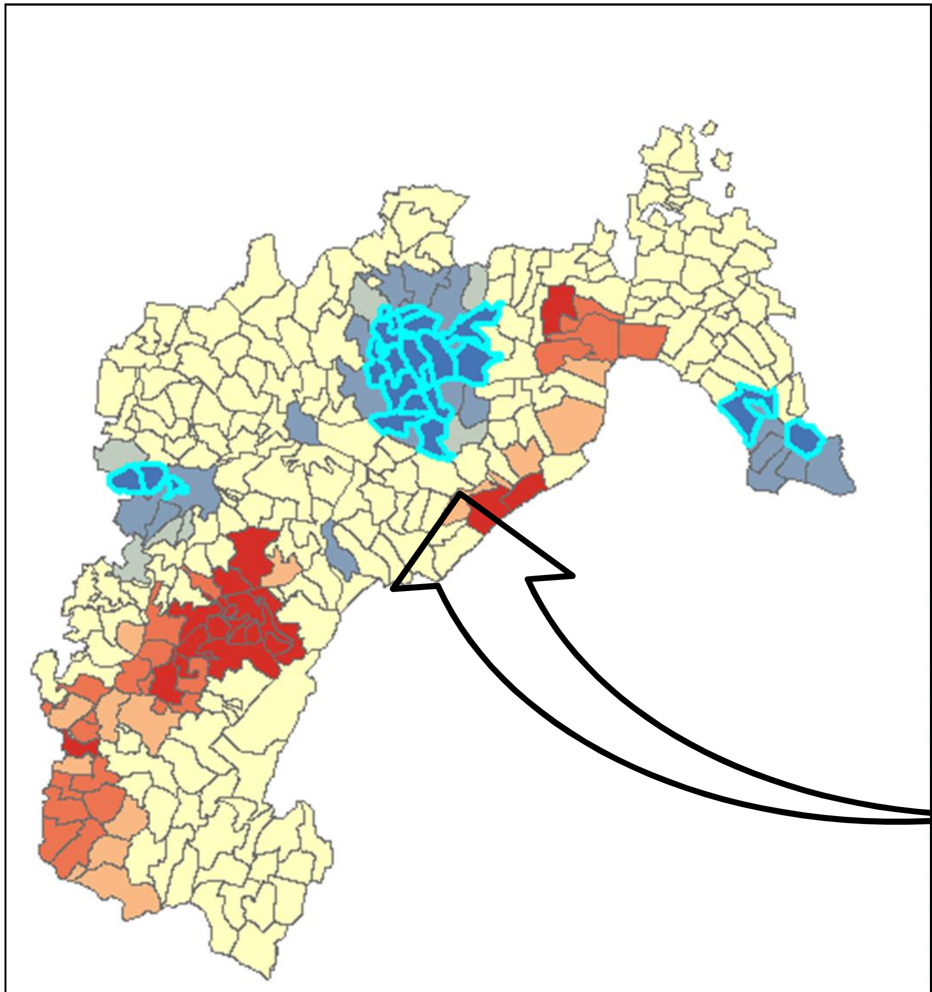
1. Z-Score statistic
2. P- Value



			ST_Perc	Shape_Length	Shape_Area	GiZScore	GiPValue
			49.407232	29004.710428	12166263.141749	4.285298	0.000018
			54.212734	32417.574101	25349397.926074	4.26253	0.00002
			69.510246	26932.823089	14077168.318564	4.110938	0.000039
			47.206704	20415.354147	13025173.112256	4.009881	0.000061
			79.835813	17846.016109	5689192.994624	3.[illegible]45844	0.00012
			69.982238	21436.555065	6582513.285992	3.601715	0.000316
			79.132131	32252.291226	15507033.616542	3.[illegible]55982	0.000377
			71.578947	17319.552104	8879928.601853	3.504818	0.000457
			65.433071	27372.964476	13186790.232074	3.45291	0.000555
			65.662799	14469.347397	9329244.971729	3.439343	0.000583
			74.401009	21475.557713	17893725.874319	3.063642	0.002187
			45.916203	25143.632213	24490633.[illegible]	2.955908	0.003117
			52.430439	16884.385176	7666229.698409	2.773227	0.00555
			48.207171	11535.754336	4020790.220803	2.751735	0.005928
			80.601915	15141.712241	10911268.851419	2.748234	0.005992
			47.826987	22862.401385	10183211.845229	2.668544	0.007618
			60.244648	22534.777021	18005655.705163	2.571326	0.010131
			79.143059	1457[illegible].687477	6690271.432707	2.519354	0.011757
			35.403283	20973.239458	16226703.202322	2.477782	0.01322
			[illegible].88947	27812.657321	22096151.790861	2.455708	0.014061
			58.885991	22308.665214	11108990.022971	2.335082	0.019539
			83.756039	18063.435892	9478294.1892	2.325976	0.02002
			54.67677	25877.301244	15915047.241227	2.299904	0.021454
			48.420374	18799.791035	14217803.737747	2.243205	0.024884
			51.725377	19318.128224	11573470.490272	2.23697	0.025288
			87.021778	19739.470256	11448223.303489	2.196074	0.028087
80	Polygon	80	36.416926	18445.78244	18910174.648378	2.145212	0.031936
106	Polygon	106	69.365976	18858.13475	14047371.826117	2.127479	0.03338

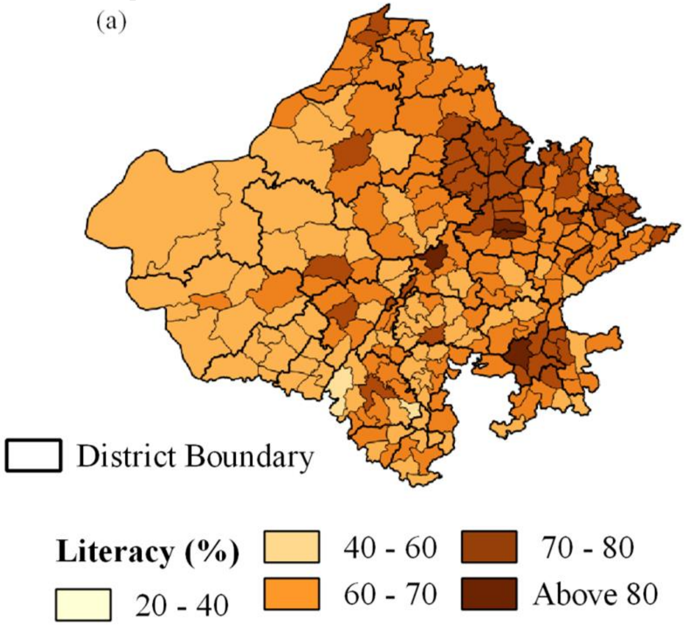
Hot Spot Analysis – Gi *

COLD SPOTS

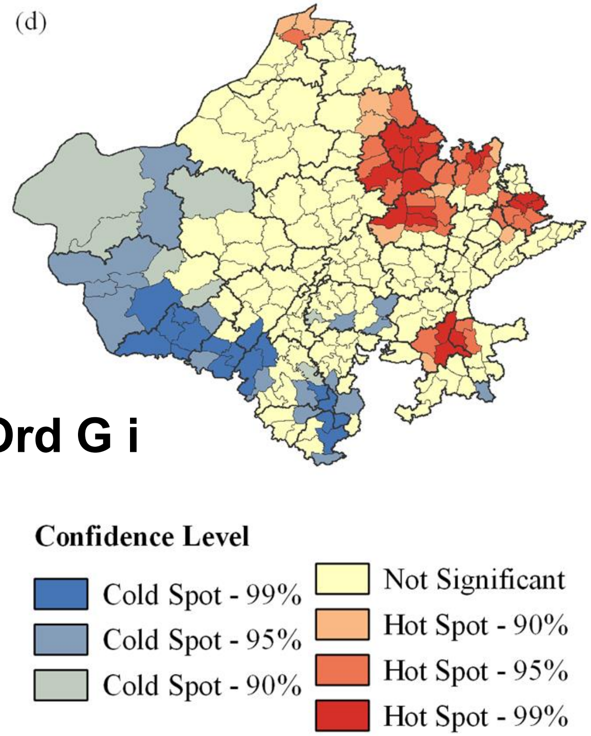


			ST_Perc	Shape_Length	Shape_Area	GiZScore	GiPValue
			2.449568	16987.06963	14488798.862213	-3.938938	0.000082
			3.501229	16651.093066	10136703.518761	-3.68088	0.000232
			6.976296	23324.298601	13313991.968128	-3.650818	0.000261
			1.772368	14155.800766	7771297.669045	-3.56045	0.00037
			7.092639	13546.792085	11249845.358022	-3.43536	0.000592
			0.580679	12178.768679	6670908.727465	-3.337357	0.000846
			1.930922	12416.074561	6526576.831567	-3.180992	0.001468
			0	18041.494005	10451209.127572	-3.155092	0.001604
			3.769075	17817.744173	8657697.442845	-2.907973	0.003638
			2.855794	15158.933978	4287036.359701	-2.837501	0.004547
			3.620743	7752.930175	3012860.569562	-2.803299	0.005058
			13.021529	17343.205498	11687254.609441	-2.802455	0.005072
			1.84551	10942.313125	6507056.383078	-2.736644	0.006207
			7.417195	26019.491328	13714340.178785	-2.678628	0.007392
			6.06142	11215.734779	7792621.980282	-2.656982	0.007884
			24.609233	28371.011594	19278469.513415	-2.640773	0.008272
			0	22597.539192	11655456.575578	-2.639362	0.008306
			5.135428	16820.441324	7942626.318896	-2.625605	0.00865
			0	17963.101143	15831125.956706	-2.620643	0.008776
179	Polygon	179	2.117877	27170.545938	21652729.107074	-2.536872	0.011185
224	Polygon	224	0.170406	19307.577992	12544165.784339	-2.532722	0.011318
176	Polygon	176	0	11959.87354	8641612.809003	-2.530141	0.011402
25	Polygon	25	0	13186.784122	9406192.220536	-2.506619	0.012189
7	Polygon	7	0.259423	16391.186078	9839672.317832	-2.487199	0.012875

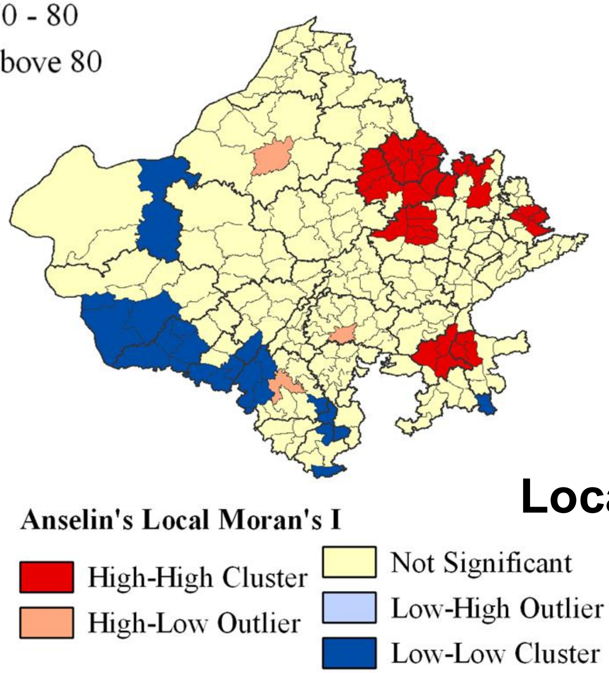
Spatial Pattern of Total Literacy in Rajasthan



Choropleth



Getis Ord G i



Local Moran's I