Using GIS Training to Address Blood Pressure Medication Adherence
Analyzing patterns in your data

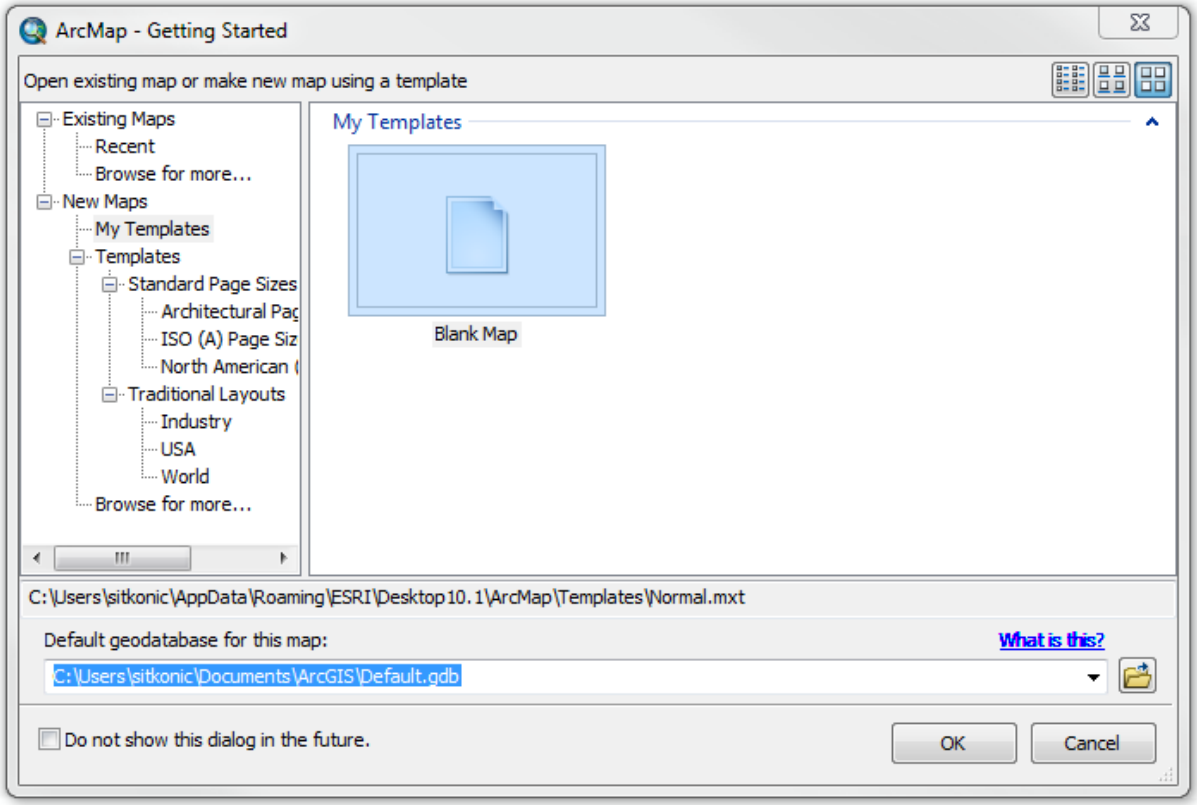
*** Files needed for exercise: *CA_BG_2010_pop.shp, CA_tract_2010.shp, CA_tract_pharm_OD_5closest.shp, and LA_greater_500Cities_tract_HBP.shp*

Goals: The goal of this exercise is to explore how to analyze patterns in your data. You will first create population weighted centroids (PwC) for census tracts in California to conceptualize where your population lives; and you are going to use recently generated population centers to inform the hotspot analysis on high blood pressure prevalence in the Greater Los Angeles Area.

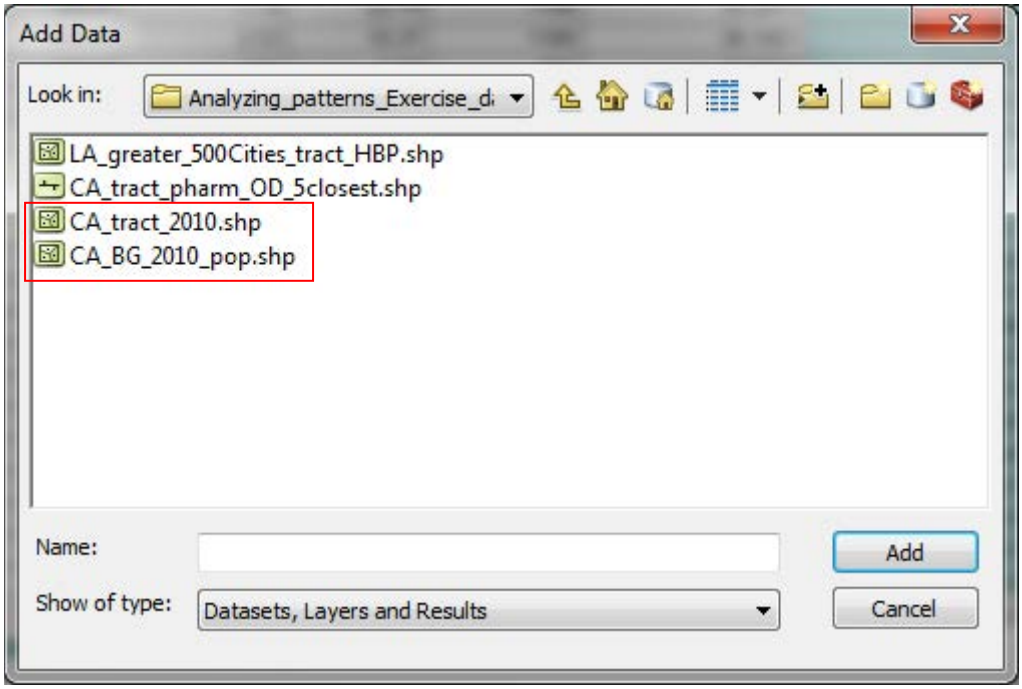
Skills: After completing this exercise, you will be able to generate population weighted centroids at different scales. You will learn how to use hotspot analysis tool provided by ArcGIS, and properly interpret the hotspot map.

Preparing Workspace

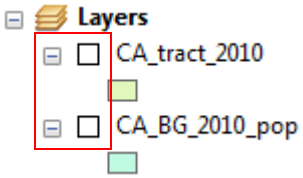
1. In the Windows Explorer Start Menu, open ArcMap.
2. In the folder tree on the left side, browse to **New Maps > My Templates** and select a **Blank Map**. Click **OK**.



3. Click the **Add Data** icon to add new data to the map.
4. Click on the **Connect to Folder** button .
5. Browse to the data folder for this exercise (*Analyzing_patterns_Exercise_data*) and click **OK**. You now have a permanent connection to that folder making it quicker to find and add data.
6. From the folder, hold Ctrl and click the shapefile *CA_tract_2010.shp* and *CA_BG_2010_pop.shp*. These files are projected census boundaries for California. And the block group file includes population counts.



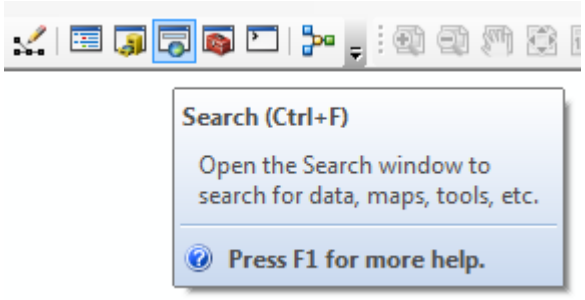
7. Once you've selected the files, click **Add** and the data will appear in the **Table of Contents** as a layer.
8. Now you have your census boundaries added into your workspace as shapefile layers. Because of the complexity of CA's boundaries, uncheck the display box before both of layers to disable the display for better performance.



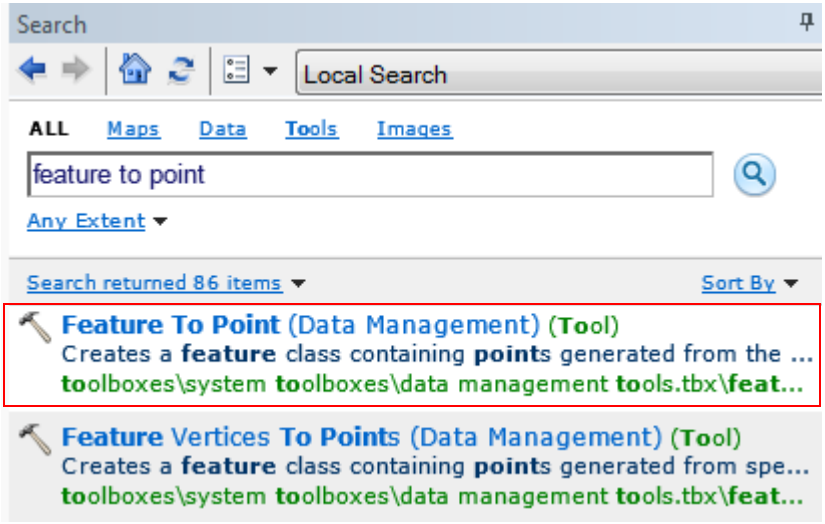
Using GIS Training to Address Blood Pressure Medication Adherence
Analyzing patterns in your data

Generating Population weighted Centroid for Census Tracts

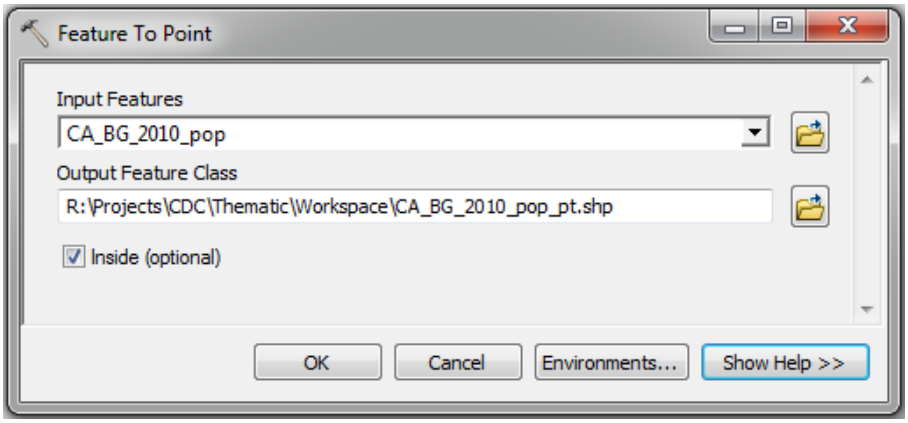
1. In order to generate the population weighted centroid for census tracts, we need sub-units of tracts with total population counts. Right click on *CA_BG_2010_pop* to **Open Attribute Table**. We have total population but no tract identifier to categorize on. We are going to use **Spatial Join** tool to resolve that.
2. Click on the **Search** icon to call out the search tool tip.



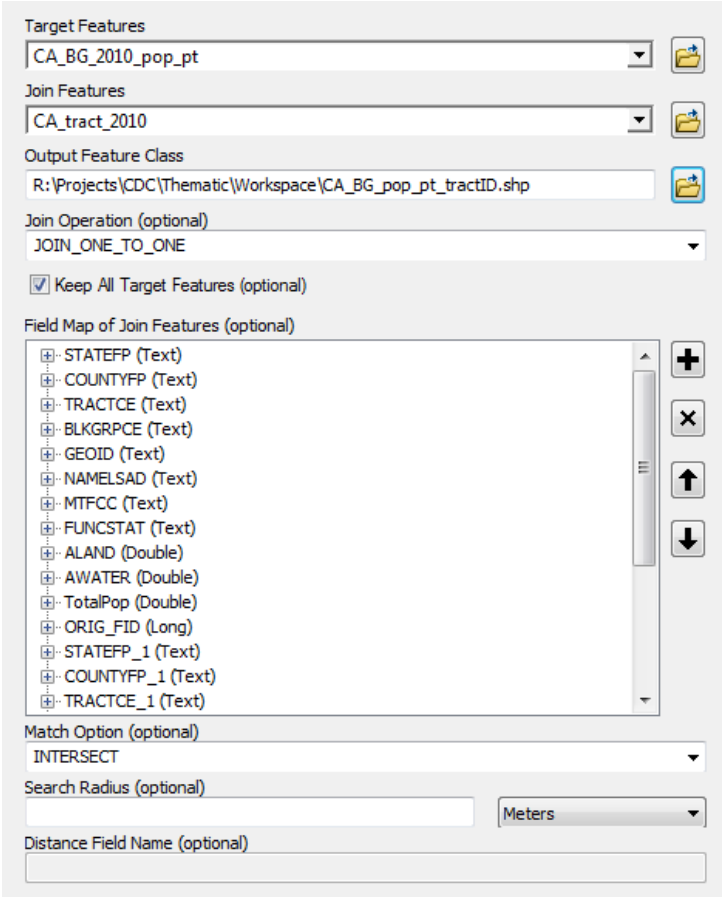
3. You want to convert all shapes into points in order to simplify the spatial relationship for spatial join. Type **feature to point** in the search box, and **select Feature To Point (Data Management)** tool in the search results.



4. Select *CA_BG_2010_pop* in the pop up window and name the output correspondingly. Please make sure to check the **Inside** option to keep the topological relation to the census tracts.

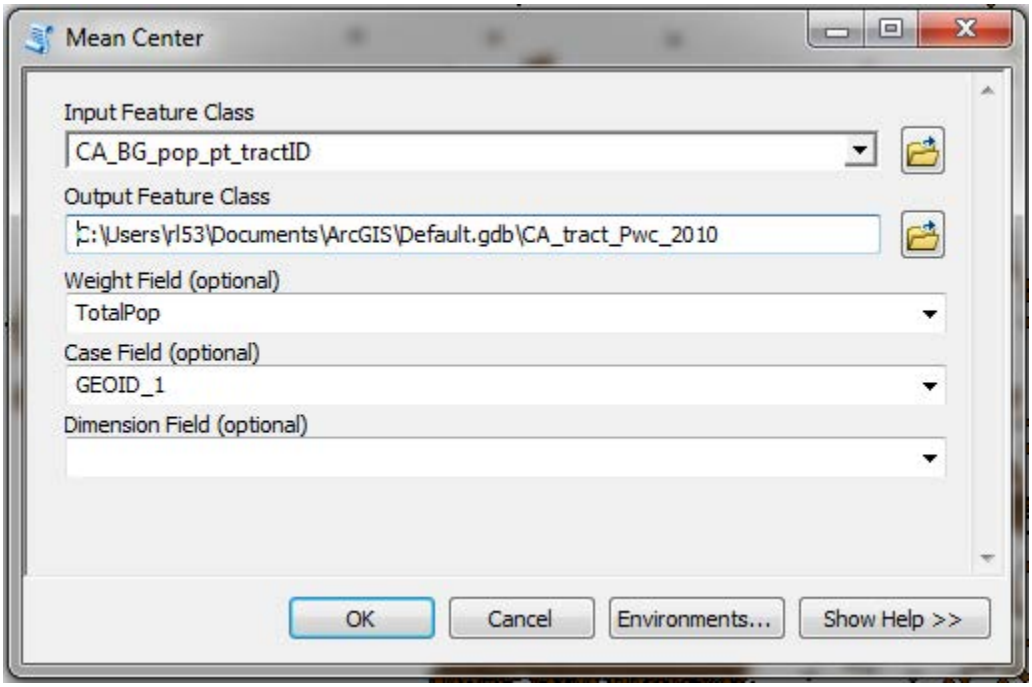


5. Use the Search tool again as in step 2 to find **Spatial Join** Geoprocessing tool under Analysis toolbox. Select the newly generated *CA_BG_2010_pop_pt* as the input dataset (because we need to attach tract ID to block groups). Use *CA_tract_2010* as Join feature. Join Operation: you want to JOIN_ONE_TO_ONE. Keep the default: keep all target features checked, and Match option: INTERSECT. Make sure you save it in your Exercise_Data folder as a shapefile.



6. Now that you have census block groups' centroid, population, and census tract ID, you can start to generate the population weighted centroid. Use the **Search** tool to find the **Mean Center tool** under **Spatial Statistics** category and click it.

7. Choose the block group points with population and tract ID as input feature data. Set **Weight Field** to be **TotalPop** and **Case Field** to be **GEOID_1** (GEOID from census tracts). Don't forget to save your output file to an accessible location with a meaningful name. Then click **OK**.

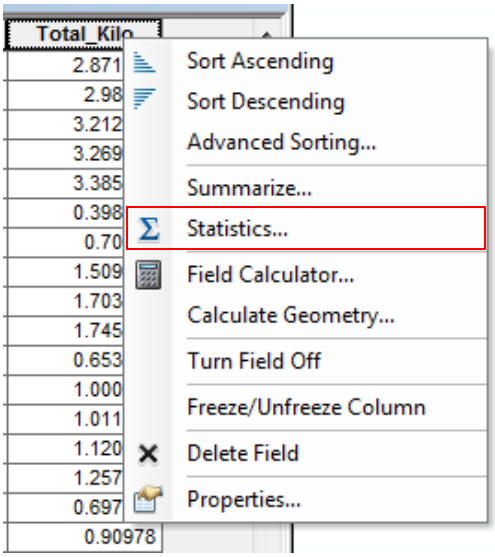


8. You will see the warning message display in the result window. It is because we didn't exclude the census tracts with zero population. So there is no population center generated for those census tracts. That won't affect our analysis.
9. Remove all dataset except CA_tract_PwC_2010 to make your workspace clean. You are going to use the newly generated population weighted centroids to run hotspot analysis.

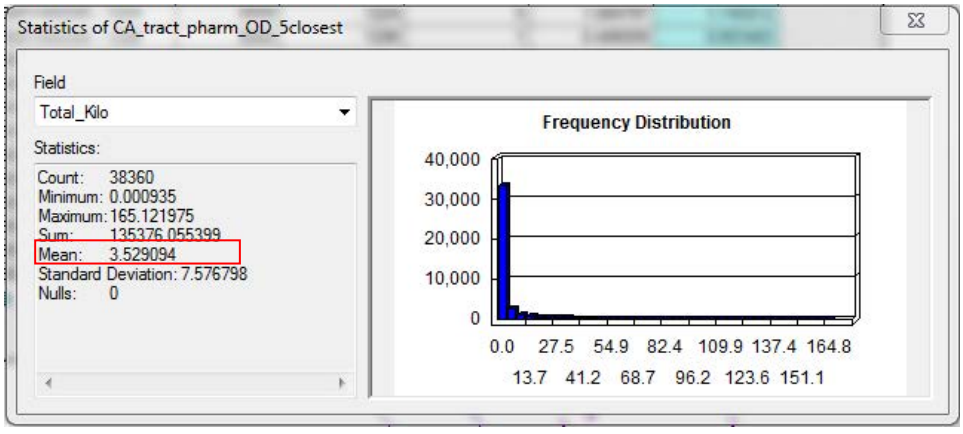
Using GIS Training to Address Blood Pressure Medication Adherence
Analyzing patterns in your data

Define your Study Scale for Hotspot analysis

1. Next, you will try to use your newly generated PwC data to define the study scale for hotspot analysis. As mentioned in the presentation, we are trying to use the average driving distance to the closest 5 pharmacies as the proximity of the community size. We can start to use the point level pharmacy data in module 1 and OD cost matrix in module 2 to generate the OD cost lines. In this exercise we provided the file *CA_tract_pharm_OD_5closest.shp* in the folder. Add *CA_tract_pharm_OD_5closest.shp* to the workspace.
2. With the help of **Statistics** tool, you can calculate the average driving distance to the nearest 5 pharmacies.
3. **Open Attribute Table** of *CA_tract_pharm_OD_5closest.shp*. Right click on the field *Total_Kilo*. This field contains the driving distance from PwC to pharmacies in kilometer. Select **Statistics** in the dropdown menu.



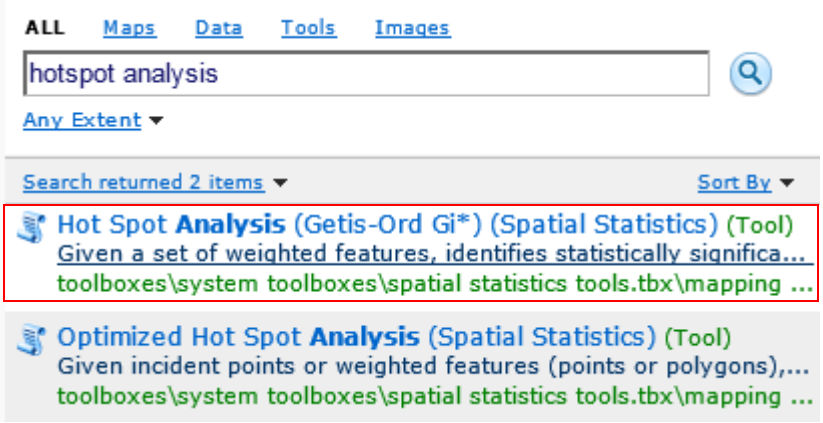
4. Find **Mean** for the Total_Kilo in the pop up. It reads 3.529 km.



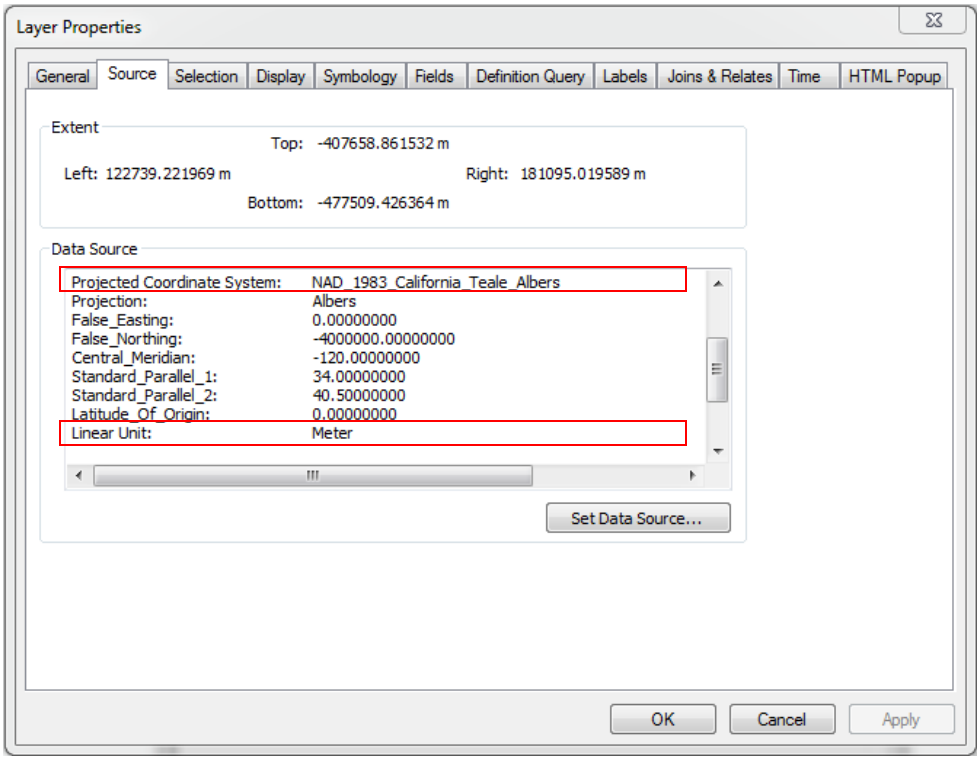
Using GIS Training to Address Blood Pressure Medication Adherence
Analyzing patterns in your data

Hotspot Analysis on Greater Los-Angeles area

1. Browse to the data folder and add *LA_greater_500Cities_tract_HBP.shp* to your workspace. This data comes from CDC's 500 cities project. It includes both model smoothed crude high blood pressure prevalence and high blood press medical adherence at census tract level.
2. Use **Search** tool to find **Hot Spot Analysis (Getis-Ord G*)** in spatial Statistics tool set; and click on it.

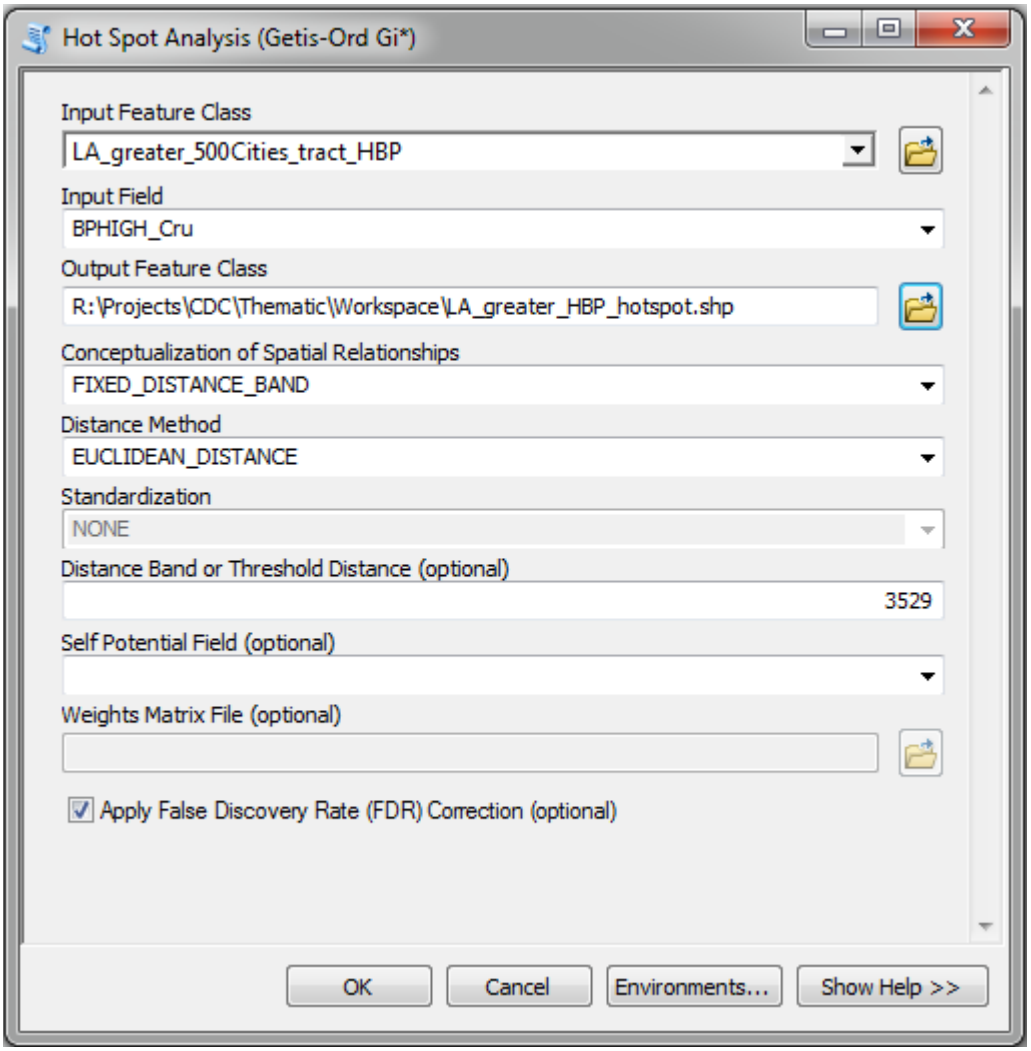


3. Before you run the **Hot Spot Analysis** tool, you want to find out the projection information of the input dataset so you can put in the number for **Distance Band** parameter.



Right click on *LA_greater_500Cities_tract_HBP* and click **Properties**. You can find coordinate system information under **Source** tab. Your data has been projected to right system and the measuring unit in your coordinate system is **Meter**. Since the number you get from last step is 3.529 km, your input for **Distance Band** should be 3529 meters.

4. Go back to **Hot Spot Analysis**, select *LA_greater_500Cities_tract_HBP* as Input Feature Class; BPHIGH_Cru (smoothed high blood pressure crude rate) as Input Field; specify the output folder and file name; choose FIXED_DISTANCE_BAND for Conceptualization of Spatial Relationships; leave Distance Method as default (EUCLIDEAN_DISTANCE); and put in 3529 meters in Distance Band. Finally, don't forget to check Apply False Discovery Rate to minimize type I error.



5. Click **OK** and you have done your first Hot Spot analysis. Try to symbolize the BPHIGH_Cru field in *LA_greater_500Cities_tract_HBP*. Please think about the following questions:

- Any similarity of two maps?
- Any difference of two maps?
- How is the hotspot map going to help you in your working environment?
- Feel free to run multiple test by using different number in Distance Band. What do you observed? How is the observation going to impact the usefulness of the map?