11 –Analyzing Spatial Patterns

Ness Sandoval
Sociology
Saint Louis University

Outline

Average Nearest Neighbor

Ripley's K function

Kernel Density Estimation

General G Statistic

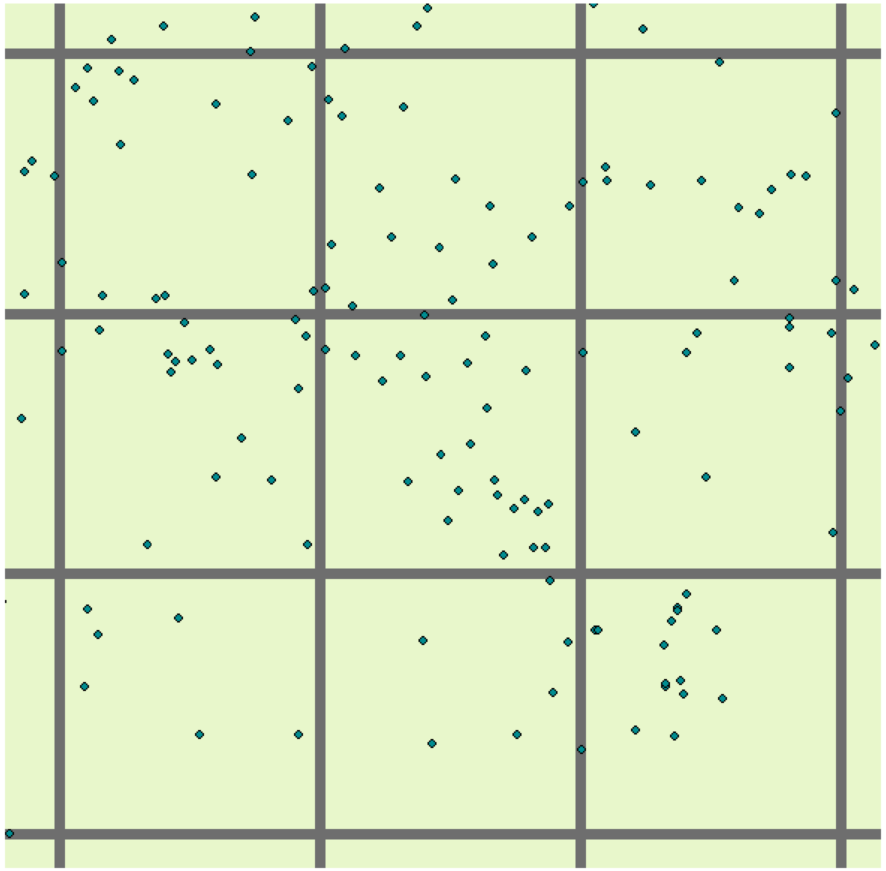
Hot-Spot Analysis Gi*

Lab Exercise

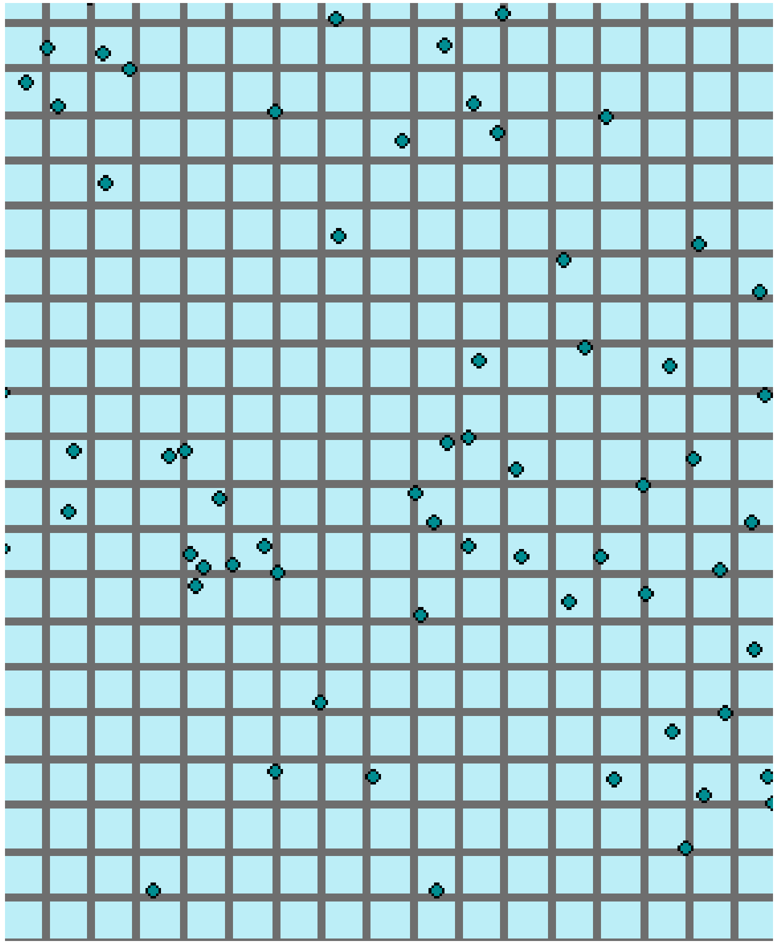
Point Pattern Analysis

- Point Association
  - Nearest Neighbor Analysis

- Point Density
  - Quadrat Analysis
  - Modifiable Areal Unit Problem



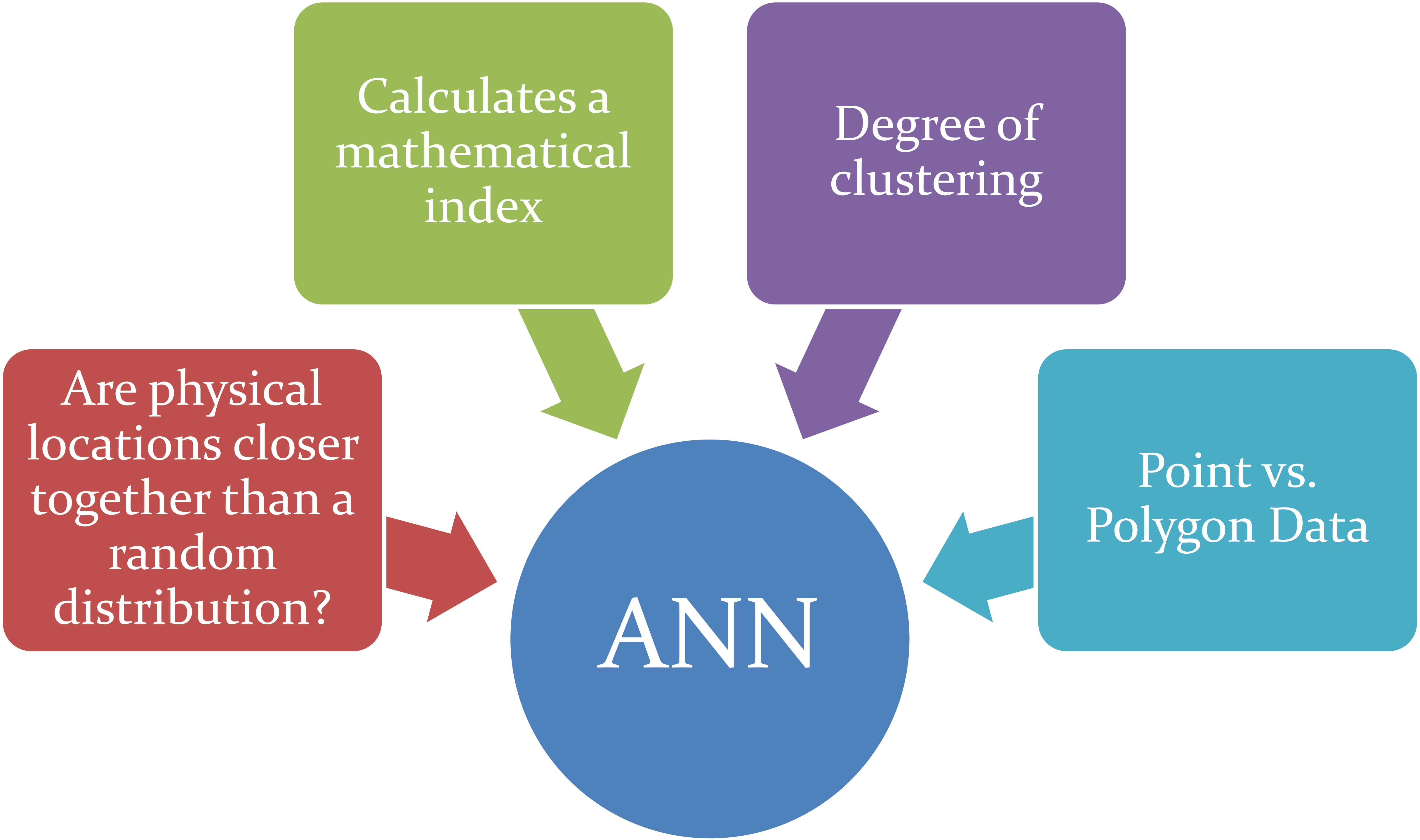
May be too big
1000 meters



May be too small
100 meters

Average Nearest Neighbor

Cluster or Dispersion



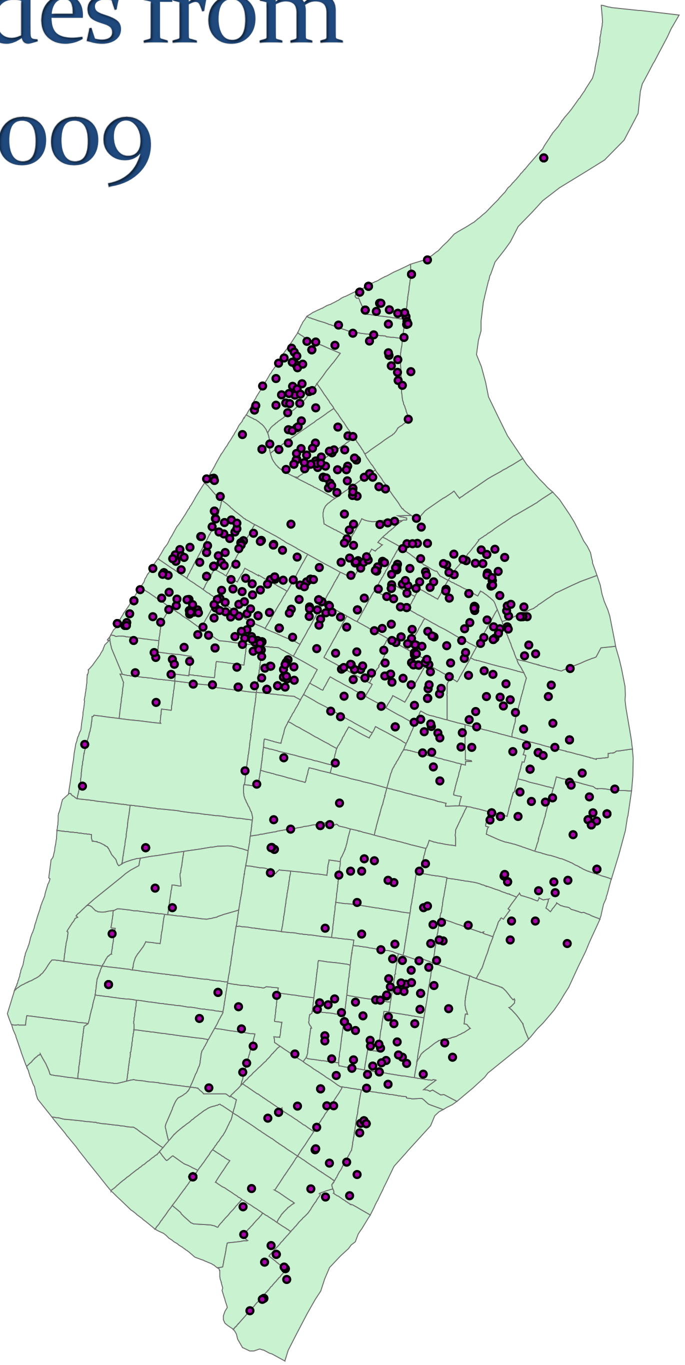
Calculates a mathematical index
Degree of clustering
Are physical locations closer together than a random distribution?
Point vs. Polygon Data
ANN

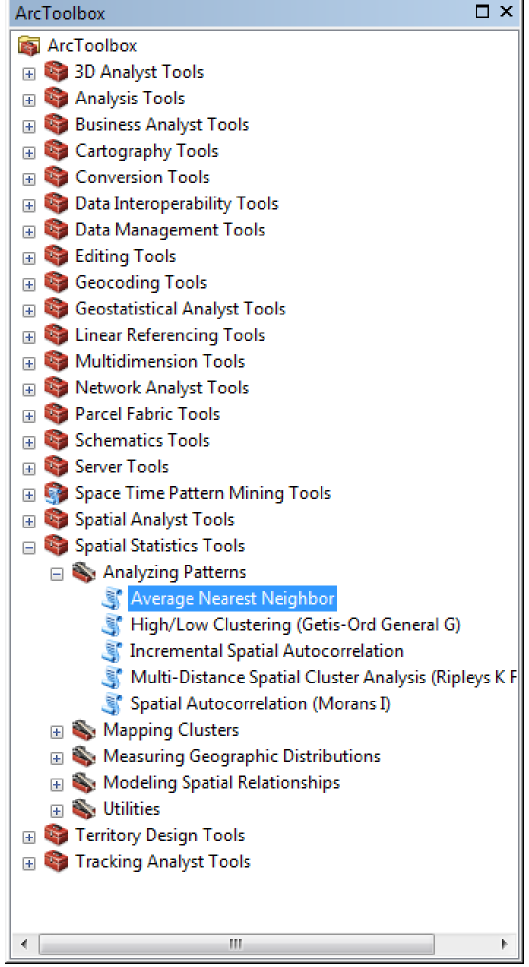
Average nearest neighbor (ANN)

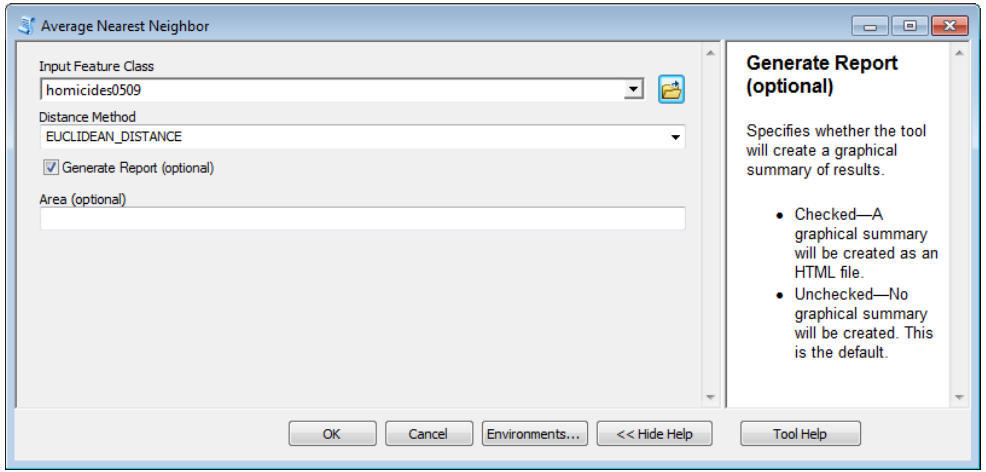
The ANN index tool will calculate a z-score which can be used to reject or fail to reject Ho.

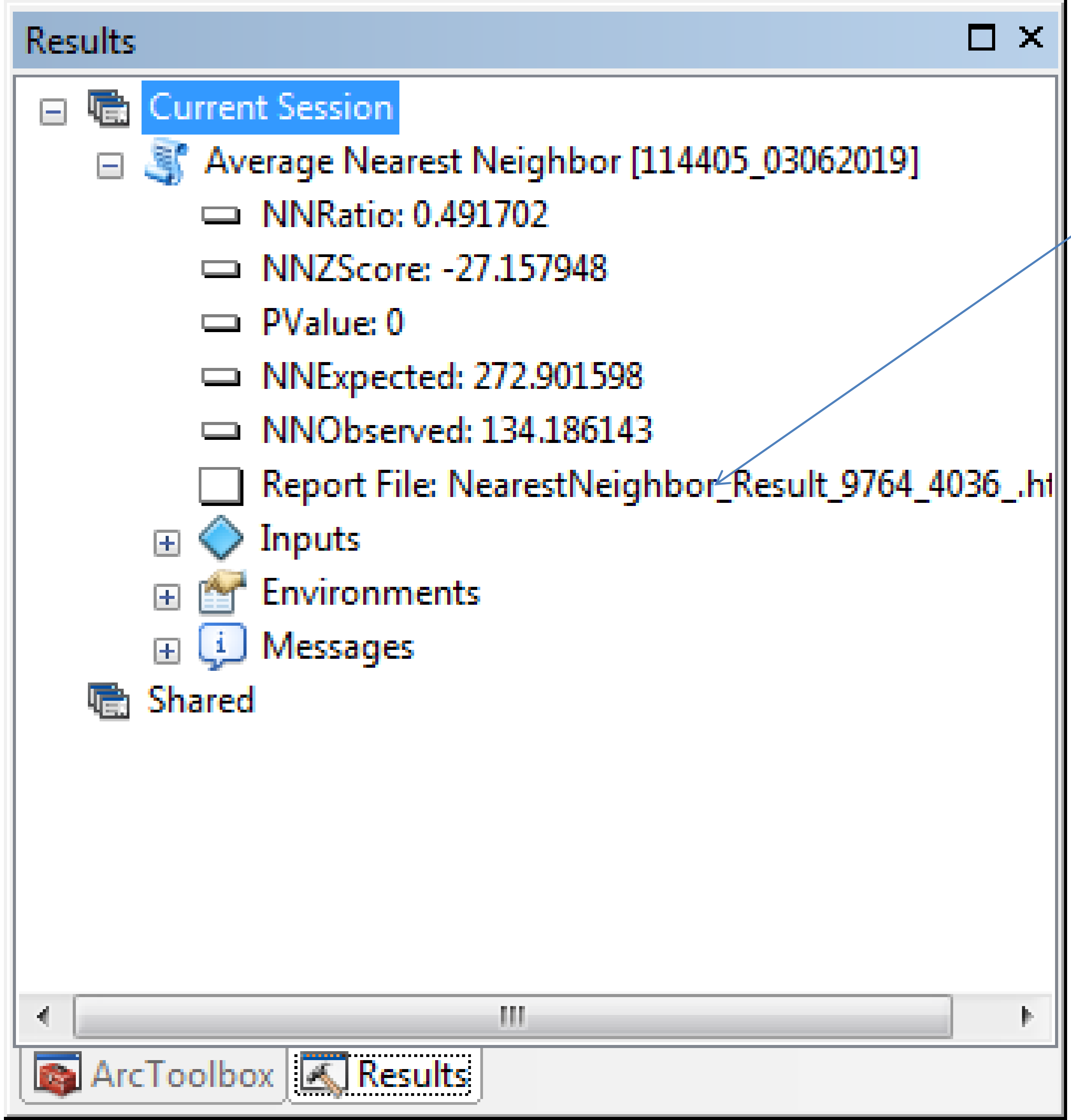
- Ho= There is no pattern
- H1=There is a pattern

Homicides from 2005 - 2009



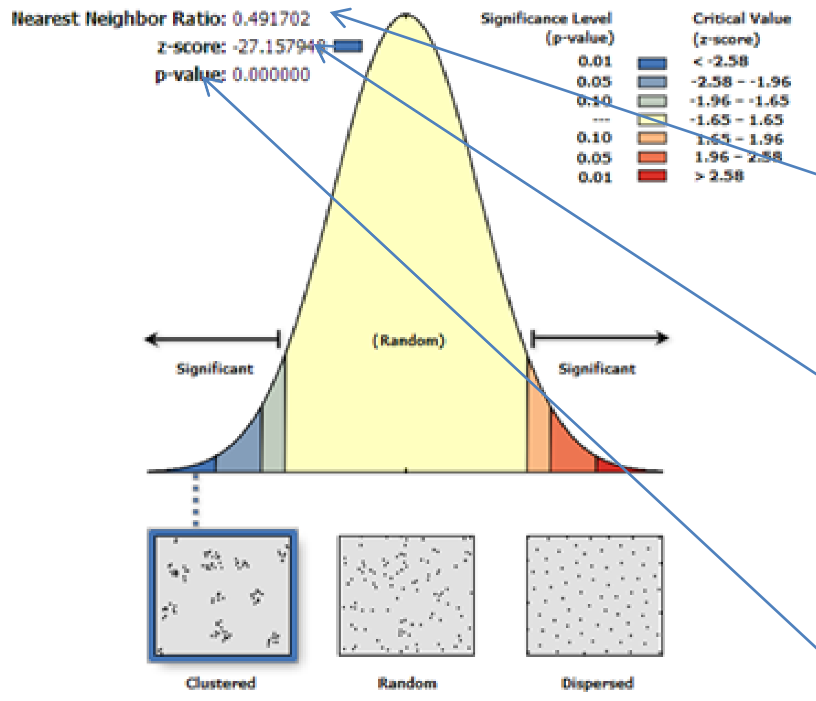






Click on the
Report File

Average Nearest Neighbor Summary



Given the z-score of -27.16, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Average Nearest Neighbor Summary

Observed Mean Distance:	134.186143 Meters
Expected Mean Distance:	272.901598 Meters
Nearest Neighbor Ratio:	0.491702
z-score:	-27.157948
p-value:	0.000000

Dataset Information

Input Feature Class:	homicides
Distance Method:	EUCLIDEAN
Study Area:	232362880.085246
Selection Set:	False

Important Points for observation

- The dataset returned an index of .49
- This tells us that the features are trending toward clustering
- The z-score is -27.16 with a significance level of 0.01. This score gives us the confidence level of rejecting the null hypothesis.

Ripley's K function

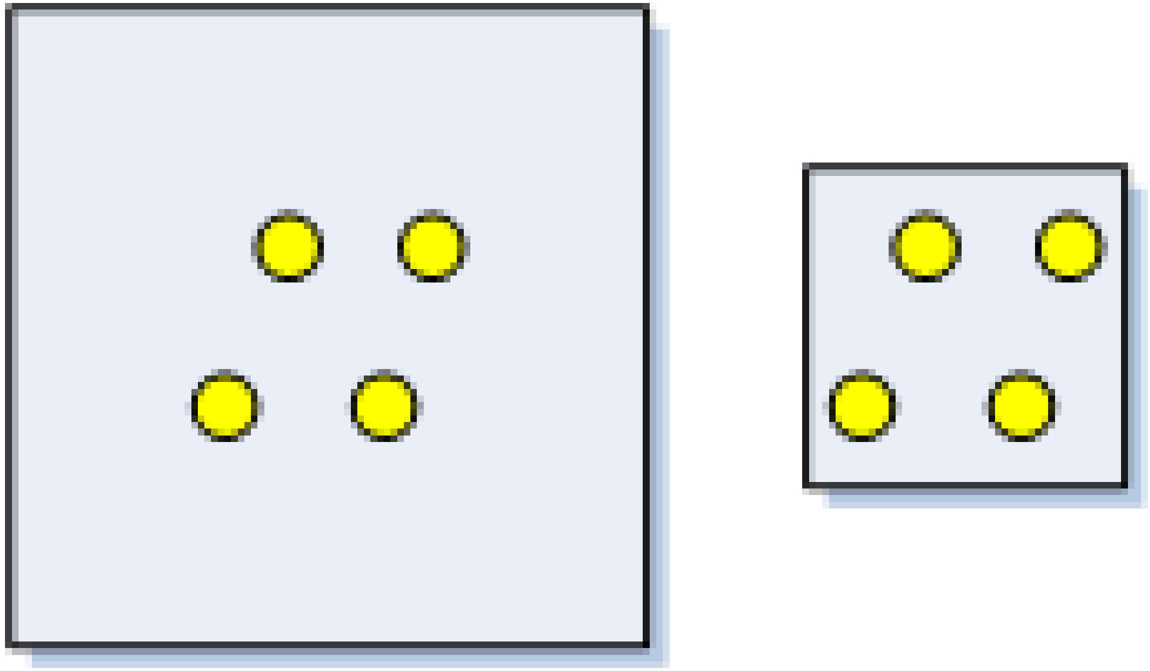
Multi-distance clustering

Ripley's K function Part 1

- Ripley's K function is a method of pattern analysis.
  - Always uses points
  - Determines if there spatial pattern that is statistically significant clustering or dispersion over a range of distances.
  - The difference between ANN, is that Ripley's K function includes all the neighboring features in the calculation, not the just the nearest one.
- The index is calculated by measuring the distance from ach feature to all the other features in the dataset.
- Points at the edge of the study may have few neighbors
- Our goal is to find the point of most clustering.

Challenges – Part 1

- The k-function statistic is very sensitive to the size of the study area.
- Identical arrangements of points can exhibit clustering or dispersion depending on the size of the study area enclosing them.
- Therefore, it is imperative that the study area boundaries are carefully considered.
- The picture to the right is a classic example of how identical feature distributions can be dispersed or clustered depending on the study area specified.



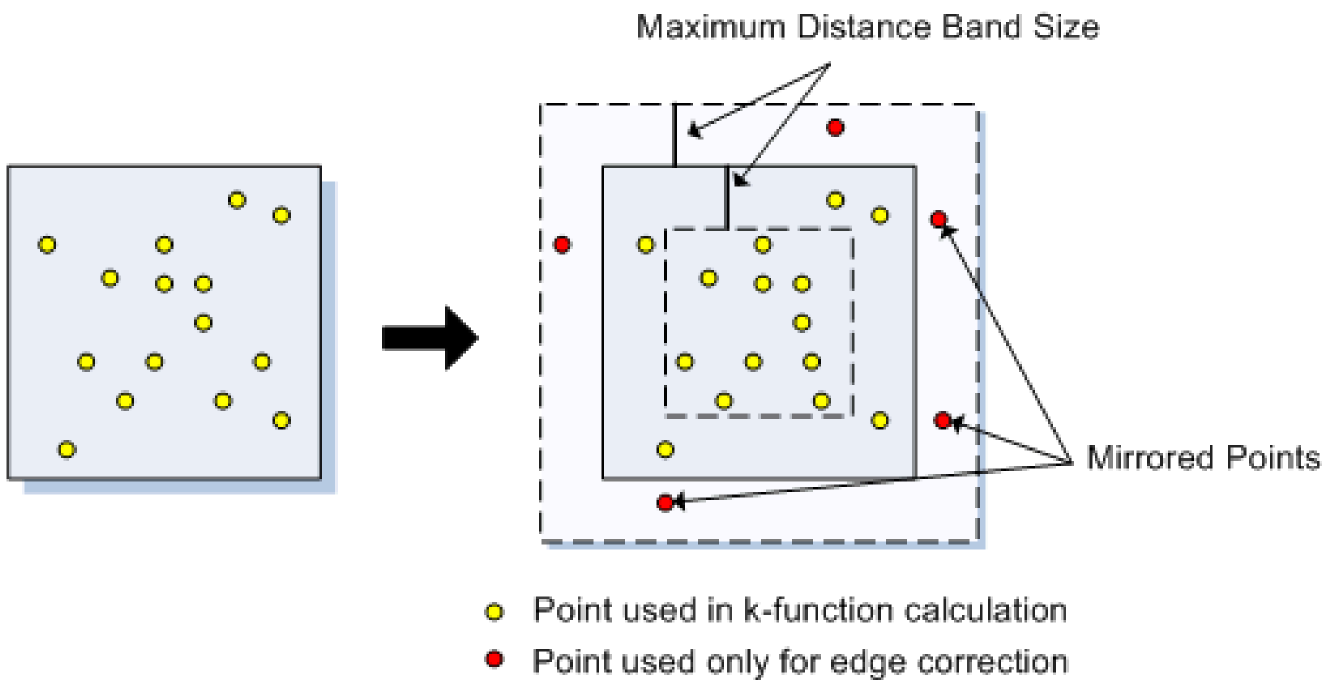
Solutions – Part 1

SIMULATE OUTER BOUNDARY VALUES

- This method creates points outside the study area boundary that mirror those found inside the boundary in order to correct for underestimates near the edges.

- Points that are within a distance equal to the maximum distance band of an edge of the study area are mirrored.

- The mirrored points are used so that edge points will have more accurate neighbor estimates.



This diagram illustrates what points will be used in the calculation and which will be used only for edge correction.

Solutions – Part 2

REDUCE ANALYSIS AREA

- This method shrinks the study area such that some points are found outside of the study area boundary.

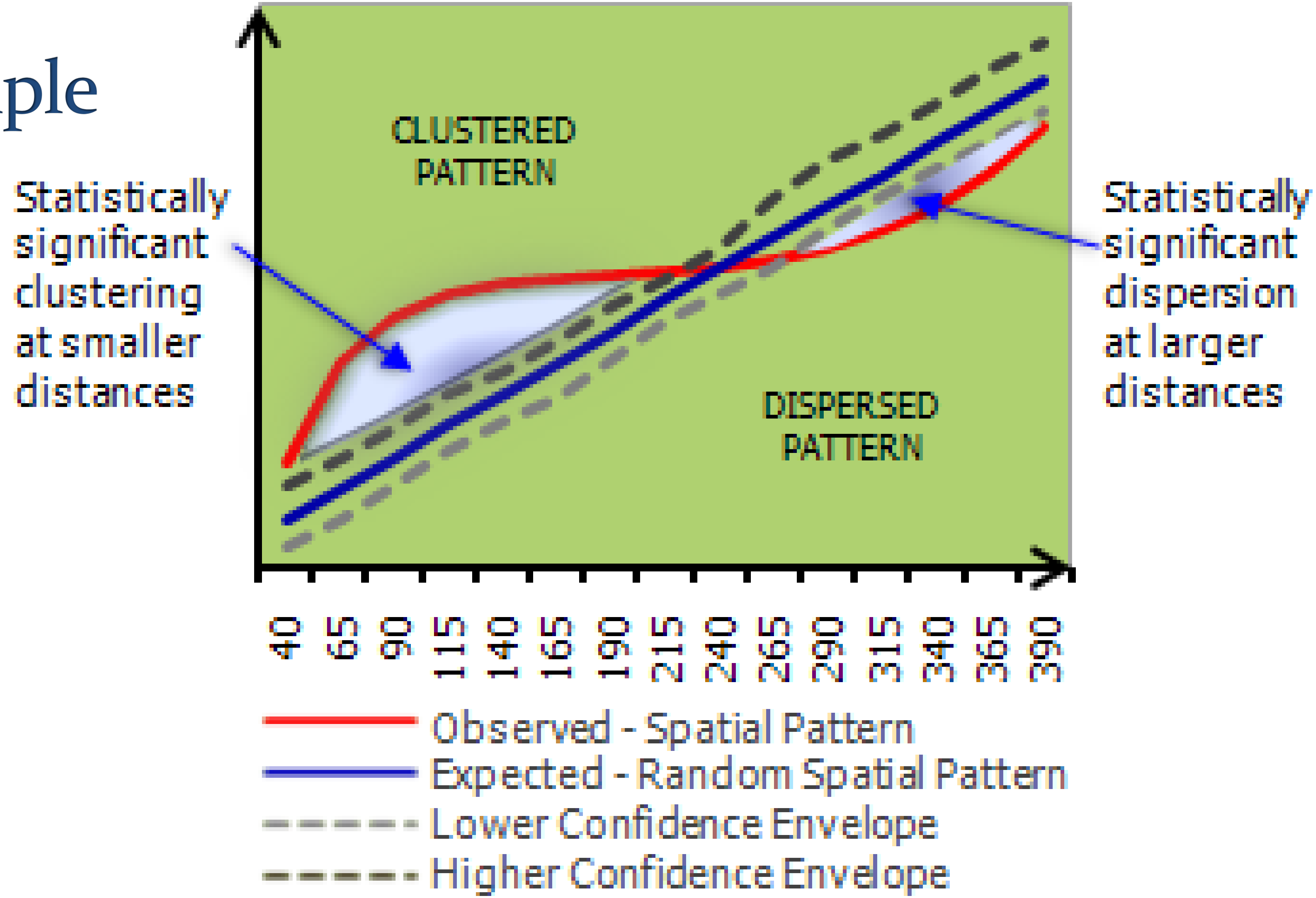
- Points found outside the study area are used to calculate neighbor counts but are not used in the cluster analysis itself.

Solutions – Part 3

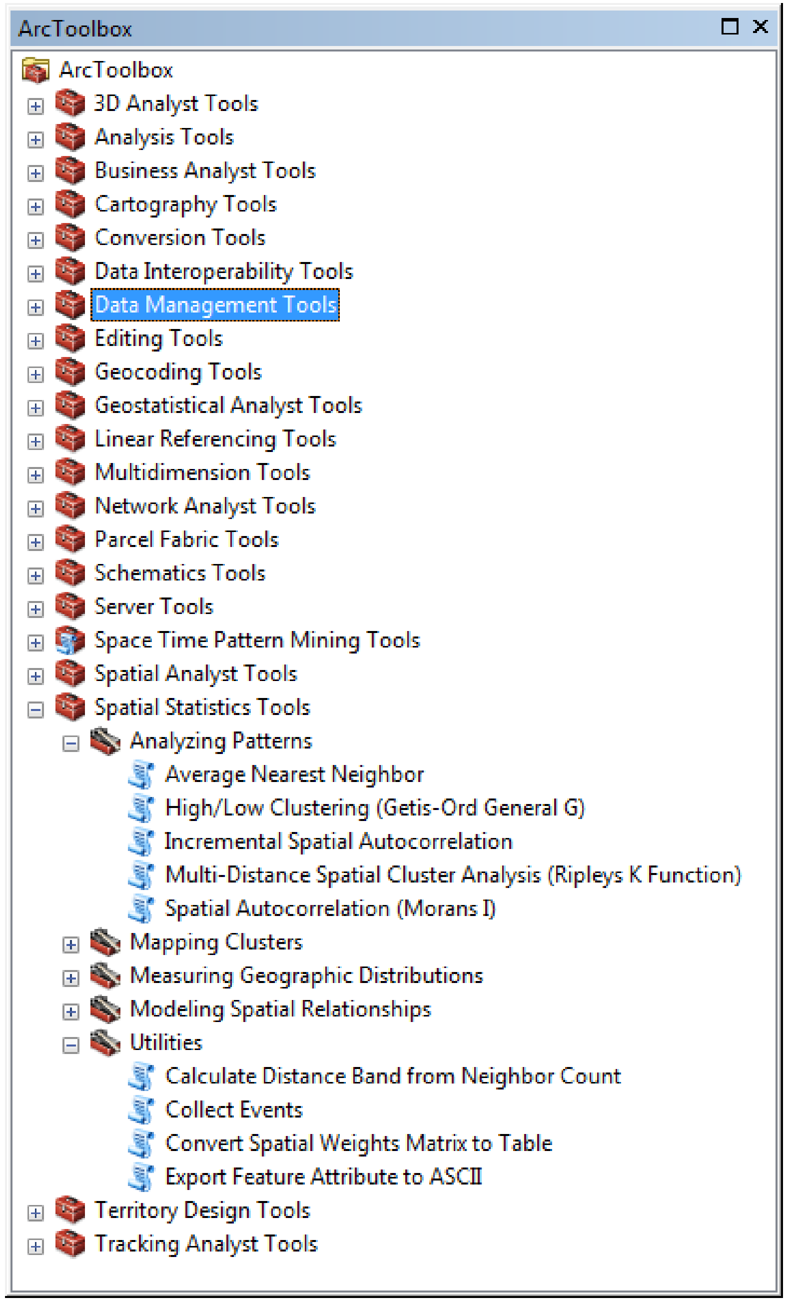
RIPLEY EDGE CORRECTION FORMULA

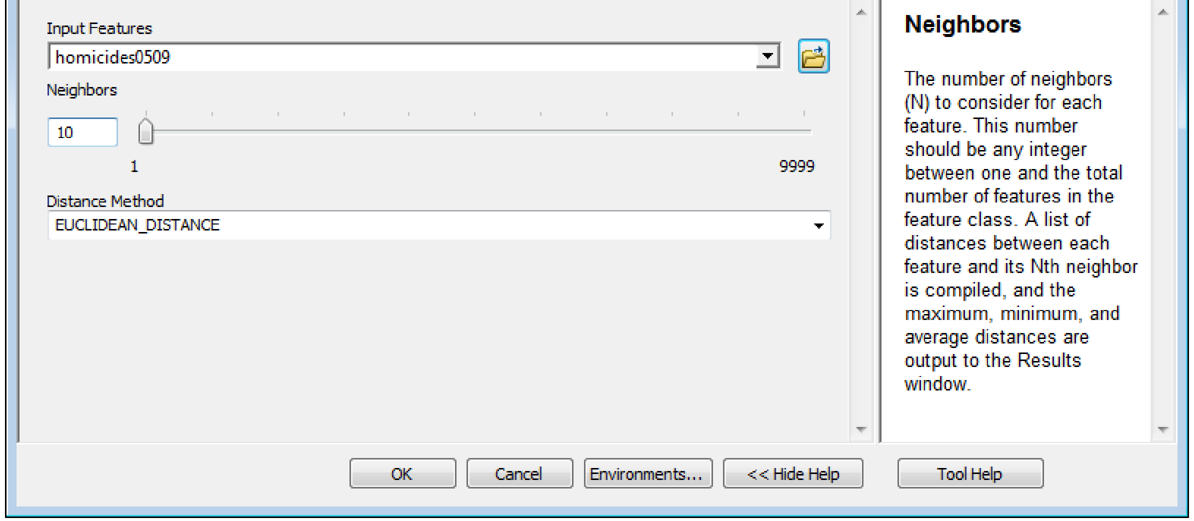
- This method checks each point's distance from the edge of the study area and its distance to each of its neighbors.
- All neighbors that are further away from the point in question than the edge of the study area are given extra weighting.
- This edge correction method is only appropriate for square or rectangular shaped study areas, or when you select MINIMUM_ENCLOSING_RECTANGLE for the Study Area Method parameter.

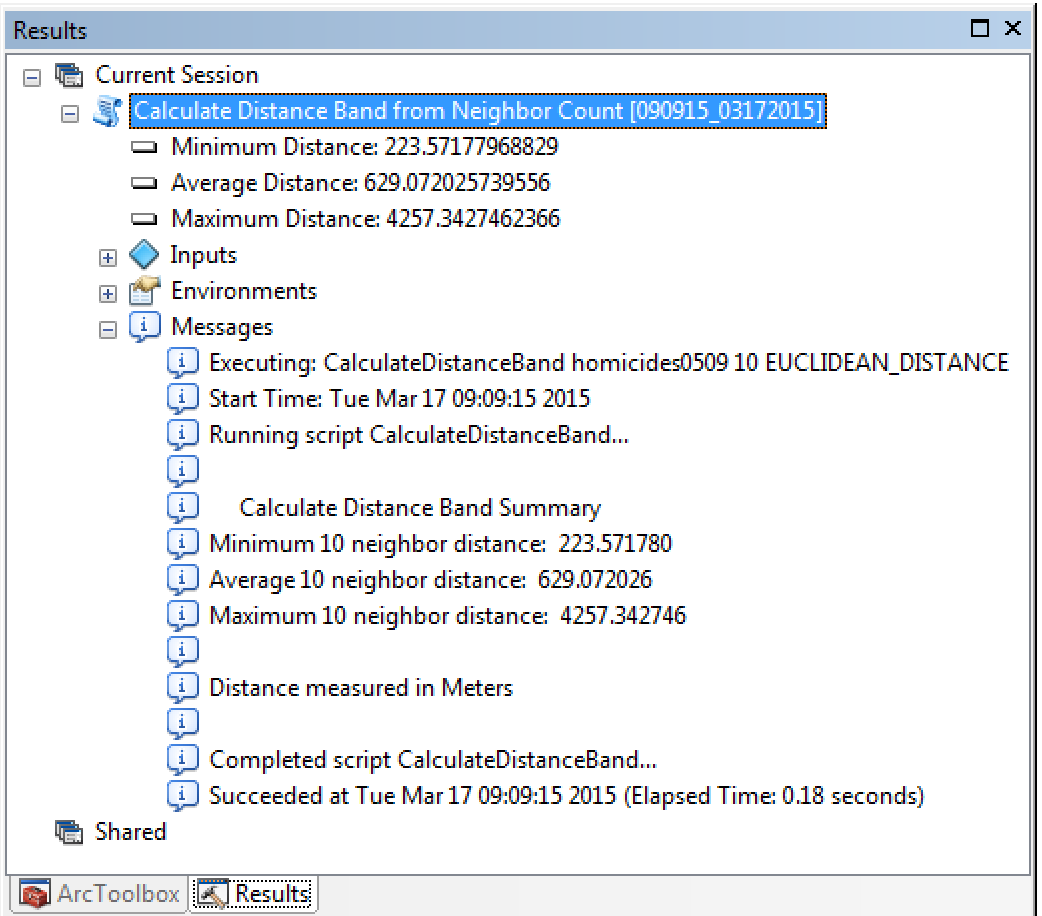
Example



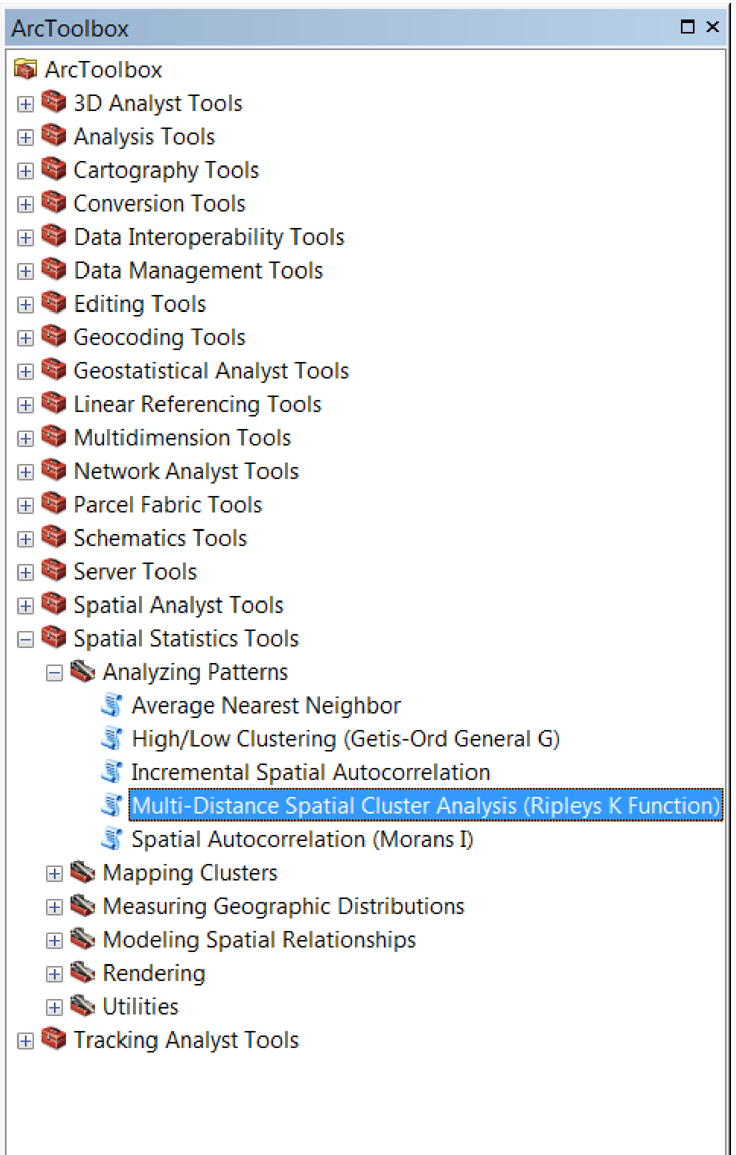
Statistically significant clustering at smaller distances
CLUSTERED PATTERN
DISPERSED PATTERN
Statistically significant dispersion at larger distances
40
65
90
115
140
165
190
215
240
265
290
315
340
365
390
Observed - Spatial Pattern
Expected - Random Spatial Pattern
Lower Confidence Envelope
Higher Confidence Envelope

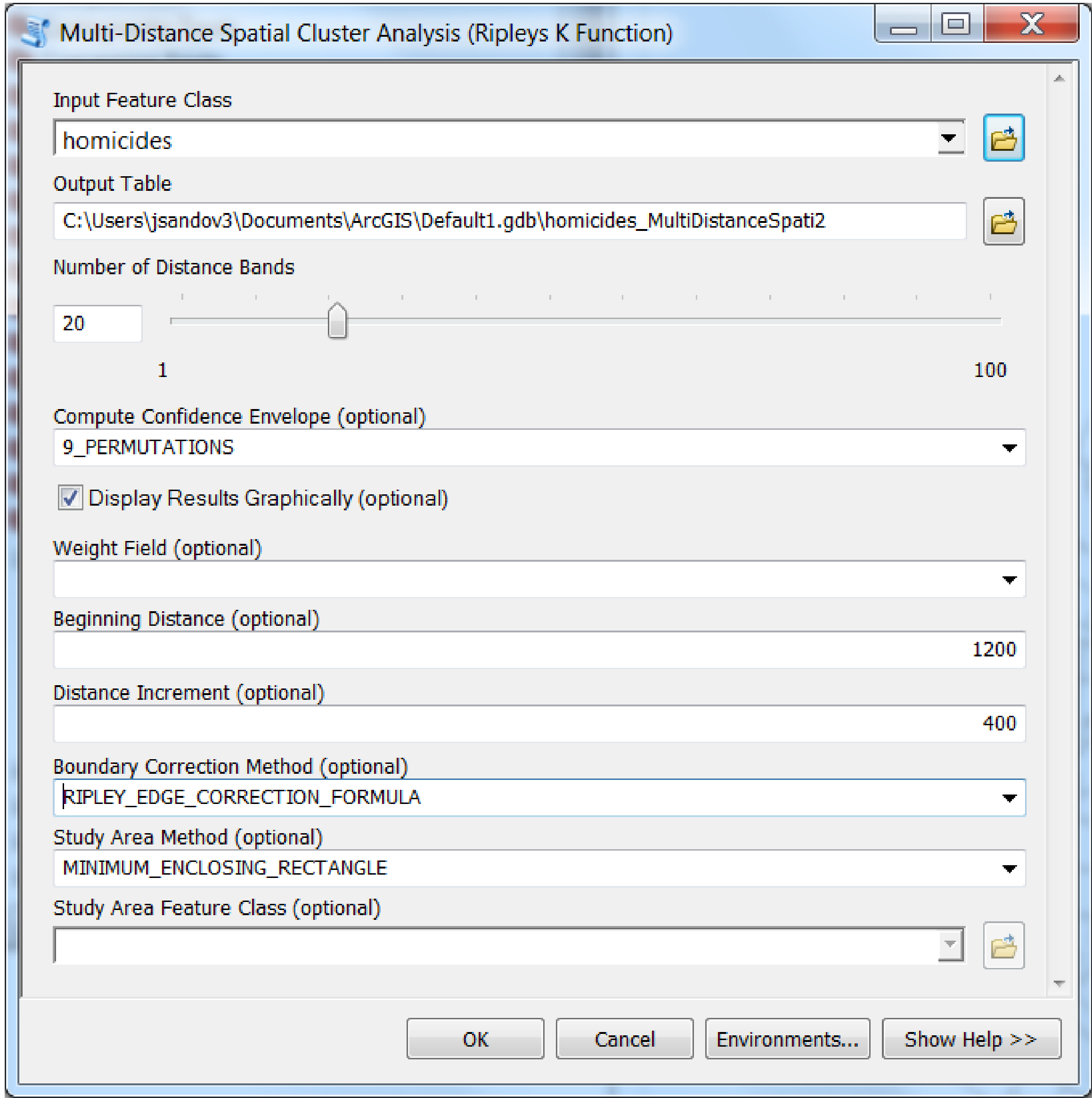

ArcToolbox
ArcToolbox
3D Analyst Tools
Analysis Tools
Business Analyst Tools
Cartography Tools
Conversion Tools
Data Interoperability Tools
Data Management Tools
Editing Tools
Geocoding Tools
Geostatistical Analyst Tools
Linear Referencing Tools
Multidimension Tools
Network Analyst Tools
Parcel Fabric Tools
Schematics Tools
Server Tools
Space Time Pattern Mining Tools
Spatial Analyst Tools
Spatial Statistics Tools
Analyzing Patterns
Average Nearest Neighbor
High/Low Clustering (Getis-Ord General G)
Incremental Spatial Autocorrelation
Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
Spatial Autocorrelation (Morans I)
Mapping Clusters
Measuring Geographic Distributions
Modeling Spatial Relationships
Utilities
Calculate Distance Band from Neighbor Count
Collect Events
Convert Spatial Weights Matrix to Table
Export Feature Attribute to ASCII
Territory Design Tools
Tracking Analyst Tools

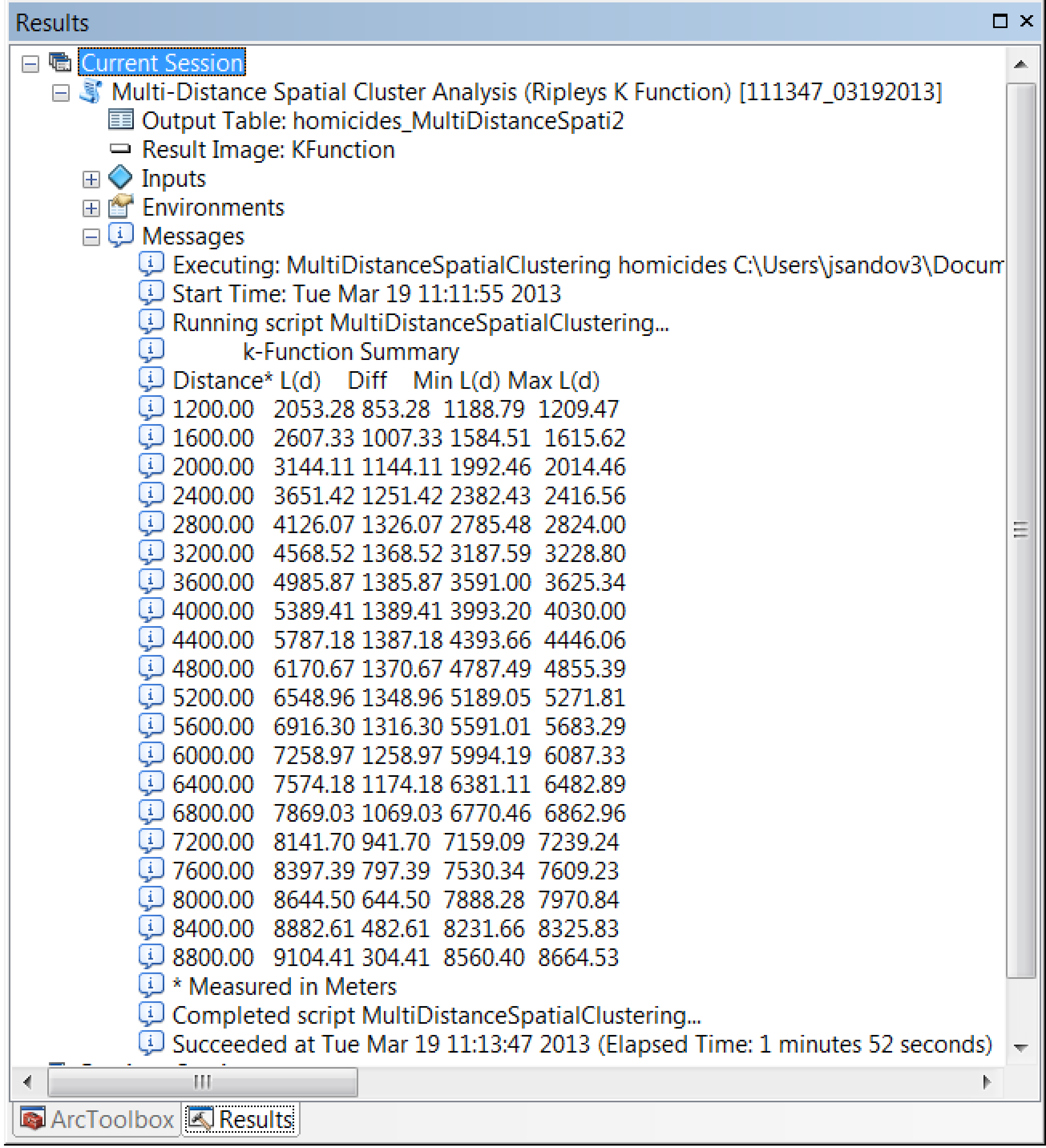

Input Features
homicides0509
Neighbors
10
1
9999
Distance Method
EUCLIDEAN_DISTANCE
Neighbors
The number of neighbors (N) to consider for each feature. This number should be any integer between one and the total number of features in the feature class. A list of distances between each feature and its Nth neighbor is compiled, and the maximum, minimum, and average distances are output to the Results window.
OK
Cancel
Environments...
<< Hide Help
Tool Help


Results
Current Session
Calculate Distance Band from Neighbor Count [090915_03172015]
Minimum Distance: 223.57177968829
Average Distance: 629.072025739556
Maximum Distance: 4257.3427462366
Inputs
Environments
Messages
Executing: CalculateDistanceBand homicides0509 10 EUCLIDEAN_DISTANCE
Start Time: Tue Mar 17 09:09:15 2015
Running script CalculateDistanceBand...
Calculate Distance Band Summary
Minimum 10 neighbor distance: 223.571780
Average 10 neighbor distance: 629.072026
Maximum 10 neighbor distance: 4257.342746
Distance measured in Meters
Completed script CalculateDistanceBand...
Succeeded at Tue Mar 17 09:09:15 2015 (Elapsed Time: 0.18 seconds)
Shared
ArcToolbox
Results

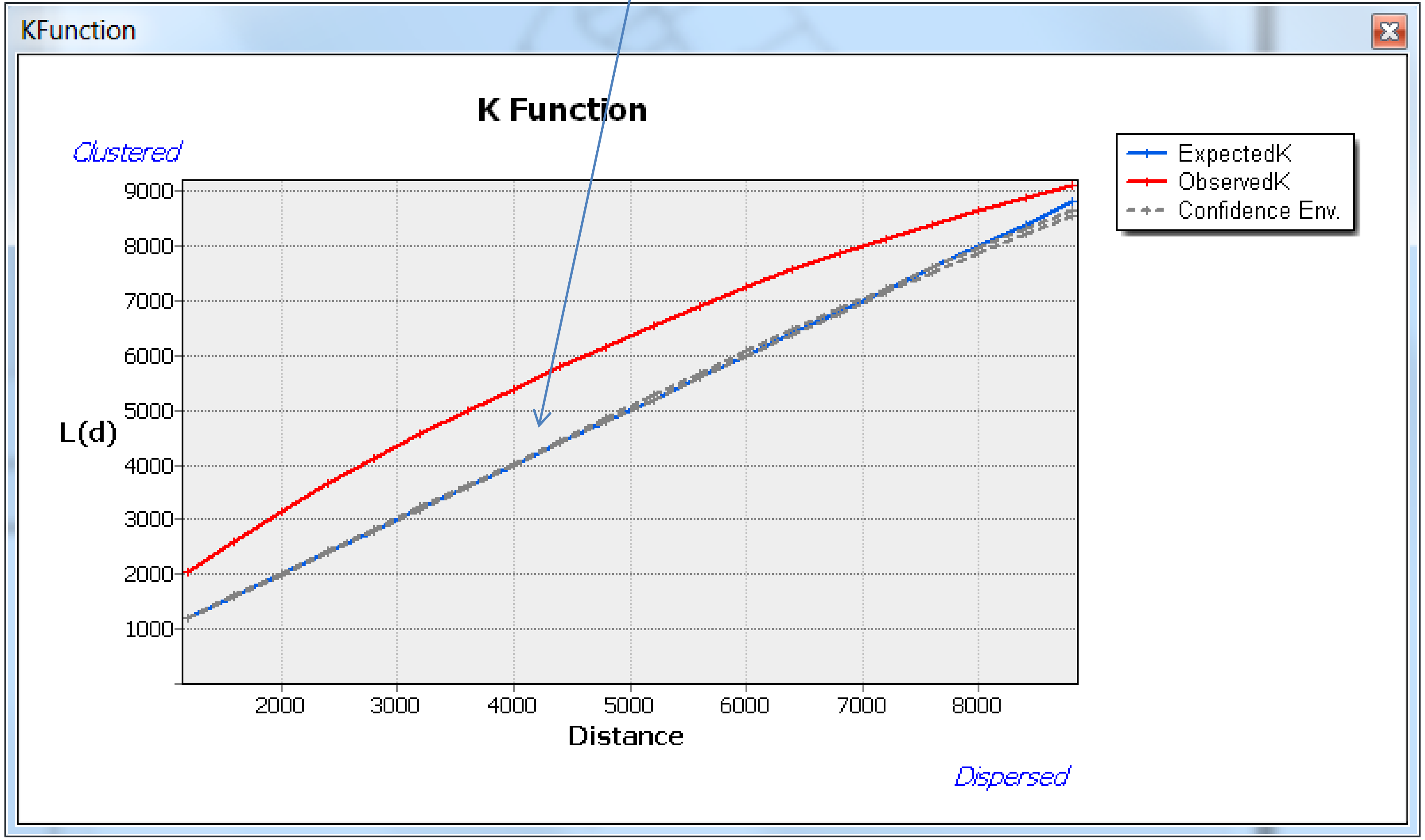
Ripley's K function Part 2

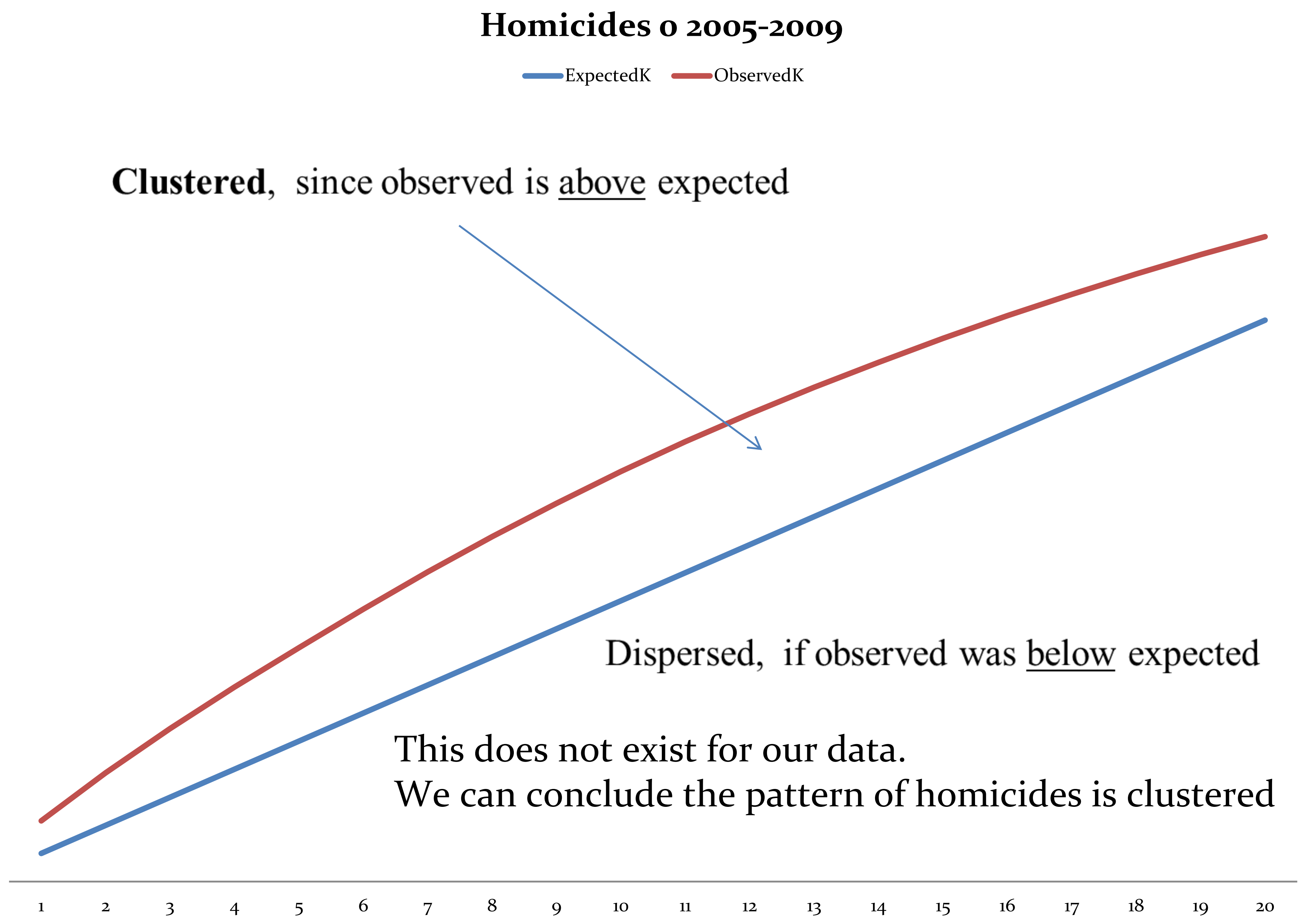





Results
Current Session
Multi-Distance Spatial Cluster Analysis (Ripleys K Function) [111347_03192013]
Output Table: homicides_MultiDistanceSpati2
Result Image: KFunction
Inputs
Environments
Messages
Executing: MultiDistanceSpatialClustering homicides C:\Users\jsandov3\Docum
Start Time: Tue Mar 19 11:11:55 2013
Running script MultiDistanceSpatialClustering...
k-Function Summary
Distance* L(d) Diff Min L(d) Max L(d)
1200.00 2053.28 853.28 1188.79 1209.47
1600.00 2607.33 1007.33 1584.51 1615.62
2000.00 3144.11 1144.11 1992.46 2014.46
2400.00 3651.42 1251.42 2382.43 2416.56
2800.00 4126.07 1326.07 2785.48 2824.00
3200.00 4568.52 1368.52 3187.59 3228.80
3600.00 4985.87 1385.87 3591.00 3625.34
4000.00 5389.41 1389.41 3993.20 4030.00
4400.00 5787.18 1387.18 4393.66 4446.06
4800.00 6170.67 1370.67 4787.49 4855.39
5200.00 6548.96 1348.96 5189.05 5271.81
5600.00 6916.30 1316.30 5591.01 5683.29
6000.00 7258.97 1258.97 5994.19 6087.33
6400.00 7574.18 1174.18 6381.11 6482.89
6800.00 7869.03 1069.03 6770.46 6862.96
7200.00 8141.70 941.70 7159.09 7239.24
7600.00 8397.39 797.39 7530.34 7609.23
8000.00 8644.50 644.50 7888.28 7970.84
8400.00 8882.61 482.61 8231.66 8325.83
8800.00 9104.41 304.41 8560.40 8664.53
* Measured in Meters
Completed script MultiDistanceSpatialClustering...
Succeeded at Tue Mar 19 11:13:47 2013 (Elapsed Time: 1 minutes 52 seconds)
ArcToolbox
Results

This indicates clustering at the low and high end



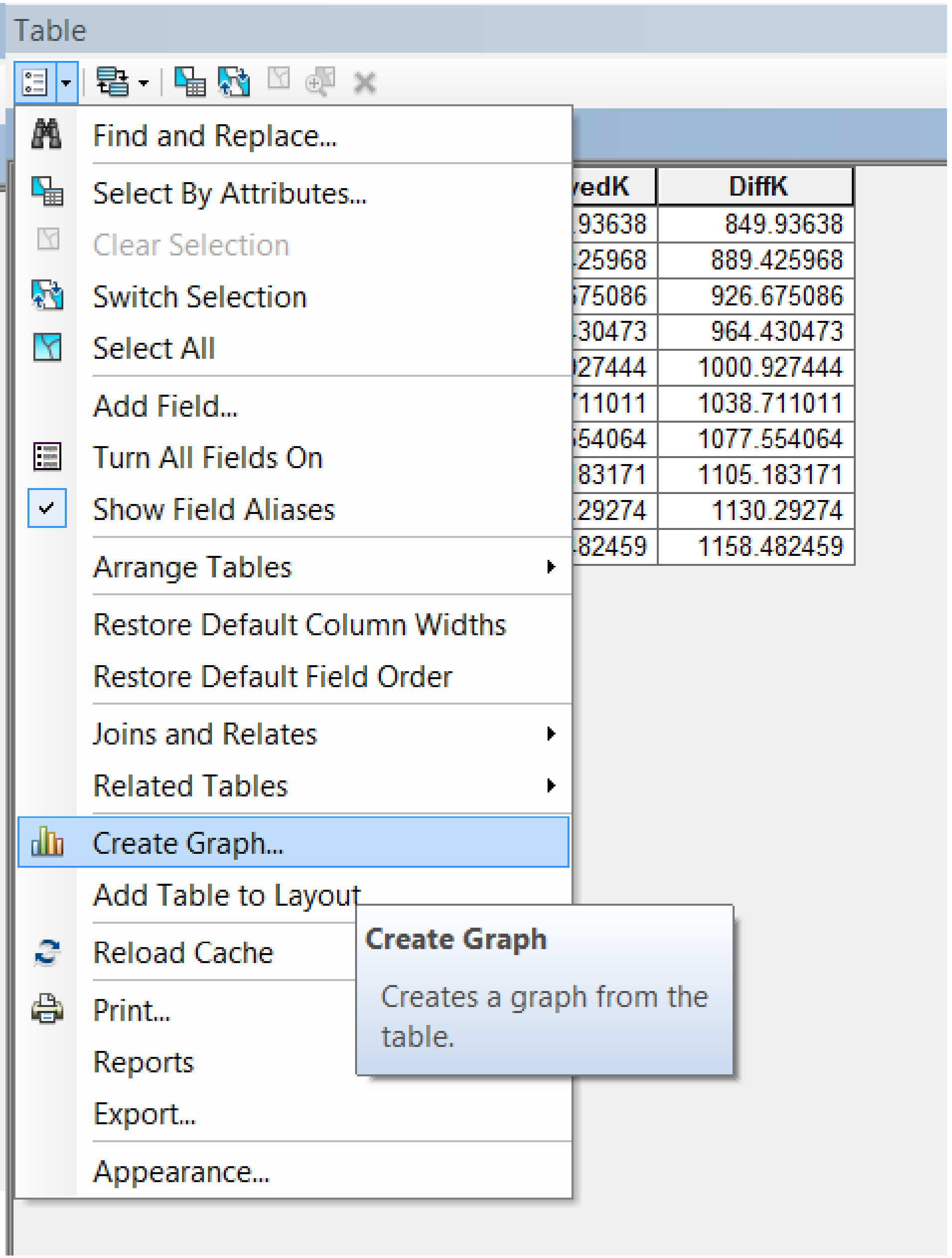

Homicides o 2005-2009
ExpectedK
ObservedK
Clustered, since observed is above expected
Dispersed, if observed was below expected
This does not exist for our data.
We can conclude the pattern of homicides is clustered
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

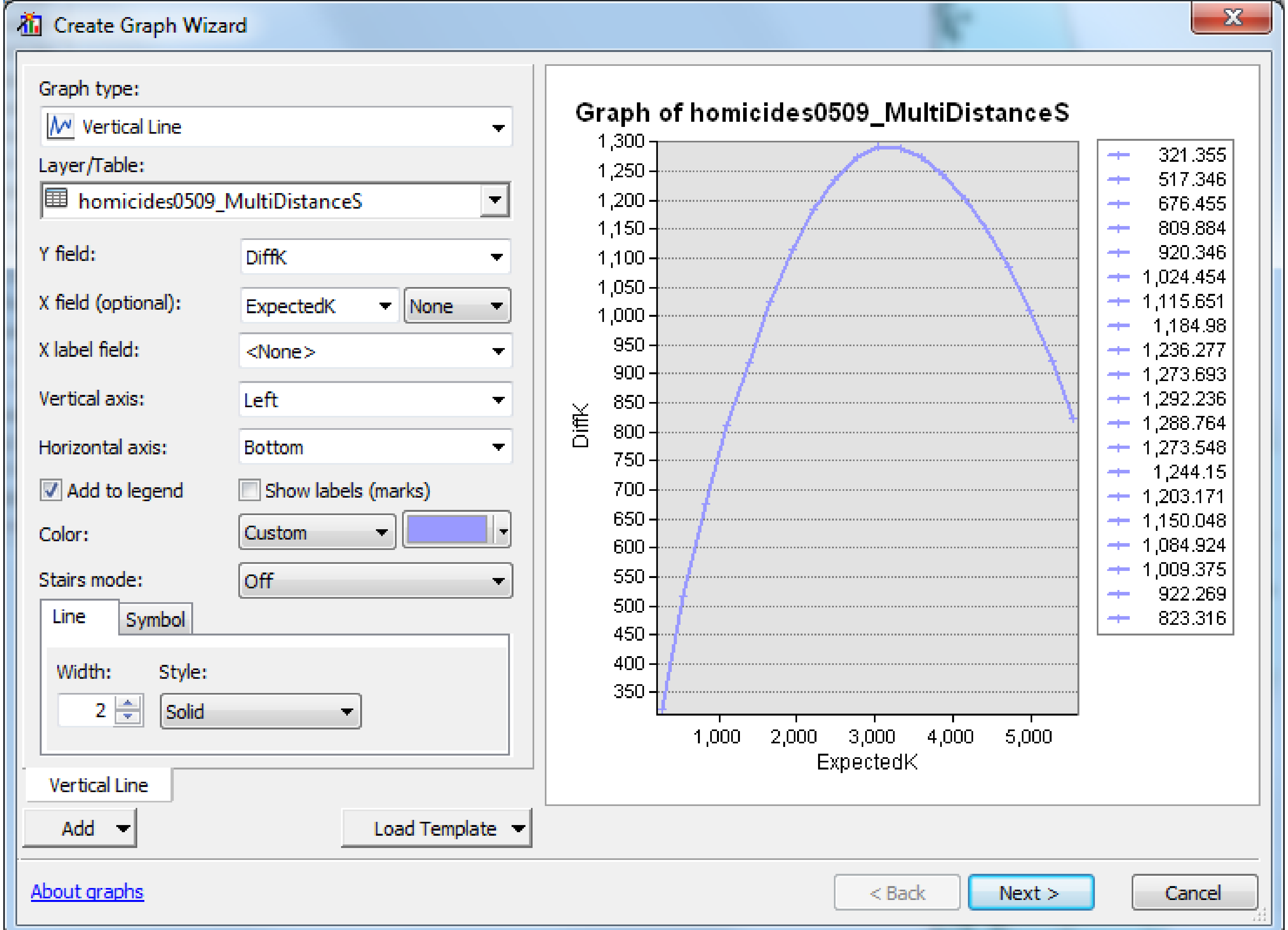
We can create a second graph – we may want to plot the difference between observed and expected k, versus distance

Table

omicides_MultiDistanceSpati2

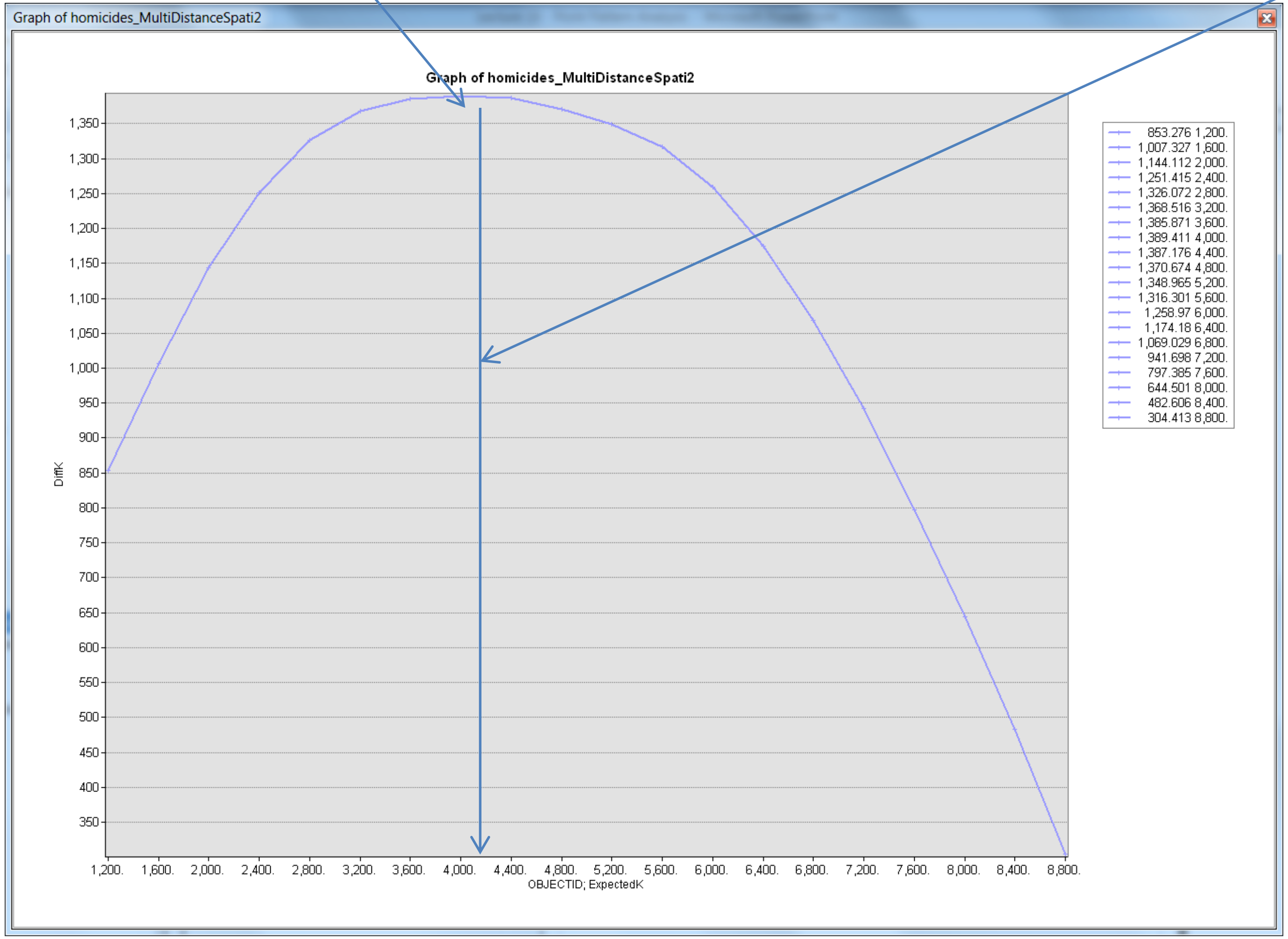
OBJECTID *	ExpectedK	ObservedK	DiffK	LwConfEnv	HiConfEnv
1	1200	2053.275681	853.275681	1188.792758	1209.467028
2	1600	2607.326824	1007.326824	1584.514082	1615.619258
3	2000	3144.11243	1144.11243	1992.46456	2014.458375
4	2400	3651.415352	1251.415352	2382.433714	2416.563197
5	2800	4126.072445	1326.072445	2785.478101	2823.998785
6	3200	4568.515832	1368.515832	3187.591722	3228.803123
7	3600	4985.871149	1385.871149	3590.997587	3625.335765
8	4000	5389.411307	1389.411307	3993.196352	4030.003285
9	4400	5787.175936	1387.175936	4393.65688	4446.055105
10	4800	6170.673691	1370.673691	4787.492673	4855.389063
11	5200	6548.964993	1348.964993	5189.053218	5271.811653
12	5600	6916.30149	1316.30149	5591.010102	5683.290435
13	6000	7258.96974	1258.96974	5994.189858	6087.328332
14	6400	7574.179736	1174.179736	6381.108861	6482.886514
15	6800	7869.028549	1069.028549	6770.461896	6862.956986
16	7200	8141.697796	941.697796	7159.090915	7239.236183
17	7600	8397.385465	797.385465	7530.337775	7609.234845
18	8000	8644.500928	644.500928	7888.277819	7970.844929
19	8400	8882.606477	482.606477	8231.663018	8325.825861
20	8800	9104.413103	304.413103	8560.399041	8664.527144




Create Graph Wizard
Graph type:
Vertical Line
Layer/Table:
homicides0509_MultiDistanceS
Y field:
DiffK
X field (optional):
ExpectedK
None
X label field:
<None>
Vertical axis:
Left
Horizontal axis:
Bottom
Add to legend
Show labels (marks)
Color:
Custom
Stairs mode:
Off
Line
Symbol
Width:
2
Style:
Solid
Vertical Line
Add
Load Template
About graphs
< Back
Next >
Cancel
Graph of homicides0509_MultiDistanceS
DiffK
ExpectedK
1,300
1,250
1,200
1,150
1,100
1,050
1,000
950
900
850
800
750
700
650
600
550
500
450
400
350
1,000
2,000
3,000
4,000
5,000
321.355
517.346
676.455
809.884
920.346
1,024.454
1,115.651
1,184.98
1,236.277
1,273.693
1,292.236
1,288.764
1,273.548
1,244.15
1,203.171
1,150.048
1,084.924
1,009.375
922.269
823.316

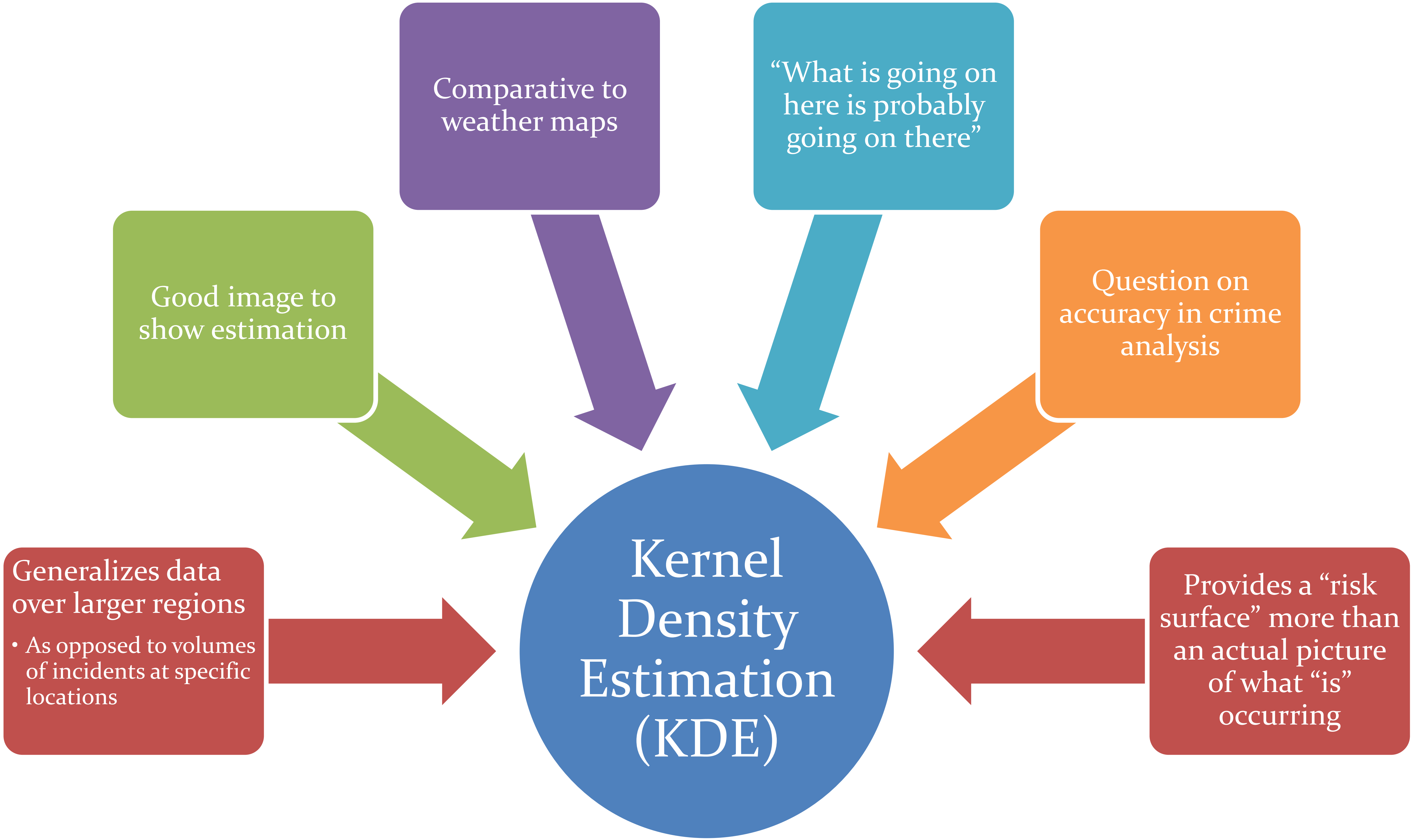
Look for the peak for most pronounced clustering

After the peak, we get less clustering



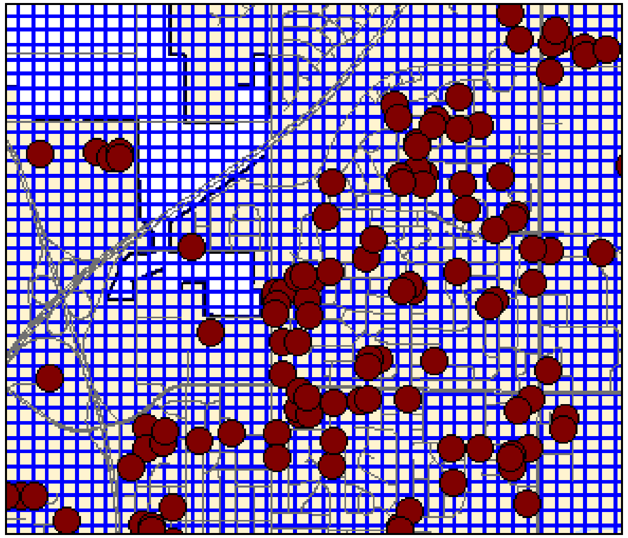
Distance between clusters is about 4,000 meters or 4 km

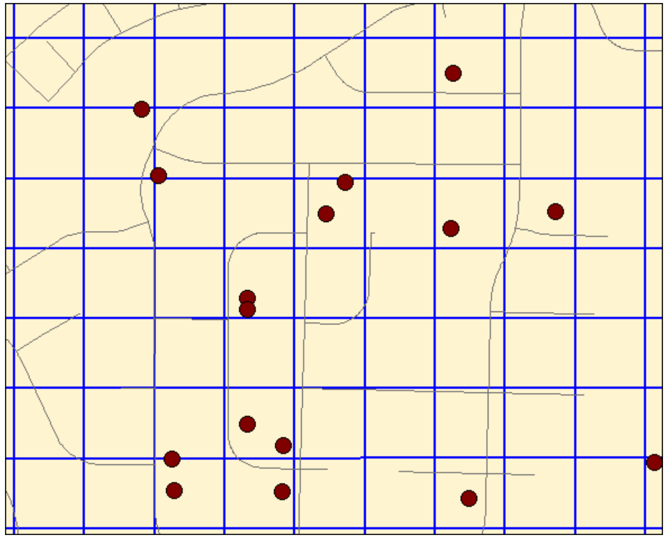
Kernel Density Estimation (KDE)


Comparative to weather maps
“What is going on here is probably going on there”
Good image to show estimation
Question on accuracy in crime analysis
Generalizes data over larger regions
• As opposed to volumes of incidents at specific locations
Kernel Density Estimation (KDE)
Provides a “risk surface” more than an actual picture of what “is” occurring

How KDE Works

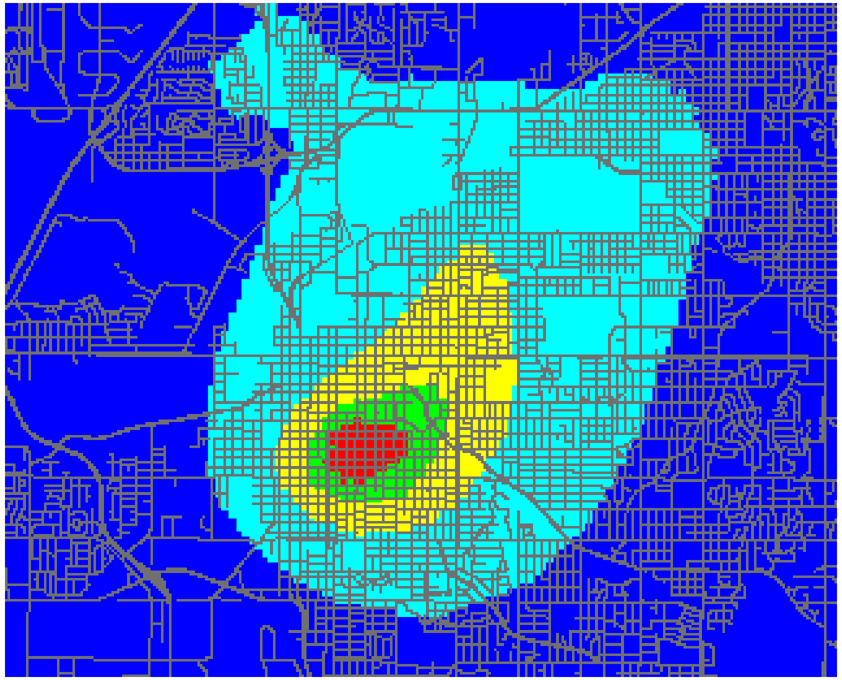
- Every point on the map has a density estimate based on its proximity to crime incidents
- Done by overlaying a grid on top of the map
  - Calculates the density estimate for the centerpoint of each grid cell
    - Number of cells in the grid is defined by the user





How KDE Works

- In a map, the grid cells are color-coded based on the density
  - Often reds for hottest area and blues for coolest
  - When making comparisons standard deviation is typically used



KDE Parameters

Many parameters involved

Analyst must use experience & judgment

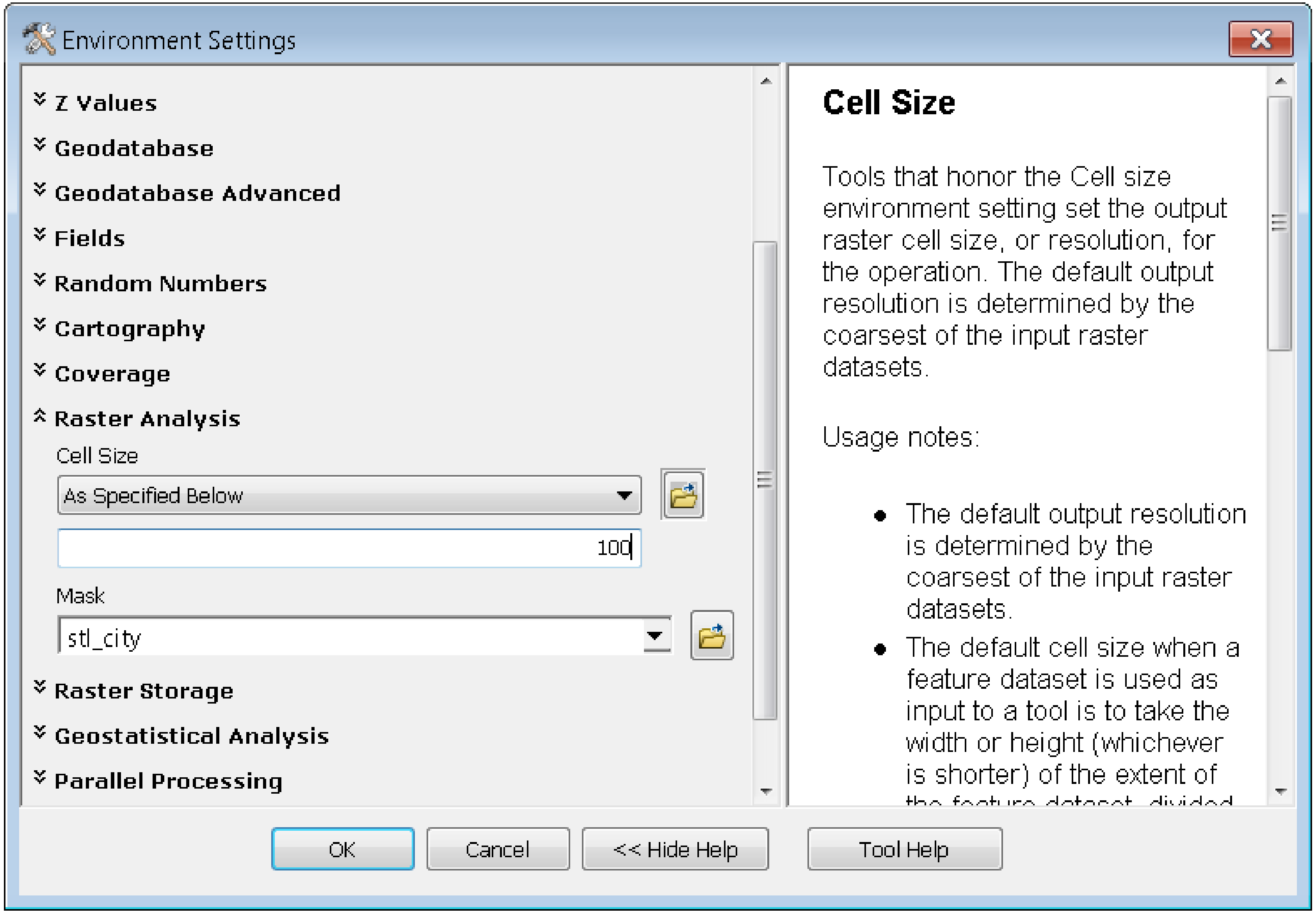
Single versus dual kernel density estimates

- Single is usually used in crime analysis
- Dual can help normalize data for population or other risk factors or calculate change from one time to the next

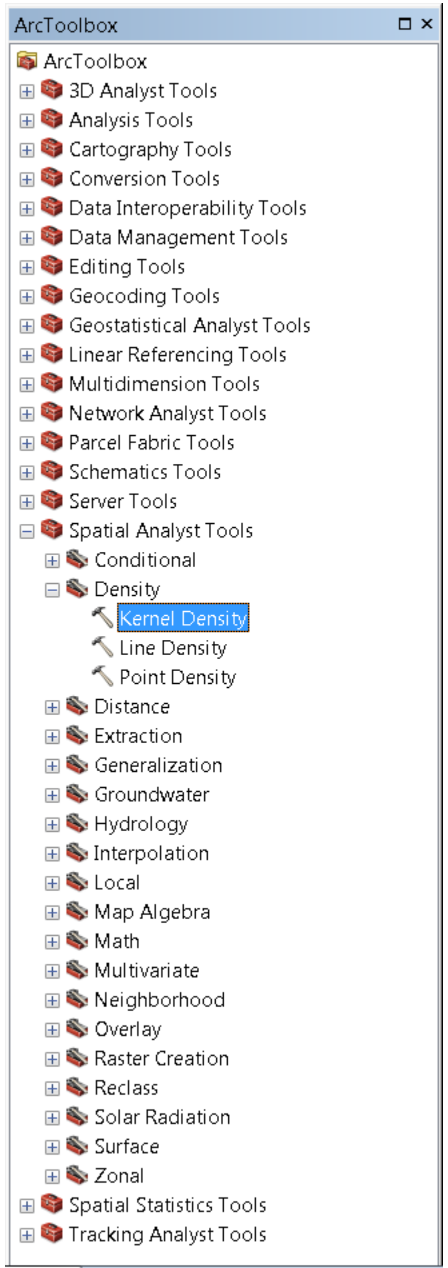
Bandwidth

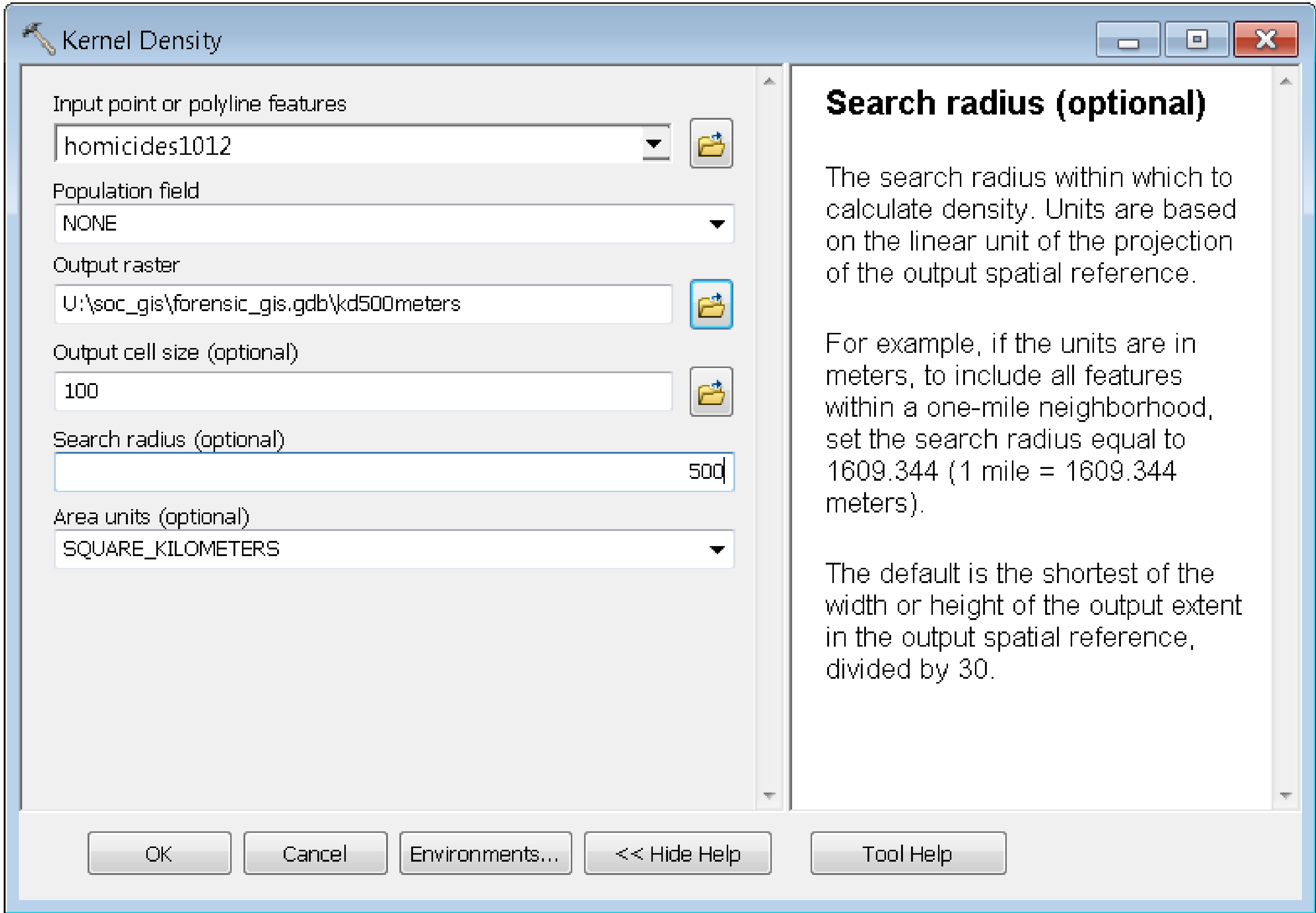
- Refers to the size of the cone; specified by user

1. Geoprocessing-> Environmental Setting->Raster Analysis->select "As specified Below ->Mask should be the city of Saint Louis. I will use 100 meters for the grid cells.

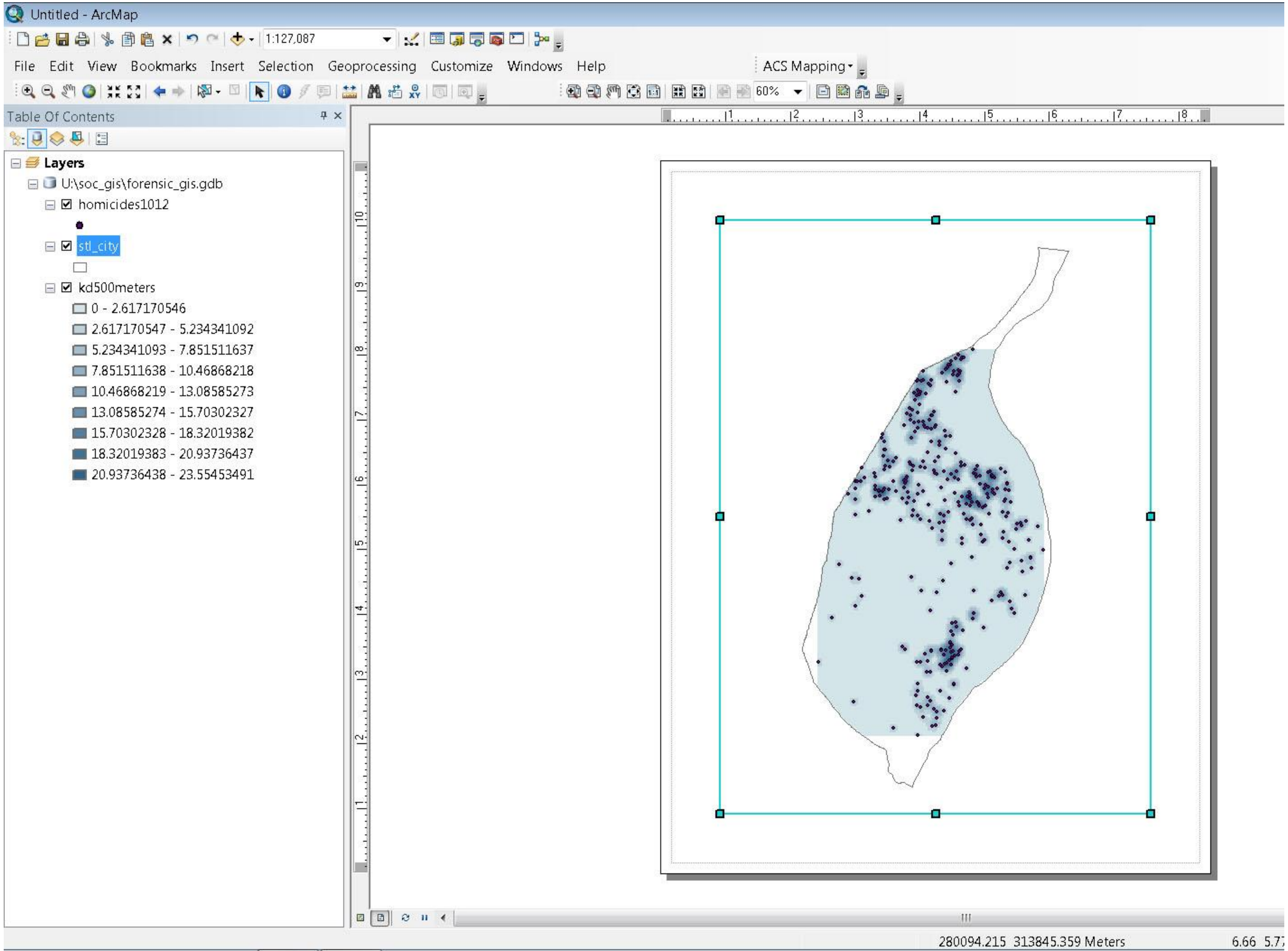


2. ArcToolbox->Spatial Analyst Tools->Density-> Kernel Density

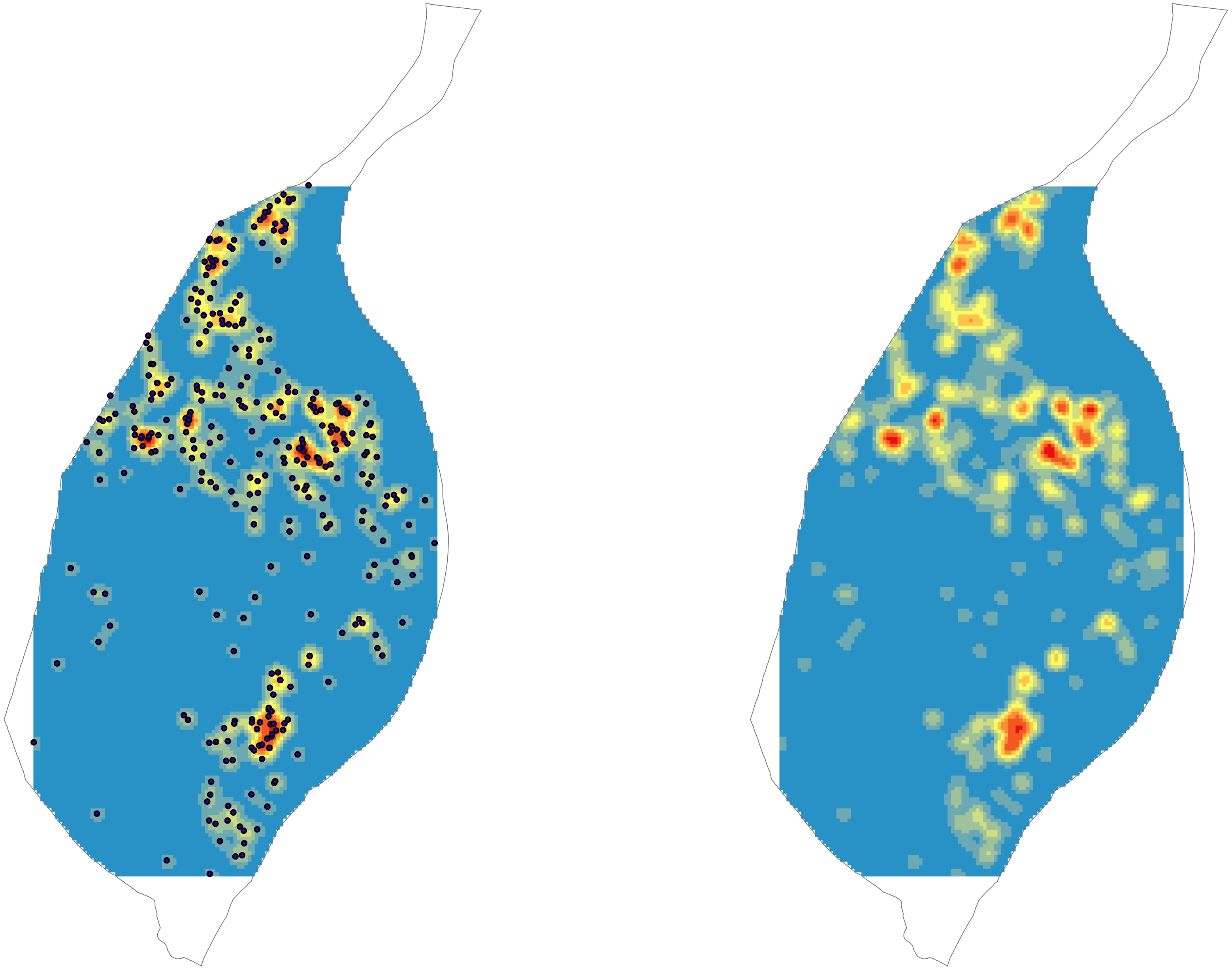




3. You should have a map that looks like this map.



4. Let's change the color ramp and remove the points.



General G Statistic

Identifying the clustering of values

General G Statistic – Part 1

- The General G (aka Getis-Ord General G) statistic looks at the similarity of the values associated with the features within a critical distance of each other. The value is exclude in the calculation – only neighbors

- The General G measures concentrations of high and low values over the entire study area.

- General G = Global Moran's I

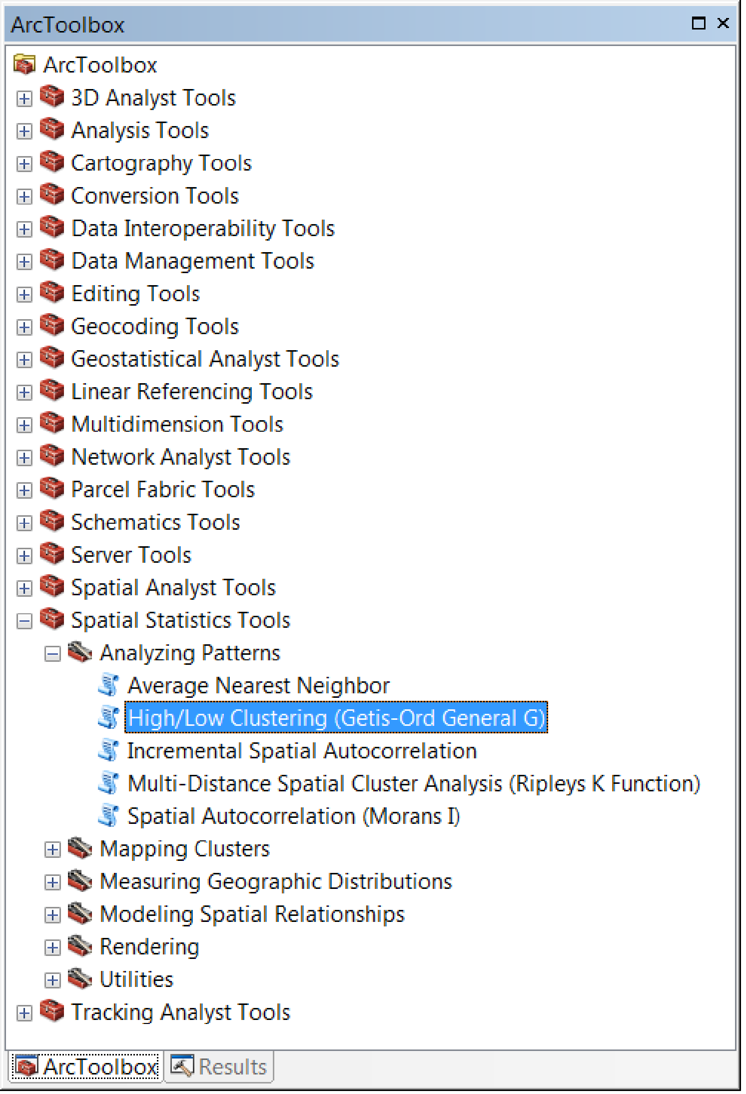
- The input feature should have an attribute demonstrating some characteristic or value associated with the location.

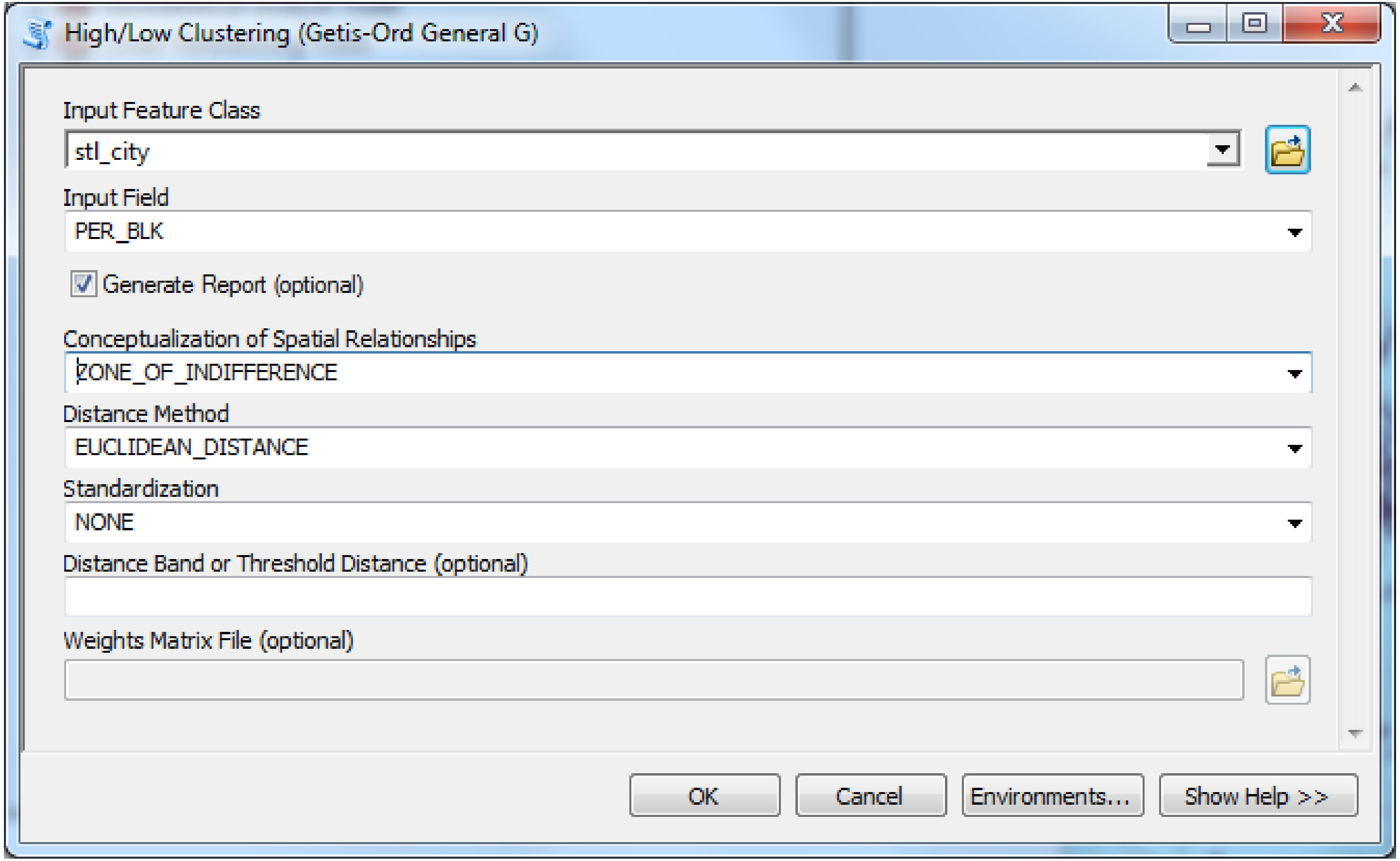
General G Statistic – Part 2

- The results will determine if the values are clustered, not the locations.

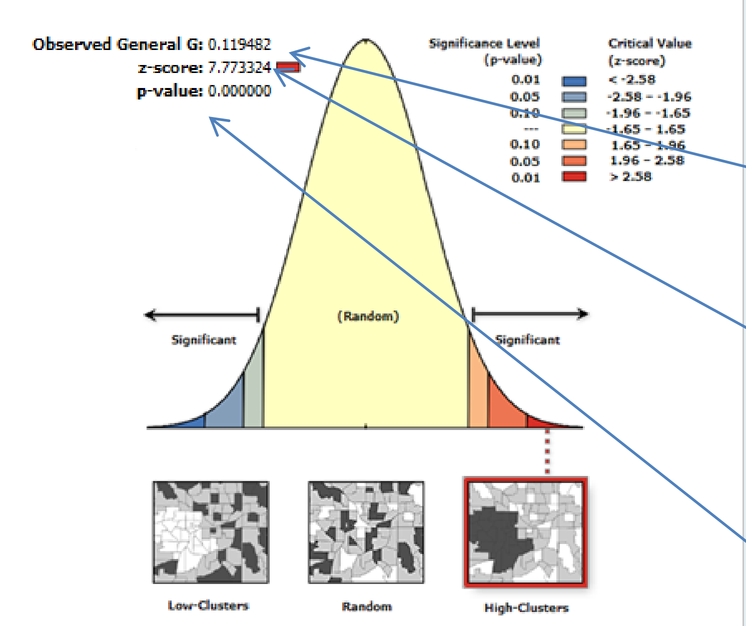
- It will distinguish between clustering of high values and low values.

- H_0=Data are randomly distributed across the study area.
- H_1 = There is clustering.


ArcToolbox
ArcToolbox
3D Analyst Tools
Analysis Tools
Cartography Tools
Conversion Tools
Data Interoperability Tools
Data Management Tools
Editing Tools
Geocoding Tools
Geostatistical Analyst Tools
Linear Referencing Tools
Multidimension Tools
Network Analyst Tools
Parcel Fabric Tools
Schematics Tools
Server Tools
Spatial Analyst Tools
Spatial Statistics Tools
Analyzing Patterns
Average Nearest Neighbor
High/Low Clustering (Getis-Ord General G)
Incremental Spatial Autocorrelation
Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
Spatial Autocorrelation (Morans I)
Mapping Clusters
Measuring Geographic Distributions
Modeling Spatial Relationships
Rendering
Utilities
Tracking Analyst Tools
ArcToolbox
Results


High/Low Clustering (Getis-Ord General G)
Input Feature Class
stl_city
Input Field
PER_BLK
Generate Report (optional)
Conceptualization of Spatial Relationships
ZONE_OF_INDIFFERENCE
Distance Method
EUCLIDEAN_DISTANCE
Standardization
NONE
Distance Band or Threshold Distance (optional)
Weights Matrix File (optional)
OK
Cancel
Environments...
Show Help >>

High-Low Clustering Report



Given the z-score of 7.77332434924, there is a less than 1% likelihood that this high-clustered pattern could be the result of random chance.

General G Summary

Observed General G:	0.119482
Expected General G:	0.085662
Variance:	0.000019
z-score:	7.773324
p-value:	0.000000

Dataset Information

Input Feature Class:	stl_city
Input Field:	PER_BLK
Conceptualization:	ZONE_OF_INDIFFERENCE
Distance Method:	EUCLIDEAN
Row Standardization:	False
Distance Threshold:	2189.7355 Meters
Weights Matrix File:	None
Selection Set:	False

Important Points for observation

- The dataset returned an index of .119482
- This tells us that there is little clustering in the features
- The z-score is 7.773324 with a significance level of 0.000. This score does gives us the confidence level of rejecting the null hypothesis.

Gi*

Hot-Spot Analysis

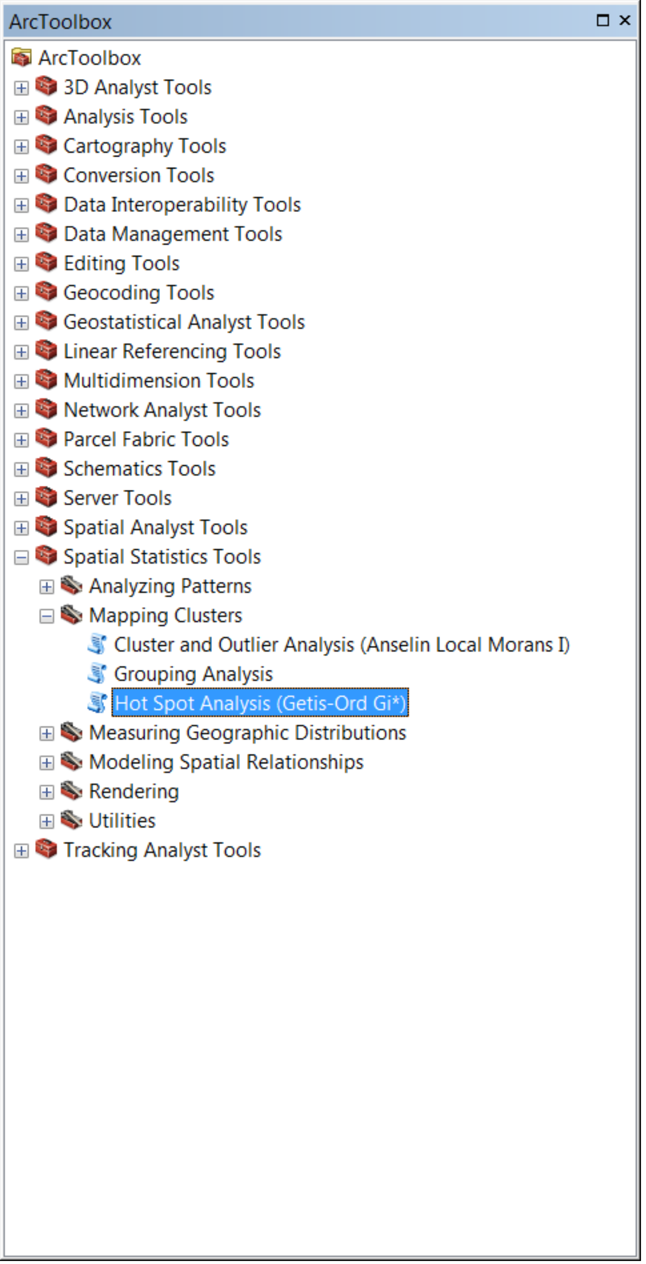
Gi*

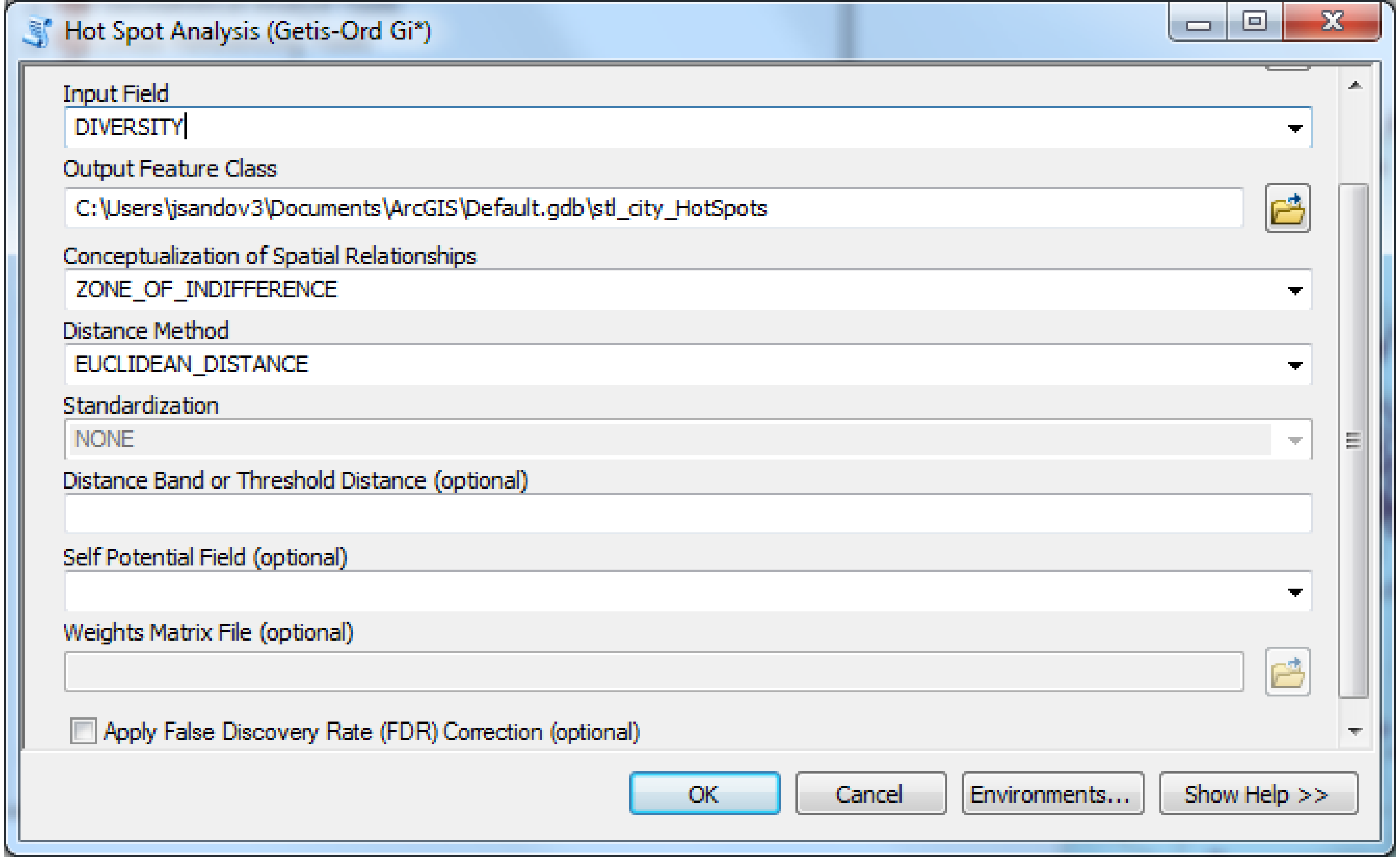
- Gi* (aka as the Gi statistic) uses both the location and the value in the pattern calculations. (i.e., more information)

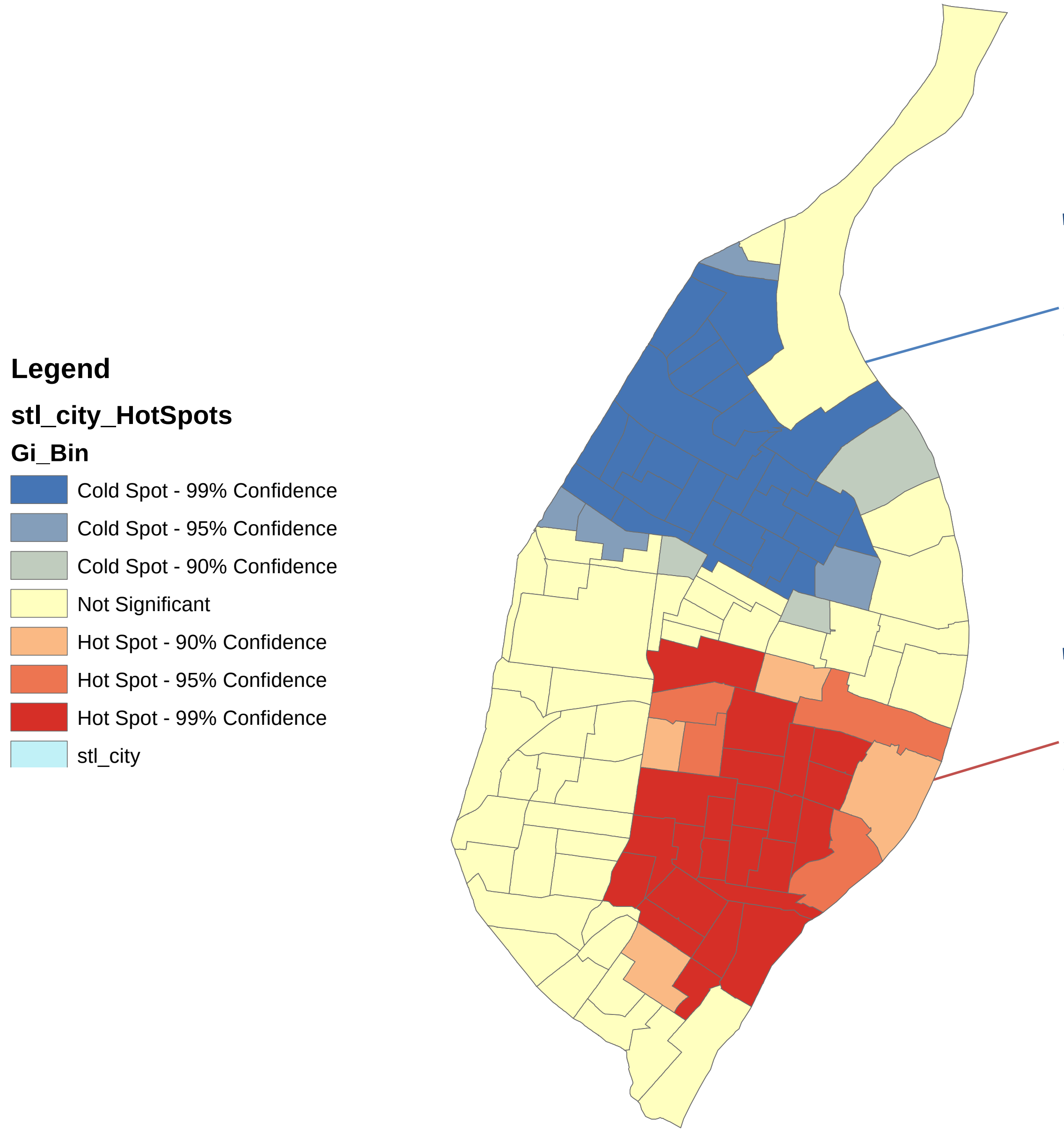
- Gi* is used to see the effect of the value field on the clustering.

- The user must select distance or neighbors.

- Gi* is similar to LISA


ArcToolbox
ArcToolbox
3D Analyst Tools
Analysis Tools
Cartography Tools
Conversion Tools
Data Interoperability Tools
Data Management Tools
Editing Tools
Geocoding Tools
Geostatistical Analyst Tools
Linear Referencing Tools
Multidimension Tools
Network Analyst Tools
Parcel Fabric Tools
Schematics Tools
Server Tools
Spatial Analyst Tools
Spatial Statistics Tools
Analyzing Patterns
Mapping Clusters
Cluster and Outlier Analysis (Anselin Local Morans I)
Grouping Analysis
Hot Spot Analysis (Getis-Ord Gi*)
Measuring Geographic Distributions
Modeling Spatial Relationships
Rendering
Utilities
Tracking Analyst Tools


Hot Spot Analysis (Getis-Ord Gi*)
Input Field
DIVERSITY
Output Feature Class
C:\Users\jsandov3\Documents\ArcGIS\Default.gdb\stl_city_HotSpots
Conceptualization of Spatial Relationships
ZONE_OF_INDIFFERENCE
Distance Method
EUCLIDEAN_DISTANCE
Standardization
NONE
Distance Band or Threshold Distance (optional)
Self Potential Field (optional)
Weights Matrix File (optional)
Apply False Discovery Rate (FDR) Correction (optional)
OK
Cancel
Environments...
Show Help >>



The blue areas represent clusters of low values

The red areas represent clusters of high values

Table

stl_city_HotSpots

	OBJECTID *	Shape *	SOURCE_ID	DIVERSITY	Shape_Length	Shape_Area	GiZScore ZOI 2189	GiPValue ZOI 2189	Gi_Bin ZOI 2189
▸	1	Polygon	1	0.38	4414.988191	910998.865291	-0.603176	0.546392	0
	2	Polygon	2	0.3	5588.341372	1676106.950701	-0.836114	0.403091	0
	3	Polygon	3	0.42	[illegible]	[illegible]	[illegible]	0.569133	0
	4	Polygon	4	0.58	6815.936063	1939468.736939	0.367454	0.71328	0
	5	Polygon	5	0.66	7487.708625	1162116.258634	1.269412	0.204294	0
	6	Polygon	6	0.75	3906.691794	831167.208584	0.243179	0.807867	0
	7	Polygon	7	0.29	4563.260523	957507.968608	-0.61976	0.535416	0
	8	Polygon	8	0.55	3533.487207	706853.680582	-2.018662	0.043522	-2
	9	Polygon	9	0.21	5516.324536	1195404.478612	-1.967682	0.049105	-2
	10	Polygon	10	0.12	4210.86845	956129.855259	-3.104829	0.001904	-3
	11	Polygon	11	0.13	4663.067288	1064700.001989	-2.837626	0.004545	-3
	12	Polygon	12	0.12	4819.317517	1017014.684177	-3.468963	0.000522	-3
	13	Polygon	13	0.15	4744.724148	1106265.984888	-4.325905	0.000015	-3
	14	Polygon	14	0.18	4575.541138	1064254.682802	-4.010201	0.000061	-3
	15	Polygon	15	0.11	4779.71489	854742.901107	-3.212202	0.001317	-3
	16	Polygon	16	0.15	4842.663499	1238223.189551	-4.549113	[illegible]	-3
	17	Polygon	17	0.12	3957.156068	655316.033691	-3.493338	[illegible]	-3
	18	Polygon	18	0.11	5608.171999	1587503.95767	-3.040845	0.002359	-3
	19	Polygon	19	0.15	4397.149981	1084963.927742	-3.493067	0.000478	-3
	20	Polygon	20	0.11	4112.381738	911653.601148	-3.50506	0.000457	-3
	21	Polygon	21	0.06	5237.164963	1164873.072028	-4.011125	0.00006	-3
	22	Polygon	22	0.21	8515.906679	3214040.746895	-3.149188	0.001637	-3
	23	Polygon	23	0.16	4775.556512	827201.840957	-2.1675	0.030197	-2
	24	Polygon	24	0.24	3652.334986	750387.253438	-1.41605	0.156761	0
	25	Polygon	25	0.11	9047.384795	3151770.078188	-2.724892	0.006432	-3
	26	Polygon	26	0.26	8342.14083	4043350.956623	-1.707945	0.087646	-1
	27	Polygon	27	0.15	5104.659371	993870.611148	-4.528333	0.000006	[illegible]
	28	Polygon	28	0.17	4277.247926	777647.756911	-4.242388	0.000022	-3
	29	Polygon	29	0.13	4271.38562	938175.793142	-4.110302	0.00004	-3
	30	Polygon	30	0.14	3923.110437	885105.582724	-4.187916	0.000028	-3
	31	Polygon	31	0.14	5034.345313	1246970.64096	-3.946875	0.000079	-3

1 (0 out of 106 Selected)

stl_city_HotSpots

Here is your value

Here is the score

Here is the significance

Here is your cluster

Lab