



# How Hot Spot Analysis (Getis-Ord $G_i^*$ ) works

ArcGIS Pro 3.6

## In this topic

1. [Calculations](#)
2. [Interpretation](#)
3. [Output](#)
4. [Hot spot analysis considerations](#)
5. [Best practices](#)
6. [Potential applications](#)
7. [Additional resources](#)

The [Hot Spot Analysis \(Getis-Ord  \$G\_i^\*\$ \)](#) tool calculates the Getis-Ord  $G_i^*$  statistic (pronounced G-i-star) for each feature in a dataset. The resultant [z-scores and p-values](#) indicate where features with either high or low values cluster spatially. This tool evaluates each feature within the context of neighboring features. A feature with a high value is interesting but may not be a statistically significant hot spot. To be a statistically significant hot spot, a feature will have a high value and be surrounded by other features with high values. The local sum for a feature and its neighbors is compared proportionally to the sum of all features. When the local sum is very different from the expected local sum, and when that difference is too large to be the result of random chance, a statistically significant [z-score](#) results. When the [FDR correction](#) is applied, statistical significance is adjusted to account for multiple testing and spatial dependency.

## Calculations

The calculations for the Getis-Ord  $G_i^*$  statistic are shown in the following image:

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - \left( \sum_{j=1}^n w_{i,j} \right)^2}{n-1}}} \quad (1)$$

where  $x_j$  is the attribute value for feature  $j$ ,  $w_{i,j}$  is the spatial weight between feature  $i$  and  $j$ ,  $n$  is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (3)$$

The  $G_i^*$  statistic is a z-score so no further calculations are required.

## Interpretation

The  $G_i^*$  statistic returned for each feature in the dataset is a z-score. For statistically significant positive z-scores, the larger the z-score, the more intense the clustering of high values (hot spots). For statistically significant negative z-scores, the smaller the z-score, the more intense the clustering of low values (cold spots). For more information about determining statistical significance and correcting for multiple testing and spatial dependency, see [What is a z-score? What is a p-value?](#)

## Output

This tool creates an output feature class with a z-score, p-value, and confidence level bin (Gi\_Bin) for each input feature.

When the tool completes, the output feature class is added to the map with rendering applied to the Gi\_Bin field.

## Hot spot analysis considerations

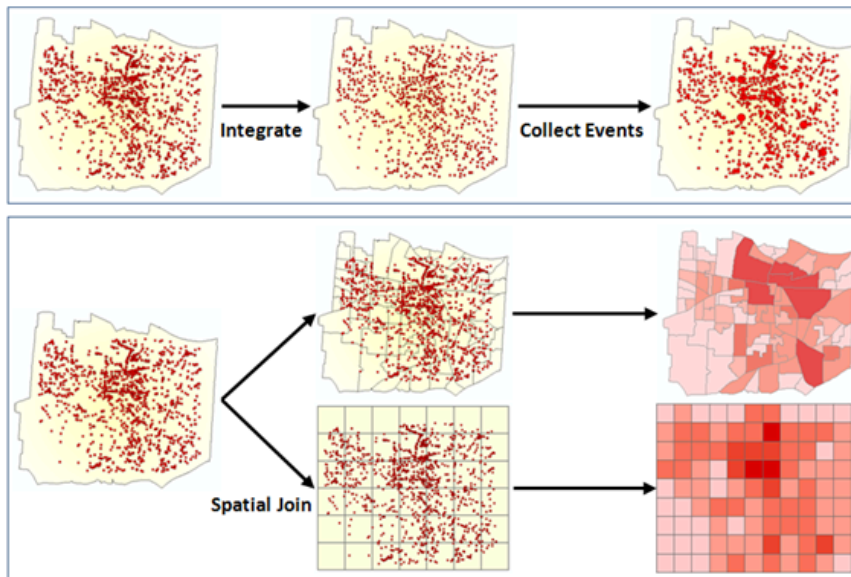
Consider the following when undertaking a hot spot analysis:

- The tool assesses whether high or low values (the number of crimes, accident severity, or dollars spent on sporting goods, for example) cluster spatially. The field containing those values is the analysis field. For point incident data, however, you may be more interested in assessing incident intensity than in analyzing the spatial clustering of any particular value associated with the incidents. In that case, aggregate the incident data before running the analysis by doing one of the following:
  - If you have polygon features for the study area, use the [Spatial Join](#) tool to count the number of events in each polygon. The resultant field containing the number of events in each polygon becomes the Input Field parameter value for the Hot Spot Analysis (Getis-Ord  $G_i^*$ ) tool.
  - Use the [Create Fishnet](#) or [Generate Tesselation](#) tool to construct a polygon grid over the point features. Then use the [Spatial Join](#) tool to count the number of events within each grid polygon. Remove any grid polygons that are outside the study area. Also, in cases where many of the grid polygons within the study area contain zeros for the number of events, increase the polygon grid size, if appropriate, or remove those zero-count grid polygons.
  - If you have a number of coincident points or points within a short distance of one another, use the [Integrate](#) tool with the [Collect Events](#) tool to snap features within a specified distance of each other together, and create a new feature class containing a point at each unique location with an associated count attribute to indicate the number of events (snapped points). Use the resultant ICOUNT field as the Input Field parameter value for the Hot Spot Analysis (Getis-Ord  $G_i^*$ ) tool.

**Note:**

If the coincident points may be redundant records, use the [Find Identical](#) tool to locate and remove the duplicates.

Examples of using the tools mentioned above to aggregate incident data are shown in the following image:



Strategies for aggregating incident data are shown.

- The recommended (and default) Conceptualization of Spatial Relationships parameter value for the Hot Spot Analysis (Getis-Ord  $G_i^*$ ) tool is Fixed distance band. The Space-Time Window, Zone of Indifference, K Nearest Neighbors, and Delaunay Triangulation options may also work well. For a discussion of best practices and strategies for determining an analysis distance value, see [Best practices for selecting a conceptualization of spatial relationships](#) and [Best practices for selecting a fixed distance band value](#). For more information about space-time hot spot analysis, see [Space-time cluster analysis](#).
- The input field determines the types of questions you can ask. If you are most interested in determining where there are lots of incidents or where high and low values for a particular attribute cluster spatially, run the Hot Spot Analysis (Getis-Ord  $G_i^*$ ) tool on the raw value incident counts. This type of analysis is particularly helpful for resource allocation problems. Alternatively (or in addition), you may be interested in locating areas with unexpectedly high values in relation to some other variable. If you are analyzing foreclosures, for example,

you may expect more foreclosures at locations with more homes (that is, you expect the number of foreclosures to be a function of the number of houses). If you divide the number of foreclosures by the number of homes, and run the Hot Spot Analysis (Getis-Ord  $G_i^*$ ) tool on this ratio, you are no longer asking where are there lots of foreclosures. Instead, you are asking where are there unexpectedly high numbers of foreclosures, given the number of homes. By creating a rate or ratio before you run the analysis, you can control certain expected relationships (for example, the number of crimes is a function of population; the number of foreclosures is a function of housing stock) and identify unexpected hot and cold spots.

## Best practices

The following are best practices for using the Hot Spot Analysis (Getis-Ord  $G_i^*$ ) tool:

- The Input Feature Class parameter value should have at least 30 features. Results aren't reliable with less than 30 features.
- Specify the appropriate Conceptualization of Spatial Relationships parameter value. For this tool, the Fixed distance band option is recommended. For space-time hot spot analysis, see [Best practices for selecting a conceptualization of spatial relationships](#).
- Specify the appropriate Distance Band or Threshold Distance parameter value.
  - All features should have at least one neighbor.
  - No feature should have all other features as neighbors.
  - Features should have about eight neighbors each, especially if the values for the Input Field parameter value are skewed. You can use the [Calculate Distance Band From Neighbor Count](#) tool to find the average distance at which each feature has eight neighbors.

## Potential applications

Applications can be found in crime analysis, epidemiology, voting pattern analysis, economic geography, retail analysis, traffic incident analysis, and demographics, for example. You can answer the following types of questions:

- Where is the disease outbreak concentrated?
- Where are kitchen fires a larger than expected proportion of all residential fires?
- Where should the evacuation sites be located?
- Where and when do peak intensities occur?
- Which locations and during which time periods should resources be allocated?

## Additional resources

For additional information regarding spatial statistics, see the following resources:

Mitchell, Andy. [The ESRI Guide to GIS Analysis, Volume 2. ESRI Press, 2005.](#)

Getis, A. and J.K. Ord. 1992. "[The Analysis of Spatial Association by Use of Distance Statistics](#)" in *Geographical Analysis* 24(3).

Ord, J.K. and A. Getis. 1995. "[Local Spatial Autocorrelation Statistics: Distributional Issues and an Application](#)" in *Geographical Analysis* 27(4).

The [spatial statistics resource page](#) has short videos, tutorials, web seminars, articles and a variety of other materials to help you get started with spatial statistics.

Scott, L. and N. Warmerdam. [Extend Crime Analysis with ArcGIS Spatial Statistics Tools](#) in ArcUser Online, April–June 2005.

---