

# Point Patterns in Forest Fires

—by Sumeeta Srinivasan

## Introduction

---

### Problem

Do the patterns of forest fires change over time? Do forest fires occur in clusters, and do the clusters change over time? Is this information useful in fighting forest fires?

This lab exercise will try to answer these questions using spatial statistics in a geographic information system (GIS) context.

### Location

Montana, United States

### Time to complete the lab

Three hours

### Prerequisites

- An understanding of basic spatial statistics
- Basic skills in the use of ArcGIS® 10
- Elementary understanding of the following: standard deviational ellipses (applied to point patterns), Moran's I to analyze point patterns, average nearest neighbor, and Ripley's K statistic, as well as kernel densities (to analyze clusters)

*Keywords: forest fire spatial patterns; kernel density estimation; Ripley's K; centroids; standard deviational ellipse; Moran's I; average nearest neighbor distance*

## Data used in this lab

- Forest fire history in Montana 2005, 2009
- Coordinate system: State plane Montana
- Datum: North American Datum 1983
- Projection: Lambert Conformal Conic
- Fire history data/metadata:  
[http://www.fs.fed.us/r1/gis/thematic\\_data/fire\\_history\\_r1\\_1985\\_2009\\_poly.htm](http://www.fs.fed.us/r1/gis/thematic_data/fire_history_r1_1985_2009_poly.htm)

## Student activity

---

In this lab exercise, you will use several spatial statistics methods to analyze forest fire patterns in Montana for two nonconsecutive years.

Average and standard deviation are two of the well-known statistical measures used to explain the distribution of nonspatial data. In the spatial context, a spatial mean or centroid and standard deviational ellipse can be calculated for point pattern data. The mean center is determined by the average x-coordinate and average y-coordinate of all the points (or feature centroids) in the study area. The standard deviational ellipse yields a new feature class centered on the mean center of all features (or of all cases when a case field is provided). The attribute values for these output ellipse polygons include the lengths of the long and short axes; the angle of orientation of the ellipse; and the case field, if specified. You can specify the axis lengths as a multiple of the standard deviations. If the feature centroids are normally distributed, one standard deviation (the default value) will cover approximately 68 percent of all input feature centroids. Three standard deviations will cover approximately 99 percent of all feature centroids.

Spatial autocorrelation is one of the most important descriptive spatial statistics. Spatial independence is an arrangement of incident locations in which there are no spatial relationships between any of the incidents. The intuitive concept is that the location of an incident (such as a fire or a crime) is unrelated to the location of any other incident, or that they are randomly placed with respect to each other. On the other hand, positive spatial autocorrelation is an arrangement of incident locations where the point locations are related. When events or people or facilities are clustered together, this arrangement indicates positive spatial autocorrelation, and Moran's I is typically used to quantify it.

The distribution of the distance to the nearest neighbor is another useful spatial statistic—in particular, the average nearest neighbor distance can be used to distinguish between clustering and dispersion. If this distance is smaller than what it would be under spatial randomness, it is interpreted as clustering. If it is larger, the distance is interpreted as dispersion. A test statistic can be constructed by taking the difference between the observed and random mean nearest neighbor

distance and standardizing by the standard error. In particular, a value less than 1 suggests clustering. The nearest neighbor index is only an indicator of first-order effects. It compares the average distance for the nearest neighbor to an expected random distance. But it does not provide any information about the second, third, or  $k$ th nearest neighbor.

The nearest neighbor distance statistics are called *first-order* statistics, since they only consider the distance to the nearest point. Second-order distance statistics consider the complete distribution of all distances in the point pattern. Ripley's K function is a second-order statistic and is essentially a test on the cumulative distribution function of the full set of interpoint distances. This distribution can be compared to a reference distribution under complete spatial randomness. A higher proportion of shorter to random distances implies clustering, whereas a higher proportion of longer to random distances implies dispersion.

Another way of visualizing the distribution of fires is through kernel density estimation. Kernel density estimation requires the use of a bandwidth. ArcMap™ software allows you to specify a kernel bandwidth as a fixed interval or to choose an adaptive interval. The bandwidth is a distance; hence, the point coordinates must be in appropriate units such as feet, meters, or miles rather than degrees (that is, universal transverse Mercator [UTM] coordinates as contrasted with latitude and longitude). Some experimentation with the choice of bandwidth is typically necessary. The kernel values are computed as totals, or absolute density (points per grid area, rescaled such that the sum over all grids adds up to the observed total); as points per areal units; or as relative density (for example, points per square mile), or probabilities.

In this exercise, you will do all the following for both the 2005 and 2009 data:

1. Look at the overall patterns using descriptive spatial statistics called the mean center and the standard deviational ellipse.
2. Use Moran's I to quantify spatial autocorrelation and see if there is global clustering in the forest fires.
3. Use point pattern statistics, such as the nearest neighbor distance and Ripley's K, to look for clustering over distance.
4. Use kernel density maps to search for possible locations of clustering based on various bandwidths.
5. Compare the results for the two years to see if there are any differences.

## Prepare your workspace

To begin, create a workspace to keep all data for this lab.

- 1 Create a *FireHistory* folder under the *MyDocuments* folder.
- 2 Create a *Data* folder under the *FireHistory* folder.

Do not use spaces in your file and folder names, because some of the spatial analysis functions in ArcGIS require that the path names do not have spaces in them.

## Collect and process data

Download, extract, and process data from the Montana website at

[http://www.fs.fed.us/r1/gis/thematic\\_data/fire\\_history\\_r1\\_1985\\_2009\\_poly.zip](http://www.fs.fed.us/r1/gis/thematic_data/fire_history_r1_1985_2009_poly.zip).

- 1 Extract the *fire\_history\_r1\_1985\_2009* geodatabase in the ZIP file you downloaded to your data workspace using file extraction software (for example, you could name it *MyDocuments\FireHistory*).

It is important to view data before moving on to the analysis.

- 2 Launch ArcCatalog™.
- 3 Examine the geodatabase in the *Data* folder.
- 4 Convert the forest fires (polygon shapes) to points for the years 2005 and 2009. In ArcToolbox™, expand *Data Management Tools* » *Features* and then double-click the *Feature to Point* tool. You should create two point shapefiles and name them appropriately, for example, *fire\_points\_2005.shp* and *fire\_points\_2009.shp*.

Note that you will need ArcGIS for Desktop Advanced for this tool to work. Alternatively, use the data supplied with this lab exercise.

## EXAMINE OVERALL PATTERNS IN FOREST FIRES

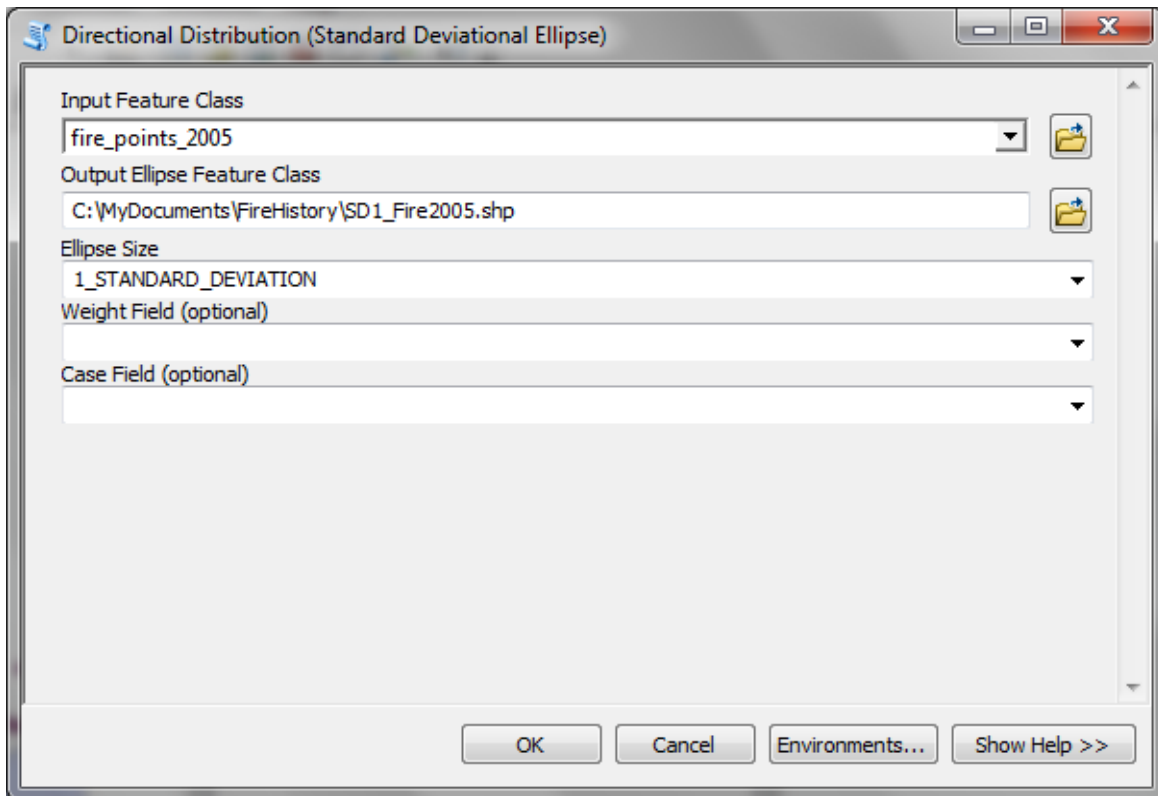
### ANALYZE

- 1 Launch ArcMap.
- 2 Add the point shapefiles for forest fires in 2005 and 2009 to your map document.
- 3 Open the attribute tables for both datasets.

Notice that the spatial distribution of forest fires in 2005 is different from that of 2009.

- 4 In ArcToolbox, expand *Spatial Statistics Tools* » *Measuring Geographic Distributions* and then double-click *Directional Distribution (Standard Deviational Ellipse)*.

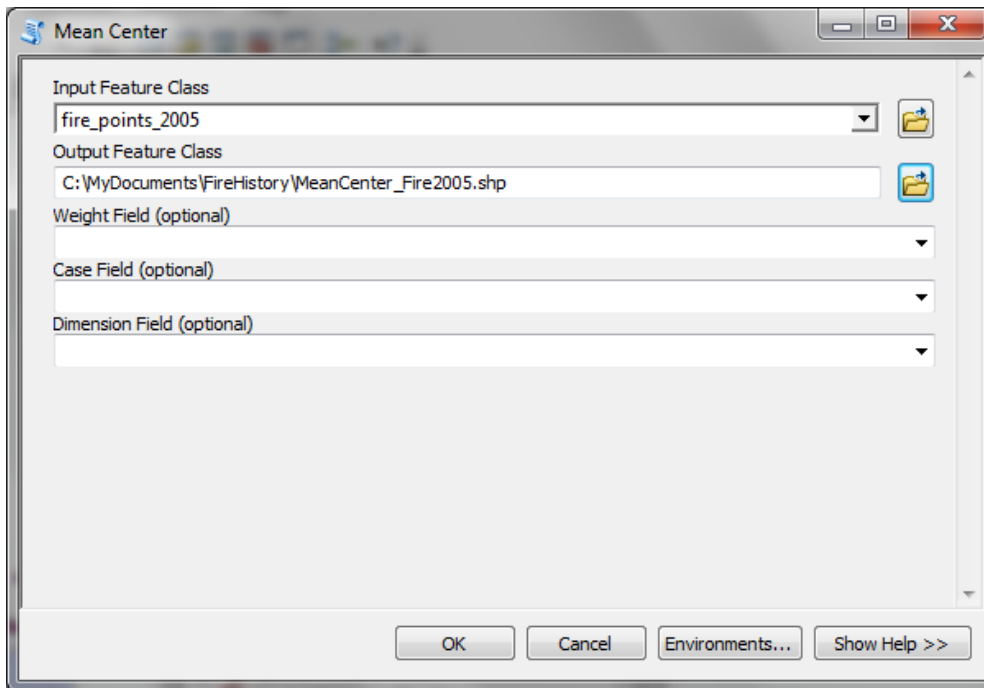
- 5 Use the shapefile for point locations of forest fires in 2005 as the input feature class. Input the parameters for output and the ellipse size as shown in the image below:



The output will automatically be added to the map document.

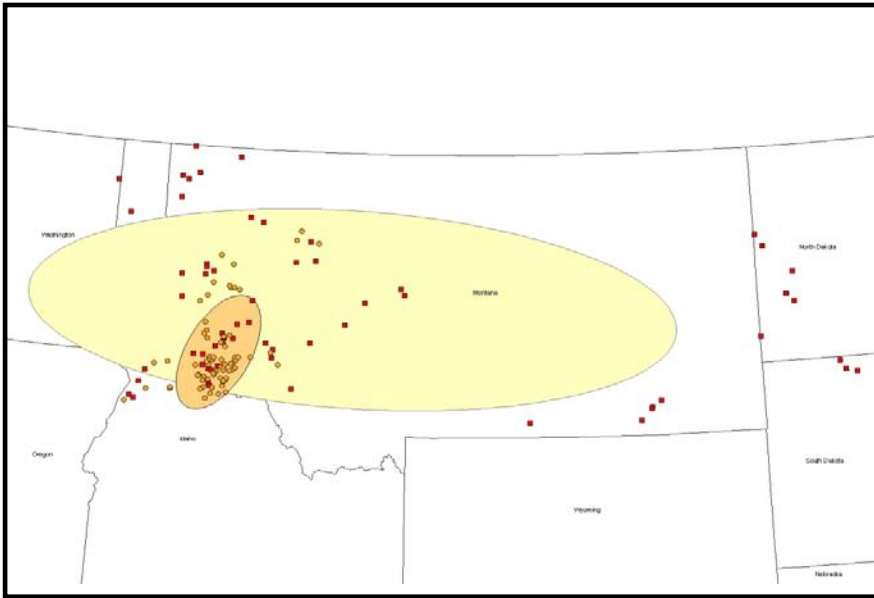
- 6 Repeat step 5 for the 2009 data.
- 7 In ArcToolbox, expand *Spatial Statistics Tools* » *Measuring Geographic Distributions* and then double-click *Mean Center*.

- 8 Input the variables as shown in the screen capture below. The resultant point feature, showing the mean center for forest fires in 2005, will be added to your map document. Repeat this for the 2009 data.



## VISUALIZE

Map 1 shows the directional ellipses for the 2005 and 2009 forest fire patterns. Note that the map uses different colors for the two years to distinguish the patterns visually. It also includes state boundaries for the United States. Also note that each fire location can be labeled for specific location names. Context is very important in discussing relative locations with planners or policy makers.



Map 1: Forest fires standard deviational ellipses

**Question 1:** *Were there any locations that had forest fires in both years? If so, where?*

**Question 2:** *Did the mean center change from 2005 to 2009? If so, where did it move, and why is this interesting?*

**Question 3:** *Did the standard deviational ellipses change from 2005 to 2009? Explain what this means about the changes in the spatial patterns of distribution of forest fires over the two years.*

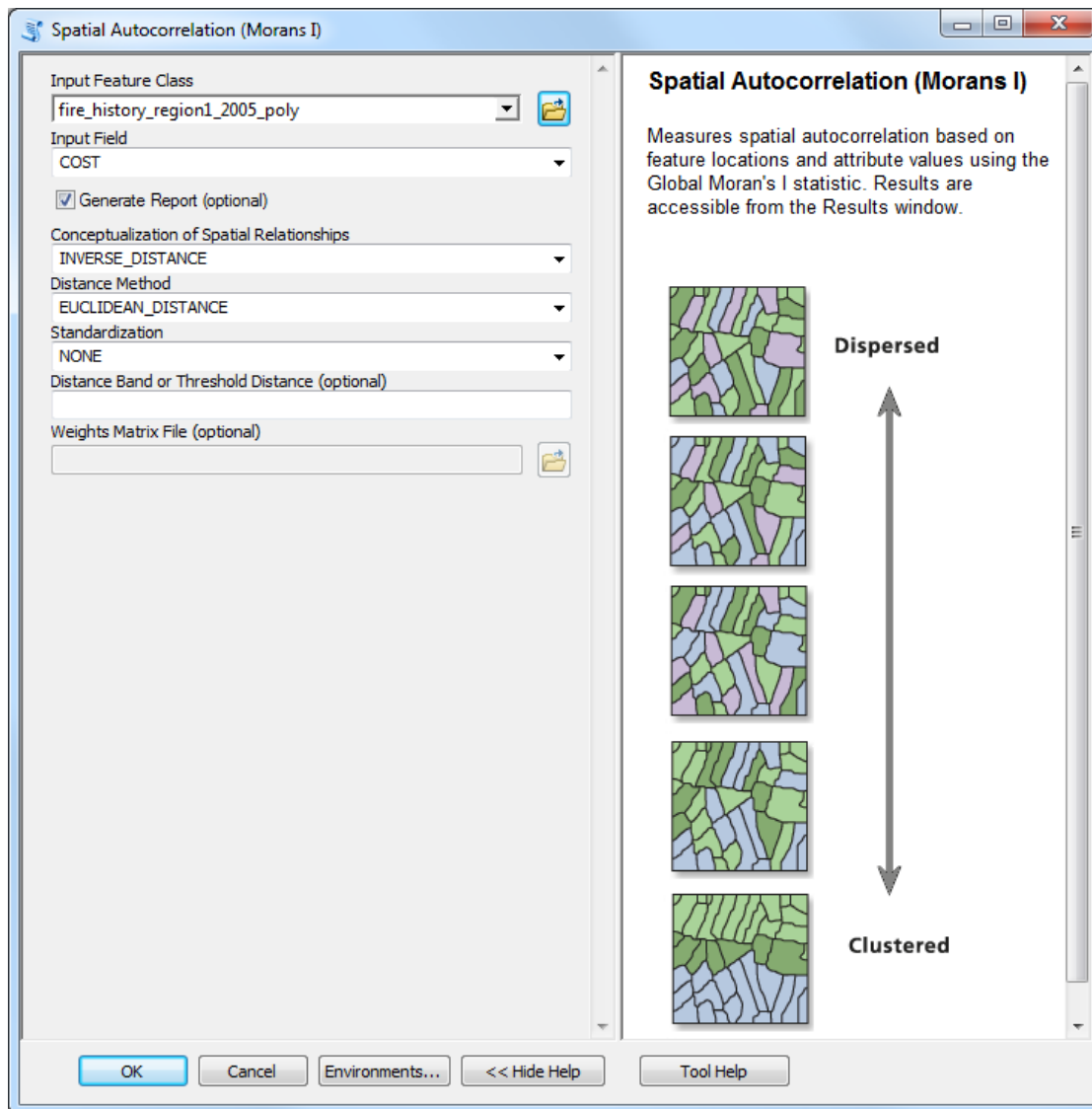
## CHECK FOR CLUSTERING

## ANALYZE

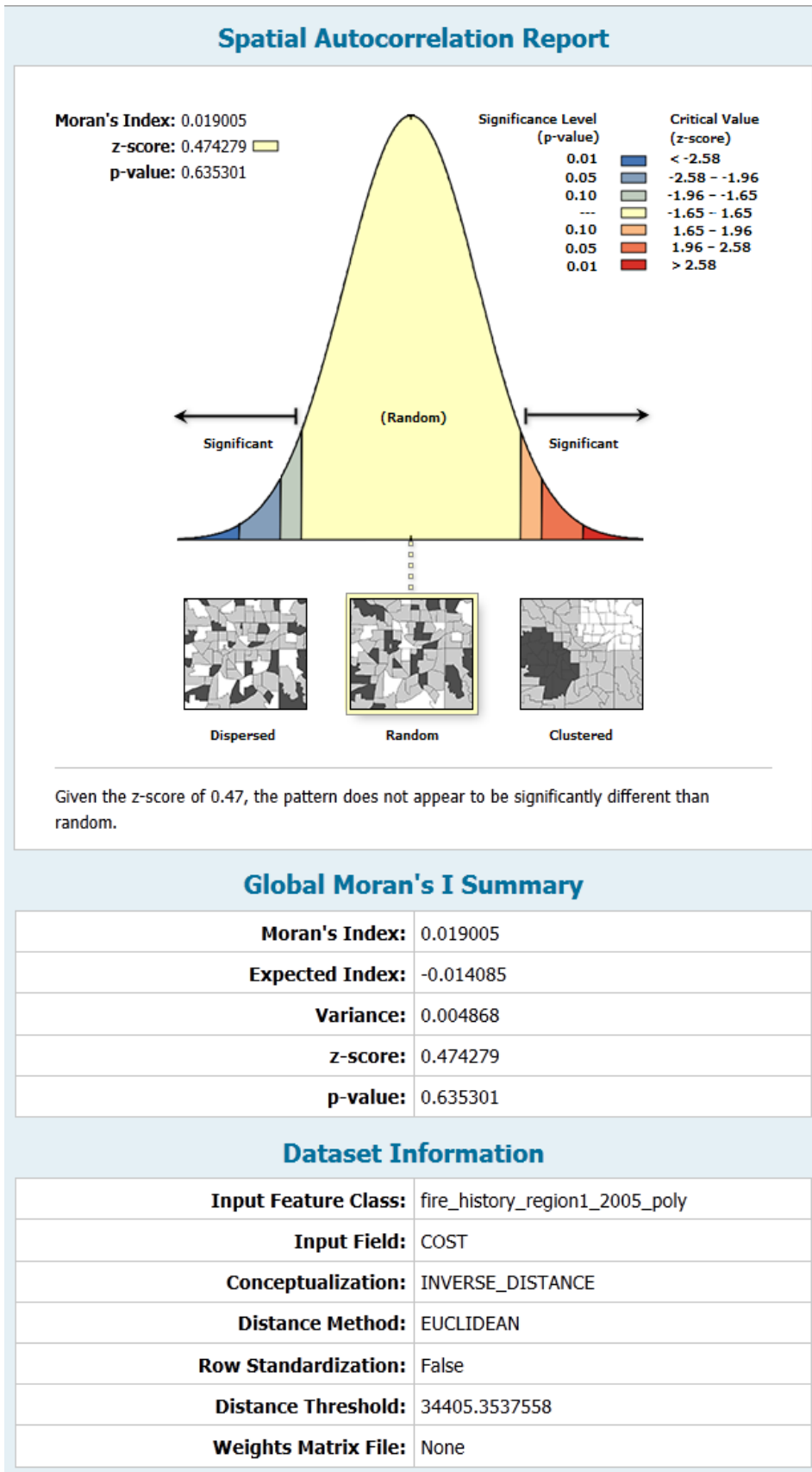
- 1 In ArcToolbox, expand *Spatial Statistics Tools* » *Analyzing Patterns* and then double-click *Spatial Autocorrelation (Morans I)*.

- 2 Use forest fires for 2005 as the input feature class. Use *COST* as the input field. Make sure that the *Generate Report (optional)* check box is checked. Note that in the previous section, you looked at the relative location of the fires without weighting by cost of the fire. You might want to repeat the previous section with *COST* as the weight field and compare the resultant ellipses and centroids.

You can view the results in ArcMap by going to the Menu bar and clicking *Geoprocessing » Results* or using an .html file that is automatically generated and can be opened in any HTML reader of your choice, such as Firefox or Internet Explorer. The .html file will be saved to your ArcGIS default directory (for example, *Libraries\Documents\ArcGIS\*).





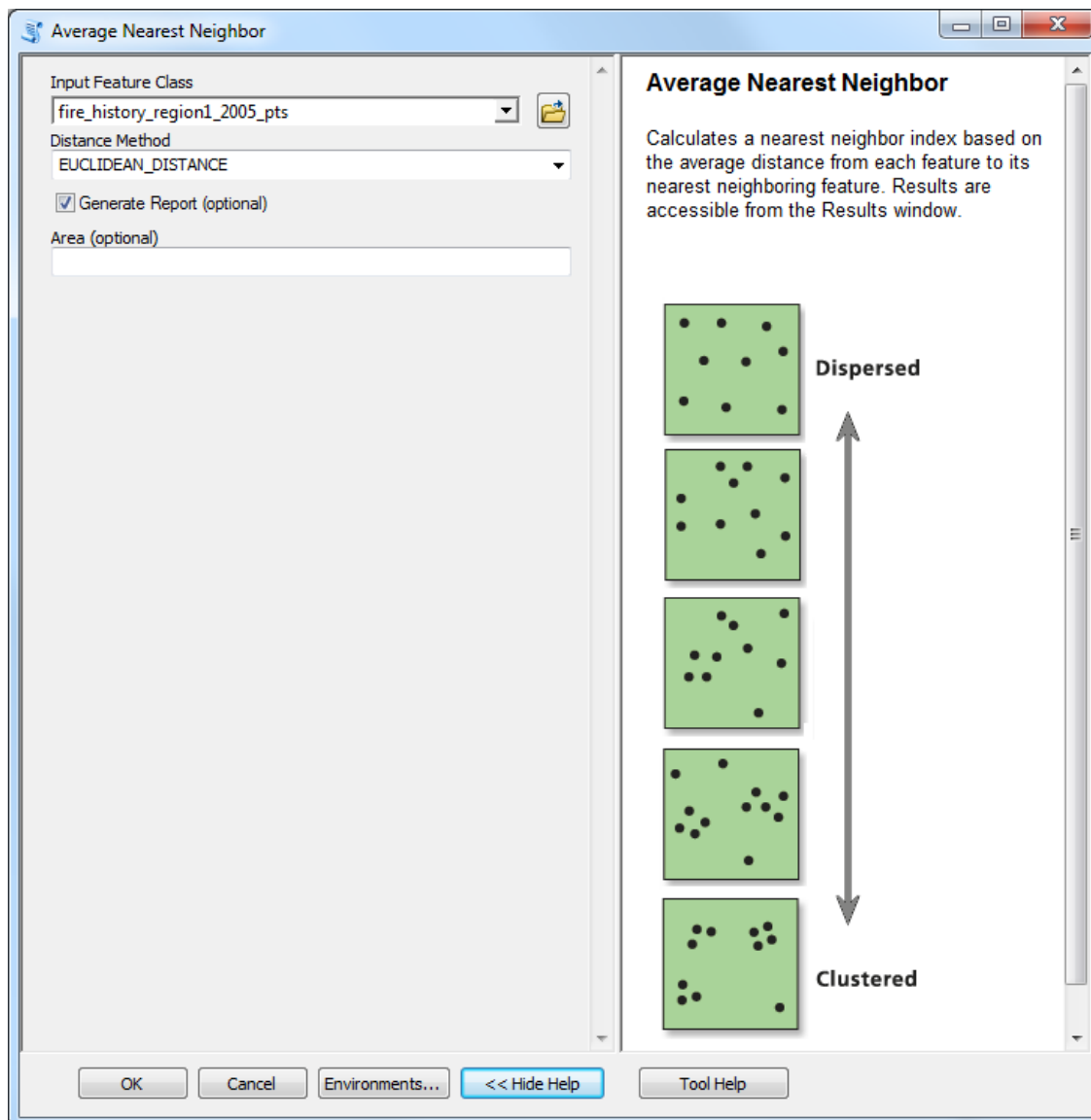


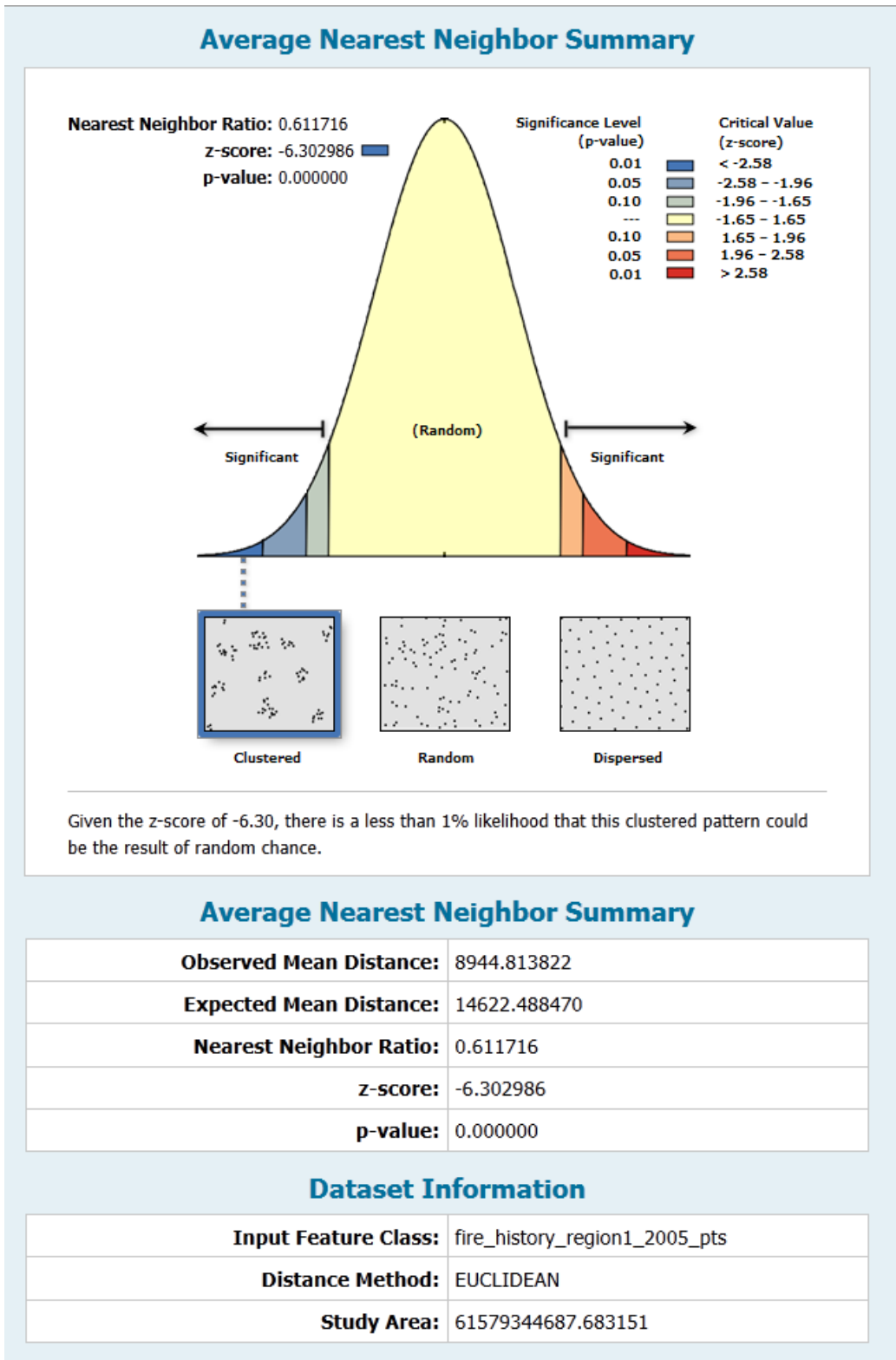
The results suggest that, globally, the fires were not clustered based on the cost of mitigation in 2005.

- 3 Repeat steps 1 and 2 for the 2009 data.

**Question 4:** What does this indicate about the clustering of the cost of fighting forest fires in 2005 and 2009?

- 4 In ArcToolbox, expand *Spatial Statistics Tools* » *Analyzing Patterns* and then double-click *Average Nearest Neighbor*. Use the shapefile for forest fires in 2005 as the input feature class and other parameters as shown in the screen capture that follows. Note that the results are not weighted by cost. Also note that the results are visible in the *Results window* as before and as an .html file.



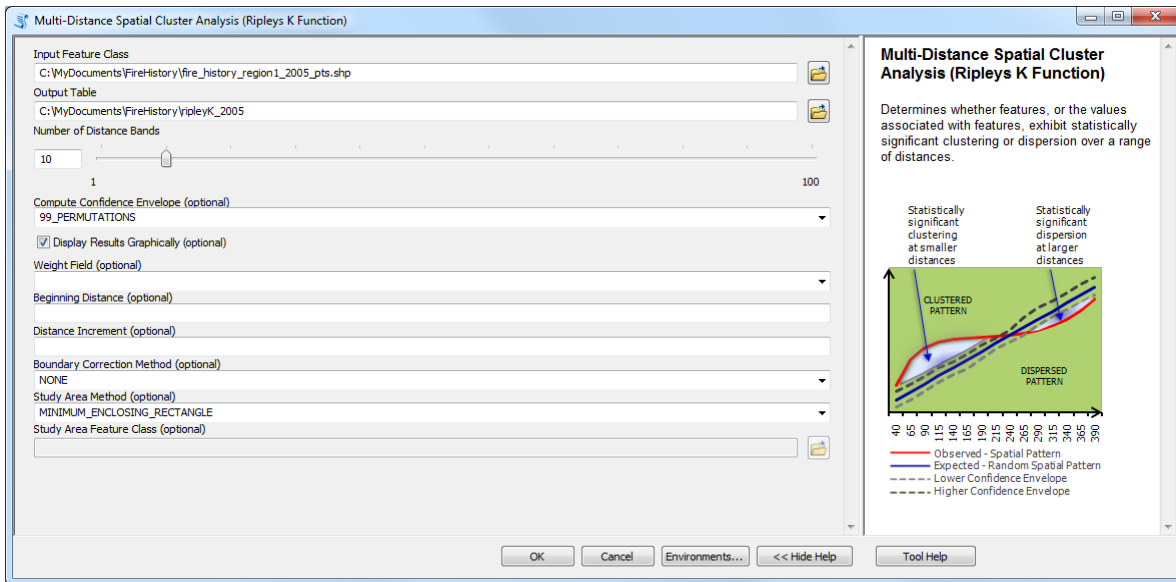


5 Repeat step 4 for 2009.

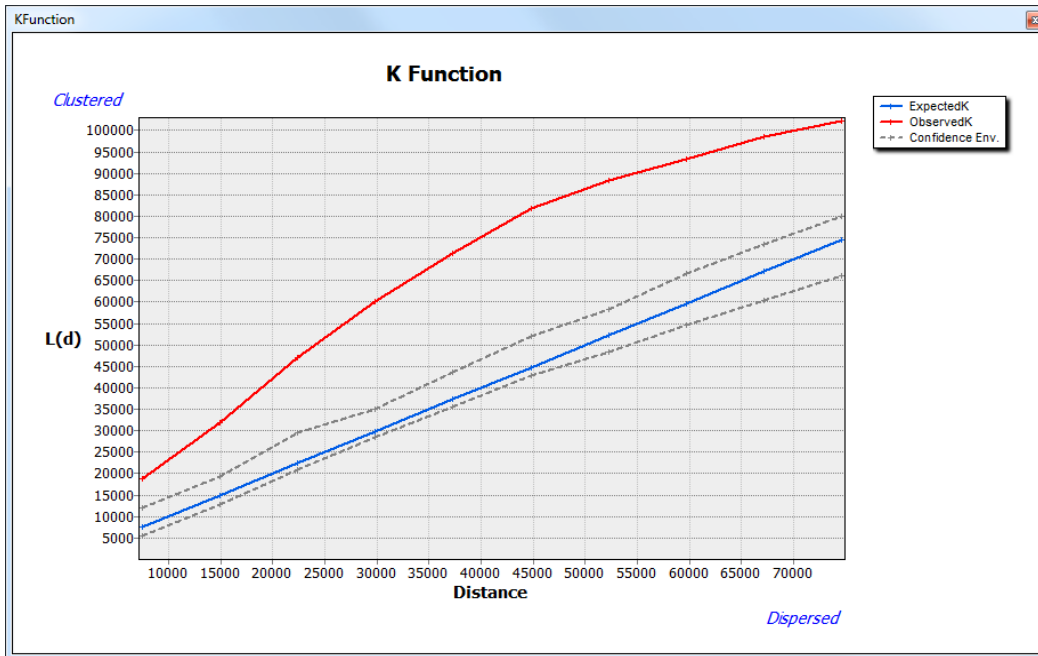
**Question 5:** *What do the nearest neighbor distances indicate about the distribution of forest fires in 2005 and 2009?*

- 6 In the same ArcToolbox location (*Spatial Statistics Tools » Analyzing Patterns*), double-click *Multi-Distance Spatial Cluster Analysis (Ripleys K Function)*. Use the forest fires in 2005 shapefile as the input feature class. Output the table to your results folder. See the screen capture that follows for the other parameters. Be sure to select *99\_PERMUTATIONS* for the confidence envelope and select the *Display Results Graphically (optional)* check box; do not use a weight field.

Note that the results are also saved in a .dbf file that you can open in any spreadsheet or statistical software to create your own graphs.



See the graphic below for the results from 2005 and compare it to the graphic shown in the preceding Ripley's K function screen capture. Note that the observed pattern is more dispersed than expected at all the distances that were plotted.



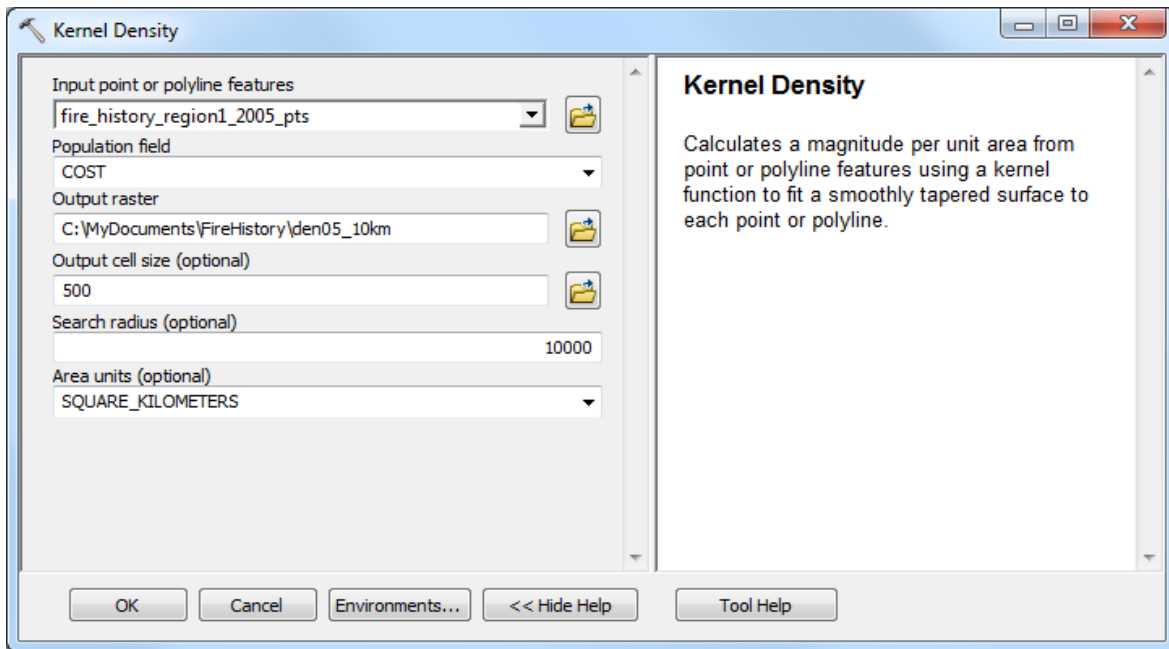
7 Repeat step 6 for 2009.

**Question 6:** What do the results suggest? At what distance does the pattern appear to be clustered, if at all? What are the differences between the results in 2005 and 2009?

Note that kernel density estimation only works with lines or points.

8 Make sure the Spatial Analyst extension has been activated (you can do this by going to the Menu bar and clicking *Customize » Extensions » Spatial Analyst*).

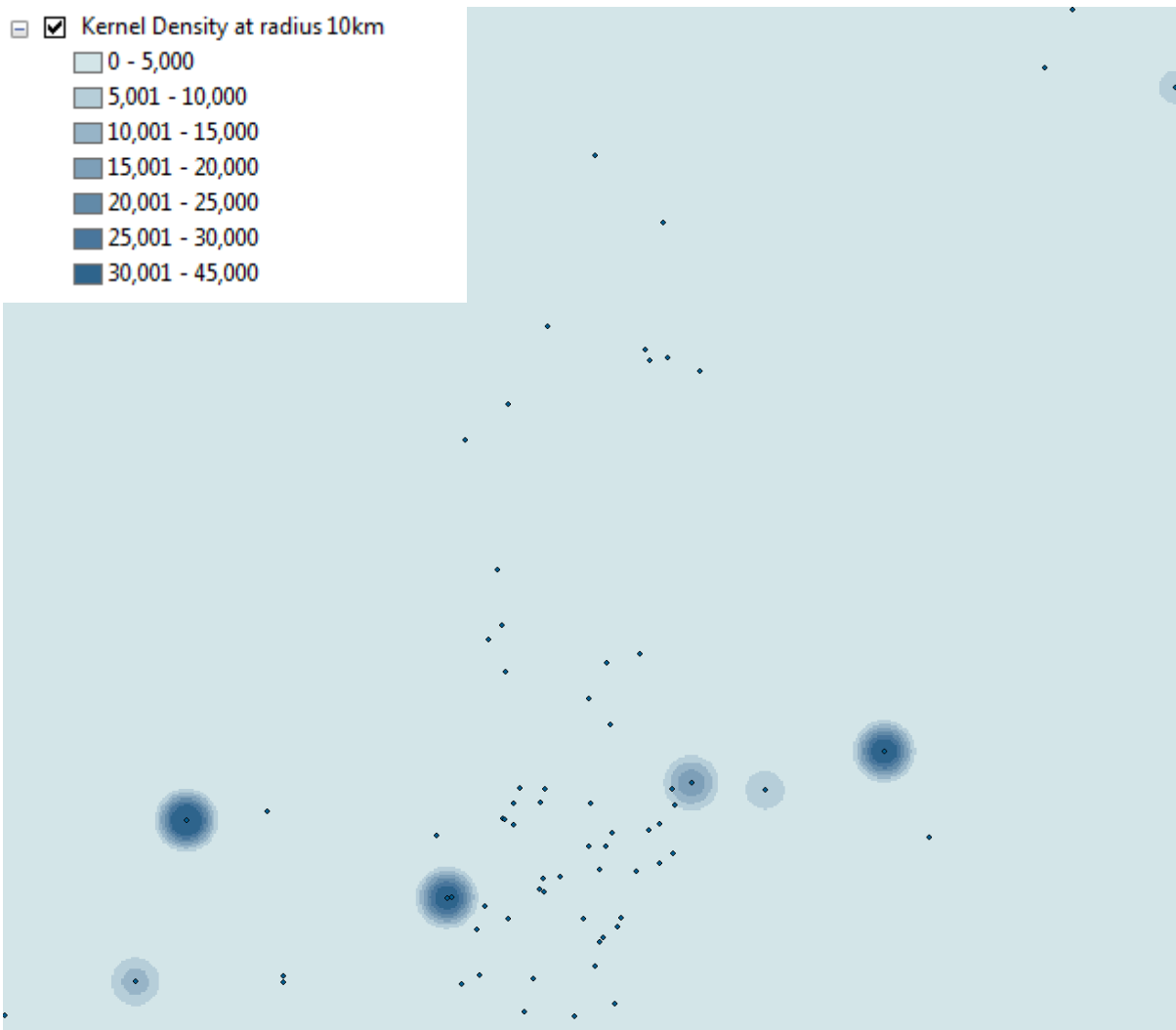
- 9 In ArcToolbox, expand *Spatial Analyst Tools » Density*, and then double-click *Kernel Density*. Use the point shapefile for forest fires as the input point features. Also, identify the name of the output density raster of fires as *fire05cost1km*. Enter the other parameters as shown in the screen capture below. Make sure you select *COST* as the population field, *500* as the output cell size, and *10000* as the search radius.



- 10 Repeat step 9 for other bandwidths (search radii). Then examine the data output by changing the symbology and classification. Try at least three to four search radii, varying between 5,000 and 20,000 meters. Also, try to generate kernel density with and without weighting by cost as the population field.
- 11 Repeat steps 9 and 10 for the 2009 data.
- 12 Add these rasters to your ArcMap document. Compare the densities for 2005 and 2009 side by side with the same search radius (bandwidth) and population field values. Change the legends so that they have the same legend classification breakpoints.

**Question 7:** *What do the results suggest? Where do the clusters occur in 2005 versus 2009? How does weighting by cost affect the patterns observed?*

## VISUALIZE



This image shows kernel density for forest fires in 2005 weighted by cost with a bandwidth (search radius) of 10 kilometers.

## Submit your work

---

Submit answers to questions 1–7 along with relevant screen captures. Describe how these maps and statistics could inform forest fire management policy.

# Credits

---

## Data

Data used in this activity courtesy of U.S. Forest Service.



# Instructor resources

---

## Additional information

This is a useful exercise to introduce basic spatial statistics and point pattern analysis to students who have had a first course in GIS and are interested in more advanced applications. It might be useful to ask students to think about this for other similar spatial point patterns in crime or pollution, for example. You could end with discussions on the theoretical implications of varying the kernel density bandwidths and what might be relevant to different kinds of applications (homicides, assaults, hazardous spills, etc). Note that all these operations could be done using CrimeStat 3.0 or later, which is publicly available at <http://www.icpsr.umich.edu/CrimeStat/download.html>.

The results from the analysis in CrimeStat require GIS software that can read shapefiles.

## Answers to questions

**Question 1:** *Were there any locations that had forest fires in both years? If so, where?*

**Answer:** Cedar Creek had fires in both years.

**Question 2:** *Did the mean center change from 2005 to 2009? If so, where did it move, and why is this interesting?*

**Answer:** Yes, the mean center changed and moved in a northern direction. This is interesting because it means that the most expensive forest fires in 2009 tended to be concentrated closer to the northern parts of the region. Also note that the mean center was in Idaho in 2005, but in 2009 it was in Montana.

**Question 3:** *Did the standard deviational ellipses change from 2005 to 2009? Explain what this means about the changes in the spatial patterns of distribution of forest fires over the two years.*

**Answer:** Yes, the ellipses change in both the spread and direction. The spread of the most expensive forest fires in 2005 was concentrated closer to the mean center and tended to be in a northeastern–southwestern direction. In 2009, the spread was wider and in a southeastern–northwestern direction from the mean center. Again, note that the bulk of the first standard deviational ellipse in 2009 was in Montana, whereas it was more equally spread across Idaho and Montana in 2005.

**Question 4:** *What does this indicate about the clustering of the cost of fighting forest fires in 2005 and 2009?*

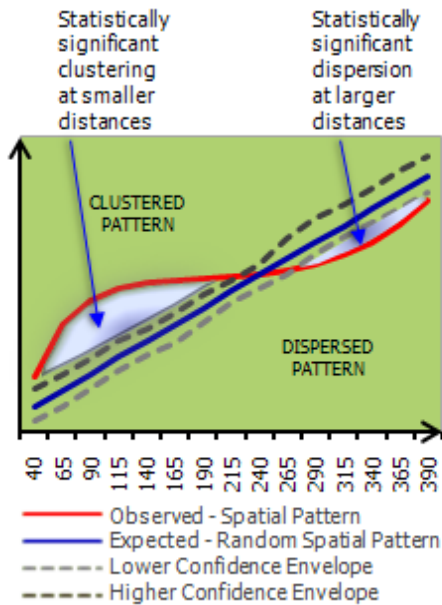
**Answer:** The more expensive fires in 2009 were in Montana, were less clustered, and covered a larger part of the region than they did in 2005. This suggests that the cost of fighting fires was less locally concentrated in 2009.

**Question 5:** *What do the nearest neighbor distances indicate about the distribution of forest fires in 2005 and 2009?*

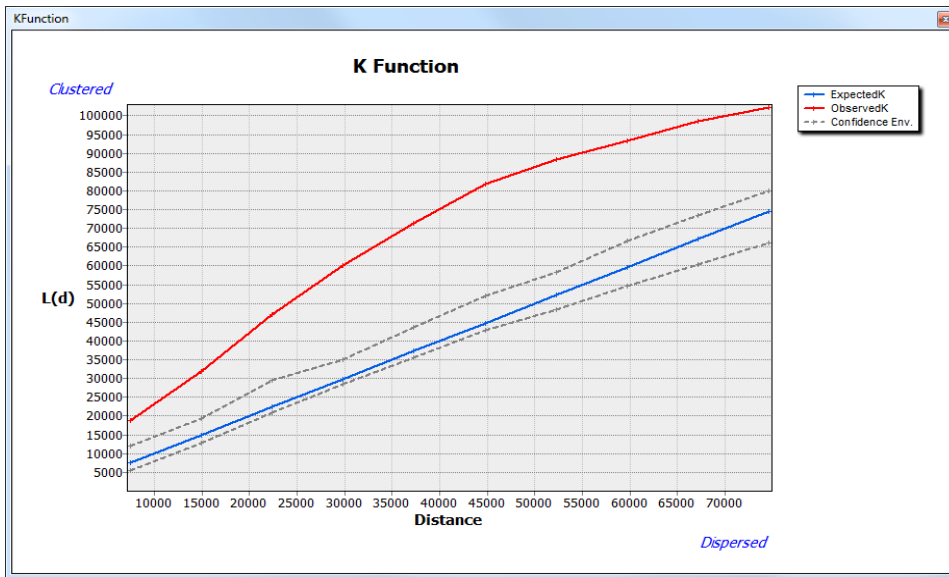
**Answer:** The nearest neighbor index is expressed as the ratio of the observed mean distance to the expected mean distance. If the index is less than 1, the pattern exhibits clustering; if the index is greater than 1, the trend is toward dispersion or competition. In this case, both patterns show clustering, although the value for 2005 is higher than for 2009, which suggests tighter clustering of the forest fire locations in 2005. Note that this measure is not weighted by cost as in the case of the mean center and standard ellipses.

**Question 6:** *What do the results suggest? At what distance does the pattern appear to be clustered, if at all? What are the differences between the results in 2005 and 2009?*

**Answer:** Compare both graphs to this picture:

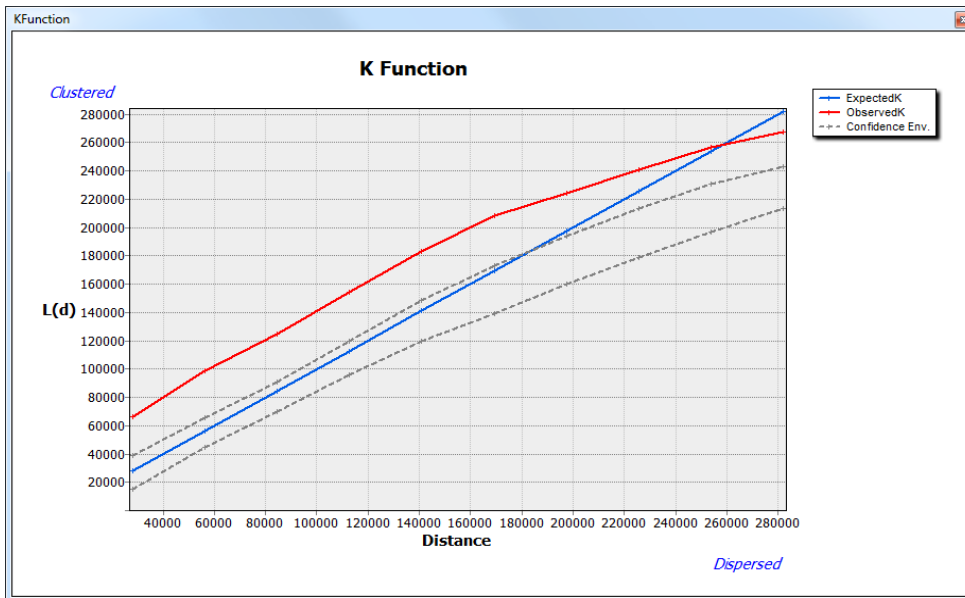


## 2005



In 2005, the graph lies above the expected Ripley's K, suggesting a highly clustered pattern. The K values, however, lie outside the confidence envelopes, so you cannot be sure that this pattern of clustering is significant.

## 2009



In 2009, the pattern is clustered until 250 km, but again, this is not significant.

Thus, based on Ripley's K, there is no significant clustering in either year.

**Question 7:** *What do the results suggest? Where do the clusters occur in 2005 versus 2009? How does weighting by cost affect the patterns observed?*

**Answer:** This will depend on the student's choice of bandwidth and should be assessed accordingly. The main intent of this question is for the student to be able to point out that the density of forest fires will vary depending on the choice of bandwidth and whether weighting by cost was used. Appropriate bandwidth for various phenomena should be discussed in class to prepare the students. For example, the appropriate bandwidth for forest fire density may depend on the types of species in the forest, rainfall in that particular year, and local microclimatic conditions favorable to forest fires.

## References and further reading

Anselin, Luc. 2003. *Point Pattern Analysis Using CrimeStat*. Urbana Champaign: University of Illinois.  
<http://geodacenter.asu.edu/system/files/points.pdf> (accessed November 2011).

Baddeley, Adrian. 2008. *Using R for Spatial Pattern Analysis*. Australia: CSIRO.  
<http://www.csiro.au/files/files/pn0y.pdf> (accessed November 2011).

Dixon, Philip M. 2002. "Ripley's K Function." In *Encyclopedia of Environmetrics* 3: 1796–1803. Chicester, UK: Wiley.  
[http://www.public.iastate.edu/~pcaragea/S40608/Notes/Dixon\\_Ripley\\_K.pdf](http://www.public.iastate.edu/~pcaragea/S40608/Notes/Dixon_Ripley_K.pdf)

Esri. "Multi-Distance Spatial Cluster Analysis (Ripley's K Function)." ArcGIS Resource Center.  
<http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/005p0000000m0000000.htm>

Fotheringham, A. Stewart, Chris Brunsdon, and Martin Charlton. 2000. *Quantitative Geography: Perspectives on spatial data analysis*. (London: Sage Publications).

Levine, Ned. 2010. *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations*. Houston, TX: Ned Levine & Associates.  
<http://www.icpsr.umich.edu/CrimeStat/>

O'Sullivan, David, and David Unwin. 2010. *Geographical Information Analysis*. (Hoboken, NJ): Wiley & Sons. Print ISBN: 9780470288573. Online ISBN: 9780470549094.