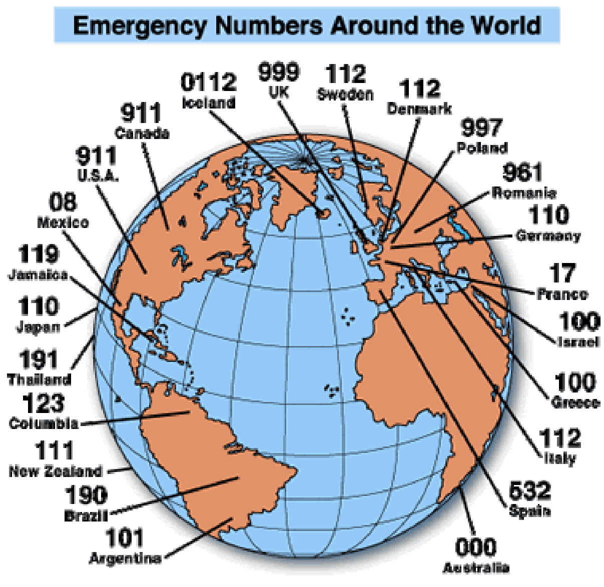
Exploring EMS 911Call data using Hot Spot Analysis

Most local, regional, and national governments, airports, train stations, and other public places have some form of emergency management and preparedness programs. It is vital for these governments and organizations to be ready for emergency situations and be in the best position to safeguard lives and property when emergencies or disasters do occur.



DISPATCH Monthly graphic

Among the analytical methods used for emergency management and preparedness are Spatial Statistics tools. These tools are especially powerful for emergency applications because of the importance of location in emergency and disaster reporting. Spatial Statistics tools look at the distribution of values associated with geographic features and make comparisons to a hypothetical random spatial distribution in order to answer location-based questions in a statistically sound way. Questions that can be addressed using Spatial Statistics tools include: Where are areas with higher than normal crime levels (hotspots)? Where are areas with higher than normal cancer or diabetes rates and what could be causing this? Where is the most accessible location for a new hospital or emergency services station?

Tutorial: Exploring EMS Call Data

Introduction:
In this tutorial, 911 Emergency call data for a portion of the Portland, Oregon metropolitan area will be investigated and analyzed. Analysis will be performed using the Spatial Statistics Hot Spot Analysis tool, which uses the Getis-Ord Gi* algorithm. This analysis will identify statistically significant spatial clusters of high and low 911 emergency calls in the study area.

Context for the analysis:
Authorities in the study area are spending a large portion of public resources responding to 911 emergency calls, and projections indicate the community will double in population in the next ten years. They want to better understand the distribution of their current calls and to identify patterns and trends to determine if more effective resource allocation is possible. Hot spot analysis will help you understand the spatial patterns in this data and can assist in finding the optimal locations for new emergency service stations based on clusters of high 911 call activity.

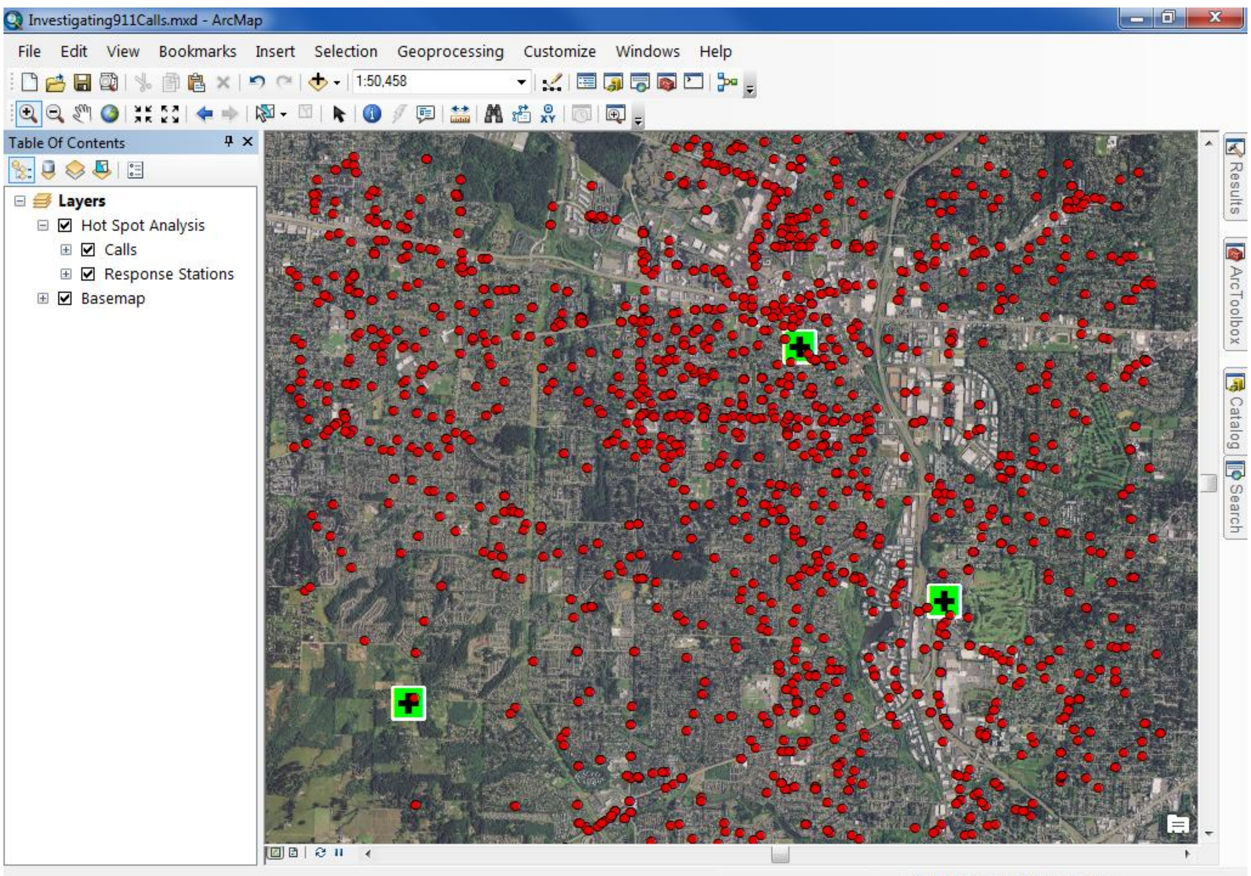
The Steps:

1. Make sure your data is projected
2. Aggregate your incident data
3. Run Hot Spot Analysis
    a. Find an appropriate scale of analysis
4. Visualize your results
5. Share your findings

Let's get started…

1. Open Investigating911Calls.mpk

In this document the Table of Contents contains several data layers for the Portland Oregon metropolitan study area, including 911 calls (Calls), emergency service stations (Response Stations) and an Imagery Basemap (requires an Internet connection).

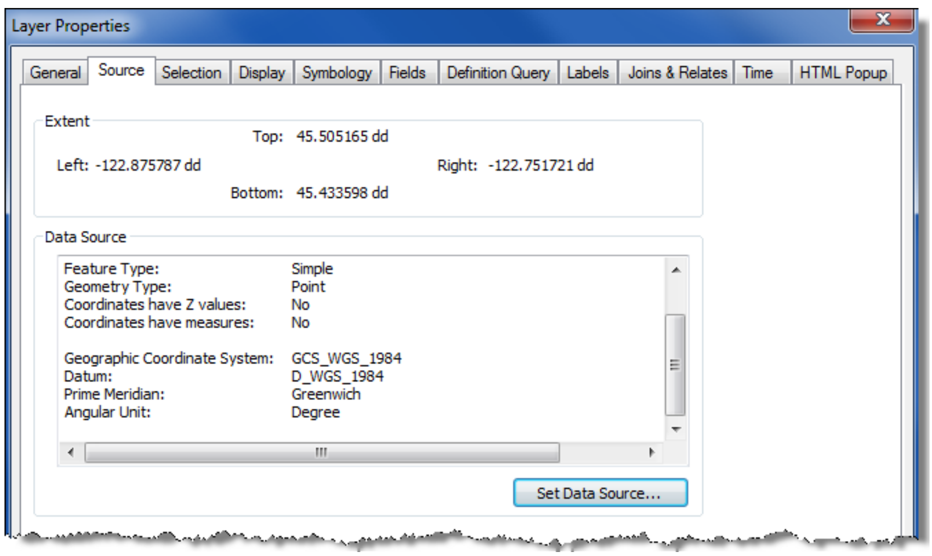


In the map each of the red points represents a single call into a 911 emergency call center. This is real data and represents over 2000 calls in the study area. Visual inspection of these call points reveals some clusters of high 911 call activity, but it's difficult to discern distinct patterns using visual analysis alone. This is where spatial statistics can help. You'll perform a hot spot analysis to identify statistically significant 911 call hot and cold spots in this study area.

Step 1: Make sure your data is projected

Note: Whenever distance is a component of your analysis, which is almost always the case with spatial statistics, project your data using a Projected Coordinate System (rather than a Geographic Coordinate System based on degrees, minutes, and seconds). In addition to choosing a Projected Coordinate System, it is important to choose a projection that makes sense for the analysis that you are doing and for your study area. Since the spatial statistics tools are using distance in many of their calculations, finding a projection that preserves either distance or area is preferred. You can learn more about projections [here](#).

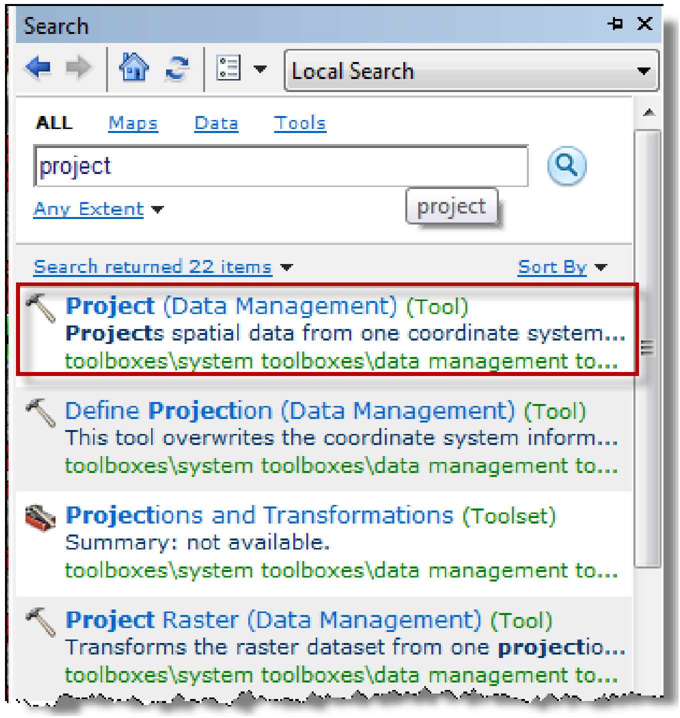
2. Check the projection of the Calls data
    a. Right-click on Calls in the Table of Contents, and choose Properties
    b. From the Source tab, notice that the data is in a Geographic Coordinate System called WGS_1984



    c. Close the Layer Properties dialog

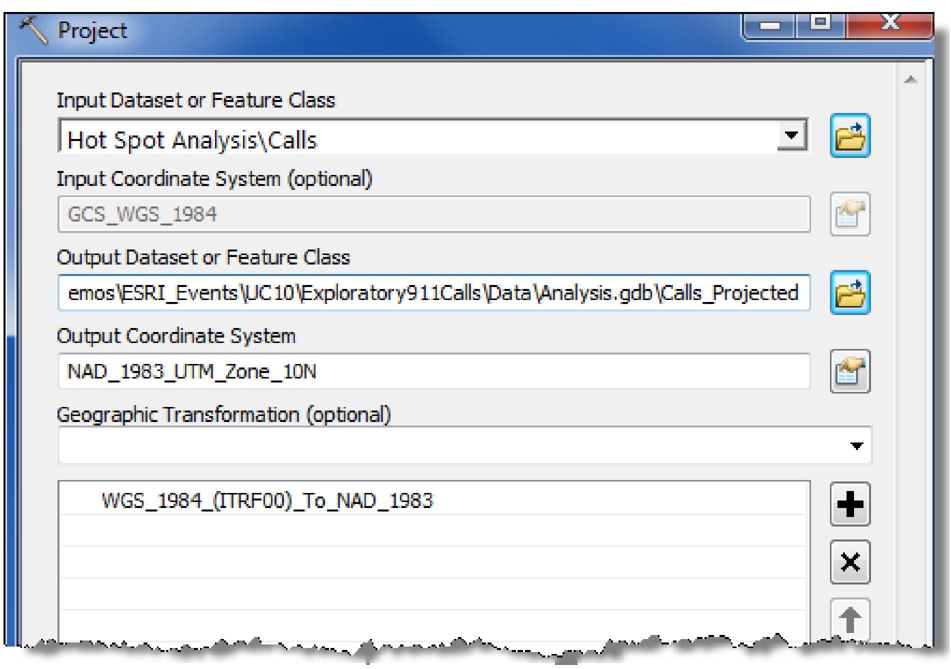
Notice that the data is using a Geographic Coordinate system instead of a Projected Coordinate system. In order to minimize the distortion of your distance calculations, you'll need to project the 911 call data.

3. Project the Calls data
    a. From the Search Window, find the 'Project' tool
    b. Click on the Project (Data Management) tool to open it

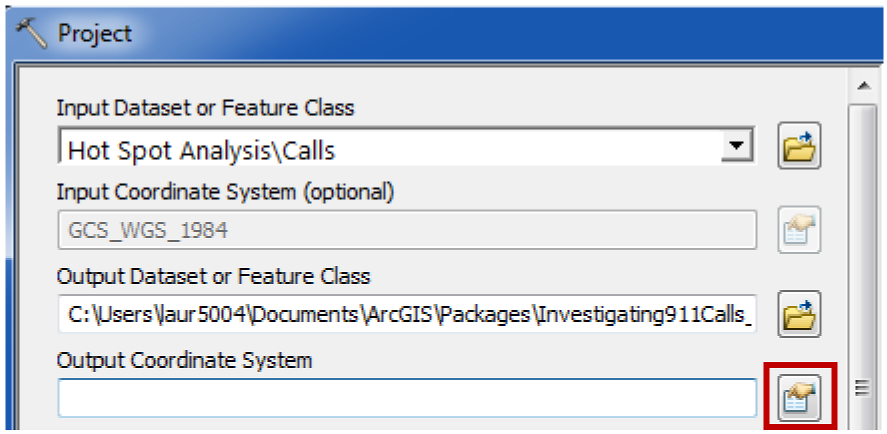


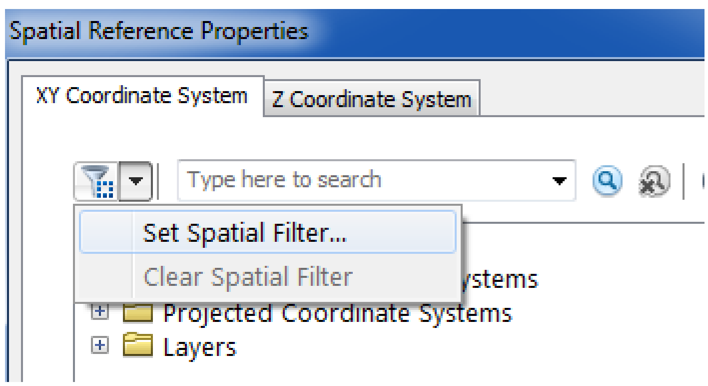
c. Fill the dialog in as follows:

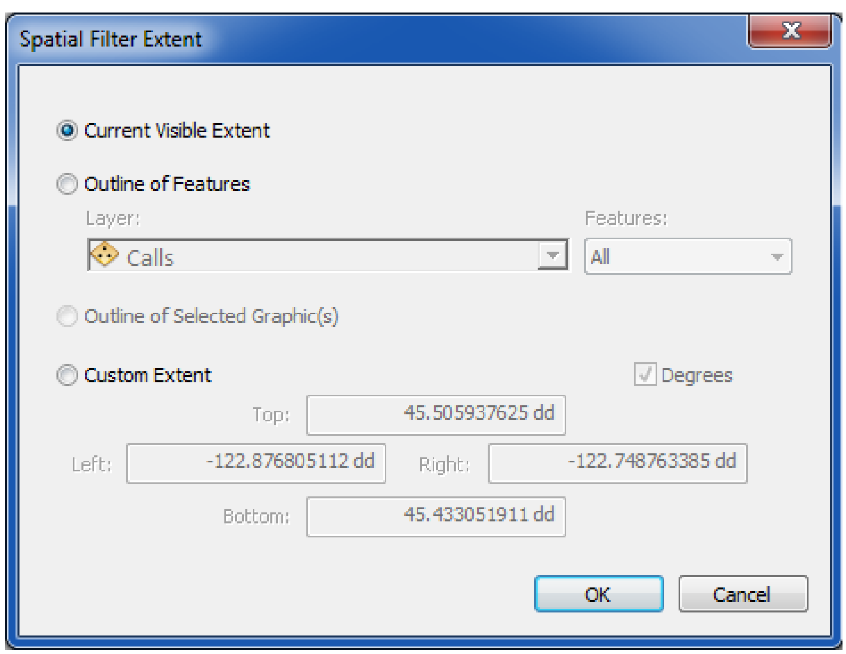
Input Dataset	Hot Spot Analysis\Calls
Input Coordinate System	GCS_WGS_1984
Output Dataset	…\Calls_Projected
Output Coordinate System	NAD_1983_UTM_Zone_10N (Projected Coordinate System\UTM\NAD1983\NAD_1983_UTM_Zone_10N)
Geographic Transformation	WGS_1984_(ITRF00)_To_NAD_1983

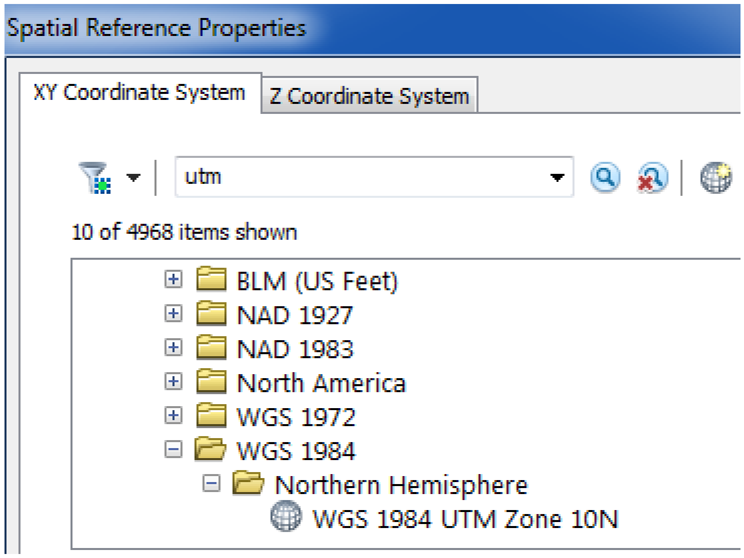


Note: New in ArcGIS 10.1, you can search for a projection that is appropriate for your location using a spatial filter. From the Spatial Reference Properties dialog you can choose to Set Spatial Filter, choose to use the Current Visible Extent, and search for a type of projection (in this case UTM would be a good search keyword). The results will include only projections that are appropriate in your study area.



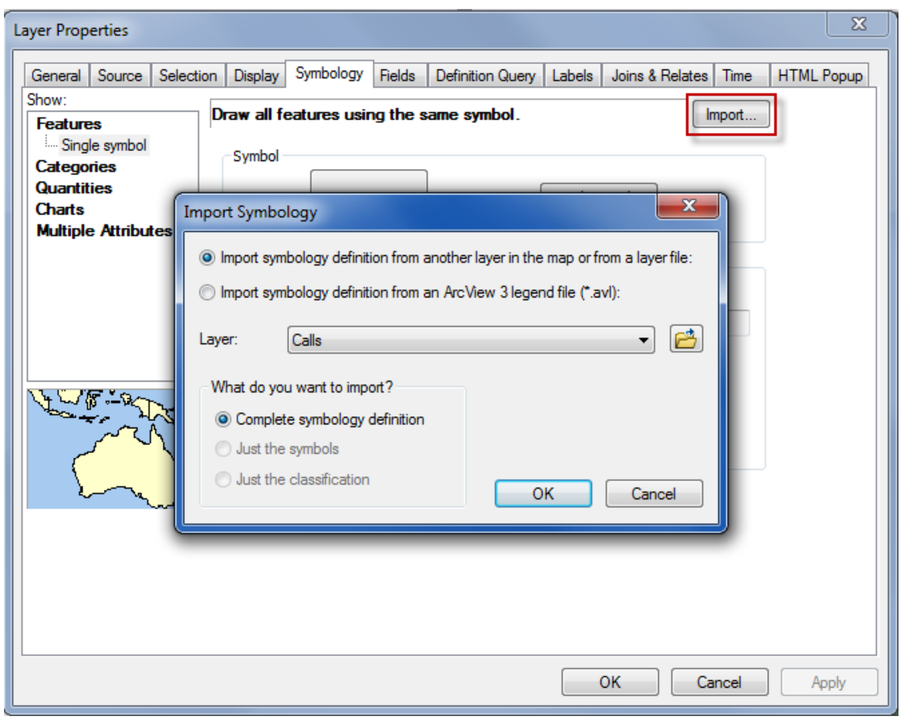






d. Click OK

4. Update the symbology
   a. Double-click on Calls_Projected in the Table of Contents, go to the Symbology Tab, and choose Import…
   b. Choose Calls, and click OK



Step 2: Aggregate your Incident Data

The hot spot analysis 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 your 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, you will need to aggregate your incident data prior to analysis. There are several ways to do this:

- If you have polygon features for your study area, you can 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 for analysis.
- Use the Create Fishnet tool to construct a polygon grid over your point features. Then use the Spatial Join tool to count the number of events falling within each grid polygon. Remove any grid polygons that fall outside your 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 prior to analysis.
- Alternatively, if you have a number of coincident points or points within a short distance of one another, you can use Integrate with the Collect Events tool to (1) snap features within a specified distance of each other together, then (2) 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 your Input Field for analysis.

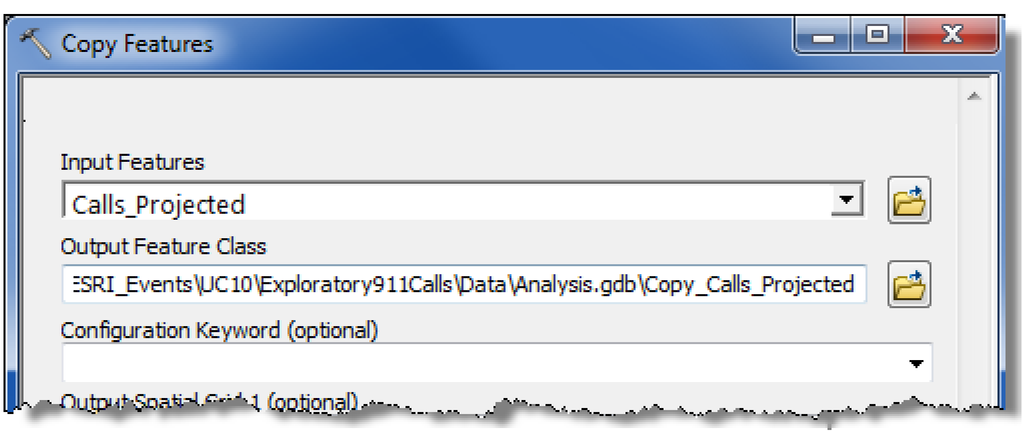
In this tutorial, we are going to use the third option, and apply the Integrate and Collect Events methods. This is a great option for our 911 data because there are many coincident and almost coincident points in our 911 data, this is a good aggregation strategy to use here.

5. Create a copy of the Calls_Projected dataset

Note: The Integrate tool modifies the input dataset by changing the locations of the input features. To preserve the original data, it is very important to make a copy of the original input data before proceeding.

   a. Search for the 'Copy Features' tool and choose Copy Features (Data Management) to open the tool
   b. Fill the dialog in as follows:

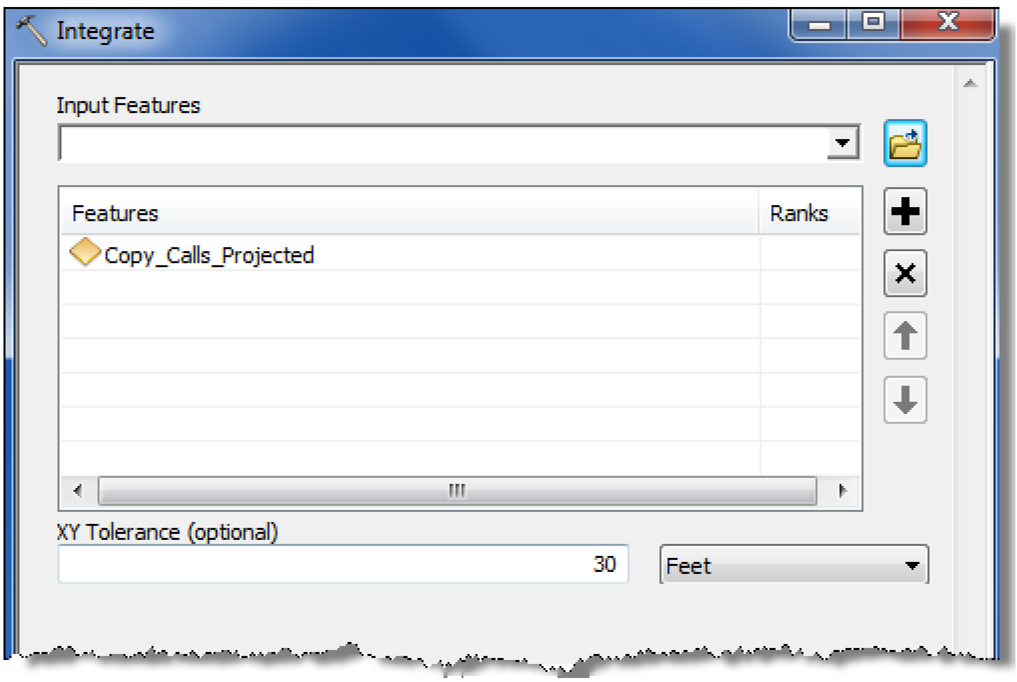
Input Feature Class	Calls_Projected
Output Feature Class	…\Copy_Calls_Projected



   c. Click OK

6. Update the symbology
    a. Double-click on Copy_Calls_Projected in the Table of Contents, go to the Symbology Tab, and choose Import…
    b. Choose Calls, and click OK
7. Run Integrate
    a. Search for the 'Integrate' tool and choose Integrate (Data Management) to open the tool
    b. Fill the dialog in as follows:

Input Features	Copy_Calls_Projected
XY Tolerance	30 Feet



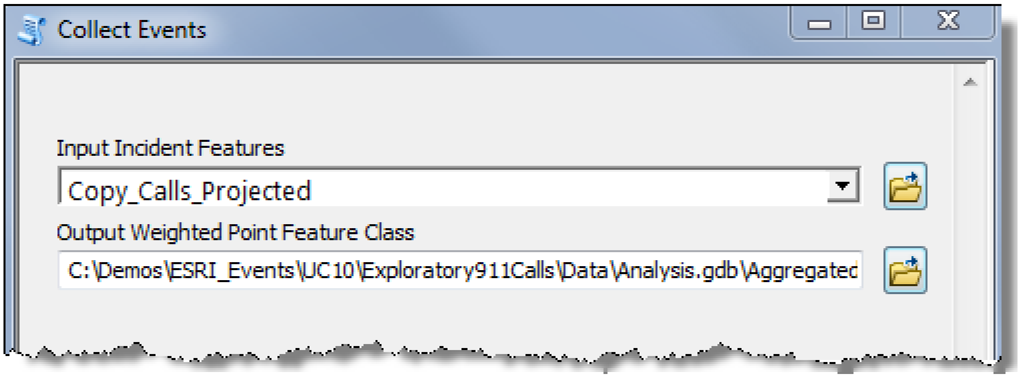
    c. Click OK

A good rule of thumb for the XY tolerance is to consider the accuracy of your data. In this case we know that it is possible that features within 30 feet of each other are likely at the same address, and simply don't have the same coordinates either from differences in geocoding or GPS device accuracy, so 30 feet is a reasonable distance to use.

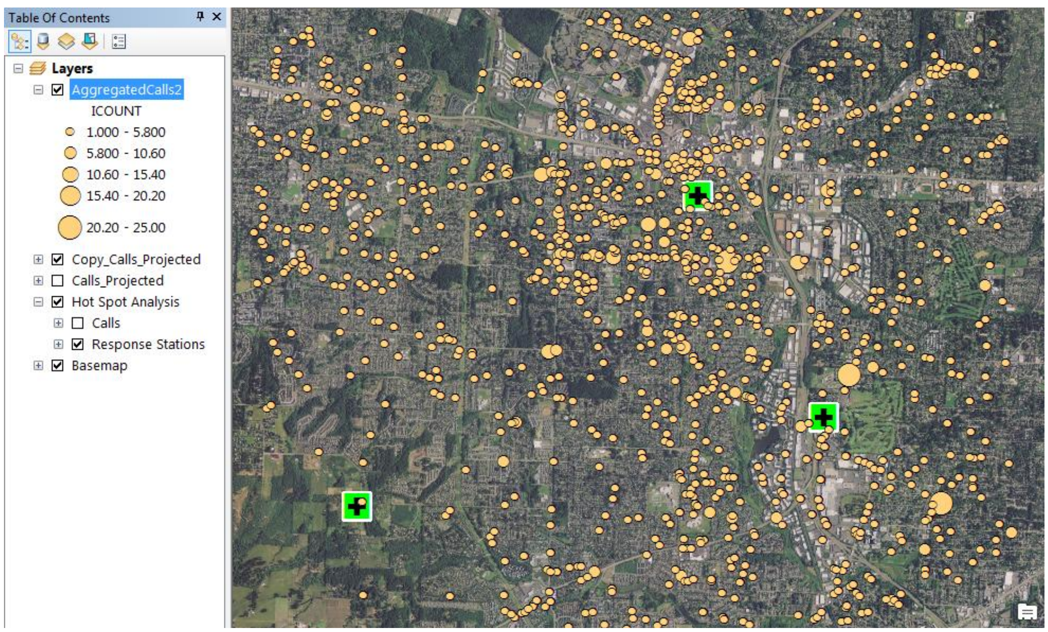
After running the Integrate tool, features within 30 feet of each other will snap to the same location creating a "stack" of coincident features with the exact same X and Y coordinates. The Collect events tool creates a single point feature at each location in the dataset with a count field reflecting the number of points found at that location. For example, if Integrate snaps 4 nearby features together, Collect Events will combine those 4 points into a single point with a count of 4.

8. Run Collect Events
    a. Search for the 'Collect Events' tool and choose Collect Events (Spatial Statistics) to open the tool
    b. Fill the dialog in as follows:

Input Incident Features	Copy_Calls_Projected
Output Weighted Point Feature Class	…\AggregatedCalls



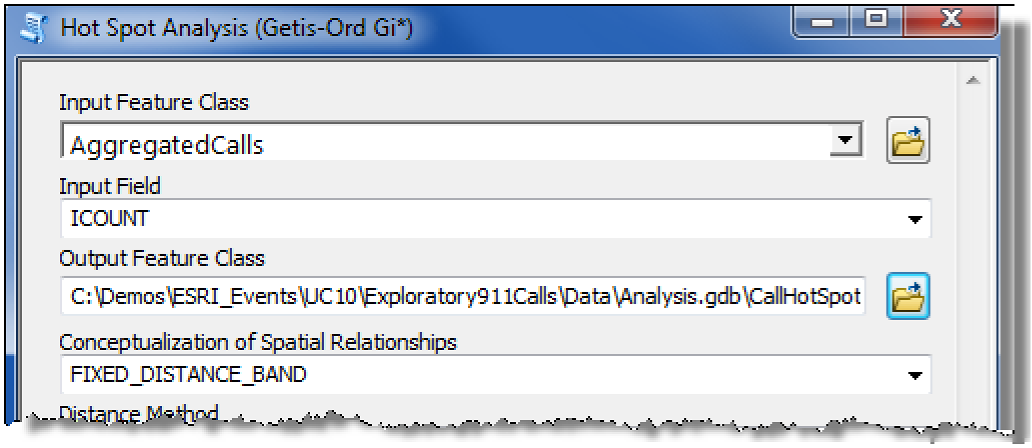
c. Click OK (Notice that the output from Collect Events is rendered with graduated circles reflecting the number of incidents at each point)



Step 3: Run Hot Spot Analysis

Now that we have our incidents aggregated, we're ready to start our Hot Spot Analysis. Let's walk through the tool and some of the important decisions that you have to make when running this kind of analysis.

9. Open the Hot Spot Analysis tool
    a. Search for the 'Hot Spot Analysis' tool and choose Hot Spot Analysis (Getis-Ord Gi*) (Spatial Statistics) to open the tool
    b. We want to use our AggregatedCalls feature class as the input
    c. We want to use the ICount as the input field, since it represents the number of 911 calls at each location
    d. Save the output to a location of your choice



The next decision that we have to make is an important one: choosing the right ***conceptualization of spatial relationships***. This is where your expertise and familiarity with your dataset and your study area become invaluable tools for decision making. The idea is that there is an interaction or influence among a feature and its neighbors…some shared commonality. The hot spot analysis tool works by looking at 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 as well. And the same goes for low values surrounded by other low values and cold spots. Because we are looking at each feature in relation to its neighboring features (because geography matters!), we have to decide what it means to be "neighbors". This is what it means to choose the conceptualization of spatial relationships. Let's walk through the options and see which one makes sense today.

For hot spot analysis the default is Fixed Distance Band, and this is a great option! Fixed Distance Band uses a critical distance to decide what neighbors to include. Because that distance is fixed, it means that the scale of analysis will not change (it will be consistent across the study area) and that is often very important. You can learn more about all of the options for the Conceptualization of Spatial Relationships here.

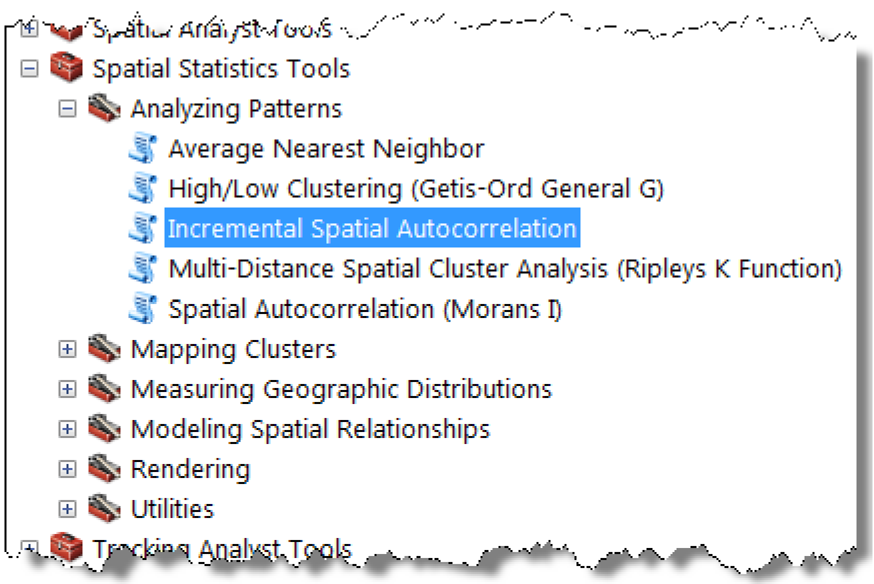
e. Keep the Conceptualization of Spatial Relationships as Fixed_Distance_Band
f. Keep Distance Method as Euclidean Distance
g. Now you need to choose a distance band. The distance you select for this parameter represents the scale of your analysis. We'll come back to the Hot Spot Analysis Tool in a minute. For now, minimize the Hot Spot Analysis tool.

Step 3a: Find an appropriate scale of analysis

There are several ways to go about choosing the right distance band, and truthfully, most of the time, there is no "right" distance band. Most of the time there are several spatial processes at play in an area, and you might find different trends depending on the scale that you choose. For instance, if you looked at 911 call hot spots at the national level you might see hot spots in several states. Alternatively, if you look at it at a neighborhood level you may find several small neighborhoods where there is a problem. For each of these analyses there would be a very different distance band that matches the scale of analysis that you are interested in, and the questions that you want to answer.

A rule of thumb is to select a distance based on what you know about the geographic extent of the spatial processes promoting clustering for the phenomena you are studying. Often, you won't know this, but if you do, you should use your knowledge to select a distance value. Suppose, for example, you know that the average journey-to-work commute distance is 15 miles. Using 15 miles for the distance band is a good strategy for analyzing commuting data. When you don't know a precise distance that makes sense for your analysis, you can use the Incremental Spatial Autocorrelation tool to help you find a distance band that reflects maximum spatial autocorrelation. Whenever you see spatial clustering on the landscape, you are seeing evidence of underlying spatial processes at work. The distance band that exhibits maximum clustering, as measured by the Incremental Spatial Autocorrelation tool, is the distance where those spatial processes are most active, or most pronounced. In this next step, you will run the Incremental Spatial Autocorrelation tool and note where the resulting z-scores seem to peak. This tool runs Spatial Autocorrelation at increasing distances and creates a graph of the z-scores at each distance. Then, you will use the distance associated with the peak value for your analysis.

10. Open Incremental Spatial Autocorrelation
    a. In the ArcToolbox window, navigate to the Spatial Statistics toolbox and expand the Analyzing Patterns toolset
    b. Double-click on the Incremental Spatial Autocorrelation tool

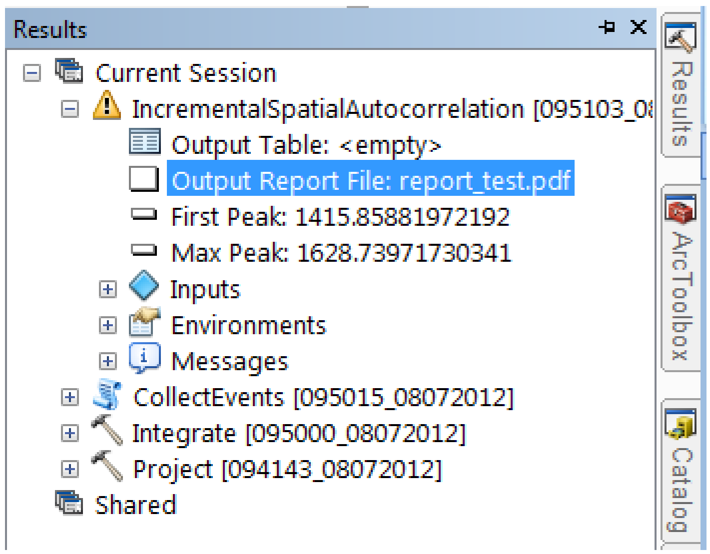


    c. Fill the dialog in as follows:

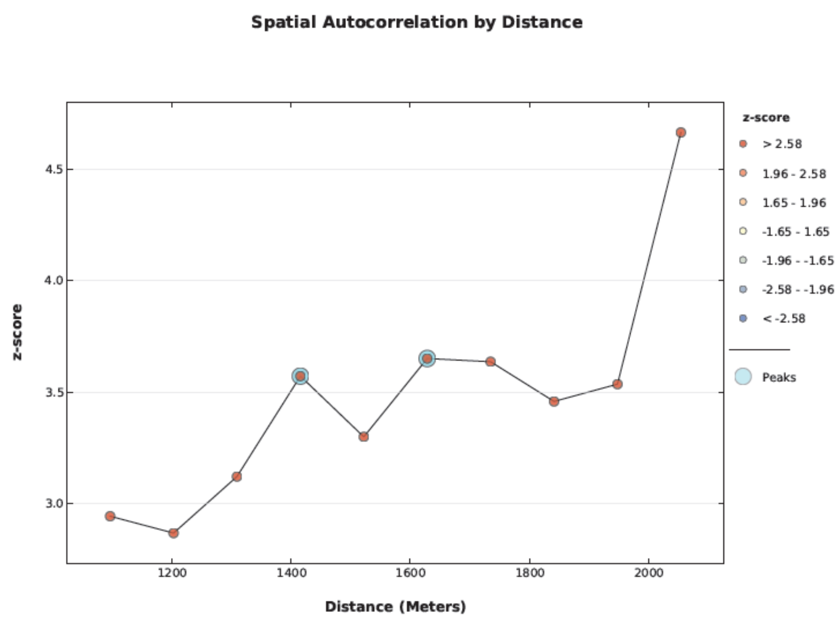
Input Feature Class	AggregatedCalls
Input Field	ICOUNT
Number of Distance Bands	10
Beginning Distance	Default
Distance Increment	Default
Distance Method	Default
Row Standardization	Default
Output Table	Default
Output Report File	…\IncrementalSA_Report.pdf

d. Click ok

e. Open the report from the Results Window by expanding the IncrementalSpatailAutocorelation result and double-clicking on the Output Report File



f. Your graph should look like this:

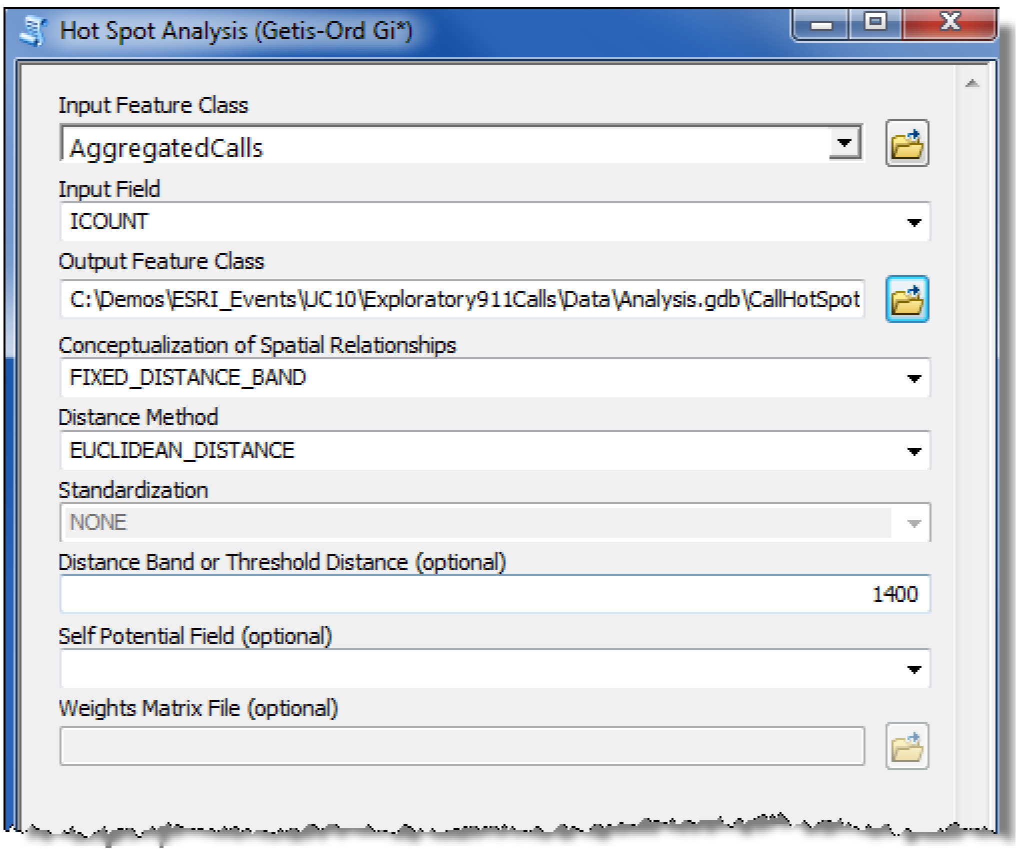


g. Notice that there are two peaks. Because the first peak is most pronounced and the distance associated with that peak more appropriately reflects the scale of analysis that we are interested in for 911 call analysis, we'll use that distance for hot spot analysis. Take note of the distance of the first Peak (about 1400m)

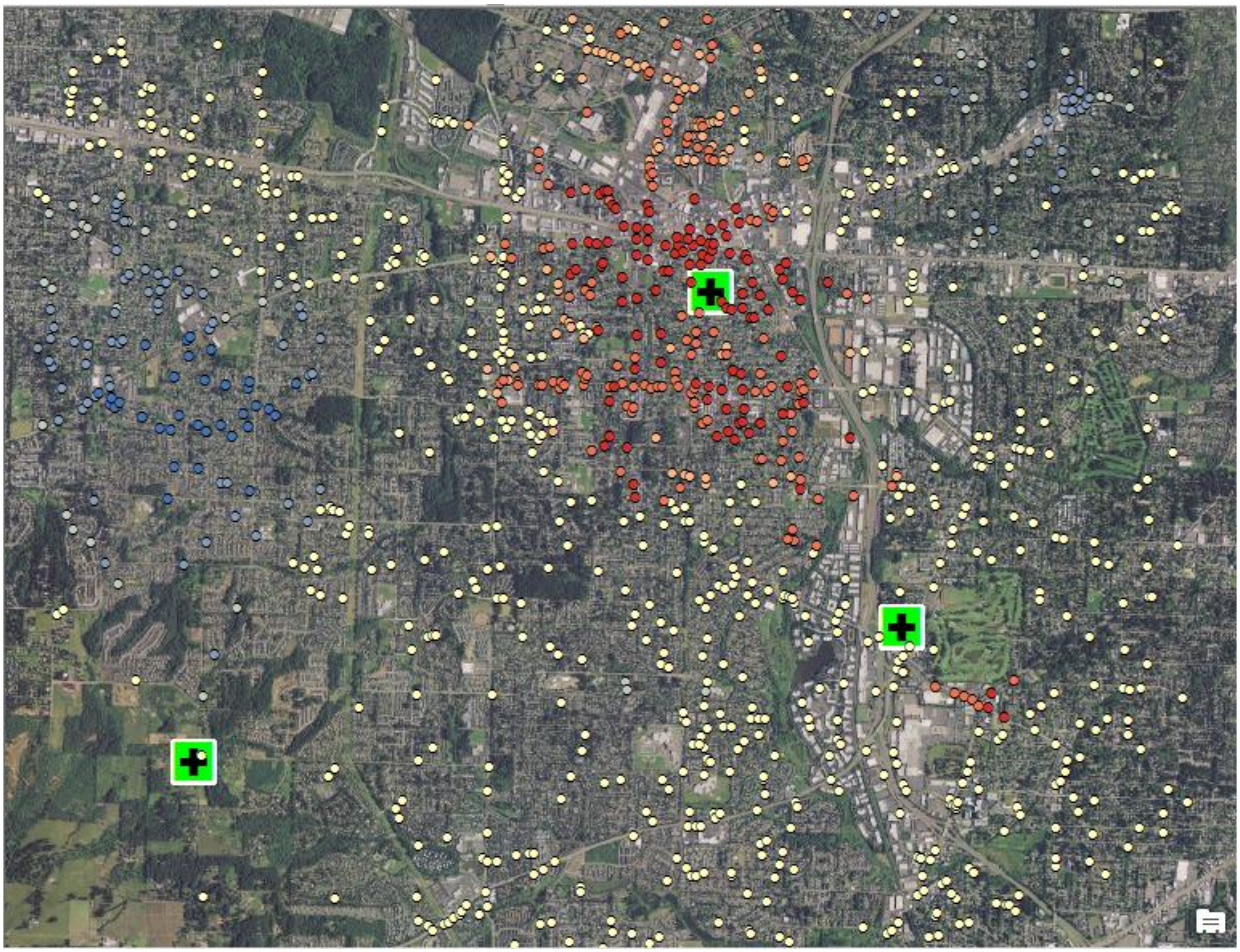
Note: There is no right or wrong distance. What's important is that you choose the distance that best reflects the questions that you're asking. If you are looking for hot spots at the local, neighborhood scale then the first peak with the smallest distance may be appropriate. If you are looking for hot spots at the regional scale then a peak with a larger distance may better reflect your question. You can learn more about Incremental Spatial Autocorrelation here.

11. Run Hot Spot Analysis (Getis-Ord Gi*) tool
    a. Un-minimize the Hot Spot Analysis tool
    b. Continue to fill in the dialog as follows:

Input Feature Class	AggregatedCalls
Input Field	ICOUNT
Output Feature Class	…\CallHotSpots
Conceptualization of Spatial Relationships	Fixed Distance Band
Distance Method	Euclidean Distance
Standardization	None
Distance Band or Threshold Distance	First peak distance from graph (1400m)
Self Potential Field	Blank
Weights Matrix File	Blank



    c. Click OK
    d. Turn off all layers except CallHotSpots, Response Stations, and Basemap



The result of the Hot Spot Analysis tool is a new feature class where every feature in your dataset is symbolized based on whether it is part of a statistically significant hot spot, a statistically significant cold spot, or is not part of any statistically significant cluster. The red areas are hot spots, or areas where high numbers of 911 calls are surrounded by other areas with high numbers of 911 calls. The blue areas are cold spots, or areas where low numbers of 911 calls are surrounded by other areas with low numbers of 911 calls. The beige areas are not part of statistically significant clusters. Statistical significance is based on p-values and z-scores that are calculated when you run Hot Spot Analysis. You can learn more about statistical significance and the output of Hot Spot Analysis here.

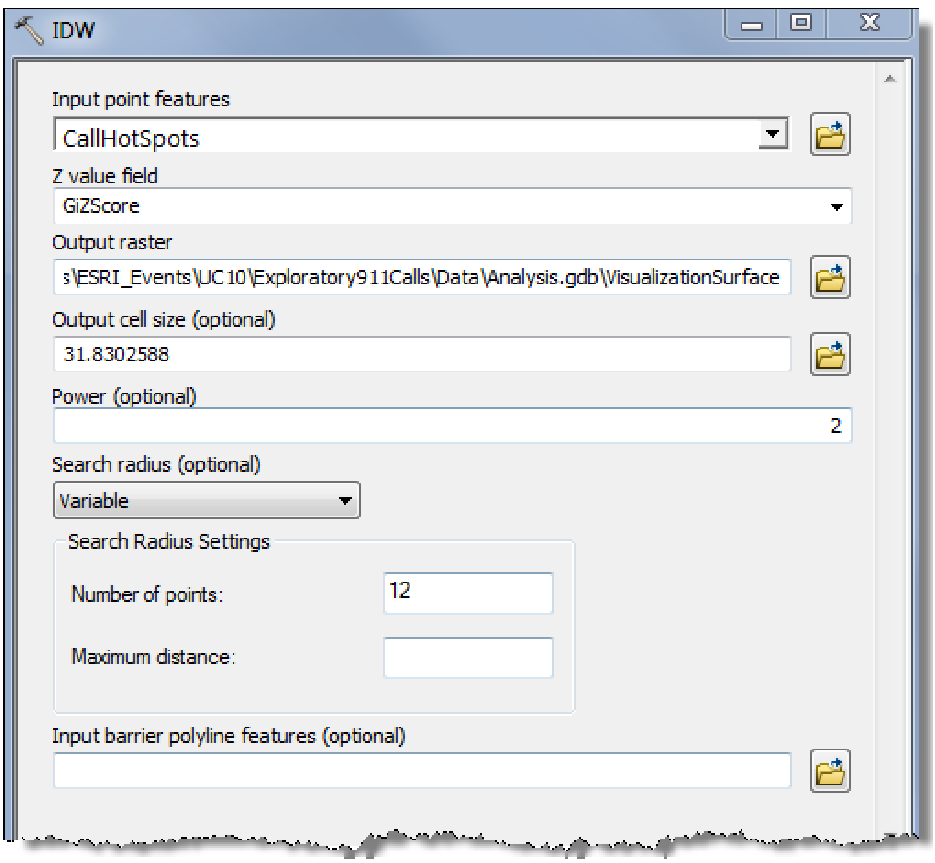
While you as the analyst understand that each location has a particular z-score and a particular p-value that determine the symbology and can guide decision-making, often you are required to share your analysis with a broader audience who may be expecting the results to look more like a continuous surface, like the heat maps people have become accustomed to in the media. In order to make the results of your analysis more approachable to a broader audience, the next step is to create a continuous surface to visualize the results. ***It is important to remember that this surface is for visualization purposes only, and the true statistical analysis happens feature by feature***. Showing both the surface and the true results of the hot spot analysis at the same time is a great way to present both the statistical results and the more approachable visualization in a defensible way.

Step 4: Visualize your results (Spatial Analyst Extension required)

There are many ways to create an interpolated surface that will effectively visualize the results of a hot spot analysis. The method that you will use in this tutorial, IDW (Spatial Analyst), interpolates a raster surface from points using an inverse distance weighted technique.

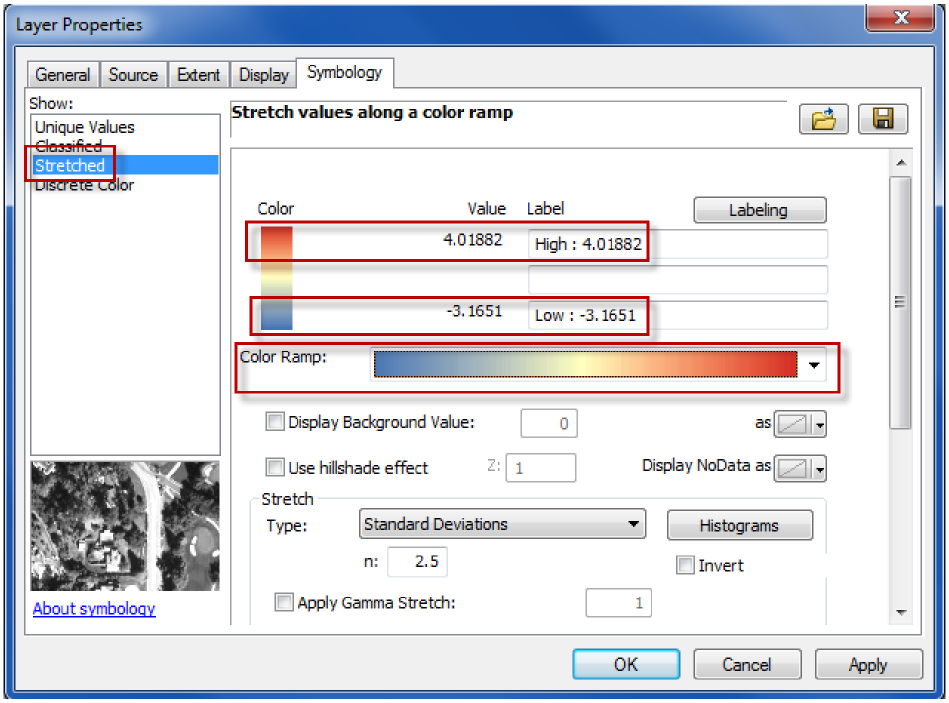
12. Create a visualization surface (using the Spatial Analyst extension)
    a. Search using the keyword 'IDW' and choose IDW (Spatial Analyst) to open the tool
    b. Fill in the dialog as follows:

Input Point Features	CallHotSpots
Z value field	GiZScore
Output Raster	…\VisualizationSurface
Output Cell Size	Default
Power	Default
Search Radius	Default
Number of Points	Default
Maximum distance	Default
Input barrier polyline features	Default

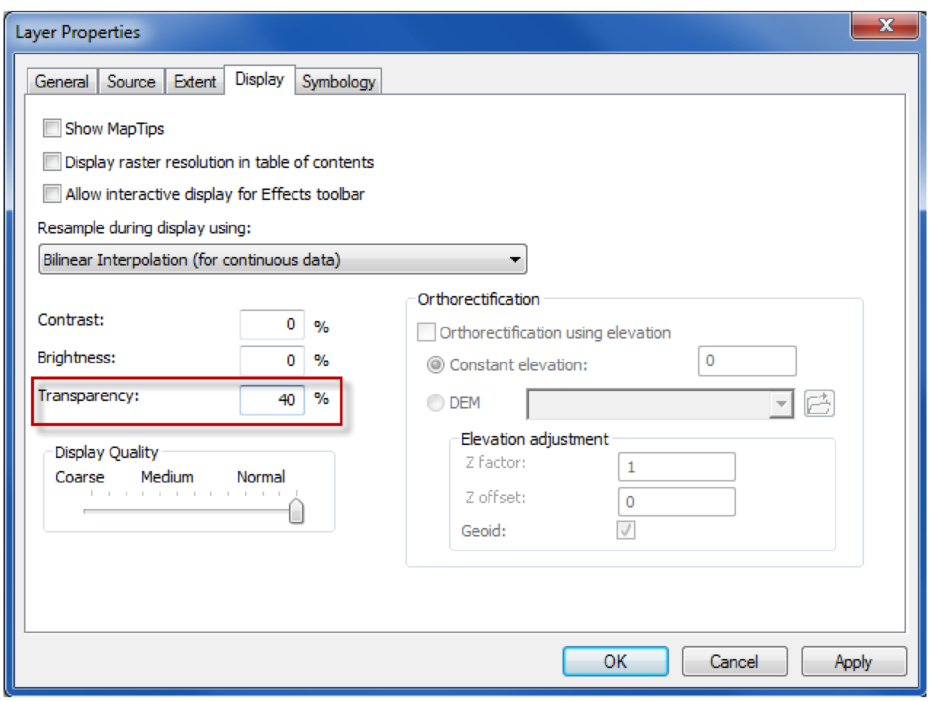


    c. Click OK
    d. Double-click on VisualizationSurface in the Table of Contents and click on the Symbology tab (we want the symbology of the raster to roughly match the red to blue symbology of our Hot Spot Analysis feature class)

e. Choose Stretched, choose the blue to red color ramp, and make sure that the high values correspond to the red colors

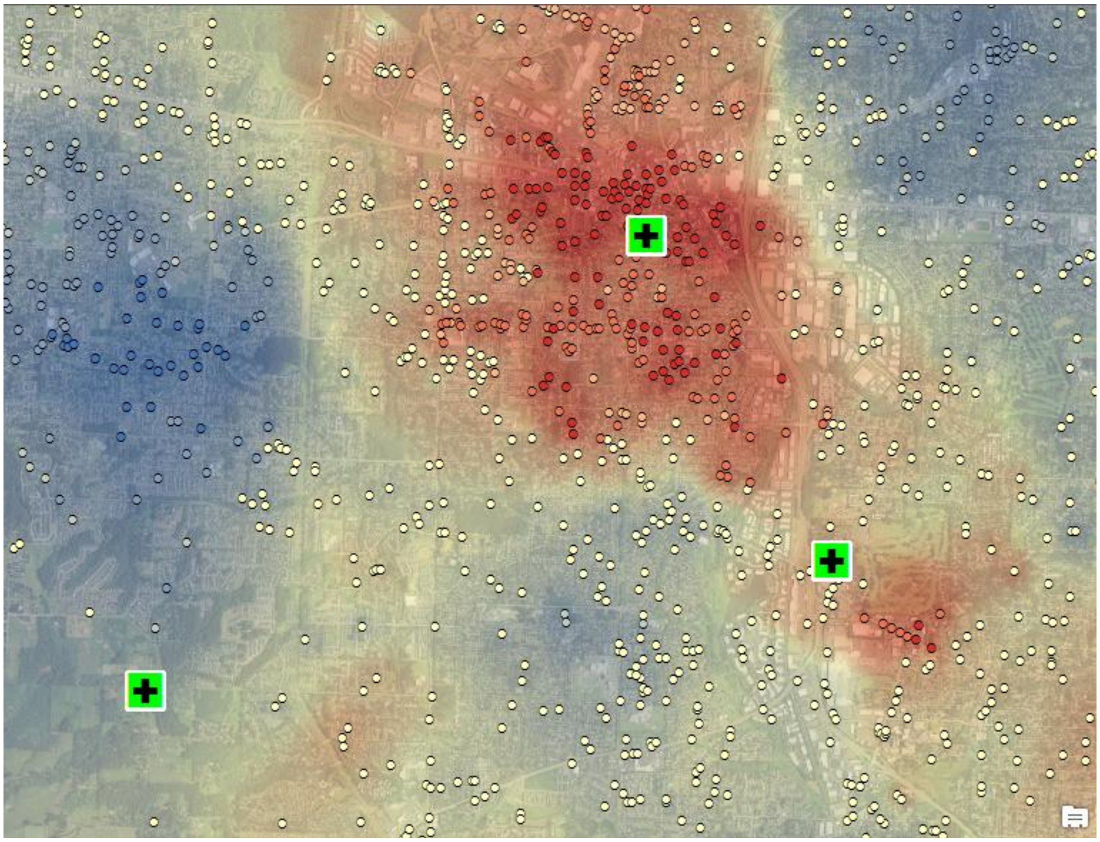


f. Click on the Display tab and give the surface 40% Transparency



g. Move VisualizationSurface just underneath CallHotSpots, leaving them both on so that you can see the statistical results and the visualization surface simultaneously

h. Move Response Stations above both the VisualizationSurface and CallHotSpots

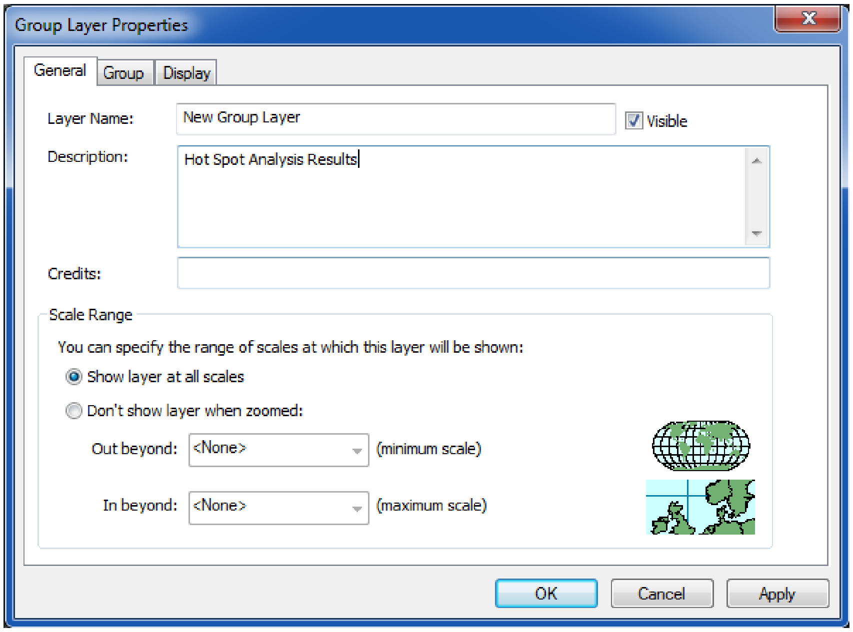


One application of spatial statistics and the Hot Spot Analysis tool is to help to resolve resource allocation problems. The Portland, OR metropolitan area authorities can use this analysis to decide if adding new stations or relocating existing stations might improve 911 call response. What would your recommendations be, based on the results of your analysis?

Now that you have finished your analysis, you will share the analysis on ArcGIS Online so that others inside of your organization and in the community can understand why and how decisions are being made.

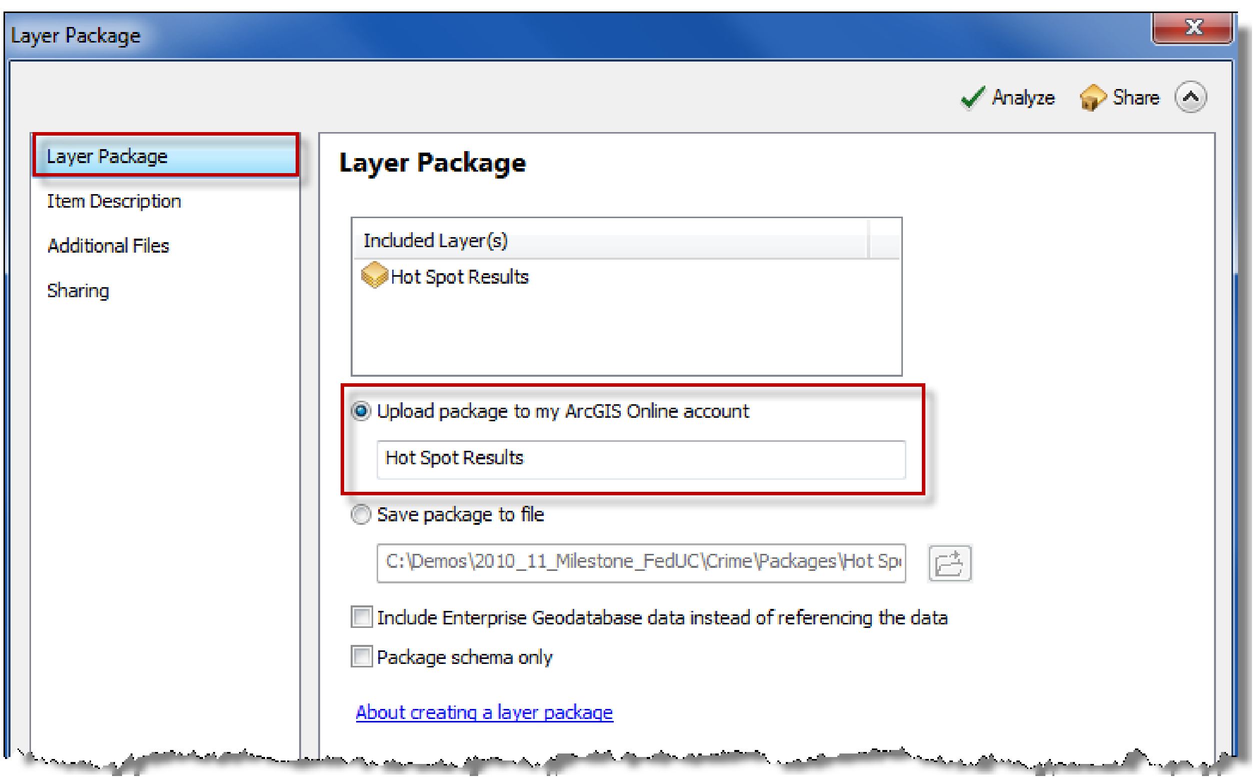
Step 5: Share your findings

13. Create a Group Layer with your results
    a. Select Response Stations, CallHotSpots, and VisualizationSurface (to select multiple items in the Table of Contents, hold down Ctrl and click each one)
    b. Right-click on the selected layers and choose Group
    c. Right-click on the newly created group, and choose Properties
    d. From the General tab, give the group layer a description

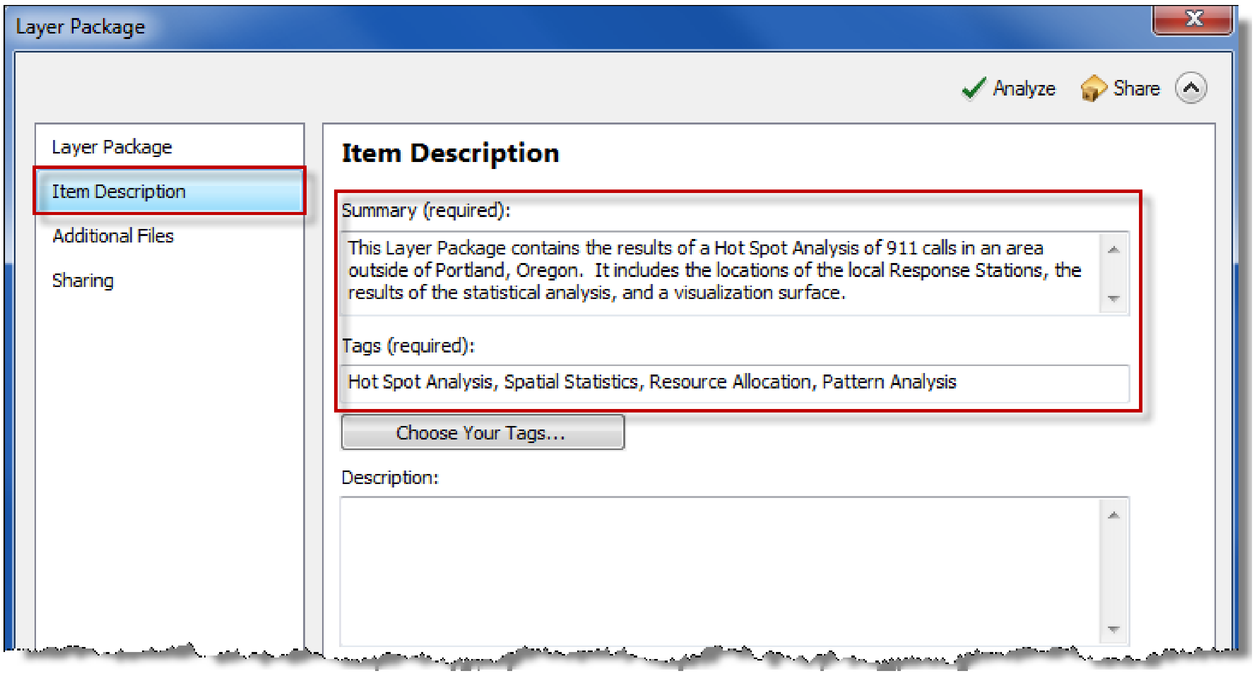


14. Share your results
    a. Right-click on the Hot Spot Results Group Layer, and choose to Create Layer Package
    b. From the Layer Package tab, choose to upload the layer package to your ArcGIS Online account

Note: If you are not already signed in, you will be prompted to sign-in to ArcGIS Online using your Esri Global Account. If you do not already have an Esri global account, or if your global account is not registered with ArcGIS Online, you can do so here.



c. From the Item Description tab, fill in the Summary and Tags, which are required for the packaging process (you may also fill out the Description, Access and Use Constraints, and Credits if you want)



d. Click Share (in the top right corner)

Conclusion

In this tutorial you have used a Hot Spot Analysis to provide decision-makers with an understanding of the spatial patterns of their 911 calls and enabled them to make more informed decisions about how to allocate resources. Along the way you have thought about the importance of projecting your data, learned how to aggregate your incident data (as well as several alternative methods of aggregation), and you've made some important decisions about the scale of your analysis. You also learned about the importance of visualizing and sharing the results of your analysis.

Now that decision-makers have information about where there are hot spots and cold spots of 911 calls in the area, the next logical question is 'Why?'. To answer those 'why' questions and start to understand what factors might be related, you may want to think about working through the Regression Analysis Tutorial.